headsep=1cm

Markov Models and Their Appliances

Ivan Zhytkevych

September 14, 2023

Contents

1 Lecture 1: Examples of Hidden Markov Models

1.1 Приклад: середньорічна температура у минулому

Example. Хотіли б визначити середню річну темепратуру в деякій області Землі, достатньо далекого минулого, про яку немає ніяких метеорологічних записів. Потрібно знайти якісь опосередковані методи. Для спрощенні будемо вважати що річна температура може бути просто високою або низькою.

Temperature: T (warm year), X (cold year).

Введемо $\{X_k\}_{k\leq 0}$ зі множиною станів $E = \{T, X\}$ де X_k температурна характеристика k-го року.

$$P\left(X_{k+1} = T \mid X_k = T\right) = 0.7.$$

$$P\left(X_{k+1} = X \mid X_k = X\right) = 0.6.$$

Введемо матрицю перехідних ймовірностей із рядками та стовпцями, що відповідають T та X температурним режимам.

$$A = \begin{array}{c c} & \begin{array}{cc} T & X \end{array} \\ \begin{array}{c} T \\ X \end{array} & \begin{pmatrix} 0,7 & 0,3 \\ 0,4 & 0,6 \end{pmatrix} \end{array} .$$

Імовірність переходу з теплого у теплий 0.7. З теплого у холодний 0.3.

Перехід схожий з умовою марковості: перехід залежит лише від температури попереднього року для температури тепершнього.

Ми шукаємо якісь опосередковані способи поміряти темепратуру. Якісь сучасні дослідження показують, що є певна кореляція між шириною кілець на спилах дерев і темепратурою якв спостерігалась у тей чи інший рік.

Тепер ми маємо спостережуваний об'єкт, а температура попередніх років не була такою. Нехай кільця на спилах дерев такі: M малі, C середні, B - великі.

$$B = \begin{pmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{pmatrix}.$$

Нехай у теплий рік ми спостерігали мале кільце з імовірністю 0.1, середнє імовірністю 0.4 та велике з 0.5.

Відповідно із таблицею і для холодного року.

Матриця B це інформація про зв'язок мід Y_k - розмір кільця в k-тий рік і X_k - температурою.

$$B_{ij} = P\left(Y_k = j \mid X_k - i\right), \qquad j \in \{M, C, B\}\ i \in \{T, X\}.$$

Щоб виправдати слово "приховані" ми кажемо, що
$\{X_k\}_{k\leq 0}$ прихована послідовність;

$\{Y_k\}_{k\leq 0}$ спостережувана.

Припустимо, що ми цікавимо конкретним чотирирічним періодом,
що $Y = (M, C, M, B)$ - наші спостереження.

За Y хочемо визначити найбільш ймовірну послідовність $X = (X_0, X_1, X_2, X_3)$.

Це є класична задача прихованих марківських моделей: за певною послідовністю випадкових величин визначити імовірну послідовність, яка відповідає певному прихованому процесу, з яким наші спостереження пов'язані у відомий нам спосіб.

Для того, щоб далі нам продовжити нам потрібно зробити ще одне припущення, що ми знаємо, чи може зробили якісь припущення про апріорний розподіл. Ввадаємо, що нам відомо, чи так було стосовно певних спостережень, що відомі такі ймовірності:

$$P\left(X_0 = T\right) = 0.6; \quad P\left(X_0 = X\right) = 0.4.$$

Таку модель позначають ще $\lambda = (\mu, A, B)$ - позначення прихованої марківської моделі.

Як би ми могли розв'язувати таку задачу?

Розлянемо скінченновимірні розподіли X_k та Y_k : $\{(X_k, Y_k)\}$

$$\begin{aligned} P_\lambda = &\left(X_0 = x_0, \ldots, X_n = x_n, Y_0 = y_0, \ldots, Y_n = y_n\right) = \\ &= P_\lambda\left(X_0 = x_0, \ldots, X_n = x_n\right) \cdot P_\lambda\left(Y_0 = y_0, \ldots, Y_n = y_n \mid \vec{X} = (x_0, \ldots, x_1)\right) \end{aligned}$$

Обчислюємо

$$\begin{gathered} P_\lambda\left(X_0 = x_0, \ldots, X_n = x_n\right) = P_\lambda\left(X_0 = x_0\right) \cdot P_\lambda\left(X_1 = x_1 \mid X_0 = x_0\right) \cdot \\ P\left(X_2 = x_2 \mid X_1 = x_1, X_0 = x_0\right) \cdot \ldots \cdot \\ P\left(X_n = x_n \mid X_{n-1} = x_{n-1}, \ldots, X_0 = x_0\right) = \end{gathered}$$

Марков міркував, що в умові не потрібні такі довгі лянцюжки іксів.

Ми припускаємо цю ланцюг марківський.

$$\begin{gathered} = P_\lambda\left(X_0 = x_0\right) \cdot P_\lambda\left(X_1 = x_1 \mid X_0 = x_0\right) \cdot P\left(X_2 = x_2 \mid X_1 = x_1\right) \cdot \ldots \cdot \\ \cdot P(X_n = x_n \mid X_{n-1} = x_{n-1}) = \\ = \mu_{x_0} \cdot A_{x_0 x_1} \cdot A_{x_1 x_2} \cdot \ldots \cdot A_{x_{n-1} x_n} \end{gathered}$$

Вважаємо, що випадкові величини $Y_0, .., Y_n$ є умовно незалежними при відомих $X_0, \ldots, X_n$

$$P_\lambda (Y_0 = y_0, \ldots, Y_n = y_n \mid X_0) \cdot P_\lambda (Y_1 = y_0 \mid X_1 = x_1) \cdot \ldots \cdot$$
$$P_\lambda (Y_n = y_n \mid X_n = x_n) =$$
$$= B_{x_0 y_0} \cdot B_{x_1 y_1} \cdot \ldots \cdot B_{x_n y_n}$$

Таким чином

$$P_\lambda (X - 0 = x_0, \ldots, X_n = x_n, Y_0 = y_0, \ldots, Y_n = y_n) =$$
$$= \underbrace{\mu_{x_0}}_{P(X_0 = x_0)} \cdot B_{x_0 y_0} \cdot A_{x_0 x_1} \cdot A_{x_1 x_2} \cdot B_{x_2 y_2} \cdot \ldots \cdot A_{x_{n-1} x_n} \cdot B_{x_n y_n}$$

На прикладі:

$$P_\lambda (X = (TTXX), Y = (MCMB)) = 0.6 \cdot 0.1 \cdot 0.7 \cdot 0.4 \cdot 0.3 \cdot 0.7 \cdot 0.6 \cdot 0.1$$

$$P_\lambda (X = (TTTT), Y = (MCMB)) =$$
$$= 0.6 \cdot 0.1 \cdot 0.7 \cdot 0.4 \cdot 0.7 \cdot 0.1 \cdot 0.7 \cdot 0.5 = 0.000412.$$

$XXXT$ найбільш ймовірна послідовність.

Нас може цікавити:

$P_\lambda (X_0 = T, Y = (MCMB)) =$
$= \sum_{x_1, x_2, x_3} P_\lambda (X_0 = T, X_1 = x_1, X_2 = x_2, X_3 = x_3, Y = (MCMB)) = 0.1887$

$$P_\lambda (X_0 = X, Y = (MCMB)) = 1 - 0.1887 = 0.81183$$

$P_\lambda (X_1 = T, Y = (MCMB)) =$
$= \sum_{x_0, x_2, x_3} P_\lambda (X_0 = x_0, X_1 = T, X_2 = x_2, X_3 = x_3) = 0.519 > 0.5;$

більш ймовірно, що рік був теплим

$P_\lambda (X_2 = T, Y = y) = 0.228 < 0.5 \Rightarrow X_2 = X$ більш ймовірно

$P_\lambda (X_3 = T, Y = y) = 0.804 > 0.5 \Rightarrow X_3 = T$ більш ймовірно

$XTXT$ - послідовність на основі "індивідуальних" ймовірностей

Обчислимо ймовірность ланцюжка $Y = (MCMB)$

$$P_\lambda (Y = (MCMB)) = \sum_{x_0, x_1, x_2, x_3} P (X = (x_0, x_1, x_2, x_3), Y = (MCMB)) =$$
$$= 0.009629$$

Ймовірність досить мала. Можливо, що модель може бути не адекватною, або ж випадок є досить рідкістним. Можливий і випадок, що ми помилилися із нашими апріорними припущеннями. Так ми наблизилися до задачі навчання: пошуку моделі при якій ймовірність випадку буде максимальною.

Прихована марківська модель це випадковий процес, який розділений на дві складові: спостережувану та приховану.

Як приклад використання такого типу моделі можна взяти коливання цін на акції, де прихований фактор може бути якимось економічним чи політничним. Змоделювати його ми не ставимо за мету, але ми маємо його враховувати для того щоб краще змоделювати ціни на акції.

Example (передача сигналу)**.** Наша задача найточніше відтворити сигнал.

$\{B_k\}_{k\geq 0}$ - послідовність в.в. зі значеннями $0, 1$;

Y_k - спостережений сигнал

Y_k співпадає з B_k з ймовірністю p, спотворення з ймовірністю $1-p$

введемо $\{\xi_k\}_{k\geq 0}$ - н.о.р., $\xi_k \sim B(p)$ тоді $X_k = B_k$;

$$Y_k = (1-\xi_k)\cdot X_k + \xi_k \cdot (1-X_k)\,.$$

Тут

$$A = \begin{pmatrix} 1-a & a \\ a & 1-a \end{pmatrix} \qquad \mu = (0.5; 0.5)\,.$$

$$B = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}.$$

$\lambda = (\mu, A, B)$ - ПММ для каналу зв'язку

Example (фінансові часові ряди)**.** Простіша модель часового ряду ринкової ціни акцій - Модель Блека-Шоуза.

S_k - ціна яку ми спостерігаємо у якийсь проміжок часу

$$S_k = \exp\left\{\mu - \frac{\sigma^2}{2} + \sigma\cdot\xi_k\right\}\cdot S_{k-1}$$

де $\xi_k \sim \mathcal{N}(0,1)$ - н.о.р

$\sigma \in \mathbb{R}$ - волатильність - характеристика мінливості ціни

μ - норма прибутку

$$M\frac{S_k}{S_{k-1}} = e^{\mu-\sigma^2/2}\cdot\varphi_{\xi_k}\left(\frac{\sigma}{i}\right) = e^{\mu}.$$

Така можель може досить добре працювати на коротких інтревалах. На великих інтервалах ціни демонструють інші властивості, що не описуються цією моделлю.

