

Глава 1. Идеи и определения

В этой главе вводятся основные определения цепей Маркова и смежные понятия.

Историческая справка

Изучение цепей Маркова возникло из фундаментальных вопросов о вероятности и зависимых событиях. В то время как ранняя теория вероятностей фокусировалась на независимых испытаниях (как в *Ars Conjectandi* Якоба Бернулли 1713 года), Андрей А. Марков (1856-1922) произвел революцию в этой области, изучая системы с памятью о текущем состоянии.

Марков, русский математик из Санкт-Петербурга, выступал за строгий и аналитический подход к теории вероятностей. Его работа возникла из научного спора с Павлом А. Некрасовым, который считал, что независимость необходима для вероятностных законов — взгляд, вытекающий из его философских воззрений. Марков распространил закон больших чисел Бернулли на зависимые величины посредством изучения цепных вероятностей. Цепи Маркова не только разрешили спор с Некрасовым, но и нашли неожиданные применения, например, в лингвистическом анализе. Так, он провёл знаменитый анализ переходов гласных и согласных в «Евгении Онегине» Пушкина, выявив зависимости в естественном языке.

Ранний пример, демонстрирующий интересное предельное поведение (подобное закону больших чисел) и воплощающий новаторские идеи Маркова, был представлен совсем юным Дьёрдем Пойа.

Пример 0.1 (Урна Пойа). Урна содержит m_1 шаров цвета 1, m_2 шаров цвета 2 и так далее до m_N шаров цвета N . На каждом шаге случайно выбирается (и извлекается из урны) шар, и его цвет фиксируется. После этого в урну добавляются ℓ шаров выбранного цвета вместе с n шарами каждого из невыбранных цветов.

Независимый случай соответствует $\ell = 1, n = 0$ (или, в более общем случае, $\ell = n + 1$). Однако элементарными методами можно вычислить предельное поведение (по мере роста числа итераций t одной и той же процедуры), которое зависит от параметров m_1, \dots, m_N, ℓ, n .

С момента своего появления цепи Маркова стали краеугольным камнем современной математики. Помимо статистики и теории вероятностей, многие (если не большинство) передовых результатов в динамических системах и эргодической теории были получены с использованием методов цепей Маркова. Прямые применения цепей Маркова встречаются в дискретных группах, комбинаторике, теориях геометрических потоков, и в целом они предоставляют существенный инструмент для интерпретации и предсказания поведения случайной и детерминированной динамики.

Неполный список тем с очень явными применениями цепей Маркова включает: процессы принятия решений, методы Монте-Карло, статистическую механику, квантовую механику, вычислительную физику, скрытые марковские модели, обучение с подкреплением, теорию очередей, анализ сетевого трафика, исследование операций, химическую кинетику, молекулярную динамику, сворачивание белков, финансовое моделирование, предсказание фондового рынка, теорию игр, экономическое моделирование, оптимизацию цепочек поставок, системы управления, робототехнику, обработку изображений, компьютерное зрение, искусственный интеллект, автономные системы, обработку естественного языка, распознавание речи,

генерацию текста, биоинформатику, геномику, моделирование заболеваний, прогнозирование климата и погоды, экологическое моделирование, популяционную динамику, эволюционную биологию, филогенетический анализ, медицинскую визуализацию, анализ сетей мозга, когнитивное моделирование, поведенческую экономику, анализ социальных сетей, поисковые системы, рекомендательные системы, анализ поведения клиентов, блокчейн, алгоритмы консенсуса, распределенные системы, отказоустойчивость, инженерную надежность, обработку сигналов, криптографию, энергетические системы, транспортные системы, системы персонализированного обучения, спортивную аналитику и прогнозирование, распределение ресурсов, кибербезопасность, обнаружение аномалий, оптимизацию траекторий.

Фундаментальные примеры

Прежде чем давать точные определения, давайте погрузимся в некоторые примеры, которые могут подсказать, какие математические объекты необходимы для правильного определения цепей Маркова, и какие задачи стоит исследовать.

