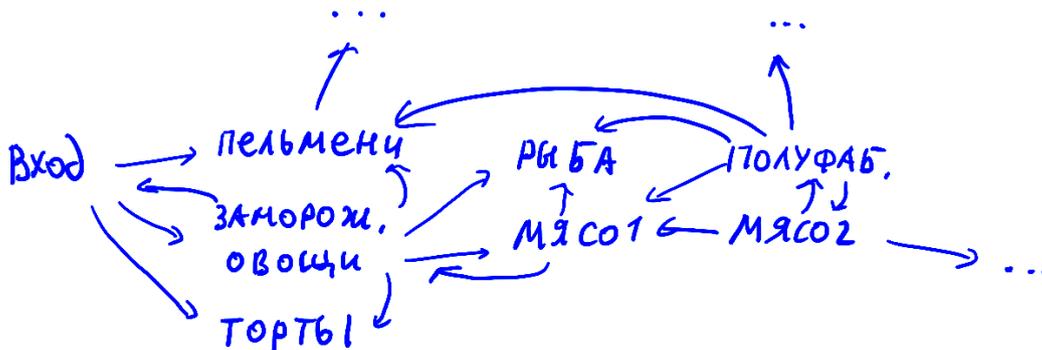


Цепи Маркова

В этой лекции мы строим ещё один простой процесс с далеко идущими последствиями. Многие с ним знакомятся даже не зная слова случайный процесс, но именно взгляд с более расширенного фундамента случайных процессов позволяет вывести понимание и приложения этой модели на принципиально новый уровень.

4.1 Конструкция цепи Маркова

Предположим, у нас есть датасет положений клиента в магазине во времени. Каждое такое наблюдение – это набор дискретных меток x_{t_1}, \dots, x_{t_k} , каждая из которых показывает ID отдела, где находился клиент в момент t_i . Мы хотели бы построить математическую модель таких данных. Это помогло бы исследовательским целям, включая в частности (но не только) тестирование новых инновационных методик, основанных на искусственном интеллекте. А ещё подобные модели – очень дешёвый способ провести предварительные тесты, связанные с АБ-тестированием: в онлайн-тестах данные добываются легко, но в оффлайне – это отдельная боль, время и риски.



Формально нас интересует $(X_t)_{t \in \mathbb{Z}_+}$ – случайный процесс на вероятностном пространстве (Ω, \mathcal{F}, P) , описывающий поведение клиента магазина. На первом этапе (но позже мы немного обобщим) мы не учитываем время, проведённое клиентом в каждом отделе, начиная с очень простого подхода. Нам, таким образом, интересует прежде всего то, как именно клиент перемещается между отделами. В данном случае моменты времени t принимают дискретные значения в $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$. Пусть $\Xi = \{1, 2, \dots, r\}$ – пространство состояний (множество значений, которые могут принимать случайные величины X_t). Т.к. Ξ – конечное пространство, логичнее всего ввести на нем наибольшую σ -алгебру $\mathcal{G} = 2^\Xi$, состоящую из всех возможных подмножеств Ξ .

Чтобы определить случайный процесс X , нам нужно задать согласованное (т.е. удовлетворяющее теореме Колмогорова) семейство совместных распределений

$$F(t_1, \dots, t_n; A_1, \dots, A_n) = P(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n), \quad A_1, \dots, A_n \in \mathcal{G}.$$

Мы будем делать это для моментов времени t_0, t_1, \dots, t_n и для одноэлементных множеств вида $A = \{a\}$. Задаём $\mu \in [0, 1]^{1 \times r}$ - это вектор-строка, описывающая распределение стартовой точки, а также $p_{ij}(t)$ - вероятности перехода из состояния i в состояние j в момент времени t .

$$\begin{aligned} \mu &= (P(X_{t_0} = 1), \dots, P(X_{t_0} = r)) \in [0, 1]^{1 \times r}, \\ p_{ij}(t) &= P(X_{t+1} = j | X_t = i). \end{aligned}$$

Поскольку p_{ij} - честные условные вероятности, сумма по j должна быть равна 1 для каждого i, t . Далее начнём с того, что определим

$$P(X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = \mu_{i_0} p_{i_0 i_1}(0) p_{i_1 i_2}(1) \dots p_{i_{n-1} i_n}(n-1).$$

Позади этой конструкции находится достаточно простая модель:

1. Старт $X_0 \sim \mu$;
2. Далее на каждом шаге t переход из состояния i в j происходит с вероятностью $p_{ij}(t)$.

