Лекция 5

Классификация состояний марковской цепи

Рассмотрим марковскую цепь $\{\xi_n,\ n \in \mathbb{N}_0\}$ с множеством состояний S. Познакомимся с *классификацией состояний* марковской цепи, предложенной Колмогоровым. Эта классификация зависит только от переходных вероятностей и не зависит от начального распределения марковской цепи, поэтому будем считать, что для одной и той же марковской цепи возможны различные начальные распределения. В случае, когда марковская цепь выходит из состояния i (это означает, что $\xi_0 = i$), будем использовать символы $\mathbf{P}^{(i)}$ и $\mathbf{E}^{(i)}$ вместо обычных $\mathbf{P}$ и $\mathbf{E}$.

Важная особенность марковской цепи состоит в наличии *вложенного* в эту цепь процесса восстановления. Пусть цепь выходит из состояния i. Рассмотрим моменты $T_1, T_2, \ldots$ первого, второго и т.д. возвращения марковской цепи в состояние i, т.е.

$$T_1 = \min\{n > 0 : \xi_n = i\},\ T_2 = \min\{n > T_1 : \xi_n = i\},\ \ldots .$$

(если при некотором $k \in \mathbb{N}$ случайная величина T_k бесконечна, то бесконечны и случайные величины $T_{k+1}, T_{k+2}, \ldots$). Рассмотрим последовательность независимых случайных величин $\tau_1, \tau_2, \ldots$, распределенных, как T_1.

Теорема 1. *Конечномерные распределения случайных последовательностей* $(T_1, T_2, T_3, \ldots)$ и $(\tau_1, \tau_1 + \tau_2, \tau_1 + \tau_2 + \tau_3, \ldots)$ *совпадают.*

Доказательство. Достаточно показать, что для произвольного $k \in \mathbb{N}$

$$(T_1, T_2, \ldots, T_k) \overset{d}{=} (\tau_1, \tau_1 + \tau_2, \ldots, \tau_1 + \ldots + \tau_k)\,. \tag{1}$$

Покажем справедливость соотношения (1) при $k = 2$. При $l_1, l_2 \in \mathbb{N}$

$$\mathbf{P}^{(i)}(T_1 = l_1, T_2 = l_1 + l_2) =$$
$$= \mathbf{P}^{(i)}\left(\xi_1 \neq i, \ldots, \xi_{l_1-1} \neq i; \xi_{l_1} = i; \xi_{l_1+1} \neq i, \ldots, \xi_{l_1+l_2-1} \neq i; \xi_{l_1+l_2} = i\right) =$$
$$= \mathbf{P}^{(i)}\left(\xi_1 \neq i, \ldots, \xi_{l_1-1} \neq i; \xi_{l_1} = i\right) \times$$
$$\times \mathbf{P}^{(i)}\left(\xi_{l_1+1} \neq i, \ldots, \xi_{l_1+l_2-1} \neq i; \xi_{l_1+l_2} = i \;\middle|\; \xi_1 \neq i, \ldots, \xi_{l_1-1} \neq i; \xi_{l_1} = i;\right) =$$
$$= \mathbf{P}^{(i)}\left(\xi_1 \neq i, \ldots, \xi_{l_1-1} \neq i; \xi_{l_1} = i\right) \times$$
$$\times \mathbf{P}\left(\xi_{l_1+1} \neq i, \ldots, \xi_{l_1+l_2-1} \neq i; \xi_{l_1+l_2} = i \;\middle|\; \xi_{l_1} = i\right)$$

(последнее равенство объясняется теоремой 1 лекции 4). Заметим, что предпоследняя вероятность по определению T_1 равна $\mathbf{P}^{(i)}(T_1 = l_1)$, а последняя вероятность в силу леммы 1 лекции 4 равна

$$\mathbf{P}\left(\xi_1 \neq i, \ldots, \xi_{l_2-1} \neq i; \xi_{l_2} = i \;\middle|\; \xi_0 = i\right) = \mathbf{P}^{(i)}(T_1 = l_2)\,.$$

