Теорија на информации

Ентропија на случаен процес

Верига на Марков

Дефиниција 1. Дискретен случаен процес X_1, X_2, … се нарекува *верига на Марков*, ако за секој $n \in N$, и за секои x_1, x_2,..., $x_{n+1} \in R_X$, точно е равенството:

$$P\{X_{n+1}= x_{n+1} | X_n= x_n, X_{n-1}= x_{n-1},\ldots, X_1= x_1\}= P\{X_{n+1}= x_{n+1} | X_n= x_n\}.$$

- Означуваме $P\{X_n= j | X_{n-1}= i\} = p_{ij}^{(n)}$.
- Ако $\{X_n | n \in N\}$ е верига на Марков, тогаш X_n се нарекува состојба на системот во момент n.

Матрица на преодни веројатности

- Веројатностите $p_{ij}^{(n)}$ се нарекуваат преодни веројатности од состојба i во состојба j во n-тиот момент на промена. Овие веројатности формираат матрица $\boldsymbol{P}^{(n)} = [p_{ij}^{(n)}]$.
- Согласно, дефиницијата на Марков процес, ако процесот има s состојби, елементите на оваа матрица ги задоволуваат следните својства:

$$0 \leq p_{ij}^{(n)} \leq 1, \quad i, j = 1, 2, \ldots, s,$$

$$\sum_{j=1}^{s} p_{ij}^{(n)} = 1, \quad i = 1, 2, \ldots, s.$$

Веројатности на состојби во момент n

- Нека $p_j^{(n)} = P\{X_n = j\}$ е веројатноста дека во n-тиот момент, системот ќе се најде во состојба j. Тогаш

$$\sum_{j=1}^{s} p_j^{(n)} = 1, \quad n = 1, 2, \ldots$$

- Со тоа е определен векторот $\boldsymbol{p}^{(n)} = (p_1^{(n)}, p_2^{(n)}, \ldots, p_s^{(n)})$.
- За $p_j^{(n)}$ добиваме:

$$p_j^{(n)} = P\{X_n = j\} = \sum_{i=1}^{s} P\{X_{n-1} = i\} P\{X_n = j \mid X_{n-1} = i\}$$

$$= \sum_{i=1}^{s} p_i^{(n-1)} p_{ij}^{(n)},$$

или во матрична форма $\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(n-1)} \boldsymbol{P}^{(n)}$, $n = 1,2,\ldots$

Веројатности на состојби во момент n

- Нека $\boldsymbol{p}^{(0)} = (p_1^{(0)}, p_2^{(0)}, \ldots, p_s^{(0)})$ е вектор на веројатности на почетната состојба на системот. Со последователна примена на $\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(n-1)} \boldsymbol{P}^{(n)}$, за $n = 1,2,\ldots$, се добива:

$$\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(n-1)} \boldsymbol{P}^{(n)} = \boldsymbol{p}^{(n-2)} \boldsymbol{P}^{(n-1)} \boldsymbol{P}^{(n)} = \ldots = \boldsymbol{p}^{(0)} \boldsymbol{P}^{(1)} \boldsymbol{P}^{(2)} \ldots \boldsymbol{P}^{(n-1)} \boldsymbol{P}^{(n)} .$$

Дефиниција 2. Веригата на Марков се нарекува *хомогена* (или *временски инваријантна*), ако условните веројатности $p_{ij}^{(n)}$ не зависат од n, т.е. за $n = 1,2,\ldots$

$$P\{X_{n+1}= j | X_n = i\} = P\{X_2 = j | X_1 = i\} = p_{ij}, \text{ за сите } i, j \in R_X.$$

- Веројатностите p_{ij} се нарекуваат преодни веројатности од состојба i во состојба j за еден чекор.
- Соодветната матрица се означува со $\boldsymbol{P} = [p_{ij}]$. Во овој случај, $\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(0)} \boldsymbol{P}^n$.