Чтобы использовать теорему Колмогорова о существовании, нужно обобщить для произвольных моментов времени. Сначала обобщим до последовательных моментов времени $t_0, t_0 + 1, \dots, t_0 + n$:

$$\begin{aligned} &P(X_{t_0} = i_{t_0}, X_{t_0+1} = i_{t_0+1}, X_{t_0+2} = i_{t_0+2}, \dots, X_{t_0+n} = i_{t_0+n}) = \\ &= \sum_{i_0, \dots, i_{t_0-1} \in \Xi} \mu_{i_0} p_{i_0 i_1}(0) p_{i_1 i_2}(1) \dots p_{i_{t_0-1} i_{t_0}}(t_0-1) p_{i_{t_0} i_{t_0+1}}(t_0) \dots p_{i_{t_0+n-1} i_{t_0+n}}(t_0+n-1). \end{aligned}$$

Чтобы обобщить до совсем произвольных моментов времени $t_1 < \dots < t_n$, нужно тоже рассмотреть вероятность для моментов $t'_0 = 0, \dots, t'_{n+k} = t_n$ и просуммировать по всем $i'_{t'_j}$ таким, что $t'_j \notin \{t_1, \dots, t_n\}$.

Для неупорядоченных моментов времени мы применяем такой же приём, как и с Винеровским процессом. Для любой перестановки $\sigma \in S_n$ отсортированного набора t_1, \dots, t_n

$$F(t_{\sigma(1)}, \dots, t_{\sigma(n)}; A_{\sigma(1)}, \dots, A_{\sigma(n)}) = F(t_1, \dots, t_n; A_1, \dots, A_n),$$

автоматически удовлетворяя перестановочное условие согласованности. Проверим второе условие согласованности:

$$\begin{aligned} &P(X_{t_1} = i_{t_1}, \dots, X_{t_n} = i_{t_n}, X_{t_{n+1}} \in \Xi) = \\ &= \sum_{i_{t_{n+1}} \in \Xi} P(X_{t_1} = i_{t_1}, \dots, X_{t_n} = i_{t_n}, X_{t_{n+1}} = i) = \\ &= \sum_{i'_{t'_1} \dots i'_{t'_{n+k}} i_{t_{n+1}}} P(X_{t_1} = i_{t_1}, \dots, X_{t_n} = i_{t_n}, X_{t_{n+1}} = i_{t_{n+1}}). \end{aligned}$$

Поскольку $p_{ij}(t)$ по конструкции честная условная вероятность и последний индекс встречается только в последнем множителе, они пропадут. Таким же образом пропадут промежуточные от $t_n + 1$ до $t_{n+1} - 1$ (если они есть).

Вывод: по теореме Колмогорова о существовании существует вероятностное пространство $(\Omega, \mathcal{F}, P_0)$ и случайный процесс X на нём с заданным семейством конечномерных распределений. Такой процесс называется *цепью Маркова* в дискретном времени. Множество значений Ξ обычно в терминологии называют множеством *состояний*.

4.2 Стохастические матрицы

Это всё выглядит достаточно громоздко, но мы сделали полезную вещь: показали, что есть бесконечная последовательность случайных величин с заданной нами структурой взаимосвязей. Иными словами, случайный процесс.