Случайные блуждания

Рассмотрим человека, путешествующего между городами по всему миру. Каждый день этот человек отправляется в аэропорт и покупает билет до пункта назначения с наилучшим прогнозом погоды на следующий день среди городов, достижимых менее чем за четыре часа пути. Если мы обозначим через X_t его положение после t дней путешествия, а через $\mathbf{X} = (X_t)_{t \in \mathbb{N}}$ — весь его путь, нас могут заинтересовать вероятностное распределение и свойства \mathbf{X} .

Пусть $p_{x,y}^t$ обозначает вероятность того, что в день t город y будет иметь наилучший прогноз погоды среди всех достижимых из x менее чем за четыре часа (в частности, это значение равно 0, если x и y находятся более чем в четырёх часах пути). Очевидно, что случайное путешествие этого человека демонстрирует следующее ключевое **Марковское свойство**: независимо от того, как город x был достигнут к моменту времени t , вероятность полететь в y на следующий день равна $p_{x,y}^t$. В вероятностных обозначениях мы можем выразить это свойство следующим образом: для любого дня t , любой пары городов x, y и любой последовательности городов x_0, x_1, \dots, x_{t-1} , посещённых до дня t :

$$\mathbb{P}(X_{t+1} = y \mid X_0 = x_0, X_1 = x_1, \dots, X_t = x) = \mathbb{P}(X_{t+1} = y \mid X_t = x) =: p_{x,y}^t \quad (1)$$

Также ясно, что если мы знаем начальное распределение μ_x , соответствующее вероятности того, что $X_0 = x$, мы можем вычислить вероятность каждого конечного пути:

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_t = x_t) = \mu_{x_0} p_{x_0, x_1}^0 \cdots p_{x_{t-1}, x_t}^{t-1} \quad (2)$$

Из этого примера мы получили все ингредиенты, необходимые для определения цепей Маркова.

- **Пространство состояний** S : Это множество городов, пространство, в котором принимает свои значения положение нашего путешественника X_t .
- **Правила перехода** $p_{x,y}^t$: Вероятности перемещения из города $x \in S$ в момент времени (день) t в город $y \in S$ в момент времени $t + 1$. Для каждых t и x , величина $p_{x,\cdot}^t \geq 0$ представляет собой вероятность по y : $\sum_y p_{x,y}^t = 1$.
- **Начальное распределение** μ : Это вероятность на пространстве состояний S , представляющая тот факт, что наш путешественник может начать в случайном месте (например, потому что он путешествовал случайным образом некоторое время до того, как мы начали наблюдение в момент времени 0).

- Как обычно, нам также нужно **вероятностное пространство** $(\Omega, \mathcal{F}, \mathbb{P})$ для реализации нашего случайного путешествия, но оно *скрыто* в обозначениях. Ω может включать гораздо больше информации, чем просто пути путешественника — например, погоду, случайные факторы, влияющие на время в пути между городами каждый день, и так далее. В этом отношении мы можем неформально рассматривать σ -алгебру \mathcal{F}_t , содержащую все события, наблюдаемые до дня t (например, погоду).

Таким образом, мы сталкиваемся с первой задачей, имеющей интуитивное решение (которое, однако, может потребовать некоторого внимания, если S не является конечным):

А. Дать математическое определение цепи Маркова как бесконечной случайной последовательности $\mathbf{X} = (X_t)_{t \in \mathbb{N}}$ так, чтобы её распределение соответствовало интуитивной идее случайных переходов по S с заданными вероятностями перехода.

Компьютерные алгоритмы

Современные компьютеры, выполняющие *любой* алгоритм, представляют собой глубокую реализацию цепей Маркова. Процессор имеет доступ к двум ресурсам:

- Память (при типичном запуске на ноутбуке это означает регистры процессора, кэш и ОЗУ; но в других ситуациях это может представлять иное хранилище), которая обычно хранит информацию в двоичной форме.
- Источник случайности (например, генератор случайных или псевдослучайных чисел), который обеспечивает независимые стохастические входные данные на каждом шаге.