Итак, при $l_1, l_2 \in \mathbb{N}$

$$\mathbf{P}^{(i)}(T_1 = l_1, T_2 = l_1 + l_2) = \mathbf{P}^{(i)}(T_1 = l_1)\,\mathbf{P}^{(i)}(T_1 = l_2) =$$

$$= \mathbf{P}(\tau_1 = l_1)\,\mathbf{P}(\tau_2 = l_2) = \mathbf{P}(\tau_1 = l_1, \tau_2 = l_2) =$$
$$= \mathbf{P}(\tau_1 = l_1, \tau_1 + \tau_2 = l_1 + l_2)\,. \qquad (2)$$

Покажем, что соотношение (2) справедливо при $l_1 \in \mathbb{N}$, $l_2 = +\infty$:

$$\mathbf{P}^{(i)}(T_1 = l_1, T_2 = +\infty) = \mathbf{P}^{(i)}(T_1 = l_1) - \mathbf{P}^{(i)}(T_1 = l_1, T_2 < +\infty) =$$

$$= \mathbf{P}^{(i)}(T_1 = l_1) - \sum_{l_2=1}^{+\infty} \mathbf{P}^{(i)}(T_1 = l_1, T_2 = l_1 + l_2) =$$

$$= \mathbf{P}(\tau_1 = l_1) - \sum_{l_2=1}^{+\infty} \mathbf{P}(\tau_1 = l_1, \tau_1 + \tau_2 = l_1 + l_2) =$$

$$= \mathbf{P}(\tau_1 = l_1) - \mathbf{P}(\tau_1 = l_1, \tau_1 + \tau_2 < +\infty) = \mathbf{P}(\tau_1 = l_1, \tau_1 + \tau_2 = +\infty)\,.$$

Наконец,

$$\mathbf{P}^{(i)}(T_1 = +\infty, T_2 = +\infty) = \mathbf{P}^{(i)}(T_1 = +\infty) =$$
$$= \mathbf{P}(\tau_1 = +\infty) = \mathbf{P}(\tau_1 = +\infty, \tau_1 + \tau_2 = +\infty)\,.$$

Таким образом, соотношение (1) при $k = 2$ доказано. При $k > 2$ соотношение (1) доказывается по индукции.

Предположим, что $\mathbf{P}^{(i)}(T_1 < +\infty) = 1$, тогда в силу теоремы 1 случайные величины $T_1, T_2 - T_1, T_3 - T_2, \ldots$ конечны, одинаково распределены и независимы. Следовательно, последовательность $T_1, T_2, T_3, \ldots$ образует случайное блуждание. Поскольку $T_1 > 0$ п.н., то величины $T_1, T_2, T_3, \ldots$ можно воспринимать как моменты восстановления. Введем соответствующий процесс восстановления:

$$\nu(0) = 0, \qquad \nu(t) = \max\{k > 0 : T_k \le t\}\,, \quad t > 0.$$

Другими словами, $\nu(t)$ совпадает с числом возвращений марковской цепи в исходное состояние, которые произошли до момента t. Процесс $\{\nu(t)\,,\ t \ge 0\}$ можно определить и в случае, когда $\mathbf{P}^{(i)}(T_1 < +\infty) < 1$, при этом его называют *обрывающимся* процессом восстановления.

Именно указанный процесс $\{\nu(t)\,,\ t \ge 0\}$ является тем самым вложенным в марковскую цепь процессом восстановления, о котором говорилось в начале лекции.

На основе случайной величины T_1 осуществляется *классификация* исходного состояния i.

1) Состояние i называется *возвратным*, если $\mathbf{P}^{(i)}(T_1 < +\infty) = 1$; другими словами, i – возвратное состояние, если цепь возвращается в это состояние с вероятностью 1.

2) Состояние i называется *невозвратным*, если $\mathbf{P}^{(i)}(T_1 < +\infty) < 1$.

3) Возвратное состояние i называется *положительно возвратным*, если $\mathbf{E}^{(i)}T_1 < +\infty$; другими словами, i – положительно возвратное состояние, если математическое ожидание времени возвращения цепи в это состояние конечно.