Для практических целей удобно произвести некоторое упрощение. Переходные вероятности $p_{ij}(t)$ удобно записать в виде матрицы $P(t) = (p_{ij}(t))_{i,j=1}^r \in \mathbb{R}^{r \times r}$. Однако не любая матрица задает семейство переходных вероятностей. Для этого нужно, чтобы её элементы $p_{ij}(t)$ в самом деле были условными вероятностями $P(X_{t+1} = j | X_t = i)$. Это значит:

1. Элементы матрицы должны быть неотрицательны:

$$p_{ij}(t) = P(X_{t+1} = j | X_t = i) \geq 0,$$

2. Сумма элементов в любой строке должна быть равна единице:

$$\sum_{j=1}^r p_{ij}(t) = P(X_{t+1} \in \Xi | X_t = i) = 1.$$

Матрицы, удовлетворяющие выше перечисленным свойствам, называются *стохастическими*.

Определение 4.1. По теореме Колмогорова начальное распределение μ и набор стохастических матриц $(P(t))_{t \in \mathbb{Z}_{\geq}}$ задают цепь Маркова, определяемую с помощью согласованного семейства совместных распределений.

Оказывается, что практически всё, что мы хотим знать про вероятности, связанные с цепью Маркова, мы можем получить из переходной матрицы и стартовых вероятностей.

Упражнение 4.1. Докажите, что если $Q \in \mathbb{R}^{r \times r}$ - стохастическая матрица, $\mu = (\mu_1, \dots, \mu_r)$ - распределение вероятностей, то μQ - также распределение вероятностей.

Упражнение 4.2. Докажите, что если $Q', Q'' \in \mathbb{R}^{r \times r}$ - стохастические матрицы, то $Q := Q' Q''$ - также стохастическая матрица.

Можем получить базовые свойства такого процесса, анализируя наше изначальное задание через конечномерные распределения. Как и в случае Винеровского процесса с независимыми приращениями, первое свойство получается, если вы присмотритесь внимательно к тому, как именно мы строили меру.

Теорема 4.1. *Для цепи Маркова, порожденной начальным распределением μ и стохастическими матрицами $P(t)$, выполняются следующие равенства:*

$$P(X_n = j | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i) = P(X_n = j | X_{n-1} = i) = p_{ij}(n-1), \quad (4.1)$$

$$P(X_{t+s} = j | X_t = i) = (P(t) \dots P(t+s-1))_{ij}, \quad (4.2)$$

$$P(X_t = j) = (\mu P(0) P(1) \dots P(t-1))_j. \quad (4.3)$$

▷

Первое свойство не так очевидно как кажется, но, как иногда говорят, в него легко поверить. Давайте докажем 4.1. Для этого запишем определение условной вероятности $P(A|B) = \frac{P(A \cap B)}{P(B)}$, а после этого используем определение совместного распределения сечений цепи Маркова.

$$\begin{aligned} P(X_k = j | X_0 = i_0, X_1 = i_1, \dots, X_{k-1} = i_{k-1}) &= \frac{P(X_0 = i_0, X_1 = i_1, \dots, X_{k-1} = i_{k-1}, X_k = j)}{P(X_0 = i_0, X_1 = i_1, \dots, X_{k-1} = i_{k-1})} = \\ &= \frac{\mu_{i_0} p_{i_0 i_1}(1) p_{i_1 i_2}(2) \dots p_{i_{n-2} i_{n-1}}(n-1) p_{i_{n-1} i_n}(n)}{\mu_{i_0} p_{i_0 i_1}(1) p_{i_1 i_2}(2) \dots p_{i_{n-2} i_{n-1}}(n-1)} = p_{i_{n-1} i_n}(n). \end{aligned}$$

Полученный факт в том числе называют ещё *Марковским свойством*.