Вводимо деякий випадковий процес $\{X_k\}_{k\geq 0}$ який має приховані фактори

$$S_k = \exp\left\{\mu(X_k) - \sigma^2(X_k)/2 + \sigma(X_k)\xi_k\right\}\cdot S_{k-1}.$$

μ та σ залежать від $\{X_k\}_{k\geq 0}$

В цьому прикладі умовні ймовірності спостережуваного $B_{xy} = P\left(Y_k = y \mid X_k = x\right)$ замінюються на умовін щільності

$$B_{xy} = P_{Y_k|X_k}\left(y|x\right).$$

Нехай $Y_k = \log \frac{S_k}{S_{k-1}} = \mu(X_k) - \sigma^2(X_k)/2 + \sigma(X_k)\xi_k$ і тоді при $X_k = x$ отримаємо

$$Y_k \sim \mathcal{N}\left(\mu(x) - \sigma^2(x)/2; \sigma^2(x)\right).$$

.

2 Lecture 2: Definition and examples of Markov Chains

Нехай $\{X_n\}_{n\geq 0}$ - послідовність в.в. заданих на одному і тому ж ймовірністному просторі і які приймають не більш ніж зліченну кількість значень.

Множину значень E називають **множиною станів** або **фазовим простором**, а її елементи позначають $i, j, k, \ldots$.

Definition 1. Послідовність в.в. $\{X_n\}_{n\geq 0}$ називають ланцюгом маркова якщо:

$$\forall n \geq 1 \quad \forall i_0, \ldots, i_n \in E$$

має місце рівність

$$P\left(X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\right) = P\left(X_n = i_n \mid X_{n-1} = i_{n-1}\right) \tag{2.1}$$

В.в X_n інтерпретують як стан системи в момент часу n.

Зауважимо, що ймовірність:

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

Тоді 2.1 має вигляд

$$P\left(X_n = i_n, \ldots, X_0 = i_0\right) = P\left(X_{n-1} = i_{n-1}, \ldots., X_0 = i_0\right) \cdot f\left(i_n, i_{n-1}, \ldots, n\right) \tag{2.2}$$

де $f\left(i_n, \ldots, i_{n-1}, n\right) = P\left(X_n = i_n \mid X_{n-1} = i_{n-1}\right)$.

Отже якщо існує детермінована функція

$$f : E \times E \times Z_+ \to [0, 1]$$

така, що виконується 2.2, то $\{X_n\}_{n\geq 0}$ утворює ланцюг Маркова.

Definition 2. Ланцюг Маркова $\{X_n\}_{n\geq 0}$ називається однорідним, якщо ймовірності

$$P\left(X_{n+1} = j \mid X_n = i\right)$$

не залежать від n.

Remark. Позначимо $p_{ij} = P\left(X_{n+1} = j \mid X_n = i\right)$

Definition 3. Ймовірності p_{ij} переходу зі стану i в стан j наз. перехідною ймовірністю зі стану i в стан j.

Матрицю $P = \left(p_{ij}\right)_{i,j\in E}$ назива.ть матрицею перехідних ймовірностей ланцюга Маркова.

Remark. Зауважимо, що матриця P є стохастичною, тобто

$$p_{ij} \geq 0 \quad \forall i, j \in E \quad \sum_{j \in E} p_{ij} = 1 \quad \forall i \in E.$$

Example. Нехай $\{\varepsilon_n\}_{n \geq 0}$ - н.о.р.,

$$P(\varepsilon_n = -1) = 1/2 = P(\varepsilon_n = 1).$$

Покладемо $X_0 = 0$, $X_{n+1} = X_n + \varepsilon_{n+1}, \quad n \geq 0$.

Переконаємося, що $\{X_n\}_{n \geq 0}$ - ланцюг Маркова.

Маємо: $\forall n \geq 1$ і $\forall i_0, \ldots, i_n \in \mathbb{Z}$:

$$\begin{gathered} P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = \\ = P(X_0 = i_0, \varepsilon_1 = i_1 - i_0, \varepsilon_2 = i_2 - i_1, \ldots, \varepsilon_n = i_n - i_{n-1},) = \\ = P(X_0 = i_0) \cdot P(\varepsilon_1 = i_1 - i_0) \cdot \ldots \cdot P(\varepsilon_n = i_n - i_{n-1}) = \\ \begin{cases} \left(\frac{1}{2}\right)^n & \text{якщо } i_0 = 0, |i_1 - i_0| \cdot \ldots \cdot |i_n - i_{n-1}| = 1 \\ 0 & \text{інакше} \end{cases} \end{gathered}$$

Знайдемо відношення:

$$\frac{P(X_0 = i_0, \ldots, X_n = i_n)}{P(X_0 = i_0, \ldots, X_{n-1} = i_{n-1})} = \begin{cases} \frac{1}{2} & \text{якщо } i_0 = 0, |i_1 - i_0| \cdot \ldots \cdot |i_n - i_{n-1}| = 1 \\ 0 & \text{інакше} \end{cases}$$

Детермінована функція $\Rightarrow \{X_n\}_{n \geq 0}$ - ланцюг Маркова.

Example (процес авторегресії порядку p)**.**

$$X_n = \sum_{k=1}^{p} a_k \cdot X_{n-k} + \varepsilon_k, \qquad n \geq p.$$

з початковою умовою $X_0 = \ldots = X_{p-1} = 0$, де $a_k \in \mathbb{R}$

$\{\varepsilon_n\}_{n \geq 0} =$ н.о.р. в.в. не залежна від $\{X_n\}_{n \geq 0}$

$\{X_n\}_{n \geq 0}$ - не є ланцюгом Маркова.

Розглянемо $\widetilde{X}_n = (X_n, X_{n+1}, \ldots, X_{n+p-1})$

$\left\{\widetilde{X}_n\right\}_{n \geq 0}$ - є ланцюгом Маркова на $\widetilde{E} = \underbrace{E \times E \times \ldots \times E}_{p}$.

2.1 Прихована Марківська Модель

Нехай $\{X_n\}_{n \geq 0}$ - послідовність в.в. на скінченній мн-ні $E = \{x_1, \ldots, x_N\}$. Нехай $\{Y_n\}_{n \geq 0}$ послідовність в.в. на скінченній мн-ні станів $F = \{y_1, \ldots, y_M\}$.

Definition 4. Послідовність $\{(X_n, Y_n)\}_{n\geq 0}$ наз. прихованою марківською моделлю, якщо

1. послідовність $\{(X_n, Y_n)\}_{n\geq 0}$ є ланцюгом Маркова на множині $E \times F$
2. $\{X_n\}_{n\geq 0}$ є ланцюгом Маркова з деяким початковим розподілом μ і матрицею перехідних ймовірностей A.
3. $\forall n \geq 1$ випадкові величини $y_0, y_1, \ldots, y_n$ є умовно незалежними при заданих $X_0, X_1, \ldots, X_n$, тобто:

$$\forall n \geq 1 \quad \forall y_0, \ldots, y_n \in F \quad \forall x_0, \ldots, x_n \in E :$$
$$P\left(Y_0 = y_0, \ldots, Y_n = y_n \mid X_0 = x_0, \ldots, X_n = x_n\right) =$$
$$= \prod_{i=0}^{n} P\left(Y_i = y_i \mid X_i = x_i\right)$$

Remark. Хоча $\{(X_n, Y_n)\}_{n\geq 0}$ і $\{X_n\}_{n\geq 0}$ є ланцюгами Марковаб послідовність $\{Y_n\}_{n\geq 0}$ може не бути ланцюгом Маркова.

Скінченновимірні розподіли $\{(X_n, Y_n)\}_{n\geq 0}$ мають такий вигляд:

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

Позначимо:

$$B_{xy} = P\left(Y_k = y \mid X_k = x\right)$$
$$A_{x_i y_j} = P\left(X_k = x_j \mid X_{k-1} = x_i\right)$$
$$\mu = \left(P\left(X_0 = x_1\right), \ldots, P\left(X_0 = x_N\right)\right)$$

Тоді:

$$P\left(Y_0 = y_0, \ldots, Y_n = y_n, X_0 = x_0, \ldots, X_n = x_n\right) =$$
$$= \mu_{x_0} \cdot \prod_{i=0}^{n-1} A_{x_i x_{i+1}} \cdot \prod_{j=0}^{n} B_{x_j y_j}$$

Позначення $\lambda = (\mu, A, B)$ використ. для позначення ПММ.

2.2 Три задачі для прихованих марківських моделей

Задача оцінювання

За заданою моделлю $\lambda = (\mu, A, B)$ і заданою послідовністю спостережень довжини T: $y = (y_0, \ldots, y_T)$, визначити ймовірність $P_{lambda}\left(Y = y\right)$, де $Y = (Y_0, \ldots, Y_T)$

Задача декодування За заданою моделлю $\lambda = (\mu, A, B)$ та за заданою послідовністю спостережнь $y = (y_0, \ldots, y_T)$ визначити найбільш імовірну послідовність станів $x = (x_0, \ldots, x_T)$ для прихованого ланцюга Маркова.

Задача навчання За заданою послідовністю спостережень $y = (y_0, \ldots, y_T)$ визначити модель $\lambda = (\mu, A, B)$, яка максимізує ймовірність $P_\lambda\left(Y = y\right)$ спостереженої послідовності.

3 Lecture 3: Три задачі для прихованих марковських моделей та їх розв'язання

4 Lecture 4: EM алгоритм для знаходження параметрів суміші розподілів

4.1 Суміш скінченної кількості ймовірністних розподілів

Example:

Нехай $Y_1, \ldots, Y_n$ — незалежні випадкові величини;
$Y_i \sim \mathcal{N}\left(\mu_i, \sigma_i^2\right), i = \overline{1, N}$. Нехай X — дискретна випадкова величина зі значеннями в множині $E = \{1, 2, \ldots, N\}$ та

$$P\left(X = i\right) = p_i, \quad i = \overline{1 = N}$$

$$\sum_{i=1}^{N} p_i = 1$$

Покладемо

$$Z = Y_X = \begin{cases} Y_1 & X = 1 \\ Y_2 & X = 2 \\ \ldots & \\ Y_n & X = N \end{cases}$$

Знайдемо щільність розподілу Z:

$$p_Z(t) = \sum_{i=1}^{N} p(X = i) \cdot P_{Z|X}(t|i) = \sum_{i=1}^{N} P\left(X = i\right) P_{Y_i}(t) =$$
$$= \boxed{\sum_{i=1}^{N} p_i \cdot \frac{1}{\sqrt{2\pi\sigma_i}} \cdot \exp\left\{-\frac{\left(t - \mu_i\right)^2}{2\sigma_i^2}\right\}}.$$

Щільність розподілу в.в. Z наз. сумішшю N нормальних розподілів $\mathcal{N}\left(\mu_1, \sigma_1^2\right), \ldots, \mathcal{N}\left(\mu_N, \sigma_N^2\right)$ зі змішувальним розподілом $(p_i)_{i=\overline{1,N}}$
Припустимо, що $\{X_n\}_{n \geq 1}$ — н.о.р.;
$X_i \sim X$
$\{Z_n\}_{n \geq 1}$ — н.о.р. в.в.: $Z_n \sim Z$

$$p_{Z_n}(t) = \sum_{i=1}^{N} P\left(X_n = i\right) \cdot p_{Z_n|X_n}(t|i)$$

$\{X_n\}_{n \geq 1}$ — є прихованою послідовністю (X — латентна змінна);
Задача: за скінченною кількість спостережень $Z_1, \ldots, Z_n$ потрібно оцінити невідомі параметри суміші, а саме: $\mu_1, \sigma_1^2, \ldots, \mu_N, \sigma_N^2$ і $p_1, \ldots, p_N$.
В теорії сигналів ці задачу можна переформулювати: визначити окремі сигнали за спостереженою суперпозицією цих сигналів.
TODO: picture Distribution for problem of mixed finite number of distributions

4.1.1 Нерівність Йенсена

Theorem 1. Нехай $g(\cdot)$ — опукла донизу функція

Опуклість: $\forall \lambda_i \geq 0,\ \sum_{i=1}^{n} \lambda_i = 1$ виконується

$$g(\lambda_1 x_1 + \lambda_2 x_2 + \ldots + \lambda_n x_n) \leq \lambda_1 g(x_1) + \ldots + \lambda_n g(x_n)$$

Нехай X — випадкова величина на $E = \{x_1, \ldots, x_n\}$ і $P(X = x_i) = \lambda_i$. Тоді

$$g(MX) \leq M(g(X)) \tag{1}$$

1 — нерівність Йенсена.