Пространство состояний S представляет возможные конфигурации памяти. На стандартном компьютере¹ мы имеем $S = \{0, 1\}^N$. Конфигурация X_t памяти в момент времени t содержит всю информацию, к которой мы можем получить доступ.

Один временной шаг может соответствовать одному тактовому циклу процессора (порядка 10^{-9} секунд) или, если мы примем упрощенную точку зрения, времени, необходимому для выполнения предопределенного блока операций в алгоритме (например, одного вычислительного шага в языке программирования высокого уровня).

Шум задается внутренним поведением устройств генерации случайных чисел как последовательность независимых случайных величин Z_t (в подходящем пространстве случайности, которое на компьютере обязательно конечно).

$$X_{t+1} = F(t, X_t, Z_t) \tag{3}$$

Марковское свойство заложено в том факте, что $F(\cdot)$ не может зависеть от прошлой истории памяти, такой как X_{t-1} , поскольку эта информация потеряна (или, точнее, то, что от неё сохранилось, закодировано в X_t).

Пример 0.2. В качестве конкретного примера рассмотрим алгоритм рандомизированной сортировки, который итерировывает следующие шаги:

1. Если список уже отсортирован, ничего не делать. Иначе:
2. С вероятностью p выполнить случайную перестановку (транспозицию) в списке.
3. С вероятностью $1 - p$ выполнить операцию случайной сортировки (транспозицию, которая упорядочивает два элемента).

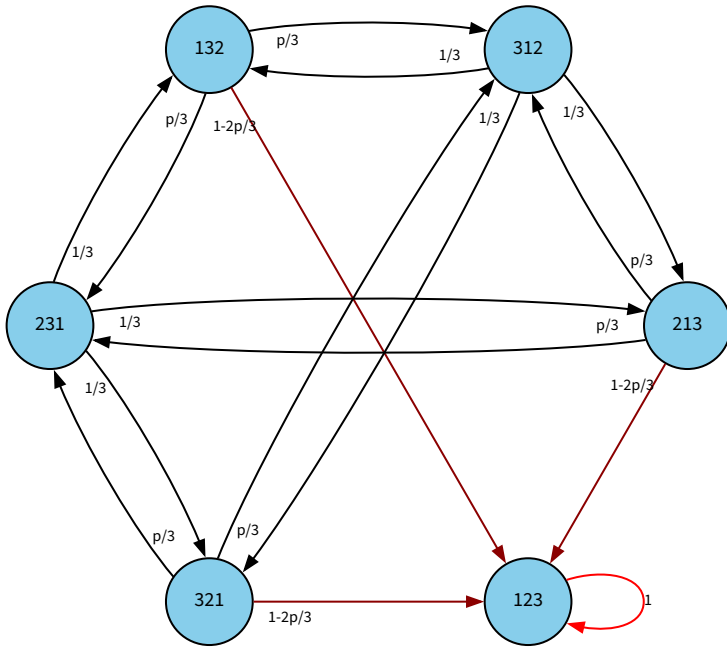


Рисунок 1

Мы легко можем видеть, что происходит на практике, если сортируем только 3 элемента с помощью этого алгоритма. Перестановки, которые находятся на расстоянии двух транспозиций от тождественной, перейдут с равной вероятностью к одному из своих соседей (независимо от того, применяется шаг 2 или шаг 3; каждый сосед здесь имеет одинаковую вероятность). Перестановки, находящиеся на расстоянии одной транспозиции от тождественной, выберут одного из своих соседей с вероятностью $p/3$ (если выбран шаг 2), но есть дополнительная вероятность $1 - p$ перейти к тождественной перестановке. На тождественной перестановке всегда применяется шаг 1.

Пример с алгоритмом наводит на интересный вопрос. Пусть $A \subset S$ представляет множество состояний памяти, определяющих условие завершения алгоритма. Обычно это очень явно, например, A состоит из состояний, где счетчик превышает определенное значение, или цикл `for` или `while` завершается. Например, в примере Пример 0.2 естественным выбором является $A = \{(1, 2, 3)\}$, синглетон отсортированного списка.