Для доказательства формулы 4.2 снова используем определение условной вероятности, после этого применим формулу полной вероятности. А в конце используем матричную запись для стохастических матриц $P(t)$, порождающих цепь Маркова.

$$\begin{aligned} P(X_{t+s} = j | X_t = i) &= \frac{P(X_{t+s} = j, X_t = i)}{P(X_t = i)} = \\ &= \frac{\sum_{i_0, i_1, \dots, i_{t-1}, i_{t+1}, \dots, i_{t+s-1} \in \Xi} P(X_0 = i_0, X_1 = i_1, \dots, X_{t-1} = i_{t-1}, X_t = i, \dots, X_{t+s-1} = i_{t+s-1}, X_{t+s} = j)}{\sum_{i_0, i_1, \dots, i_{t-1} \in \Xi} P(X_0 = i_0, X_1 = i_1, \dots, X_{t-1} = i_{t-1}, X_t = i)} = \\ &= \frac{\sum_{i_0, i_1, \dots, i_{t-1} \in \Xi} \mu_{i_0} p_{i_0 i_1}(0) \dots p_{i_{t-1} i}(t-1) \sum_{i_{t+1}, \dots, i_{t+s-1} \in \Xi} p_{i i_{t+1}}(t) \dots p_{i_{t+s-1} j}(t+s-1)}{\sum_{i_0, i_1, \dots, i_{t-1} \in \Xi} \mu_{i_0} p_{i_0 i_1}(1) \dots p_{i_{t-1} i}(t-1)} = \end{aligned}$$

$$\begin{aligned}
&= \sum_{i_{t+1}, \dots, i_{t+s-1} \in \Xi} p_{i_{t+1}i_{t+1}}(t+1)p_{i_{t+1}i_{t+2}}(t+2)\dots p_{i_{t+s-2}i_{t+s-1}}(t+s-1)p_{i_{t+s-1}j}(t+s-1) = \\
&= (P(t)P(t+1)\dots P(t+s-1))_{ij}.
\end{aligned}$$

Аналогично доказывается утверждение 4.3: используем формулу полной вероятности, а после этого используем определение совместного распределения цепи Маркова, а также матричную запись для стохастических матриц $P(t)$.

$$\begin{aligned}
P(X_t = j) &= \sum_{i_0, i_1, \dots, i_{t-1} \in \Xi} P(X_0 = i_0, X_1 = i_1, \dots, X_{t-1} = i_{t-1}, X_t = j) = \\
&= \sum_{i_0, i_1, \dots, i_{t-1} \in \Xi} \mu_{i_0} p_{i_0 i_1}(1) \dots p_{i_{t-2} i_{t-1}}(t-1) p_{i_{t-1} j}(t) = (\mu P(1) P(2) \dots P(t))_j.
\end{aligned}$$

□

4.3 Однородная цепь Маркова

Отдельного внимания заслуживает случай, когда переходная матрица не зависит от времени.

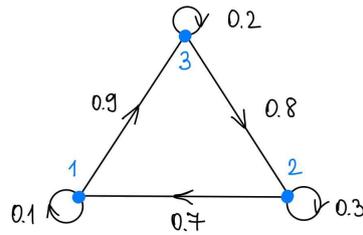
Определение 4.2. *Цепь Маркова называется однородной, если стохастические матрицы, ее определяющие, не зависят от времени, т.е. $P(t) = P, \forall t \in \mathbb{N}$.*

Такую цепь Маркова можно представить в виде ориентированного взвешенного графа $G = (V, E, W)$. Множество вершин графа V соответствует пространству состояний $\Xi = \{1, 2, \dots, r\}$ цепи Маркова. Ребро графа $e_{ij} \in E$ имеет вес W_{ij} равный вероятности перехода p_{ij} из состояния i в состояние j . Любая реализация случайного процесса, соответствующего цепи Маркова, представляется как путь в графе. Следовательно, однородную цепь Маркова можно представить как распределение вероятностей на пространстве путей в полученном графе.

Пример 4.1. *Марковской цепи с переходными вероятностями*

$$P = \begin{pmatrix} 0.1 & 0.9 & 0 \\ 0 & 0.2 & 0.8 \\ 0.7 & 0 & 0.3 \end{pmatrix}$$

соответствует следующий граф:



Упражнение 4.3. Для однородной цепи Маркова, порожденной начальным распределением μ и стохастической матрицей P найдите вероятность перехода за s шагов $P(X_{t+s} = j | X_t = i)$ и распределение вероятностей для X_t , используя доказанные выше факты.