При $g(x) = -\log x$ (опукла донизу функція) отримаємо

$$\log(M(X)) \geq M(\log X)$$

або ж

$$\log\left(\sum_{i=1}^{n} \lambda_i x_i\right) \geq \sum_{i=1}^{n} \lambda_i \log(x_1)$$

і ця нерівність перетворюється в рівність лише при $x_1 = x_2 = \ldots x_n$.

4.1.2 Відстань Кульбака-Лейблера між розподілом (Kullback-Leibler devirgence)

Definition 5. Нехай є $E = \{1, 2, \ldots, N\}$ і f і g — два розподіли ймовірності на E:

$$f = \left(f(1), \ldots, f(N)), \quad \sum_{i=1}^{N} f(i) = 1\right)$$

$$g = \left(g(1), \ldots, g(N)), \quad \sum_{i=1}^{N} g(i) = 1\right)$$

Тоді їхньою відносною ситропією або ж відстанню Кульбака-Лейблера між f і g наз. віличина

$$D(f|g) = \sum_{i=1}^{N} f(i) \cdot \log \frac{f(i)}{g(i)}$$

Тут $0 \cdot \log \frac{0}{g(i)} = 0,\ f(i) \cdot \log \frac{f(i)}{0} = \infty$

Remark. Відстань Кульбака-Лейблера не є метричною відстаню, оскільки $D(f|g) \neq D(g|f)$.

Example: Інформаційне наповнення розподілу

X — випадкова величина на $E = \{1, \ldots, N\}$ з розподілом $f = (p_1, \ldots, p_N)$ і Y — в.в. з рівномірним розподілом на E, $g = \left(\frac{1}{N}, \ldots, \frac{1}{N}\right)$. Тоді

$$D(f|g) = \sum_{i=1}^{N} p_i \log \frac{p_i}{\frac{1}{N}} = \sum_{i=1}^{N} p_i \log N p_i =$$

$$= \log N + \underbrace{\sum_{i=1}^{N} p_i \log p_i}_{-H(X)} = \log N - H(X)$$

де $H(X)$ — ентропія в.в. X.

Example: Відстань К.-Л. між двома розподілами Бернуллі

$X \sim B(p_1), Y \sim B(p_2)$

$$f = (1 - p_1, p_1), \qquad g = (1 - p_2, p_2)$$

Тоді

$$D(f|g) = (1 - p_1) \log \frac{1 - p_1}{1 - p_2} + p_1 \cdot \log \frac{p_1}{p_2} =$$
$$= -(1 - p_1) \log(1 - p_2) - p_1 \cdot \log p_2 - h(p_1)$$

де $h(p) = -p \log p - (1 - p) \log(1 - p)$ — функція ентропії.

Lemma 1. Для довільних розподілів f, g на $E = \{1, \ldots, N\}$ величина $D(f|g)$ є невід'ємною

$$D(f|g) \geq 0$$

Proof. Нехай X — випадкова величина з розподілом f. Тоді:

$$D(f|g) = \sum_{i=1}^{N} f(i) \log \frac{f(i)}{g(i)} =$$

$$= M\left[\log \frac{f(x)}{g(x)}\right] = -M\left[\log \frac{g(x)}{f(x)}\right] \overset{\text{н. Йенсена}}{\geq} -\log M\left[\frac{g(x)}{f(x)}\right] =$$

$$= -\log \sum_{i=1}^{N} f(x_i) \frac{g(x_i)}{f(x_i)} = -\log \underbrace{\sum_{i=1}^{N} g(x_i)}_{=1} = -\log 1 = 0.$$

□

4.2 Постановка задачі

Нехай $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ — незалежні в.в. на $E \times Y$, де $E = \{1, .., N\}$. Позначимо

$$\alpha_j = P(X_i = j), \qquad j = \overline{1, N}$$

і $\forall y \in Y$ нехай $\forall k = \overline{1, n}$

$$P(y|\varphi_j) = P_{Y_k|X_k}(y|j)$$

умовна щільність розподілу Y_k при заданому X_k, φ_j — вектор параметрів, від якого цей умовний розподіл залежить.

В першому прикладі

$$p\left(y|\varphi_j\right) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left\{-\frac{\left(y-\mu_j\right)^2}{2\sigma_j^2}\right\} \qquad \varphi_j = \left(\mu_j, \sigma_j^2\right)$$

Покладемо

$$\theta = (\alpha_1, \dots, \alpha_N; \varphi_1, \dots, \varphi_N)$$

як вектор невідомих параметрів.

Задача: за спостереженням $Y_1, \dots, Y_n$ побудувати оцінку для θ.

$$\theta^*_{ML} = \underset{\theta}{\operatorname{argmax}}\, p\left(y_1, \dots, y_n|\theta\right)$$

Позначимо: $y_1^n = (y_1, \dots, y_n)$, $x_1^n = (x_1, \dots, x_n)$

Запишемо функцію правдоподібності $p(y_1^n|\theta)$:

$$\begin{aligned} P\left(y_1^n|\theta\right) &\overset{\text{незал.}}{=} p(y_1|\theta) \cdot \ldots \cdot p(y_n|\theta) = \\ &= \sum_{i=1}^N \alpha_i p(y_1|\varphi_i) \cdot \sum_{i=1}^N \alpha_i p(y_2|\varphi_i) \cdot \ldots \cdot \sum_{i=1}^N \alpha_i p(y_n|\varphi_i) = \\ &= \prod_{k=1}^n \underbrace{\sum_{i=1}^N \alpha_i p(y_k|\varphi_i)}_{\equiv f(y_k|\theta)} = \prod_{k=1}^n f(y_k|\theta). \end{aligned}$$

Розглянемо апостеріорну щільність:

$$p(x_1^n|y_1^n, \theta) = \frac{p\left(x_1^n, y_1^n|\theta\right)}{p\left(y_1^n|\theta\right)}$$

З наших припущень випливає:

$$p\left(x_1^n, y_1^n|\theta\right) \overset{\text{незал.}}{=} \prod_{k=1}^n p\left(x_k, y_k|\theta\right) = \prod_{k=1}^n \alpha_{x_k} \cdot p\left(y_k|\varphi_{x_k}\right)$$

$$p\left(y_1^n|\theta\right) = \prod_{k=1}^n f\left(y_k|\theta\right)$$

Тоді

$$p\left(x_1^n|y_1^n, \theta\right) = \prod_{k=}^n \frac{\alpha_{x_k} \cdot p\left(y_k|\varphi_{x_k}\right)}{\sum_{j=1}^N \alpha_j \cdot p(y_k|x_j)} = \prod_{k=1}^n \frac{\alpha_{x_k} p(y_k|\varphi_{x_k})}{f(y_k|\theta)}$$

Звідси

$$\log p(y_1^n|\theta) = \log p(x_1^n, y_1^n|\theta) - \log p(x_1^n|y_1^n, \theta) = \sum_{k=1}^n \log\left[p(y_k|\varphi_{x_k})\alpha_{x_k}\right] - \log p(x_1^n|y_1^n, \theta) \qquad (2)$$

4.3 Квазі-log правдоподібність

Припустимо, що $\theta^{(t)} = \left(\alpha_1^{(t)}, \ldots, \alpha_N^{(t)}; \varphi_1^{(t)}, \ldots, \varphi_N^{(t)}\right)$ — деяке наближення θ.

Ідея: покращити $\theta^{(t)}$, наблизивши до θ^*_{ML}.

Домножимо 2 на $p\left(x_1^n | y_1^n, \theta^{(t)}\right)$ і підсумуємо за всіма x_1^n. Оскільки $\sum\limits_{x_1^n} p\left(x_1^n \mid y_1^n, \theta^{(t)}\right) = 1$, то отримаємо

$$\log p\left(y_1^n | \theta\right) = \sum_{x_1^n} p\left(x_1^n | y_1^n, \theta^{(t)}\right) \cdot \log p\left(x_1^n, y_1^n | theta\right) - \sum_{x_1^n} p\left(x_1^n | y_1^n, \theta^{(t)}\right) \cdot \log p\left(x_1^n | y_1^n, \theta\right)$$

Введемо допоміжну функцію

$$Q(\theta, \theta^{(t)}) := \sum_{x_1^n} p\left(x_1^n | y_1^n, \theta^{(t)}\right) \cdot \log p\left(x_1^n, y_1^n | \theta\right)$$

$Q\left(\theta | \theta^{(t)}\right)$ — умовне математ. сподівання:

$$Q\left(\theta | \theta^{(t)}\right) = M\left[\log p\left((X_1, Y_1), \ldots, (X_n, Y_n) | \theta\right) \mid Y_1^n = y_1^n, \theta = \theta^{(t)}\right]$$

В цих позначеннях для кожного θ отримаємо рівність:

$$\log p(y_1^n | \theta) - \log p(y_1^n | \theta^{(t)}) =$$

$$= Q(\theta | \theta^{(t)}) - Q(\theta^{(t)} | \theta^{(t)}) + \sum_{x_1^n} p(x_1^n | y_1^n, \theta) \cdot \log \frac{p(x_1^n | y_1^n, \theta^{(t)})}{p(x_1^n | y_1^n, \theta)} \geq$$

відстань Кульбака-Лейблера між $p(\cdot | y, \theta^{(t)})$ і $p(\cdot | y, \theta) \Rightarrow$ це величина невід'ємна

$$\geq Q(\theta | \theta^{(t)}) - Q\left(\theta^{(t)} | \theta^{(t)}\right)$$

Якщо означимо

$$\theta^{(t+1)} = \operatorname*{argmax}_{\theta} Q\left(\theta | \theta^{(t)}\right)$$

то отримаємо нову оцінку $\theta^{(t+1)}$, для якої

$$\log p\left(y_1^n | \theta^{(t+1)}\right) \geq \log p\left(y_1^n | \theta^{(t+1)}\right)$$

а отже ми покращили $\theta^{(t)}$ в тому сенсі, що збільшили значення $Q(\theta | \theta^{(t)})$.

