Лекція А

ЛІНГВІСТИЧНЕ МОДЕЛЮВАННЯ

1. Лінгвістичне моделювання

1.1. Формальна граматика

1.2. Форми реалізації формальних граматик

1.3. Приховані марковські моделі

2. ДАНІ ДЛЯ МОДЕЛЮВАННЯ

Вхідними даними для моделювання є часовий ряд довжиною M:

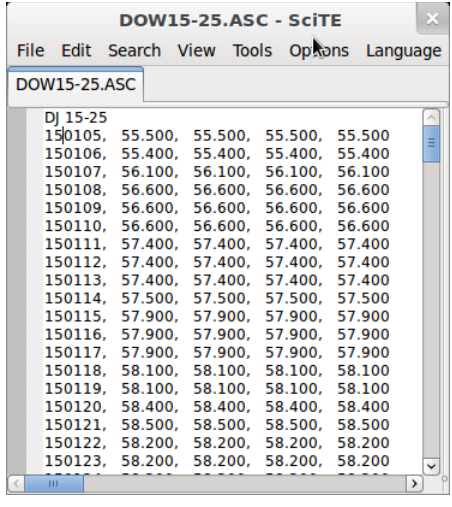
$X=\{x_i\}^{M}_{i=1}=\{x_1,x_2,...,x_i,...,x_M\}$,

який описує деякий динамічний процес.

Для розрахунково-графічної роботи пропонується для моделювання використання часових рядів економічно-фінансової природи — біржеві індекси, котирування акцій, курси валют тощо.

Конкретні дані у файлі студент бере у викладача при отриманні завдання на розрахунково-графічну роботу.

Так, наприклад, студент отримує файл щоденних змін біржового індексу Доу-Джонс с 1915 по 1925 рік.



Перший стовпчик це дата - “150105” - 5 січня 1915 року.

Для моделювання береться другий стовпчик — значення індексу — 55.500.

3. ОСНОВНІ КРОКИ ВИКОНАННЯ РОЗРАХУНКОВО-ГРАФІЧНОЇ РОБОТИ

1. Побудова на основі часового ряду $X=\{x_i\}^{M}_{i=1}$ (який видається викладачем) різницевих рядів X^1, X^2,... :

$$x^1_i=x_{i+1}-x_i$$

$$x^2_i=x^1_{i+1}-x^1_i$$

...

2. Сортуємо ряд за зростанням $X^1 \rightarrow X^{s1}$.

3. Знаходимо max (X^{s1}) та min(X^{s1}).

4. Розбиваємо відрізки [min(X^{s1}),0] та [0,max (X^{s1})] на N відрізків за правилами інтервалізації, відповідно до свого варіанту. N змінюється від 10 до 33 (в залежності від алфавіту, який буде обраний на етапі лінгвістизації).

Варіант	Правило інтервалізації 1	Правило інтерваліза
1	Рівнозначні інтервали	Розподіл Гауса
2	Рівноймовірний розподіл	Розподіл Лапласа
3	Біномальний розподіл	Розподіл Пуасона
4	Логарифмічний розподіл	Бета-розподіл
5	Рівнозначні інтервали	Розподіл Дирихле
6	Рівноймовірний розподіл	Розподіл Стьюдента
7	Біномальний розподіл	Розподіл Гауса
8	Логарифмічний розподіл	Розподіл Лапласа
9	Рівнозначні інтервали	Розподіл Пуасона
10	Рівноймовірний розподіл	Бета-розподіл

11	Біномальний розподіл	Розподіл Дирихле
12	Логарифмічний розподіл	Розподіл Стьюдента
13	Рівнозначні інтервали	Розподіл Стьюдента
14	Рівноймовірний розподіл	Розподіл Дирихле
15	Біномальний розподіл	Бета-розподіл
16	Логарифмічний розподіл	Розподіл Пуасона

Розбиття на інтервали відбувається таким чином, щоб кількість елементів різнецевого (1-ї або 2-ї різниці) ряду в кожний інтервал потрапляла у відповідності до певного розподілу. Тобто частота попадання елементів до інтервалу [a,b] дорівнювала теоретичній ймовірності

$$P\{x \in [a,b]\} = F(b) - F(a) \quad ,$$

де F — функція відповідного розподілу.

В результаті отримуємо дві множини інтервалів:

1) $I_{0,1}=[a_0,a_1]$, $I_{1,2}=[a_1,a_2]$,..., $I_{N-2,N-1}=[a_{N-2},a_{N-1}]$, $I_{N-1,N=}[a_{N-1},a_N]$, де $a_0=\min(X^{s1})$, $a_N=0$;

2) $J_{0,1}=[b_0,b_1]$, $J_{1,2}=[b_1,b_2]$,..., $J_{N-2,N-1}=[b_{N-2},b_{N-1}]$, $J_{N-1,N=}[b_{N-1},b_N]$, де $b_0=0$, $b_N=\max(X^{s1})$

Обираємо алфавіт потужності 2N відповідно до обраного N. Якщо, наприклад N=26, то можна за основу взяти алфавіт A={a,b,...,z,A,B,C,...,Z}, dim(A)=26x2=52, A= $\{\alpha_i\}_{i=1}^{2N}$.