Преодни веројатности за n чекори

- Во хомогена верига на Марков, означуваме:

$$p_{ij}(n) = P\{X_n = j \mid X_0 = i\}.$$

- Тоа е веројатноста дека системот ќе помине од состојба i во состојба j за n чекори.

Равенства на Чепман-Колмогоров

$$p_{ij}(n+1) = P\{X_{n+1} = j \mid X_0 = i\}$$

$$= \sum_{k=1}^{s} P\{X_{n+1} = j, X_n = k \mid X_0 = i\}$$

$$= \sum_{k=1}^{s} P\{X_n = k \mid X_0 = i\} P\{X_{n+1} = j \mid X_n = k, X_0 = i\}$$

$$\overset{B.M.}{=} \sum_{k=1}^{s} P\{X_n = k \mid X_0 = i\} P\{X_{n+1} = j \mid X_n = k\}$$

$$= \sum_{k=1}^{s} p_{ik}(n) p_{kj}$$

Матрица на преодни веројатности за n чекори

- Согласно равенствата на Чепмен-Колмогоров за хомоген Марков процес, имаме:

$$p_{ij}(n+1) = \sum_{k=1}^{s} p_{ik}(n)\, p_{kj}, \quad i, j = 1, 2, \ldots, s, \;\; n = 1, 2, \ldots$$

- Овој систем равенства може да се запише во матрична форма

$$\boldsymbol{P}(n+1) = \boldsymbol{P}(n) \cdot \boldsymbol{P}, \text{ за секој } n = 1,2,\ldots$$

- Оттука, се добива дека $\boldsymbol{P}(n) = \boldsymbol{P}^n$.
- Според ова, секоја конечна хомогена верига на Марков е определена, ако е познат векторот $p^{(0)}$ на веројатности на почетната состојба и матрицата $\boldsymbol{P} = [p_{ij}]$ на преодни веројатности за еден чекор.

n-димензионална распределба

- Нека $\{X_i \mid i = 1,2,\dots\}$ е хомогена верига на Марков. Тогаш за распределбата на случајниот вектор $(X_0, X_1,\dots,X_n)$ се добива:

$$p(x_0, x_1,\dots,x_n) = P\{X_0 = x_0, X_1 = x_1,\dots, X_n = x_n\}$$

$$= P\{X_0 = x_0\}\cdot P\{X_1 = x_1 \mid X_0 = x_0\}\cdot P\{X_2 = x_2 \mid X_1 = x_2, X_0 = x_0\}\cdot\dots$$

$$\dots\cdot P\{X_n = x_n \mid X_{n-1} = x_{n-1},\dots, X_0 = x_0\}$$

$$= P\{X_0 = x_0\}\cdot P\{X_1 = x_1 \mid X_0 = x_0\}\cdot P\{X_2 = x_2 \mid X_1 = x_1\}\cdot\dots\cdot P\{X_n = x_n \mid X_{n-1} = x_{n-1}\}$$

$$= p_{x_0}\cdot p_{x_0 x_1}\cdot p_{x_1 x_2}\cdot\dots\cdot p_{x_{n-1} x_n}.$$

- Ако е можно со позитивна веројатност од произволна состојба на веригата да се стигне во која било друга состојба со конечен број чекори, тогаш велиме дека веригата е иредуцибилна.

Стационарна распределба

- Случаен процес е (строго) стационарен, ако заедничката распределба на било кое подмножество случајни променливи е инваријантна во однос на транслација на временскиот индекс, т.е. во дискретен случај

$$P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} = P\{X_{1+l} = x_1, X_{2+l} = x_2, \ldots, X_{n+l} = x_n\}.$$

- Може да се покаже точноста на следната теорема:

  Теорема 1. Конечна хомогена верига на Марков е строго стационарен процес, акко

$$\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(0)},$$

  за секој $n = 1, 2, \ldots$

- Значи, верига на Марков е стационарен процес, ако почетната состојба е распределена согласно стационарната распределба.