Аналогия с графами здесь не случайна: в неориентированном графе с весами рёбер 1 матричная степень A^n матрицы смежности даёт в ячейке $(A^n)_{ij}$ количество путей $i \rightarrow j$ длины n . В случае же переходной матрицы за счёт стохастичности мы получаем в $(P^n)_{ij}$ вероятность пройти путь $i \rightarrow j$ ровно за n шагов при условии старта в i .

4.4 Эргодическая теорема

Вернемся к нашему примеру с отделами магазина. Предположим, покупатель первый раз пришел в магазин, он еще не привык к расположению отделов, поэтому перемещается по ним случайным образом. Однако с течением времени (если он достаточно долго ходит) он в теории изучит расположение отделов и привыкнет, забывая о том, откуда он вообще пришёл (это похоже на психологическую драму...). Если мы предполагаем что такое явление может иметь место, то в теории можно посмотреть и оценить, как часто он посещает каждый из отделов, один из вариантов – вычислить вероятности $\lim_{s \rightarrow \infty} P(X_s = j)$. Насколько вообще это может быть разумным подходом? И вообще существуют ли такие пределы и от чего они зависят? Этими вопросами в частности занимается эргодическая теория, очень известный раздел динамических систем, который в частности (но не только) в качестве примеров рассматривает цепи Маркова.

Другое интересное применение такого предела – исследование самых популярных веб-страниц, либо авторов в графах цитирования (вершины представляют собой статьи или авторов, а ребро проходит, если один цитирует другого). В частности, на вычислении такого предельного распределения (на миллионах и более вершин) основан знаменитый алгоритм PageRank, где есть *random surfer*, пользователь, который случайно нажимает по ссылкам и исследует интернет, забывая о предыдущих посещённых страницах. Главный вклад авторов этого подхода – необычное приложение, но ещё и метод вычисления, который реализуется как поиск собственного вектора с собственным значением 1 через степенной метод. Итерационный метод был выбран потому, что всю матрицу хранить и использовать дорого, но если её представлять как разреженную, то хранить можно эффективно и вычислить умножение такой матрицы на вектор можно достаточно быстро.

Не обошлось и без регуляризации, чтобы гарантировать выполнение условий, которые мы обсуждаем ниже.

Оказывается, что при некоторых предположениях такой предел существует и более того, не зависит от начального распределения μ . То есть, клиент, как и свойственно в принципе людям, через долгое время вообще забывает, с какого отдела он начинал изучать магазин.

Определение 4.3. *Стохастическая матрица P называется эргодической, если существует такое T , что $(P^{(T)})_{ij} = p_{ij}^{(T)} > 0$ для всех i, j .*

Заметим, что поскольку множество возможных состояний $\Xi = 1, 2, \dots, r$ конечно, можно найти такое α что $(P^{(T)})_{ij} = p_{ij}^{(T)} \geq \alpha$ для всех i, j .

Условие эргодичности не очень удобно проверять, но есть эквивалентные ему условия, которые можно несложно доказать.

Упражнение 4.4. *Докажите, что стохастическая матрица P эргодическая тогда и только тогда, когда*

1. Из любой вершины в любую есть путь конечной длины с ненулевой вероятностью;
2. Не существует периодических состояний. Периодом состояния i называют

$$T = \text{GCD} \{t : P(X_t = i | X_0 = i) > 0\},$$

если $T = 1$, то состояние аperiodично, иначе периодично с периодом T .

Интересный факт про периодичность – это то, что период T не означает, что можно вернуться за T шагов с ненулевой вероятностью. Периодичность означает лишь, что если возвращение происходит, то только во времена кратные T в силу особенной структуры переходов.