Відсортуємо символи алфавіту у наступному порядку: $\alpha_1 = z$, $\alpha_2 = y$, . . . , $\alpha_{N-1} = b$, $\alpha_N = z$, $\alpha_{N+1} = A$, $\alpha_{N+2} = B$, . . . , $\alpha_{2N-1} = Y$, $\alpha_{2N} = Z$.

5. Побудувати відображення L: $X^1 \rightarrow Y$ за такими правилами:

$$L(x_i)=\begin{Bmatrix} \alpha_j & якщо & x_i \in I_{j-1,j} \\ \alpha_{N+j} & якщо & x_i \in J_{j-1,j} \end{Bmatrix}$$

Застосувати відображення L до елементів ряду X1. В результаті отримуємо ряд:

$L(x^1_1), \ldots, L(x^1_M)$.

6. Будуємо матрицю передування для прихованої марковської моделі.

Множина станів - це обраний нами алфавіт.

	z	...	a	A	...	Z
z						
...						
a						
A						
...						
Z						

Рис.1. Матриця передування

Для кожної пари станів, наприклад <d,S> підраховуємо $v_{d,S}$ скільки разів вона зустрічається в лінгвістичному ланцюжку $L(x^1_1), \dots, L(x^1_M)$. Поділивши $v_{d,S}$ на загальну кількість входжень літери “d” w_d отримуємо частоту переходів зі стану “d” в стан “S”:

$$\nu(d \to S) = \frac{v_{d,S}}{w_d} \quad .$$

7. Знайти в лінгвістичному ланцюгу повтори переходу від двох, трьох та більше станів.
8.

9. Побудувати розширену матрицю, додавши до станів варіанти двох, трьох та більше станів, що зустрічаються в нашому лінгвістичному ланцюгу.

10. Побудувати візуальне відображення матриці, замінивши частоти пофарбуванням клітинки таблиці матриці у кольори від білого до чорного в залежності від значення частоти.

	1	2	
	1	3	
	4	5	6
	6	7	8

Рис.2. Кольорове відображення елементів матриці передування

10.Побудувати по розширеній матриці передування правила ймовірнісної граматики:

	...	Z	...
...	...	...	...
SaX	...	0,5	...
...	...	...	...
...	...	...	...

Рис.3. Розширена матриця передування

Тобто для кожної ненульової клітинки (див. приклад на рис.3) будується правило наступного вигляду:

$$Sax \xrightarrow{0.5} Z$$

Алфавіт та правила передування утворюють лінгвістичну модель ряду X^1.

11.

12. Ту саму процедуру побудови лінгвістичної моделі повторюємо для інших різниць ряду X - X^2, X^3, X^4, X^5, X^6.
13.

14. Будуємо лінгвістичну модель за п.1-11 для алфавіту потужності — 10, 15, 20, 26.

15. Зробити аналіз відмінностей результатів лінгвістичного моделювання одного й того ж самого чисельного ряду, які виникають при двох різних правилах інтервалізації (при незмінному алфавіті та його потужності).

Програмна реалізація повинна давати можливість зміни вхідного числового ряду, зміни алфавіту та його потужності, а також результатів лінгвістичного моделювання на друк (екран, файл) лінгвістичного ланцюга, розширеної матриці передування та правил передування.

Програмна реалізація здійснюється в системі MathCAD та на мові програмування, обраній студентом особисто.

4. ЗМІСТ ЗВІТУ

Звіт повинен мати наступні розділи:

1. Завдання на графічно розрахункову роботу.
2. Опис вхідних даних, природи їх, стислий опис предметної області, особливості їх отримання.
3. Теоретичні відомості про розподіл ймовірностей з відповідного варіанту завдання на РГР, а також алгоритми, функції та особливості реалізації цих розподілів ймовірностей. Повинні бути приведені попередні розрахунки для формування розподілів ймовірностей.

4. Хід виконання розрахункової роботи з виведенням відповідних проміжних та кінцевих результатів у вигляді таблиць, графіки тощо.
5. Аналіз результатів моделювання.

Додаток 1. Програма на MathCAD.

1) Лістінг
2) Скріншоти виконання

Додаток 2. Програма на обраній студентом мові програмування

1) Лістінг
2) Скріншоти виконання

Звіт роздруковується, а електронна його версія та відповідні програми надсилаються за 3 дні до захисту РГР на e-mail викладача.

Під час захисту РГР студент демонструє розроблені ним програмні засоби та робить відповідні пояснення.