4.4 Крок E і крок M

EM-алгоритм має такий вигляд:

1. **Початок**: нехай $\theta^{(t)} = \left(\alpha_1^{(t)}, \ldots, \alpha_N^{(t)}; \varphi_1^{(t)}, \ldots, \varphi_N^{(t)}\right)$ — деяке наближення до θ^*_{ML}
2. **Крок E:** обчислюємо умовне математичне сподівання

$$Q(\theta | \theta^{(t)}) = \sum_{x_1^n} p\left(x_1^n | y_1^n, \theta^{(t)}\right) \cdot \log p\left(x_1^n, y_1^n | \theta\right)$$

3. **Крок** M**:** визначаємо $\theta^{(t+1)}$ як
$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmax}}\, Q\left(\theta|\theta^{(t)}\right)$$

4. **Покласти** $\theta^{(t+1)} \mapsto \theta^{(t)}$ і повторитиб починаючи з кроку E.

Питання, які виникають:

1. Чи збігається ця процедура?
2. Чи збігається до лок./глоб. максимуму ф. правдоподібності.

4.5 Збіжність

Нехай $\mathcal{L}(\theta) = \log p\left(y_1^n|\theta\right)$, $\theta^{(t)}$ — умовне мат. сподівання з EM-алгоритму.

Ми довели, що $\left\{\mathcal{L}\left(\theta^{(t)}\right)\right\}$ є монотонною.

Якщо $\left\{\mathcal{L}\left(\theta^{(t)}\right)\right\}$ є обмеженою (на додачу до монотонності) , то $\mathcal{L}\left(\theta^{(t)}\right) \to \mathcal{L}^*,\ t \to \infty$

Але звідси ще не випливає, що $\mathcal{L}^*$ є лок./глоб. максимумом для $\mathcal{L}\left(\theta\right)$. Більше того, якщо навіть $\mathcal{L}\left(\theta^{(t)}\right) \underset{t\to\infty}{\to} \mathcal{L}^*$, то це не означає, що $\theta^{(t)} \underset{t\to\infty}{\to} \theta^*_{ML}$

Умови для збіжності EM-алгоритму:

1. $\theta \in C \subseteq \mathbb{R}^n$
2. $\mathcal{L}(\theta)$ — неперервна і диференційована у внутрішності C.
3. $\theta^*_{ML} \in \inf(C)$
4. $Q(a|b)$ є неперервною за $a, b \in C$
5. $\left\{\mathcal{L}\left(\theta^{(t)}\right)\right\}_{t=0}^{\infty}$ — обмежена послідовність для довільного $\theta^{(t)}$

Remark. Ці умови (в основному) виконуються для розподілів, що належать до експоненційної сім'ї розподілів.

4.6 Явна форма кроків E, M

$$\begin{aligned}
Q\left(\theta, \theta^{(t)}\right) &= M\left[\log p\left((X_1, Y_1), \ldots, (X_n, Y_n)|\theta\right)|Y_1^n = y_1^n, \theta = \theta^{(t)}\right] \overset{\text{незал.}}{=} \\
&\overset{\text{незал.}}{=} M\left[\sum_{k=1}^{n} \log p\left((X_k, Y_k)|\theta\right)|Y_1^n = y_1^n, \theta = \theta^{(t)}\right] \overset{\text{лінійність}}{=} \\
&\overset{\text{лінійність}}{=} \sum_{k=1}^{n} M\left(\log p\left((Y_k, Y_k)|\theta\right)|Y_1^n = y_1, \theta = \theta^{(t)}\right) = \\
&= \sum_{k=1}^{n}\sum_{j=1}^{N} \log\left[\alpha_j p\left(y_k|\varphi_j\right)\right] \cdot p\left(x_k|Y_1^n = y_1^n, \theta = \theta^{(t)}\right) = \\
&= \sum_{j=1}^{N}\sum_{k=1}^{n} \underbrace{\log\left[\alpha_j \cdot p(y_k|\varphi_j)\right]}_{\text{на суму двох}} \cdot \frac{\alpha_j^{(t)} \cdot p\left(y_k|\varphi_j^{(t)}\right)}{f\left(y_k|\theta^{(t)}\right)}.
\end{aligned}$$

Звідси

$$Q\left(\theta|\theta^{(t)}\right) = \sum_{j=1}^{N}\left[\sum_{k=1}^{n} \frac{p(y_k|\varphi_j^{(t)}\alpha_j^{(t)}}{f(y_k|\theta^{(t)}}\right] \cdot \log\alpha_j + \sum_{n=j=1}^{N}\sum_{k=1}^{n} \log\left[p\left(y_k|\varphi_j\right)\right] \cdot \frac{p\left(y_k|\varphi_j^{(t)}\right) \cdot \alpha_j^{(t)}}{f\left(y_k|\theta^{(t)}\right)}.$$

Максимізація дає таке

$$\alpha_j^{(t+1)} = \frac{1}{n} \sum_{k=1}^{n} \underbrace{p\left(y_k | \varphi_j^{(t)}\right) \cdot \alpha_j^{(t)}}_{\equiv \omega_j\left(y_k, \theta^{(t)}\right)} = \frac{1}{n} \sum_{k=1}^{n} \omega_j \left(y_k, \theta^{(t)}\right)$$

$$\varphi_j^{(t+1)} = \underset{\varphi_j}{\operatorname{argmax}} \sum_{j=1}^{N} \log \left[p \left(y_k | \varphi_j\right)\right] \cdot \omega_j \left(y_k, \theta^{(t)}\right)$$

Задача може мати єдиний розв'язок, який можна знайти аналітично (для "хороших" розподілів).

5 Lecture 5: Ланцюги Маркова. Означення та основні властивості.

Нехай $\{X_n\}_{n \geq 0}$ — послідовність в.в. заданих на не більш ніж зліченній множині $E = \{1, 2, \ldots\}$.

Definition 6. $\{X_n\}_{n \geq 0}$ утворює ланцюг маркова з початковим розподілом $\lambda = (\lambda_i)_{i \in E}$ та матрицю похідних ймовірностей $P = (p_{ij})_{i,j \in E}$ якщо:

1. $P\left(X_0 = i\right) = \lambda_i, \ i \in E$
2. $P\left(X_{n+1} = j | X_n = i_n, \ldots, X_0 = i_0\right) = P\left(X_{n+1=j} \mid X_n = i_n\right) = p_{i_n, j}$

$$\forall n \geq 0 \text{ та } \forall i_0, i_1, \ldots, i_n, j \in E$$

Remark. Розглядаємо однорідні ланцюги Маркова, для яких перехідні ймовірності

$$P\left(X_{n+1} = j | X_n = i\right) = p_{ij}$$

не залежить від n.

Це означає, що $P\left(X_1 = j, X_0 = i\right) = P\left(X_2 = j, X_1 = i\right) = \ldots$

Матриця $P = (p_{ij})_{i,j \in E}$ є стохастичною

$$p_{ij} \geq 0 \quad \text{та} \quad \sum_{j \in E} p_{ij} = 1$$

Lemma 2. Нехай E та F — дві зліченні (скінченні) множини і нехай

$$f : \mathbb{N} \times E \times F \to E$$

Нехай $X_0, Y_1, \ldots$ — незалежні в сукупності в.в. такі що X_0 приймає значення в E; Y_n приймають значення в F. Нехай $\{X_n, n \geq 1\}$ означено наступним чином:

$$X_{n+1} = f\left(n, X_n, Y_{n+1}\right), \quad n \in \mathbb{N}$$

тоді $\{X_n\}_{n \geq 0}$ — ланцюг Маркова.

Написати доведення

Example (Підкидання грального кубика). X_n — кількість шісток, які випали до n-го підкидання.

$$X_{n+1} \begin{cases} X_n & \text{якщо в } (n+1)\text{-му підкиданні не випала шістка} \\ X_n + 1 & \text{якщо шістка випала у } (n+1)\text{-му підкиданні} \end{cases}$$

Тоді $X_{n+1} = X_n + Y_{n+1}$, де $\{Y_n\}_{n\geq 0}$ — н.о.р. в.в. з

$$P\left(Y_n = 0\right) = \frac{5}{6}; \quad P\left(Y_n = 1\right) = \frac{1}{6}$$

$\{X_n\}_{n\geq 0}$ — ланцюг Маркова.

Множина станів $E = \{0, 1, 2, \ldots\}$.

Перехідні ймовірності:

$$p_{ij} = P\left(X_{n+1} = j | X_n = i\right) = \begin{cases} 5/6 & j = i \\ 1/6 & j = i + 1 \\ 0 & \text{інакше} \end{cases}$$

Example (Гіллястий процес Ватсона-Гальтона). Нехай спостерігається деяка популяція частинок, еволюція яких така: кожна частинка незалежно від інших в деякі моменти (дискретні) часу $n = 0, 1, 2, \ldots$ перетворюється на i частинок з ймовірністю $p_i : \sum_{i=0}^{\infty} p_i = 1$.

Стан системи: X_n — кількість частинок в момент часу n (в n-му поколінні).

$$X_{n+1} = Y^{(1)} + Y^{(2)} + \ldots + Y^{(X_n)}$$

де $Y^{(i)}$ — кількість нащадків i-ї частинки з попереднього покоління.

Example (Модель Еренфестів дифузії молекул газу). Є дві урни A та B, в яких знаходяться N частинок. В кожен момент часу навмання вибирається число з $\{1, \ldots, N\}$ і частинка з відповідним номером перекладається в іншу урну.