Пример 4.2. *Эргодической матрицей будет, например,*

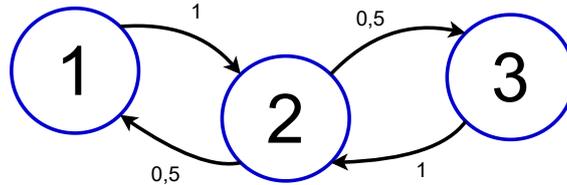
$$\begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix}.$$

Менее очевидный пример – это

$$\begin{bmatrix} 2/3 & 0 & 1/3 \\ 0 & 2/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix},$$

в таком графе уже будут петли.

Неэргодические цепи можно легко получать, если, например, в графе будет не 1 связная компонента, а несколько, либо если есть периодичное состояние. Есть гораздо менее очевидные примеры вроде такого:



Можете ли вы привести свой нетривиальный пример неэргодичной цепи?

Если матрица эргодическая, то для неё верна эргодическая теорема.

Теорема 4.2. (Эргодическая теорема) Пусть цепь Маркова порождена эргодической матрицей P . Тогда верны следующие утверждения:

1. распределение вероятностей $P(X_s = j)$ для случайной величины X_s сходится к π независимо от начального распределения μ , т.е. $\lim_{s \rightarrow \infty} (\mu P^{(s)})_j = \pi_j$.
2. распределение π единственно и инвариантно относительно стохастической матрицы P , т.е. $\pi P = \pi$.
3. переходные вероятности за s шагов сходятся к распределению π , т.е. $\lim_{s \rightarrow \infty} p_{ij}^{(s)} = \pi_j$.

Определение 4.4. Распределение π из предыдущей теоремы называется инвариантным или стационарным распределением эргодической цепи Маркова.

Обратите внимание (это можно от противного сделать), что если цепь эргодическая, то инвариантное распределение единственно. Существует много разных доказательств эргодической теоремы, мы попробуем это сделать используя теорию метрических пространств и сжимающих отображений, потому что в более общем случае Марковских процессов рассуждение очень похожее и с похожим инструментарием. Мы используем общие аргументы, которые приходят из функционального анализа, но если вы до этого момента не были с ним знакомы, для вас всё, что нужно ниже. Если вы знаете уже, что такое метрика и полное метрическое пространство, можете немного пролистать.

Начнём с того, что вспомним (или построим) фундамент, заодно освоив новый и широко используемый инструмент.

Определение 4.5. Метрикой на пространстве \mathcal{X} называется $d : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}_{\geq 0}$, удовлетворяющая свойствам

1. $d(x, y) \geq 0$ в общем, но $d(x, y) = 0$ тогда и только тогда, когда $x = y$.
2. Симметричность: $d(x, y) = d(y, x)$.
3. Неравенство треугольника: $d(x, y) \leq d(x, z) + d(y, z)$.

По своей сути метрика – это расстояние между точками. Если на \mathcal{X} можно определить метрику d , то (\mathcal{X}, d) называют *метрическим пространством*.

Второй кирпичик – это критерий Коши, который мы все помним из классики матанализа: числовая последовательность (x_n) сходится тогда и только тогда, когда

$$\forall \varepsilon > 0 \exists T > 0 : \forall k, m > T |x_k - x_m| < \varepsilon.$$

В случае метрических пространств это условие выглядит похоже:

$$\forall \varepsilon > 0 \exists T > 0 : \forall k, m > T d(x_k, x_m) < \varepsilon.$$

Но не всегда верен критерий Коши. Если верен, то такое пространство называют *полным метрическим пространством*.

К счастью для нас, рассматриваемое нами пространство – это по сути множество конечномерных векторов, про которое с L^p метрикой известно, что оно полное и это показывается через критерий Коши для числовых последовательностей. В качестве метрики возьмём L^1 расстояние с небольшой поправкой:

$$d(\mu', \mu'') = \frac{1}{2} \sum_{i=1}^r |\mu'_i - \mu''_i|.$$

Отображение $A : \mathcal{X} \rightarrow \mathcal{X}$ *сжимающее*, если есть $q < 1$ такой, что для всех $x, y \in \mathcal{X}$ расстояние $d(Ax, Ay) < qd(x, y)$. В качестве отображения у нас выступает стохастическая матрица.