Если стоимость (например, фактическое физическое время) перехода из состояния $x \in S$ в $y \in S$ равна $c(x, y)$, то мы хотим оценить $\sum_{t=0}^{\tau_A-1} c(X_t, X_{t+1})$, где $\tau_A \in \mathbb{N}$ — случайное время, равное первому моменту t , такому что X_t касается A (в Пример 0.2 — первый момент, когда список отсортирован). Итак, некоторые примеры интересных вопросов для исследования:

В. Оценить распределение $\sum_{t=0}^{\tau_A-1} c(X_t, X_{t+1})$. Например, для $c \equiv 1$, это равно количеству шагов, необходимых для достижения A . Классическая задача (для заданного целевого множества A) — минимизировать $\mathbb{E}[\tau_A]$ по всем выборам вероятностей перехода (алгоритмам) с некоторым заданным свойством. С. Оценить вероятность достижения множества A (например, решения нашей задачи, отсортированного списка), прежде чем будет достигнуто множество B (например, переполнение).

Упражнение 0.1. Проверьте, что если $\mathbf{X} = (X_t)_{t \in \mathbb{N}}$ получена как в Уравнение 3, для заданной функции F и последовательности $(Z_t)_{t \in \mathbb{N}}$ независимых случайных величин, то свойство Уравнение 1 выполняется.

¹ В упрощенных моделях (например, если мы не хотим рассматривать ошибки округления), память можно рассматривать как произведение дискретных и непрерывных моделей.

Финансовые модели

Цепи Маркова широко используются в финансовом моделировании для представления стохастической эволюции цен активов, процентных ставок и других экономических переменных. Простым, но показательным примером является модель с дискретным временем для цены акции.

Пусть Y_t представляет цену акции в момент времени t (где t может представлять дни, недели или какой-либо другой дискретный временной интервал). Стандартный подход состоит в том, чтобы предположить случайную динамику для доходности R_t акции. Это означает

$$Y_{t+1} = Y_t(1 + R_t)$$

где для простоты можно предположить, что R_t может принимать только конечное число значений, скажем r_1, r_2, \dots, r_n . Например, мы можем иметь $r_1 = -0.02, r_2 = 0.01$ и $r_3 = 0.05$, представляющие соответственно для каждого периода времени потерю 2%, прибыль 1% и прибыль 5%. Вероятности перехода R_t тогда описывали бы правдоподобие каждой доходности при условии текущей доходности, но разумно моделировать их как зависящие от цены акции Y_t :

$$p_{i,j}(x) = \mathbb{P}(R_{t+1} = r_j \mid R_t = r_i, Y_t = x)$$

В этом случае R_t не является цепью Маркова в том смысле, что Уравнение 1 не выполняется (для R_t). Однако легко проверить, что пара $X_t = (Y_t, R_t)$ действительно удовлетворяет Уравнению 1, поэтому X_t является цепью Маркова.

Этот пример показывает, что с точки зрения моделирования интересная часть заключается в том, чтобы фактически обнаружить, какие переменные составляют цепь Маркова в данной случайной динамике. Для описанной здесь модели хорошо известной задачей справедливой оценки является следующая:

D. Найти вероятностную меру \mathbb{Q} , которая абсолютно непрерывна относительно исходной вероятности \mathbb{P} и такова, что Y_t является мартингалом.

Языковые модели

Языковая модель может быть формально определена как распределение вероятностей на последовательностях токенов (слов, или символов, или строк символов). Модели цепей Маркова предоставляют фундаментальную основу для этих распределений посредством так называемых n -граммных приближений. Очевидно, слишком наивно полагать, что одиночные токены обладают марковским свойством. Например, давайте возьмем отдельные слова в качестве токенов. Рассмотрим предложение «I want to learn stochastic processes. I will start reading about»; существует высокая вероятность того, что следующими словами будут «Markov Chains». Это, безусловно, не так, если мы должны сделать вывод о продолжении предложения только по слову «about». Другими словами, обуславливание по нескольким токенам перед текущим изменяет распределение вероятностей иначе, чем обуславливание по одному прошлому токenu — прямое нарушение Уравнение 1.