X_n — кількість частинок в A в момент n.

$$X_{n+1} = \begin{cases} X_n - 1 & A \to B \\ X_n + 1, B \to A & \end{cases}$$

Припустимо, що $X_n = i$.

$$p_{i,i-1} = \frac{i}{N}; \; p_{i,i+1} = \frac{N-1}{N}; \; p_{ij} = 0 \text{ при } j \neq i \pm 1$$

$$P = \begin{bmatrix} 0 & 1 & 0 & \ldots & & 0 \\ \frac{1}{N} & 0 & 1 - \frac{1}{N} & \ldots & & 0 \\ 0 & \frac{2}{N} & 0 & 1 - \frac{2}{N} & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \ldots & \ldots & \ldots & \ldots & 1 & 0 \end{bmatrix}$$

$$P(i \to i+1) > P(i \to i-1) \text{ при } i < \frac{N}{2}$$

$$P(i \to i+1) < P(i \to i-1) \text{ при } i > \frac{N}{2}$$

5.1 Властивості ланцюга Маркова

5.1.1 Скінченновимірні розподіли

1.

$$P\left(X_0=i_0, x_1=i_1, \ldots, X_n=i_n\right)=$$
$$=P\left(X_0=i_0\right) \cdot P(X_1=i_1|X_0=i_0) \cdot P\left(X_2=i_2|X_1=i_1, X_0=i_0\right) \times$$
$$\times P\left(X_n=i_n|X_{n-1}=i_{n-1}, \ldots, X_0=i_0\right)=$$
$$=P(X_0=i_0) P\left(X_1=i_1|X_0=i_0\right) \cdot \ldots \cdot P\left(X_n=i_n|X_{n-1}=i_{n-1}\right)=\boxed{\lambda_{i_0} \cdot p_{i_0 i_1} \cdot \ldots \cdot p_{i_{n-1} i_n}}.$$

2.

$$P\left(X_n=j|X_0=i\right)=\frac{P\left(X_n=j, X_0=i\right)}{P(X_0=i)}$$

$$P\left(X_n=j, X_0=i\right)=\sum_{i_1, i_2, \ldots, i_{n-1} \in E} P\left(X_n=j, X_{n-1}=i_{n-1}, \ldots, X_1=i_1, X_0=i\right)=$$
$$\sum_{i_1, \ldots, i_{n-1} \in E} \lambda_i \cdot p_{i i_1} \cdot p_{i_1 i_2} \cdot * p_{i_{n-1} j}=\boxed{\lambda_i \cdot\left(P^n\right)_{ij}}.$$

Таким чином:

$$P\left(X_n=j|X_0=i\right)=\frac{\lambda_i \cdot\left(P_{ij}^n\right)}{\lambda_i}=\left(P_{ij}^n\right)$$

Отже

$$\boxed{p_{ij}^{(n)}=P\left(X_n=j|X_0=i\right)=\left(P^n\right)_{ij}}$$

3.

$$P\left(X_n=j\right)=\sum_{i \in E} P\left(X_0=i\right) \cdot P\left(X_n=j|X_0=i\right)=$$
$$\sum_{i \in E} \lambda_i \cdot\left(P^n\right)_{ij}.$$

Оскільки $\left(P^n\right)_{ij}=P\left(X_n=j|X_0=i\right)$ то

$$\sum_{j \in E}\left(P^n\right)_{ij}=\sum_{j \in E} P\left(X_n=j|X_0=i\right)=1$$

Отже P^n — стохастична для довільного n.

Theorem 2 (Марковська властивість)**.** Нехай $\{X_n\}_{n \geq 0}$ — ланцюг Маркова (λ, P).

Тоді

$$\forall M \geq 1 \text{ та } \forall i \in E$$

при умові, що $X_m=i$ послідовність $\{X_{m+n}, n \geq 0\}$ є ланцюгом Маркова з початковим розподілом

$$\delta_i=(0 \ldots 0,1,0 \ldots 0)$$

і матрицею перехідних ймовірностей P.

Більше того: розподіл в.в. $X_{m+1}, X_{m+2}, \ldots$ при $X_m=1$ не залежить від $X_0, \ldots, X_{m-1}$.

Proof. Достатньо довести, що

$$P\left(X_{m+1}=j_1,\dots,X_{m+k}=j_k|X_m=i,X_{m-1}=i_{m-1},\dots,X_0=i_0\right)=\\ =P\left(X_1=j_1,\dots,X_k=j_k|X_0=i\right).$$

$$\begin{aligned}P\left(X_{m+1}=j_1,\dots,X_{m+k}=j_k|X_m=i,\dots,X_0=i_0\right)=\\ =\frac{P\left(X_0=i_0,\dots,X_{m-1}=i_{m-1},X_m=i,\dots,X_{m+k}=j_k\right)}{P\left(X_0=i_0,\dots,X_m=i\right)}=\\ =\frac{\left[\lambda_0 p_{i_0i_1}\cdot\ldots\cdot p_{i_{m-1}i}\right]\cdot p_{ij_1}\dots p_{j_{k-1}j_k}}{\lambda_0 p_{i_0i_1}\dots p_{i_{m-1}i}}=\\ =p_{ij_1}\cdot\ldots\cdot p_{j_{k-1}j_k}=P\left(Y_1=j_1,\dots,Y_k=j_k|Y_0=i\right).\end{aligned}$$

де $\{Y_k\}_{k\geq 0}$ — ланцюг Маркова (δ,P), де $\delta=(0\dots0,1,0\dots0)$. ∎

Example.

$$E=\{1,2\};\quad P=\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{3} & \frac{2}{3}\end{pmatrix},\quad \lambda=\left(\frac{1}{2},\frac{1}{2}\right)$$

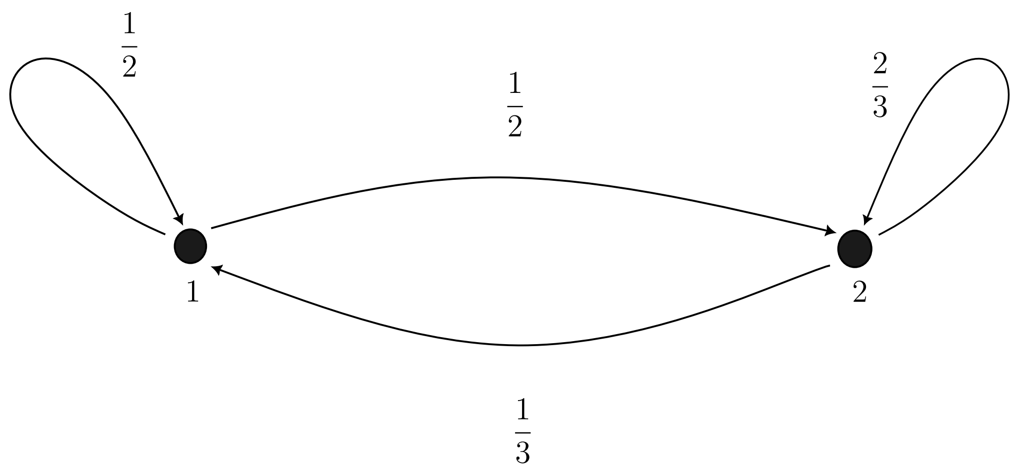

Figure 1: Example: states transition

$$P\left(X_0=1,X_1=2\right)=P\left(X_0=1\right)\cdot P\left(X_1=2|X_0=1\right)=\\ =\lambda_1\cdot p_{12}=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}.$$

$$P\left(X_3=1|X_1=2\right)=p_{21}^{(2)}=\left(p^2\right)_{21}$$

$$p^2=\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{3} & \frac{2}{3}\end{pmatrix}\cdot\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{3} & \frac{2}{3}\end{pmatrix}=\begin{pmatrix}\cdot & \cdot\\ \boxed{\frac{7}{18}} & \cdot\end{pmatrix}.$$

$$P\left(X_3=1\right)=\sum_{i\in E}P\left(X_0=i\right)\cdot P\left(X_3=1|X_0=i\right)=\frac{1}{2}*p_{11}^{(3)}+\frac{1}{2}p_{21}^{(3)}=\frac{1}{2}\left((p^{(3)})_{11}+(p^{(3)})_{21}\right).$$

5.2 Класифікація станів ланцюга Маркова

Definition 7. Стан i назхивається несуттєвим якщо

$$\exists j\in E\ \ (j\neq i):\ p_{ij}^{(m)}>0\ \text{ при } m\geq 1 \text{ але } p_{ji}^{(n)}=0\ \ \forall n\geq 1$$

Example. Ланцюг, в якого всі стани є несуттєвими.

$$E=\{1,2,3,\dots\}$$

$$p_{ij}=\begin{cases}p & j=i\\ 1-p & j=i+1\\ 0 & \text{інакше}\end{cases}.$$

Remark. Якщо E — скінченна, то всі стани не можуть бути несуттєвими.

Надалі вважаємо, що E — множина суттєвих станів.

Definition 8. Стан $j\in E$ наз. **досяжним** зі стану $i\in E$ $(i\to j)$ якщо $\exists m>0: p_{ij}^{(m)}>0$.

Стан i,j називається **сполучним** $(i\leftrightarrow j)$, якщо $\exists m_1,m_2>0: p_{ij}^{(m_1)}>0,\ p_{ji}^{(m_2)}>0$.

Remark. Відношення $\leftrightarrow$ є відношенням еквівалетності: симметричне, рефлексивне та транзитивне.

Proof. Справді: якщо $i\leftrightarrow j, j\leftrightarrow i$, то

$$\exists m_1,m_2>0:\quad p_{ij}^{(m_1)}>0,\ \ p_{ji}^{(m_2)}>0$$

$$\exists r_1,r_2>0:\quad p_{jk}^{(r_1)}>0,\ \ p_{kj}^{(r_2)}>0$$

∎

Тоді

$$\begin{aligned}p_{ik}^{(m_1+r_1)}&=\sum_{l\in E}P\left(X_{m_1+r_1}=k,X_{m_1}=l|X_0=i\right)=\\&=\sum_{l\in E}P\left(X_{m_1}=l|X_0=i\right)\cdot P\left(X_{m_1+r_1}=k|X_{m_1}=l,X_0=i\right)=\\&=\sum_{l\in E}p_{il}^{(m_1)}\cdot p_{lk}^{(r_1)}\geq p_{ij}^{(m_1)}p_{jk}^{(r_1)}>0.\end{aligned}$$

Аналогічно

$$p_{ki}^{(m_2+r_2)}=\sum_{l\in E}p_{kl}^{(m_2)}p_{li}^{(r_2)}\geq p_{kj}^{(m_2)}p_{ji}^{(r_2)}>0$$

Тоді $E=\bigsqcup_{i=1}^{S}E_i$, де $E_1,\dots,E_S$ — неперетинні класи сполучних станів.

Definition 9. Ланцюг, всі стани якого утворюють один клас сполучних станів, називається нерозкладним.

Example.

$$E = \{1, 2, 3, 4\}$$

$$P = \begin{pmatrix} 0 & 0.5 & 0.5 & 0 \\ 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0.5 & 0.5 \end{pmatrix}$$

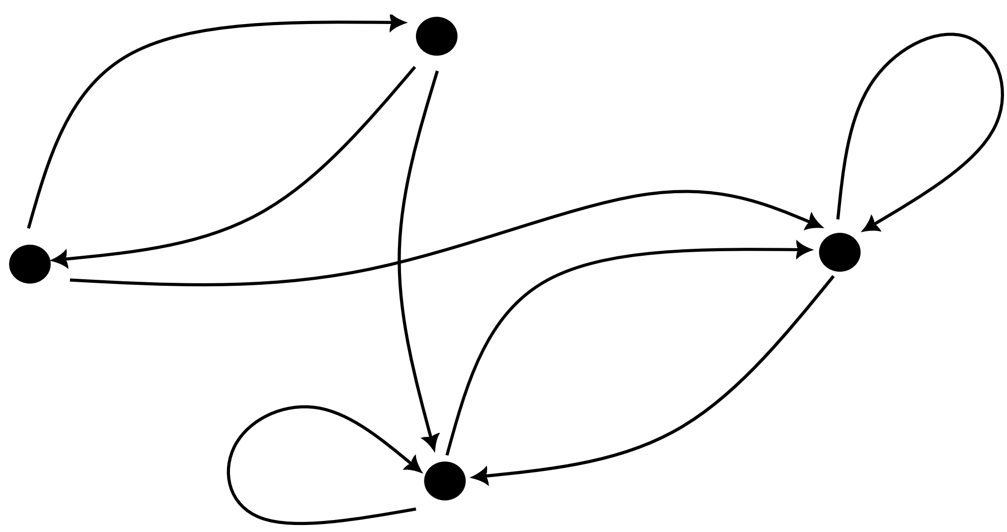

Figure 2: Приклад нерозкладного ланцюга Маркова

Охарактеризуємо стани:

$$1 \to 3 \text{ але } 3 \not\to 1 \Rightarrow \text{ 1 несуттєвий}$$

$$2 \to 4 \text{ але } 4 \not\to 2 \Rightarrow \text{ 2 несуттєвий}$$

3, 4 — суттєві стани. $\widetilde{E} = \{3, 4\}$ — клас сполучних станів.