Определување на стационарна распределба

- Бидејќи $\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(n-1)}\boldsymbol{P}$, означувајќи $\boldsymbol{p}^{(n)} = \boldsymbol{p}^{*}$, за секој $n = 1,2,\ldots$, равенството добива облик:

$$\boldsymbol{p}^{*} = \boldsymbol{p}^{*}\,\boldsymbol{P}.$$

- Според Теорема 1, конечна хомогена верига на Марков е стационарна, ако векторот $\boldsymbol{p}^{(0)} = (p_1^{(0)}, p_2^{(0)},\ldots, p_s^{(0)})$ на веројатности на почетната состојба е решение на претходната равенка.
- Таа равенка е, всушност, хомоген систем од s линеарни равенки:

$$p_j^{*} = \sum_{k=1}^{s} p_k^{*} p_{kj}, \qquad j = 1,\ldots,s.$$

- За да решението на системот биде распределба на веројатност, мора да биде задоволен условот: $\sum_{j=1}^{s} p_j^{*} = 1.$

Стационарна распределба

Дефиниција. Секоја распределба $\mathbf{p}^* = (p_1^*, p_2^*, \dots, p_s^*)$ која го задоволува системот равенки

$$p_j^* = \sum_{i=1}^{s} p_i^* p_{ij}$$

се нарекува *стационарна распределба на веригата*, а веројатностите p_j^* се нарекуваат *стационарни* или *финални веројатности.*

- Да воочиме дека горниот систем линеарни равенки може да има и повеќе решенија, т.е. да постојат повеќе стационарни распределби.

Регуларност

Дефиниција. Една матрица $\boldsymbol{P}$ се нарекува *регуларна*, ако постои природен број k, така што сите елементи во матрицата $\boldsymbol{P}^k$ се строго позитивни.

Теорема. Ако матрицата на преодни веројатности $\boldsymbol{P}$ на една верига на Марков е *регуларна*, тогаш постои единствена стационарна распределба.

Пример 1

- Се разгледува верига на Марков со две состојби, зададена со матрица на преодни веројатности:

$$\boldsymbol{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}.$$

- Нека стационарната распределба е определена со вектор $p^* = (p_1^*, p_2^*)$, каде p_i^* е стационарната веројатност на i-тата состојба, i =1,2. За определување на овие веројатности се добива систем во матрична форма:

$$\begin{bmatrix} p_1^* & p_2^* \end{bmatrix} = \begin{bmatrix} p_1^* & p_2^* \end{bmatrix} \cdot \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix},$$

Пример 1: продолжение

- или во развиена форма:

$$\begin{cases} \alpha p_1^* = \beta p_2^* \\ p_1^* + p_2^* = 1 \end{cases}.$$

- Решенија на системот се:

$$p_1^* = \frac{\beta}{\alpha + \beta}, \qquad p_2^* = \frac{\alpha}{\alpha + \beta}.$$

- Ако веригата на Марков има почетна состојба распределена согласно стационарната распределба, тогаш резултантниот процес е стационарен.

Пример 1: продолжение

- Ентропијата на состојбата X_n во момент n е:

$$H(X_n) = H\left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right).$$

- Но, ова не е рата со која ентропијата расте за $H(X_1, X_2, \ldots, X_n)$. Зависноста помеѓу X_i ќе има свое влијание.

Рата на ентропија

- Ако е дадена низа од n случајни променливи, тогаш се поставува прашањето:
  - Како се менува ентропијата со растење на n?
- Затоа се дефинира рата на ентропија како рата (брзина) на менување.
- Дефинираме две величини:

$$H_n = \frac{1}{n} H(X_1, X_2, \ldots, X_n)$$

$$h_n = H(X_n \mid X_{n-1}, X_{n-2}, \ldots, X_1)$$

- Всушност, H_n може да се толкува како просечна ентропија по буква во n-члена порака, а h_n како условна ентропија на n-тиот симбол кога се познати претходните $(n-1)$ симболи.