Чтобы подойти к этой проблеме, мы можем сделать n -граммные приближения. Пусть V будет нашим словарем (множеством возможных токенов). Модель цепи Маркова n -го порядка предполагает, что вероятность каждого токена зависит только от предыдущих n токенов: для $t \geq n, w, w_1, \dots, w_t \in V$

$$\mathbb{P}(W_{t+1} = w \mid W_1 = w_1, \dots, W_t = w_t) = \mathbb{P}(W_{t+1} = w \mid W_{t-n+1} = w_{t-n+1}, \dots, W_t = w_t) \quad (4)$$

В этом случае, чтобы получить цепь Маркова, мы можем выбрать $S = V^n$, другими словами, марковское свойство Уравнение 1 восстанавливается, как только мы рассматриваем кортежи из n токенов в качестве состояний. Однако в практических случаях, например, когда мы хотим вывести следующий токен из предыдущей строки токенов, это не очень эффективно, если мы не возьмем n огромным. Если мы читаем детектив

ный роман и читаем фразу «Детектив понял, что преступником является», следующее слово может очень сильно зависеть от какой-то детали, написанной в начале книги. Поэтому n может потребоваться длиной с наш текстовый корпус. Математически это заставляет нас вступить на опасную территорию, где $S = V^{\mathbb{N}}$, так что мы имеем несчётное пространство состояний, и к нашему списку добавляется дополнительная задача:

Е. Расширить математические определения цепей Маркова, когда S является произвольным измеримым пространством.

На практике, с другой стороны, мы все еще можем взять n очень большим, и в этом случае нам остаётся сложнейшая задача оценки вероятностей перехода для кортежей из n токенов по нашему корпусу текста. Современные нейронные языковые модели расширяют эту концепцию за счет распределенных представлений, которые могут улавливать зависимости большей дальности.

https://youtu.be/KZeIEiBrT_w

Цепи Маркова: Определения

Давайте углубимся в математическую теорию. В дальнейшем $(\Omega, \mathcal{F}, \mathbb{P})$ — заданное [вероятностное пространство](#).

Определение 0.1 (Цепь Маркова на счётном пространстве состояний). Пусть S — непустое конечное или счётное множество. **Цепь Маркова** $\mathbf{X} = (X_t)_{t \in \mathbb{N}}$ — это последовательность случайных величин $X_t: \Omega \rightarrow S$, такая что для $t \in \mathbb{N}$ и $x \in S$

$$\mathbb{P}(X_{t+1} = x \mid (X_s)_{s \leq t}) = \mathbb{P}(X_{t+1} = x \mid X_t) \quad (5)$$

S называется **пространством состояний** цепи Маркова. Величины $p_{x,y}^t := \mathbb{P}(X_{t+1} = y \mid X_t = x)$ называются **вероятностями перехода**, а Уравнение 5 называется **Марковским свойством**.

Цепь Маркова называется **однородной** (или однородной по времени), если вероятности перехода $p_{x,y}^t \equiv p_{x,y}$ не зависят от временного параметра t .

Упражнение 0.2. Проверьте, что вероятности перехода действительно являются вероятностями в следующем смысле: для каждого $t \geq 0$, $x \in S$, $p_{x,\cdot}^t$ определяет вероятность на S :

- $p_{x,y}^t \geq 0$.
- $\sum_{y \in S} p_{x,y}^t = 1$.

Определение 0.2 (Пространство Скорохода на счётном пространстве состояний). Пусть S — конечное или счётное множество, и пусть $(p_{x,y}^t)_{t \geq 0; x,y \in S}$ — вероятности перехода. **Пространство Скорохода**, связанное с $(p_{x,y}^t)$, — это пространство путей

$$D(S) := \left\{ \mathbf{x} \in S^{\mathbb{N}} : p_{x_t, x_{t+1}}^t > 0, \text{ для всех } t \in \mathbb{N} \right\}$$

$D(S)$ естественным образом наделяется σ -алгеброй, которую оно наследует как подпространство $S^{\mathbb{N}}$.