Definition 10. Стан $i \in E$ має період $d = d(i)$, якщо:

1. $p_{ii}^{(n)} > 0$ лише для тих n, які кратні d
2. d — найбільше з чисел, які володіють властивістю 1.

$$d = GCD\{n : p_{ii}^{(n)} > 0\}$$

Definition 11. Якщо $d = 1$, то стан називається **аперіодичним**.

Definition 12. Якщо **всі стани** нерозкладного ланцюга Маркова є аперіодичними, то сам **ланцюг** називається **аперіодичним**.

6 Lecture 6: Час і ймовірність досягнення

Нехай $A \subset E$ — деяка підмножина мн-ни станів.

Позначимо

$$\tau^A = inf\ \{n \geq 0 : X_n \in A\}$$
$$\tau^A = \infty,\ \text{якщо}\ \{n \geq 0 : X_n \in A\} = \varnothing$$

τ^A — момент 1-го досягнення множини A.

Розглянемо

$$h^A = P\left(\tau^A < \infty | X_0 = i\right)$$

це ймовірність того, що стартувавши з i, ланцюг коли-небудь потрапить в A.

$$K_i^A = M\left[\tau^A | X_0 = i\right] = \sum_n n \cdot P\left(\tau^A = n | X_0 = i\right) + \infty \cdot P\left(\tau^A = \infty | X_0 = i\right)$$

середній час на потрапляння в A.

Theorem 3. При заданій множині $A \subset E$ величини h_i^A та K_i^A є мінімальними розв'язками таких систем лінійних рівнянь.

$$\begin{cases} h_i^A = 1, & i \in A \\ h_i^A = \sum\limits_{j \in E} p_{ij} \cdot h_j^A, & i \notin A \end{cases}$$

$$\begin{cases} K_i^A = 0 & i \in A \\ K_i^A = 1 + \sum\limits_{j \in E} p_{ij} \cdot K_j^A & j \notin A \end{cases}$$

(якщо $g_i \geq 0$ — інший розв'язок, то $g_i \geq h_i^A,\ i \in E$).

Proof. Якщо $i \in A$, то $\tau^A = 0$ і тоді $h_i^A = 1$.

Якщо $i \notin A$, то $\tau^A \geq 1$ і тоді

$$\begin{aligned} h_i^A &= P\left(\tau^A < \infty | X_0 = i\right) = \sum_{j \in E} P\left(\tau^A < \infty, X_1 = j | X_0 = i\right) = \\ &= \sum_{j \in E} P\left(X_1 = j | X_0 = i\right) \cdot P\left(\tau^A < \infty | X_1 = j, X_0 = i\right) = \\ &= \sum_{j \in E} p_{ij} \cdot P\left(\tau^A < \infty | X_0 = j\right) \\ &= \sum_{j \in E} p_{ij} \cdot h_j^A. \end{aligned}$$

Якщо $i \in A$, то $\tau^A = 0 \Rightarrow K_i^A = 0$.

Якщо $i \notin A$, то $\tau^A \geq 1$ і

$$\begin{aligned}
K_i^A &= M\left[\tau^A | X_0 = i\right] = \\
&= \sum_{j\in E} M\left[\tau^A, X_1 = j | X_0 = i\right] = \\
&= \sum_{j\in E} P\left(X_1 = j | X_0 = i\right) M\left[\tau^A | X_1 = j, X_0 = i\right] = \\
&= \sum_{j\in E} p_{ij} \cdot \left[M\left[\tau^A | X_0 = j\right] + 1\right] = \\
&= \sum_{j\in E} p_{ij} \cdot \left(K_j^A + 1\right).
\end{aligned}$$

□

Example.

$$E = \{1, 2, \dots, 7\}$$

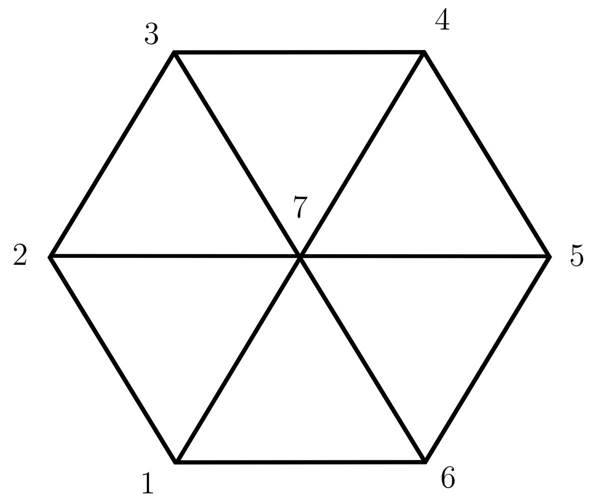

Figure 3: Example to the theorem 3. States

$$X_0 = 1$$

З кожної вершини частинка з ймовірністю $\frac{1}{3}$ переходить в одну з трьох сусідніх вершин (з'єднаниз з нею ребром).

Яка ймовірність того, що частинка повернеться в стан 1 раніше ніж побуває в стані 7?

$$h_i = P\,(\text{ частинка досягне 1 раніше за 7 } | X_0 = i) \quad u = \overline{1,7}$$

Хочемо знайти h_1.

Маємо такі рівняння:

$$
\begin{aligned}
h_1 &= \frac{1}{3} \cdot h_2 + \frac{1}{3} \cdot h_6 + \frac{1}{3} \cdot 0 = \frac{1}{3}h_2 + \frac{1}{3}h_6 + = \frac{2}{3}h_2 \\
h_2 &= \frac{1}{3} \cdot 1 + \frac{1}{3}h_3 + \frac{1}{3} \cdot 0 = \frac{1}{3} + \frac{1}{3}h_3 \\
h_3 &= \frac{1}{3}h_2 + \frac{1}{3}h_4 \\
h_4 &= \frac{1}{3}h_3 + \frac{1}{3}h_5 = \frac{2}{3}h_3.
\end{aligned}
$$

Тоді:

$$
\begin{aligned}
h_3 &= \frac{1}{2}h_2 + \frac{2}{9}h_3 \\
h_3 &= \frac{2}{7}h_2 \\
h_2 &= \frac{1}{7}h_2 + \frac{1}{3} \Rightarrow h_2 = \frac{7}{18}.
\end{aligned}
$$

Тоді $h_1 = \frac{2}{3} \cdot h_2 = \frac{7}{27}$.

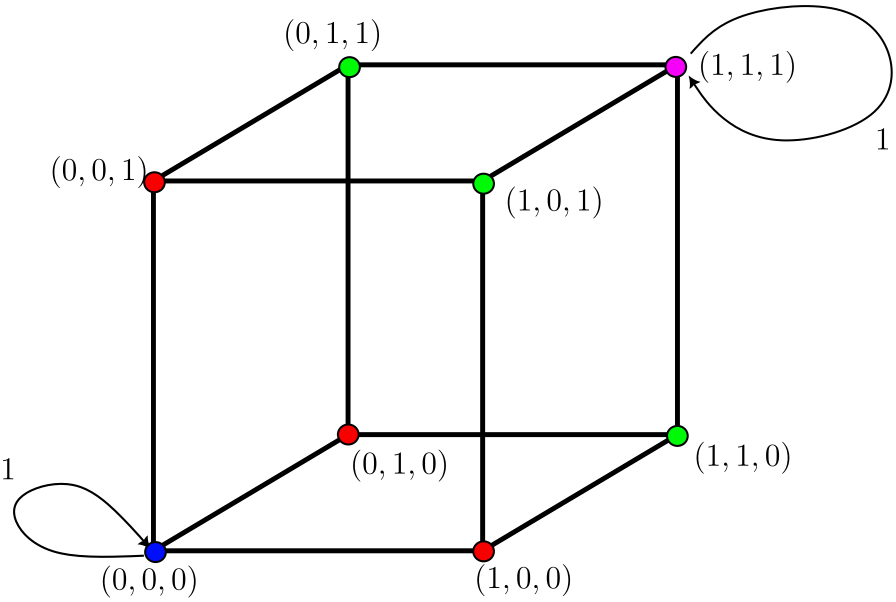

Figure 4: Example 2 to the theorem 3

Example.

$$X_0 = (0, 0, 1)$$

З ймовірністю $\frac{1}{3}$ з кожної вершини частинка переходить в одну з трьох сусідніх.

Будемо вважати, що стани $(0, 0, 0)$ та $(1, 1, 1)$ є поглинальними станами.

Яка ймовірність того, що ланцюг буде поглинуто саме в стані $(0, 0, 0)$? Можна переформулювати у питання: ланцюг потрапить у стан $(1, 1, 1)$ раніше, ніж в стан $(0, 0, 0)$.