4) Возвратное состояние i называется *нуль возвратным*, если $\mathbf{E}^{(i)}T_1 = +\infty$.

Поскольку случайная величина T_1 является целочисленной, то ее распределение является арифметическим. Пусть d – максимальный шаг этого распределения. Ясно, что d совпадает с НОД носителя распределения случайной величины T_1.

5) Возвратное состояние i называется *периодическим*, если $d > 1$; число d называется *периодом*; другими словами, i – периодическое состояние с периодом d, если возвращение в это состояние может происходить только в моменты времени, кратные d (и d – наибольшее из чисел, обладающих этим свойством).

6) Возвратное состояние i называется *непериодическим*, если $d = 1$.

Пусть i – возвратное состояние. Тогда, как уже сказано, $\{\nu(t), \ t \geq 0\}$ – процесс восстановления. Напомним, что функция

$$U(t) = 1 + \mathbf{E}^{(i)}\nu(t) = 1 + \sum_{k=1}^{\infty} \mathbf{P}^{(i)}(T_k \leq t), \quad t \geq 0,$$

называется функцией восстановления для процесса $\{\nu(t), \ t \geq 0\}$. Ее локальный аналог – это последовательность

$$u_n = \sum_{k=1}^{\infty} \mathbf{P}^{(i)}(T_k = n), \quad n \in \mathbb{N}. \tag{3}$$

Ясно, что u_n – вероятность того, что n является моментом восстановления, а в рассматриваемом случае – моментом возвращения в состояние i. Следовательно, при $n \in \mathbb{N}$

$$u_n = \mathbf{P}^{(i)}(\xi_n = i) = p_{ii}(n). \tag{4}$$

и при $N \in \mathbb{N}$

$$U(N) = 1 + \sum_{n=1}^{N} u_n = 1 + \sum_{n=1}^{N} p_{ii}(n). \tag{5}$$

Наша задача – дать критерии введенных ранее понятий в терминах переходных вероятностей.

Теорема 2 (критерий возвратности). *Состояние i марковской цепи $\{\xi_n\}$ является возвратным тогда и только тогда, когда ряд $\sum_{n=1}^{\infty} p_{ii}(n)$ расходится.*

Доказательство. Ввиду (5) ряд $\sum_{n=1}^{\infty} p_{ii}(n)$ расходится тогда и только тогда, когда $\lim_{N\to\infty} U(N) = +\infty$. Пусть цепь выходит из состояния i. Если i – возвратное состояние, то все случайные величины $T_1, T_2, \ldots$ конечны, поэтому

$$\lim_{N\to\infty} \nu(N) = +\infty,$$

где, напомним, $\{\nu(t), \ t \geq 0\}$ – процесс восстановления, построенный по последовательности $\{T_1, T_2, \ldots\}$. Следовательно, по теореме о монотонной сходимости

$$\lim_{N\to\infty} U(N) = 1 + \lim_{N\to\infty} \mathbf{E}^{(i)}\nu(N) = 1 + \mathbf{E}^{(i)} \lim_{N\to\infty} \nu(N) = +\infty.$$

Пусть теперь состояние i невозвратно. Тогда $\mathbf{P}^{(i)}\left(T_1=+\infty\right):=p>0$. Пусть случайное событие A_k, $k\in\mathbb{N}$, означает, что случайные величины $T_1,\ldots,T_{k-1}$ конечны, а $T_k=+\infty$. Положим $q=1-p$. Очевидно, что

$$\mathbf{P}^{(i)}\left(A_k\right)=\mathbf{P}\left(\tau_1<+\infty,\ldots,\tau_{k-1}<+\infty,\tau_k=+\infty\right)=q^{k-1}p,\qquad k\in\mathbb{N}.$$