Для технической стороны нам понадобится несколько фактов, но они все возникают из того, что нам понадобится показать сжимающее отображение. А из него при помощи критерия Коши мы докажем сходимость последовательности

$$\mu_{t+1} = \mu_t P$$

со стартом $\mu_0 = \mu$.

Лемма 4.3. Пусть Q стохастическая матрица, тогда для любых $\mu', \mu'' \in \mathcal{P}$

$$d(\mu'Q, \mu''Q) \leq (1 - \alpha)d(\mu', \mu'')$$

и при этом $\alpha \leq 1$.

▷

Пусть $\mu', \mu'' \in \mathcal{P}$. Обозначим через \sum^p суммирование по тем индексам, для которых слагаемые положительны.

Пусть Q - стохастическая матрица. Тогда $\mu'Q, \mu''Q$ будут распределениями вероятностей. Используя доказанные выше технические утверждения, запишем оценку на $d(\mu'Q, \mu''Q)$:

$$d(\mu'Q, \mu''Q) = \frac{1}{2} \sum_{i=1}^r |\mu'_i - \mu''_i|.$$

Главный трюк здесь – это представить сумму модулей в удобном виде. Например, так:

$$0 = \sum_{i=1}^r \mu'_i - \sum_{i=1}^r \mu''_i = \sum_{i=1}^r p(\mu'_i - \mu''_i) - \sum_{i=1}^r p(\mu''_i - \mu'_i).$$

Отсюда мы получаем

$$\sum_{i=1}^r p(\mu'_i - \mu''_i) = \sum_{i=1}^r p(\mu''_i - \mu'_i),$$

а ещё

$$d(\mu', \mu'') = \frac{1}{2} \sum_{i=1}^r p(\mu'_i - \mu''_i) + \frac{1}{2} \sum_{i=1}^r p(\mu''_i - \mu'_i) = \sum_{i=1}^r p(\mu'_i - \mu''_i).$$

Давайте разовьём эту тему явно подставив то, с чего мы начинали:

$$d(\mu'Q, \mu''Q) = \sum_{i=1}^r p((\mu'Q)_i - (\mu''Q)_i) = \sum_{i=1}^r p\left(\sum_{j=1}^r \mu'_j q_{ji} - \sum_{j=1}^r \mu''_j q_{ji}\right) = \sum_{i \in I} \sum_{j=1}^r (\mu'_j - \mu''_j) q_{ji}.$$

Множеством I обозначим множество индексов слагаемых выше, которые суммируются в \sum^p . В такой нотации уже можно поменять порядок суммирования, тогда получим

$$= \sum_{j=1}^r (\mu'_j - \mu''_j) \sum_{i \in I} q_{ji} \leq \sum_{j=1}^r p(\mu'_i - \mu''_i) \sum_{i \in I} q_{ji}.$$

Заметим ещё, что

$$\sum_{i \in I} q_{ji} = 1 - \sum_{i \notin I} q_{ji} = 1 - \alpha,$$

тогда если вспомнить про

$$d(\mu', \mu'') = \sum_{j=1}^r p(\mu'_i - \mu''_i),$$

то в финале получим

$$d(\mu'Q, \mu''Q) \leq (1 - \alpha)d(\mu', \mu'').$$

□

Параметр α здесь очень важен: в самом общем случае стохастической матрицы Q мы можем подтвердить только, что $\alpha \leq 1$, но если мы знаем, что все $q_{ij} > 0$, то $\alpha < 1$. Этот факт мы эксплуатируем ниже, чтобы собрать доказательство эргодической теоремы.