Введемо класи станів (пофарбоно на рисунку):

$$
\begin{aligned}
E_1 &= \{(1,1,1)\} \\
E_2 &= \{(0,0,0)\} \\
E_3 &= \{(0,1,1),(1,0,1),(1,1,0)\} \\
E_4 &= \{(0,0,1),(0,1,0),(1,0,0)\}.
\end{aligned}
$$

Укрупнення станів

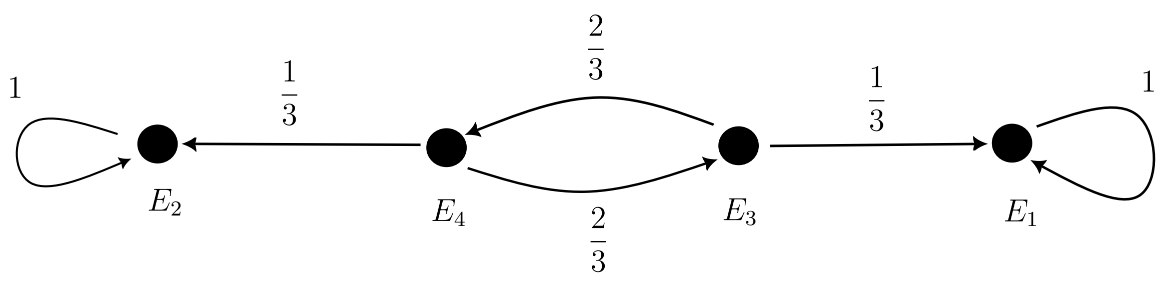

Figure 5: Example 2 to the theorem 3 states

$$
\begin{aligned}
h_i = P\,(\text{ ланцюг потрапить в (0,0,0) раніше, ніж в (1,1,1) } |X_0 = i) \\
i \in \{\underset{3}{E_1}, \underset{0}{E_2}, \underset{2}{E_3}, \underset{1}{E_4}\}.
\end{aligned}
$$

$$
\begin{aligned}
h_0 &= 1 \\
h_3 &= 0
\end{aligned}
$$

$$
\begin{cases}
h_1 = \frac{1}{3} \cdot 1 + \frac{2}{3} \cdot h_2 \\
h_2 = \frac{2}{3} \cdot h_1 + \frac{1}{3} \cdot 0
\end{cases}
\Rightarrow
\begin{aligned}
h_2 &= \frac{2}{3} h_1 \\
h_1 &= \frac{1}{3} + \frac{4}{9} h_1
\end{aligned}.
$$

$$
\Rightarrow \frac{5}{9} h_1 = \frac{1}{3}
$$

$$
h_1 = \frac{3}{5}; \qquad h_2 = \frac{2}{5}
$$

7 Рекурентність ланцюгів Маркова

7.1 Строго марківська властивість

Definition 13. Випадкова величина T, яка залежить від ланцюга $X_0, X_1, \ldots$ і приймає значення дискретні $n = 0, 1, 2, \ldots$ називається **моментом зупинки** якщо подія $\{T = n\}$ описується лише в термінах випадкових величин $X_0, \ldots, X_n$ і не залежить від в.в. $X_{n+1}, X_{n+2}, \ldots$.

Example. 1. τ^A — момент зупинки. $\{\tau^A = n\} = (X_0 \notin A, X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A)$ — ланцюг вперше в момент часу n потрапив в множину A.

2. $\mathcal{L}^A$ — останній момент перебування в A після потрапляння в неї.

$$\mathcal{L}^A = \min_n \{n > \tau^A : X_n \in A, X_{n+1} \notin A\}$$

Отже подія $\{\mathcal{L}^A = n\}$ залежить від $X_{n+1} \Rightarrow \mathcal{L}^A$ не є моментом зупинки.

Theorem 4 (Строго марковська властивість)**.** Нехай $(X_n)_{n\geq 0}$ — ланцюг Маркова (λ, P) Нехай T— момент зупинки. Тоді при умові, що $T < \infty$ і $X_T = i$ послідовність $(X_{T+n})_{n\geq 0}$ утворює ланцюг маркова (δ_i, P). Зокрема, $X_{T+1}, X_{T+2}, \ldots$ не залежить від $X_0, \ldots, X_{T-1}$.

Proof. Нехай $T = \tau^i$ — момент першого досягнення стану i. *Хочемо показати:*

$$P\left(X_{T+1} = j_1, \ldots, X_{T+m} = j_m\right) | X_T = i, T < \infty = P\left(X_1 = j_1, \ldots, X_m = j_m | X_0 = i\right)$$

$$\forall j_1, \ldots, j_m, i \in E$$

$$\begin{aligned}
&P\left(X_{T+1} = j_1, \ldots, X_{T+m} = j_m | X_T = i, T < \infty\right) = \\
&= \frac{P\left(X_{T+1} = j_1, \ldots, X_{T+m} = j_m, X_t = i, T < \infty\right)}{P\left(X_T = i, T < \infty\right)} = \\
&= \frac{\sum_{k=0}^{\infty} P\left(X_{k+1} = j_1, \ldots, X_{k+m} = j_m, X_k = i, T = k\right)}{P\left(X_T = i, T < \infty\right)} = \\
&= \frac{\sum_{k=0}^{\infty} P\left(X_k = i, T = k\right) \cdot P\left(X_{k+1} = j_1, \ldots, X_{k+m} = j_m | X_k = i, T = k\right)}{P\left(X_T = i, T < \infty\right)} = \qquad \text{марк. власт. } k\text{-детерм.} \\
&= \frac{\sum_{k=0}^{\infty} P\left(X_k = i, T = k\right) \cdot P\left(X_1 = j_1, X_m = j_m | X_0 = i\right)}{P\left(X_T = i, T < \infty\right)} = \\
&= P\left(X_1 = j_1, \ldots, X_m = j_m | X_0 = i\right) \cdot \frac{P\left(X_T = i, T < \infty\right)}{P\left(X_T = i, T < \infty\right)} = \\
&= P\left(X_1 = j_1, \ldots, X_m = j_m | X_0 = i\right).
\end{aligned}$$

□

Example (Частково спостережувані процеси)**.** $\{X_n\}_{n\geq 0}$ — ланцюг Маркова

1. Спостереження доступні лише тоді, коли

$$X_{n+1} \neq X_n$$

2. Спостереження доступні лише тоді, коли

$$X_n \in A \subset E$$

7.2 Рекурентність: означення та властивості

Example. $E = \{1, 2, 3, \dots, \}$; $p_{ij} = p < 1$, $p_{i,i+1} = 1 - p$; $X_0 = 0$

Всі стани не суттєві

Після виходу з нуля ланцюг більше ніколи в нього не повернеться.

Example (Симметричне випадкове блукання).

$$X_0 = 0 \quad X_{n+1} = X_n + \varepsilon_{n+1}$$

де $P(\varepsilon_n = \pm 1) = \dfrac{1}{2}$; $\{\varepsilon_n\}$ — н.о.р.

$$p = P(X_n = 0 \text{ для деякого } n \geq 1 \,|X_0 = 0)$$

Позначимо $h_m = P(X_n = 0 \text{ для деякого } n \geq 0 \,|X_0 = m)$ та $h_0 = 1$.

$h_m = \dfrac{1}{2} h_{m+1} + \dfrac{1}{2} h_{m-1}$ — формула повної ймовірності.

$h_m = a \cdot m + b$

$h_0 = 1$, $0 \leq h_m \leq 1 \Rightarrow a = 0, b = 1$

Отже $h_m = 1 \ \ \forall m$. Тоді

$$p = \frac{1}{2} h_1 + \frac{1}{2} h_{-1} = 1$$

Висновок: вип. симетр. блукання з ймовірністю 1 повернеться в 0 (коли-небудь).

Definition 14. Стан $i \in E$ називається **рекурентним**, якщо ймовірність повернення в стан і після виходу з нього дорівнює 1.

$$P(\exists n \geq 1 : X_n = i | X_0 = i) = 1$$

Стан i наз. **нерекурентним**, якщо

$$P(X_n = i \text{ для деякого } n \geq 1 \mid X_0 = i) < 1$$

Позначимо як момент 1-го попадання в стан i.

$$T_i = \min\{n \geq 1 : X_n = i\}$$

Remark. T_i — момент зупинки, а отже марковський момент часу.

Дамо альтернативні означення рекурентності та нерекурентності:

Definition 15. Стан $i \in E$ називається рекурентним, якщо

$$P(T_i < \infty \mid X_0 = i) = 1$$

а відповідно для нерекурентного

$$P(T_i < \infty \mid X_0 = i) < 1$$

Введемо позначення

$$\mathcal{N}_i = \sum_{n=0}^{\infty} \mathbb{1}(X_n = i)$$

загальна кількість відвідувань стану i, включаючи $n = 0$.

Corollary 1.

$$P\left(\mathcal{N}_i > K \mid X_0 = i\right) = \left(f_i\right)^K, \quad K = 0, 1, 2, \ldots$$

де $f_i = P\left(T_i < \infty \mid X_0 = i\right) \equiv P_i\left(T_i M < \infty\right)$

Proof. $K = 0:$

$$P_i(\mathcal{N}_i > 0) = P\left(\mathcal{N}_i > 0 \mid X_0 = i\right) = 1$$

$$\left(f_i\right)^0 = 1$$

$\Rightarrow$ рівність виконується ☐

---

8 Lecture 7: Час і ймовірність досягнення

Нехай маємо $\left(X_n\right)_{n\geq 1}$ — Ланцюг Маркова на $E = \{1, 2, 3, \ldots\}$ з початковим розподілом $\lambda = \left(\lambda_i\right)_{i\in E}$ та матрицею перезіжних ймовірностей p.

Theorem 5. Якщо стан $i \in E$ є рекурентним і $i \leftrightarrow j$ (стани i, j є сполучними), то стан j теж є рекурентним.

Якщо стан j не є рекурентним, то

$$\forall i \in E \sum_{n=0}^{\infty} p_{ij}^{(n)} < \infty$$

та

$$p_{ij}^{(n)} \to_{n\to\infty} 0$$

Proof. Оскільки $i \leftrightarrow j$, то $\exists k, m > 0$ так що

$$p_{ij}^{(k)} > 0, \quad p_{ji}^{(m)} > 0$$

Тоді

$$p_{jj}^{(k+n+m)} \geq p_{ji}^{(k)} \cdot p_{ii}^{(n)} \cdot p_{ij}^{(m)}$$

Якщо

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$$

, то $\sum_{n=0}^{\infty} p_{jj}^{(n)} = +\infty$.

2. Позначимо

$$T_j = \min\{n \geq 1 : X_n = j\}$$

як момент 1-го попадання в стан j після $n = 0$.

І нехай $f_{ij}^{(n)} = P\left(T_j = n \mid X_0 = i\right)$ — ймовірність того, що ланцюг вперше попаде в j в момент часу n, за умови, що $X_0 = i$.

Тоді за формулою повної ймовірності

$$\begin{aligned}
p_{ij}^{(n)} &= P\left(X_n = j | X_0 = i\right) = \\
&= \sum_{k=1}^{n} P\left(X_n = j, T_j = k | X_0 = i\right) = \\
&= \sum_{k=1}^{n} P\left(T_j = k | X_0 = i\right) \cdot P\left(X_n = j | T_j = k, X_0 = i\right) = \text{ Строго марківська властивість} \\
&= \sum_{k=1}^{n} P\left(T_j = k | X_0 = i\right) \cdot P\left(X_{n-k} = j | X_0 = j\right) = \\
&= \sum_{k=1}^{n} f_{ij}^{(k)} \cdot p_{jj}^{(n-k)}
\end{aligned}$$

.

Звідси

$$\begin{aligned}
\sum_{n=0}^{\infty} p_{ij}^{(n)} = \sum_{n=0}^{\infty} \sum_{k=1}^{n} f_{ij}^{(k)} \cdot p_{jj}^{(n-k)} =& \\
&= \sum_{k=1}^{\infty} f_{ij}^{(k)} \cdot \sum_{n=k}^{\infty} p_{jj}^{(n-k)} = \\
&= \underbrace{\sum_{k=1}^{\infty} f_{ij}^{(k)}}_{f_i = \sum_{k=1}^{\infty} f_{ij}^{(k)} = P(\text{ потрапити коли-небудь з i в j }) \leq 1} \cdot \sum_{m=0}^{\infty} p_{jj}^{(m)} = \\
&= f_{ij} \cdot \sum_{m=0}^{\infty} p_{jj}^{(m)}
\end{aligned}$$

.