Рата на ентропија

Дефиниција 1. Рата на ентропија на случајниот процес $\{X_i \mid i=1,2,\ldots\}$ се дефинира со:

$$H = \lim_{n\to+\infty} H_n = \lim_{n\to+\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n),$$

ако лимесот постои.

Примери на случајни процеси и нивните рати на ентропија

- *Генератор на случајни броеви* (ГСБ). Нека ГСБ има m можни еднакввоверојатни излези, т.е. може да генерира било која буква од $\{1,2,\ldots, m\}$ со иста веројатност.
  - Тогаш ГСБ може да генерира m^n можни пораки со должина n, и сите ќе бидат генерирани со еднаква веројатност.
  - Оттука, $H(X_1, X_2,\ldots,X_n) = \log m^n = n \log m$
  - Просечната ентропија по буква ќе биде:

$$H_n = \log m.$$

Примери на случајни процеси и нивните рати на ентропија

- *Низа од независни и еднакво распределени случајни променливи*. Нека X_1, X_2,… се независни и еднакво распределени случајни променливи. Тогаш

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \ldots, X_1)$$

$$\overset{\text{нез.}}{=} \sum_{i=1}^{n} H(X_i)$$

$$\overset{\text{едн.}}{=} nH(X_1)$$

- Оттука, $H_n = H(X_1)$.

Примери на случајни процеси и нивните рати на ентропија

- *Низа од независни, но не еднакви случајни променливи*. Во овој случај,

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \ldots, X_1) \overset{\text{нез.}}{=} \sum_{i=1}^{n} H(X_i),$$

  но, $H(X_i)$ не се еднакви.
- Затоа може да се избере низа распределби за X_1, X_2,… такви што $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(X_i)$ не постои.
- На пример, нека низата се состои од бинарни случајни променливи, каде $p_i = P\{X_i = 1\}$ не се константи, туку се функции од i, зададени со:

Продолжение на претходниот пример

$$p_i = \begin{cases} 0,5, & \text{ако } 2k < \log\log i \le 2k+1 \\ 0, & \text{ако } 2k+1 < \log\log i \le 2k+2 \end{cases}, \quad \text{за} \quad k = 0,1,2,\ldots$$

- Во овој случај, може да се генерира експоненцијално долга низа каде $H(X_i) = 1$, а после неа да следува експоненцијално долга низа каде $H(X_i) = 0$.
- Значи, просечната вредност на $H(X_i)$ ќе осцилира помеѓу 0 и 1, па лимес од неа нема да постои.
- Затоа, H за овој процес не е дефинирана.

Друга мерка за ратата на ентропија

- Ќе дефинираме друга мерка за ратата на ентропија со

$$h = \lim_{n\to+\infty} h_n = \lim_{n\to+\infty} H(X_n \mid X_{n-1}, X_{n-2}, \ldots, X_1).$$

ако лимесот постои.

- Ќе покажеме дека $H = h$, но најпрво ќе покажеме кога лимес од $H(X_n \mid X_{n-1}, \ldots, X_1)$ постои.

Егзистенција на *h* за стационарен процес

Теорема 1. За стационарен случаен процес, низата условни ентропии $H(X_n \mid X_{n-1},\ldots, X_1)$ е опаѓачка и има лимес h.

Доказ:

$$H(X_{n+1} \mid X_n,\ldots,X_2,X_1) \overset{\text{условна ентропија}}{\leq} H(X_{n+1} \mid X_n,\ldots,X_2) \overset{\text{стац.}}{=} H(X_n \mid X_{n-1},\ldots,X_1)$$

Оттука, низата $H(X_n \mid X_{n-1},\ldots, X_1)$ е опаѓачка низа од ненегативни броеви, т.е. опаѓачка и ограничена од долу со 0, па таа е конвергентна, т.е. h постои.