События $A_1,A_2,\ldots$ образуют полную группу событий, поскольку они попарно несовместны и $\sum_{k=1}^{\infty}\mathbf{P}\left(A_k\right)=\sum_{k=1}^{\infty}q^{k-1}p=1$. Если произошло событие A_k, то $\lim_{N\to\infty}\nu\left(N\right)=k-1$. Следовательно,

$$\mathbf{E}^{(i)}\lim_{N\to\infty}\nu\left(N\right)=\sum_{k=1}^{\infty}\left(k-1\right)\mathbf{P}^{(i)}\left(A_k\right)=\sum_{k=1}^{\infty}\left(k-1\right)q^{k-1}p<+\infty.$$

Откуда по теореме о монотонной сходимости

$$\lim_{N\to\infty}U\left(N\right)=1+\lim_{N\to\infty}\mathbf{E}^{(i)}\nu\left(N\right)=1+\mathbf{E}^{(i)}\lim_{N\to\infty}\nu\left(N\right)<+\infty.$$

Теорема доказана.

Замечание 1. Ясно, что если исходное состояние i марковской цепи является возвратным, то число возвращений цепи в это состояние бесконечно. Из доказательства теоремы 2 следует, что если исходное состояние i марковской цепи является невозвратным, то число возвращений цепи в это состояние будет конечно, поэтому соответствующий вложенный процесс восстановления и называют обрывающимся.

Теорема 3. *Возвратное состояние i является положительно возвратным тогда и только тогда, когда* $\limsup_{n\to\infty}p_{ii}\left(n\right)>0$.

Доказательство. Пусть цепь выходит из состояния i. Положим $a=\mathbf{E}^{(i)}T_1$. По локальной теореме восстановления существует

$$\lim_{n\to\infty\ (n\in d\cdot\mathbb{N})}u_n=\pi_i,$$

где $\pi_i=d/a$, где, напомним, d – максимальный шаг распределения случайной величины T_1. Следовательно, по формуле (4)

$$\lim_{n\to\infty\ (n\in d\cdot\mathbb{N})}p_{ii}\left(n\right)=\pi_i.$$

Если i – положительно возвратное состояние, то $a<+\infty$ и, значит, $\pi_i>0$. Следовательно, $\limsup_{n\to\infty}p_{ii}\left(n\right)>0$.

Если i – нуль возвратное состояние, то $a=+\infty$ и, значит, $\pi_i=0$. Кроме того, возвращение марковской цепи в состояние i может происходить только в моменты времени, кратные d, поэтому $p_{ii}\left(n\right)=0$ при $n\notin d\cdot\mathbb{N}$. Следовательно, $\limsup_{n\to\infty}p_{ii}\left(n\right)=0$. Теорема доказана.

Теорема 4. *Возвратное состояние i является периодическим с периодом $d>1$ тогда и только тогда, когда*

$$\text{НОД}\left\{n>0:p_{ii}\left(n\right)>0\right\}=d. \tag{6}$$

Доказательство. Покажем, что соотношение (6) справедливо при любом значении d (это и будет означать справедливость теоремы). Пусть цепь выходит из состояния i. Ввиду (3) и (4)

$$p_{ii}(n) = \sum_{k=1}^{\infty} \mathbf{P}^{(i)}(T_k = n).$$

Таким образом, $p_{ii}(n) > 0$ тогда и только тогда, когда $\mathbf{P}^{(i)}(T_k = n) > 0$ при некотором $k \in \mathbb{N}$. Пусть D_k – носитель распределения случайной величины T_k. Из сказанного следует, что

$$\{n \in \mathbb{N} : p_{ii}(n) > 0\} = D, \tag{7}$$

где $D = \bigcup_{k=1}^{\infty} D_k$.

Так как случайная величина T_1 имеет арифметическое распределение с максимальным шагом d, то при каждом $k \in \mathbb{N}$ случайная величина T_k также имеет арифметическое распределение с шагом d, следовательно, $D_k \subset \{0, d, 2d, \ldots\}$ и, значит, $D \subset \{0, d, 2d, \ldots\}$, поэтому

$$\text{НОД}(D) \geq d.$$

С другой стороны,

$$\text{НОД}(D) \leq \text{НОД}(D_1) = d.$$

Итак,

$$\text{НОД}(D) = d. \tag{8}$$

Из (7) и (8) следует соотношение (6). Теорема доказана.