Якщо стан j — нерекурентний, то $\sum\limits_{n=0}^{\infty} p_{jj}^{(n)} < +\infty$.

Тоді і $\sum\limits_{n=0}^{\infty} p_{ij}^{(n)} < +\infty$, а тому і $p_{ij}^{(n)} \to 0 mn \to \infty$. (необхідна умова збіжності ряду). □

Consequence. 1. Всі стани скінченного нерозкладного ланцюга Маркова є рекурентними.

2. Всі стани нерозкладного ланцюга Маркова є рекурентними або нерекурентними одночасно.
3. Стан i несуттєвий $\Rightarrow$ стан i нерекурентний.

Proof. Якщо стан i — несуттєвий, то $\exists j \in E$ та $\exists m > 0$ такі, що

$$p_{ij}^{(m)} > 0 \text{ але } p_{ji}^{(n)} = 0 \quad \forall n$$

Нехай

$$A = \{N_i = +\infty\}, \qquad \text{де } N_i \text{ кількість відвідувань стану і}$$

$$B = \{X_m = j\}$$

Тоді $A \cap B = \varnothing$ і тому $P\left(A \cap B\right) = 0$

З іншого боку:

$$0 < P_i(B) = P\left(X_m = j|X_0 = i\right) = p_i\left(A \cap B\right) + p_i\left(\overline{A} \cap B\right) =$$

$$= p_i(\overline{A} \cap B) \leq p_i(\overline{A})$$

Звідси

$$p_i(A) = 1 - p_i(\overline{A}) < 1 \Rightarrow p_i(A) = 0 \Rightarrow p_i\left(N_i = +\infty\right) = 1$$

$$\Rightarrow i \text{ — нерекурентний}$$

□

9 Нульова та додатня рекурентність

Питання: *яким є середній час очікування повернення в деякий рекурентний стан?*

Нехай $T_i = \min\{n \geq 1 : X_n = i\}$ — момент 1-го потрапляння в стан $i \in E$.

Definition 16. Рекурентний стан $i \in E$ називається додатнім рекурентним, якщо

$$M_i T_i = M\left[T_i | X_0 = i\right] < \infty \qquad \left(\frac{1}{M_i T_i} > 0\right)$$

Рекурентний стан $i \in E$ називається рекурентним нульовим, якщо

$$M_i T_i = +\infty \qquad \left(\frac{1}{M_i T_i} = 0\right)$$

Example (Симметричне випадкове блукання на $\mathbb{Z}$)**.** $(X_n)_{n\geq 0}$ — симметричне випадкове блукання:

$$X_0 = 0; \quad X_{n+1} = X_n + \varepsilon_{n+1}, \quad \{\varepsilon_n\}_{n\geq 1} \text{ — н.о.р.}$$

Стан 0 є рекурентним.

Нехай $g(x)$— середній час до потрапляння в 0 з точки $x \in \mathbb{Z}$. За формулою повної ймовірності:

$$g(x) = \frac{1}{2}g(x+1) + \frac{1}{2}g(x-1) + 1$$

Розв'язок цього рівняння слід шукати у вигляді

$$g(x) = ax^2 + bx + c$$

Але шукати в такому вигляді не будемо, а натомість пепишемо рівняння в іншому вигляді:

$$g(x+1) - g(x) = (g(x) - g(x-1)) - 2 =$$

$$= (g(x-1) - g(x-2) - 2) - 2 = (g(x-1) - g(x-2)) - 4 = \ldots$$

Якби $g(x) < \infty$, то права частина стане від'ємною за скінченну кількісьб ітерацій. Зліва стоїть невід'ємна величина. Це протиріччя доводить, що $g(x) = +\infty \ \forall x \in \mathbb{Z}/\ \{0\}$ є нескінченністю.

Висновок: симметричне випадкове блукання є рекурентним нульовим.

Theorem 6. Якщо $i \in E$ є рекурентним додатнім і $j \leftrightarrow i$, то j теж є рекуретним додатнім.

Proof. Припустимо, що $X_0 = i$. Введемо такі моменти часу

$$\tau_0 = 0$$

$$\tau_1 = \min\{n \geq 1 : X_n = i\} \text{ — момент першого повернення в i}$$

$$\vdots$$

$$\tau_{k+1} = \min\{n > \tau_k : X_n = i\} \text{ — момент k+1 повернення в i}$$

Нехай $Y_n = \tau_n - \tau_{n-1}$ — час між послідовними поверненнями в i.

Зі строго марківської властивості слідує, що

$$\{Y_n\}_{n\geq 1} \text{ — н.о.р. в.в.}$$

$$MY_n = MY_1 = M\tau_1 < \infty$$

Розглянемо "екскурсії":

$$\xi^{(0)} = (X_n, 0 \leq n < \tau_1)$$
$$\xi^{(1)} = (X_n, \tau_1 \leq n < \tau_2)$$
$$\vdots$$
$$\xi^{(k)} = (X_n, \tau_k \leq n < \tau_{k+1})$$

еволюція $\{X_n\}_{n\geq 0}$ між послідовними потрапляннями в i.

Зі строго марковської властивості випливає, що

$$\{\xi^{(k)}\}_{k\geq 0} \text{ — незалежні і однаково розподілені}$$

Нехай $j \leftrightarrow i$. Оскільки i — рекурентний, то j — також рекурентний. Отже стан j відвідується нескінченно часто з ймовірністю 1.

Звідси випливає, що в нескінченно багатьох екскурсіях буде відвідувати стан j з ймовірністю 1.

малюнок

Нехай $f_{ij} = P\left(T_j < \infty | X_0 = i\right)$. Ця ймовірність є однаковою для всіх екскурсій.

Нехай R_1 та R_2 — індекси 1-ї та 2-ої по рахунку екскурсій, в яких відбулося відвідування стану j.

Тоді що ми можемо сказати про середній час відвідування стану j?

$$\begin{aligned} M_j T_t &\leq M \sum_{i=1}^{R_2 - R_1 + 1} Y_i = \\ &= MY_1 \cdot M\left(R_2 - R_1 + 1\right) = \\ &= MY_1 \left(1 + M\left(R_2 - R_1\right)\right). \end{aligned}$$

Оскільки

$$P(R_2 - R_1 = k) = P(\text{ в } (k-1) \text{ екскурсіях потрапляння в } j \text{ не було, а в } k \text{ відбулося }) =$$

$$= (1 - f_{ij})^{k-1} \cdot f_{ij}, \quad k = 1, 2, \ldots$$

$$R_2 - R_1 \sim Geom\left(f_{ij}\right) \Rightarrow M\left(R_2 - R_1\right) = \frac{1}{f_{ij}}$$

Таким чином

$$j \leftrightarrow i \Rightarrow f_{ij} > 0; \qquad M_i t_i < \infty \Rightarrow M_j T_j < \infty$$

Тобто, стан j — рекурентний додатній. □

Corollary 2. Всі стани нерозкладного рекурентного ланцюга Маркова є додатними або нульовими одночасно.

10 Інваріантний розподіл ланцюга Маркова

Definition 17. Вектор $\pi = (\pi_i)_{i \in E}$ називається інваріантним розподілом ланцюга $(X_n)_{n \geq 0}$ якщо

$$\begin{cases} \pi = \pi P \\ \pi_i \geq 0 \\ \sum\limits_{i \in E} \pi_i = 1 \end{cases}$$

Corollary 3. Якщо інваріантний розподіл співпадає з початковим, то

$$\forall i \in E \;\; \forall n \geq 0 \;\; P\left(X_n = i\right) = \pi_i$$

Proof. Справді

$$\begin{aligned} P\left(X_n = i\right) &= \sum_{n=j \in E} P\left(X_0 = j\right) \cdot p_{ji}^{(n)} = \\ &= \sum_{n=j \in E} \pi_j \left(p^n\right)_{ji} = \\ &= \left(\pi p^n\right)_i = \\ &= \left(\underbrace{\pi \cdot p}_{\pi} \cdot p^{n-1}\right) = \\ &= \left(\pi p^{n-1}\right)_i = \left(\pi p\right)_i = \pi_i. \end{aligned}$$

□

Corollary 4. Нехай E — скінченна множина. Припустимо, що для деякого $i \in E$:

$$\forall j \in E \quad p_{ij}^{(n)} \to \pi_j, \quad n \to \infty$$

Тоді $\pi = (\pi_i)_{i \in E}$ — інваріантний розподіл.

Proof. Перевіримо, що $\sum\limits_{i\in E}\pi_i = 1$:

$$\sum_{i\in E}\pi_i = \sum_{i\in E}\lim_{n\to\infty} p_{ij}^{(n)} = \text{E — скінч.}$$

$$\lim_{n\to\infty}\sum_{i\in E} p_{ij}^{(n)} = \lim_{n\to\infty} 1 = 1$$

Далі

$$\begin{aligned}\pi_j = \lim_{n\to\infty} p_{ij}^{(n)} &= \lim_{n\to\infty}\sum_{k\in E} p_{ik}^{(n-1)}\cdot p_{kj} = \\ &= \sum_{k\in E}\underbrace{\lim_{n\to\infty} p_{ik}^{(n-1)}}_{\pi_k}\cdot p_{kj} = \\ &= \sum_{k\in E}\pi_k\cdot p_{kj}.\end{aligned}$$

$$\Rightarrow \pi = \pi P$$

□

Example.

$$p = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

$$0<\alpha<1 \qquad 0<\beta<1$$

Система рівнянь для знаходження інваріантного розподілу:

$$\begin{aligned}\pi_1 &= (1-\alpha)\,\pi_1 + \beta\pi_2 \\ \pi_2 &= \alpha\pi_1 + (1-\beta)\pi_2 \\ \pi_1+\pi_2 &= 1.\end{aligned}$$

існує єдиний розв'язок:

$$\pi_1 = \frac{\beta}{\alpha+\beta}$$

$$\pi_2 = .$$

Example (інваріантних розподілів безліч)**.**

$$p = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Ланцюг розкладний

$$\begin{cases}\pi_1 = \pi_1 \\ \pi_2 = \pi_2 \\ \pi_1+\pi_2 = 1\end{cases}$$

$$\pi = (\alpha, 1-\alpha),\ \text{де } 0\le a\le 1$$

є безліч інваріантних розподілів.

Example (інваріантний розподіл не існує). $(X_n)_{n\geq 0}$ — симметричне випадкове блукання

$$\pi_k = \frac{1}{2}\pi_{k-1} + \frac{1}{2}\pi_{k+1}$$

рівняння арифметичної прогрессії

$$\pi_k = A \cdot k + b$$

Але $0 \leq \pi_k \leq 1$. тому $A = 0, \pi_k = b = const$

Задовольнити умову $\sum\limits_{k=-\infty}^{\infty} \pi_k = 1$ неможливо.

Інваріантного розподілу не існує.

11 Lecture 8: Граничні теореми для ланцюгів Маркова

11.1 Умови існування інваріантного розподілу