▷ (Доказательство эргодической теоремы)

1. Для начала докажем, что распределение для X_s сходится при $s \rightarrow \infty$. Пусть $\mu_0 = \mu$ - распределение X_0 , тогда $\mu_s := \mu_0 P^s$ - распределение для X_s . Будем доказывать сходимость через проверку критерия Коши, для этого попробуем посмотреть на

$$d(\mu^s, \mu^{s+t}) = d(\mu P^{(s)}, \mu P^{(s+t)}).$$

Для любого $\varepsilon > 0$ должен существовать $T_0 > 0$ такой, что для любых $t > T_0, s > 0$ расстояние $d(\mu^s, \mu^{s+t}) < \varepsilon$. Отсюда в силу полноты метрического пространства мы получим, что существует предел π , то есть $d(\mu_t, \pi) \rightarrow 0$ при $t \rightarrow \infty$.

Чтобы было $\alpha < 1$ в условии, полученном выше, достаточно потребовать $q_{ij} > 0$ для всех i, j . Условие эргодичности говорит, что такая матрица есть и это P^T .

Если мы выберем $T_0 = Tn$, то из полученной нами оценки будет

$$d(\mu^t, \mu^{t+s}) \leq (1 - \alpha)^n d(\mu P^{(t-nT)}, \mu P^{(s+t-nT)}) \leq (1 - \alpha)^n.$$

Тогда для точности возьмём n такой, чтобы $(1 - \alpha)^n < \varepsilon$.

2. Давайте проверим, что $\pi P = \pi$.

$$\pi P = \lim_{s \rightarrow \infty} \mu P^{(s)} P = \lim_{s \rightarrow \infty} \mu P^{(s+1)} = \pi.$$

3. Теперь давайте докажем, что такое π единственно.

Пусть $\pi_1 P = \pi_1, \pi_2 P = \pi_2$. Тогда

$$d(\pi_1, \pi_2) = d(\pi_1 P^{(T)}, \pi_2 P^{(T)}) \leq (1 - \alpha) d(\pi_1, \pi_2).$$

Т.е. $d(\pi_1, \pi_2) = 0$, что означает, что π_1 и π_2 совпадают и распределения $\mu^s = \mu P^{(s)}$ сходятся к π независимо от начального распределения μ .

4. Попробуйте доказать последнее сами, это как в пункте 2, надо только подобрать удачно стартовое распределение.

□

Пример 4.3. Классическими примерами неэргодических цепей Маркова являются несвязные цепи Маркова (число сообщающихся классов больше одного) и цепи с циклами.

На этом мы завершаем наш короткий экскурс в цепи Маркова, но мы ещё к ним не один раз вернёмся при решении задач, основываясь на том базисе, что у нас уже есть.

Литература

- [1] Alfred Cowles 3rd. Can stock market forecasters forecast? *Econometrica*, 1(3):309–324, 1933.
- [2] Alfred Cowles 3rd and Herbert E. Jones. Some a posteriori probabilities in stock market action. *Econometrica*, 5(3):280–294, 1937.
- [3] Louis Bachelier. Théorie de la spéculation. *Annales scientifiques de l'École Normale Supérieure*, 3e série, 17:21–86, 1900.
- [4] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.
- [5] Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.
- [6] L. Isserlis. On a Formula for the Product-Moment Coefficient of any Order of a Normal Frequency Distribution in any Number of Variables, November 1918.
- [7] M. G. Kendall and A. Bradford Hill. The analysis of economic time-series-part i: Prices. *Journal of the Royal Statistical Society. Series A (General)*, 116(1):11–34, 1953.
- [8] Paul A. Samuelson. Proof that properly discounted present values of assets vibrate randomly. *The Bell Journal of Economics and Management Science*, 4(2):369–374, 1973.
- [9] T. Tao. *An Introduction to Measure Theory*. Graduate studies in mathematics. American Mathematical Society, 2011.
- [10] Holbrook Working. A random-difference series for use in the analysis of time series. *Journal of the American Statistical Association*, 29(185):11–24, 1934.
- [11] Ширяев А.Н. Булинский А.В. *Теория случайных процессов*. М:ФИЗМАТЛИТ, 2005.
- [12] Синай Я.Г. Коралов Л.Б. *Теория вероятностей и случайные процессы*. МЦНМО, 2013.
- [13] А.Н. Ширяев. *Основы стохастической финансовой математики*. МЦНМО, 2016.