Определение 1. Два состояния $i, j \in S$ марковской цепи называются *сообщающимися*, если существуют такие $n_1, n_2 \in \mathbb{N}$, что $p_{ij}(n_1) > 0$ и $p_{ji}(n_2) > 0$.

Установим *свойство солидарности* сообщающихся состояний марковской цепи.

Теорема 5. *Пусть i и j – сообщающиеся состояния марковской цепи. Тогда, если i – возвратное (невозвратное) состояние, то и j – возвратное (невозвратное) состояние; если i – положительно (нуль) возвратное состояние, то и j – положительно (нуль) возвратное состояние; если i – непериодическое (периодическое с периодом d) состояние, то и j – непериодическое (периодическое с периодом d) состояние.*

Доказательство. Существуют такие $n_1, n_2 \in \mathbb{N}$, что $p_{ij}(n_1) > 0$ и $p_{ji}(n_2) > 0$. В силу уравнения Колмогорова-Чепмена при $n \in \mathbb{N}$

$$p_{ii}(n_1 + n_2 + n) \geq p_{ij}(n_1)\, p_{jj}(n)\, p_{ji}(n_2).$$

Аналогично,

$$p_{jj}(n_1 + n_2 + n) \geq p_{ji}(n_2)\, p_{ii}(n)\, p_{ij}(n_1).$$

Следовательно, при $n > n_1 + n_2$

$$c^{-1} p_{ii}(n_1 + n_2 + n) \geq p_{jj}(n) \geq c\, p_{ii}(n - n_1 - n_2),$$

где постоянная $c := p_{ij}(n_1)\, p_{ji}(n_2)$ положительна и не зависит от n. Но это означает, что оба ряда $\sum_{n=1}^{\infty} p_{ii}(n)$ и $\sum_{n=1}^{\infty} p_{jj}(n)$ расходятся или сходятся одновременно. Значит, состояния i, j возвратны или невозвратны одновременно. Аналогично проверяются остальные утверждения теоремы.

Следующий результат дополняет теорему солидарности. Пусть марковская цепь $\{\xi_n,\ n \in \mathbb{N}_0\}$ стартует из состояния i. Обозначим $T_{i,j}$ *момент первого достижения* состояния j этой цепью.

Лемма 1. *Пусть i и j – сообщающиеся состояния. Если i – возвратное состояние, то п.н. $T_{i,j} < +\infty$.*

Доказательство. Пусть (как и прежде) $T_1, T_2, \ldots$ – моменты первого, второго и т.д. возвращения цепи в состояние i. Участки марковской цепи между двумя последовательными моментами возвращения в состояние i назовем *циклами* (*первый* цикл – между моментами $T_0 = 0$ и T_1, *второй* цикл – между моментами T_1 и T_2 и т.д.). Ясно, что циклы независимы и одинаково распределены.

Введем случайные события $B_k = \{\exists n : T_{k-1} < n \leq T_k,\ \xi_n = j\}$, где $k \in \mathbb{N}$. В силу сказанного эти события независимы и равновероятны. Поскольку i, j – сообщающиеся состояния, то $\mathbf{P}^{(i)}(B_1) := p > 0$ (докажите сами).

Введем случайные события $A_k = \{T_{k-1} < T_{i,j} \leq T_k\}$, где $k \in \mathbb{N}$. Ясно, что

$$A_k = \overline{B_1} \ldots \overline{B_{k-1}} B_k,$$

поэтому $\mathbf{P}^{(i)}(A_k) = q^{k-1}p$ при $k \in \mathbb{N}$, где $q = 1 - p$. События $A_1, A_2, \ldots$ образуют полную группу событий, поскольку они попарно несовместны и

$$\sum_{k=1}^{\infty} \mathbf{P}^{(i)}(A_k) = \sum_{k=1}^{\infty} q^{k-1}p = 1.$$

Если произошло событие A_k, то $T_{i,j} \leq T_k < +\infty$. Итак, $\mathbf{P}^{(i)}(T_{i,j} < +\infty) = 1$. Лемма доказана.