Средина на Cesaro

Теорема 2. (Средина на Cesaro) Ако $a_i \to a$ и $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$ тогаш и $b_n \to a$.

Доказ: Бидејќи $a_i \to a$, за секој $\varepsilon > 0$, постои природен број $N(\varepsilon)$, така што за секој $n \geq N(\varepsilon)$, важи $|a_n - a| \leq \varepsilon$. Оттука,

$$|b_n - a| = \left|\frac{1}{n}\sum_{i=1}^{n}(a_i - a)\right| \leq \frac{1}{n}\sum_{i=1}^{n}|a_i - a| = \frac{1}{n}\sum_{i=1}^{N(\varepsilon)}|a_i - a| + \frac{1}{n}\sum_{i=N(\varepsilon)+1}^{n}|a_i - a|$$

$$\leq \frac{1}{n}\sum_{i=1}^{N(\varepsilon)}|a_i - a| + \frac{n - N(\varepsilon)}{n}\varepsilon \leq \frac{1}{n}\sum_{i=1}^{N(\varepsilon)}|a_i - a| + \varepsilon.$$

Бидејќи, првиот член тежи кон 0, кога $n \to \infty$, може да се направи $|b_n - a| \leq 2\varepsilon$, со избор на доволно големо n. Така, $b_n \to a$.

За стационарен процес $H = h$

Теорема 3. За стационарен случаен процес, лимесите за H и h постојат и се еднакви, т.е. $H = h$.

Доказ: Според верижното правило:

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} = \frac{1}{n}\sum_{i=1}^{n} H(X_i \mid X_{i-1}, \ldots, X_1),$$

т.е. ратата на ентропија е средина од условните ентропии. Но, условните ентропии тежат кон h, кога $n \to \infty$. Според Теорема 2 (Cesaro), добиваме:

$$H = \lim_{n\to\infty} \frac{H(X_1, X_2, \ldots, X_n)}{n} = \lim_{n\to\infty} H(X_n \mid X_{n-1}, \ldots, X_1) = h.$$

Рата на ентропија за стационарна верига на Марков

- За стационарна верига на Марков, ратата на ентропија е дадена со:

$$H = h = \lim_{n\to\infty} H(X_n \mid X_{n-1},\dots,X_1) = \lim_{n\to\infty} H(X_n \mid X_{n-1}) = H(X_2 \mid X_1),$$

каде условната ентропија е пресметана со користење на дадената стационарна распределба. Овој резултат е изразен во следната

Теорема 4. Нека $\{X_i \mid i = 1,2,\dots\}$ е стационарна верига на Марков со стационарна распределба $\boldsymbol{p}^*$. Тогаш ратата на ентропија е:

$$H = -\sum_i p_i^* \sum_j p_{ij} \log p_{ij}.$$

Доказ:

$$H = H(X_2 \mid X_1) = \sum_i p_i^* H(X_2 \mid X_1 = i) = -\sum_i p_i^* \sum_j p_{ij} \log p_{ij}.$$

Пример

- За верига на Марков со две состојби зададена со матрица на преодни веројатности:

$$\boldsymbol{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix},$$

утврдивме дека стационарната распределба е

$$p_1^* = \frac{\beta}{\alpha+\beta}, \qquad p_2^* = \frac{\alpha}{\alpha+\beta}.$$

Оттука, ратата на ентропија е

$$H = \frac{\beta}{\alpha+\beta} H(1-\alpha, \alpha) + \frac{\alpha}{\alpha+\beta} H(\beta, 1-\beta).$$

Забелешка

- Ако веригата на Марков е иредуцибилна и непериодична, тогаш таа има единствена стационарна распределба и која и да е почетната распределба, таа ќе тежи кон стационарната кога $n \to +\infty$.
- Во тој случај, иако почетната распределба не е стационарна, ратата на ентропија се пресметува со истите формули како и претходно.