Определение 0.3. Для S и $p_{x,y}^t$, определенных выше, **вероятности перехода за несколько шагов** определяются для $s \leq t$ как

$$p_{x,y}^{(s,t)} := \sum_{z_{s+1}, \dots, z_{t-1} \in S} p_{x, z_{s+1}}^s p_{z_{s+1}, z_{s+2}}^{s+1} \dots p_{z_{t-1}, y}^{t-1} \quad (6)$$

в частности, $p_{x,y}^{(t,t+1)} = p_{x,y}^t$ и $p_{x,y}^{(t,t)} = \delta_x(\{y\})$.

Если $p_{x,y}^t \equiv p_{x,y}$ однородна по времени, то $p^{(s,t)}$ зависит только от $t - s$, и поэтому мы обозначаем в этом случае $p_{x,y}^{(t)} \equiv p^{(s,s+t)}$ независимо от $s \geq 0$.

Упражнение 0.3. Убедитесь, что вероятности перехода за несколько шагов действительно соответствуют своему названию, а именно, если \mathbf{X} — цепь Маркова с вероятностями перехода $(p_{x,y}^t)$, то

$$\mathbb{P}(X_t = y \mid X_s = x) = p_{x,y}^{(s,t)}$$

Упражнение 0.4. Убедитесь, что вероятности перехода за несколько шагов удовлетворяют уравнениям Чепмена-Колмогорова (также известным как полугрупповое свойство): для $s \leq t \leq u$

$$p_{x,z}^{(s,u)} = \sum_{y \in S} p_{x,y}^{(s,t)} p_{y,z}^{(t,u)}$$

Пусть S — конечное или счётное множество², и \mathbf{X} — Марковский процесс с:

- начальным распределением $\mu \in \mathcal{P}(S)$, вероятностной мерой на S .
- вероятностями перехода $(p_{x,y}^t)_{t \geq 0; x, y \in S}$.

Тогда закон распределения \mathbf{X} на $D(S)$ однозначно определяется μ и $(p_{x,y}^t)$, поскольку

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \dots, X_t = x_t) = \mu_{x_0} p_{x_0, x_1}^0 \cdots p_{x_{t-1}, x_t}^{t-1}$$

для всех $t \geq 0$ и $x_0, \dots, x_t \in S$. Согласно [теореме Колмогорова](#), это сразу определяет единственную вероятностную меру на $D(S)$.

Примечание. В общем случае мы считаем, что вероятности перехода зафиксированы раз и навсегда, тогда как начальное условие может меняться. Поэтому удобны следующие обозначения:

- Для $x \in S$, \mathbb{P}_x обозначает вероятность $\mathbb{P}_x(\cdot) = \mathbb{P}(\cdot \mid X_0 = x)$ (где мы неявно предположили $\mathbb{P}(X_0 = x) > 0$).
- Более обще, для вероятностной меры μ на S

$$\mathbb{P}_\mu := \sum_x \mu_x \mathbb{P}_x$$

Это следует интерпретировать в линейном смысле, например

$$\mathbb{E}_\mu[f(X_1, X_2)] = \sum_x \mu_x \mathbb{E}_x[f(X_1, X_2)]$$

Заметим, в частности, что $\mathbb{P}_{\delta_x} \equiv \mathbb{P}_x$. Смысл этого обозначения в том, что обычно нас интересует только закон распределения \mathbf{X} , и при \mathbb{P}_μ , \mathbf{X} является цепью Маркова с теми же вероятностями перехода (что и при исходной \mathbb{P}), и начальным распределением μ . Точнее, если мы начнем с цепи Маркова $\mathbf{X}: \Omega \rightarrow S^{\mathbb{N}}$ над вероятностным пространством $(\Omega, \mathcal{F}, \mathbb{P})$, и для простоты примем $\mathbb{P}(X_0 = x) > 0$ для всех $x \in S$, то для каждой вероятности $\mu \in \mathcal{P}(S)$ мы можем рассмотреть пространство $(\Omega, \mathcal{F}, \mathbb{P}_\mu)$, где \mathbf{X} имеет те же вероятности перехода, но начальное распределение μ .

Примечание. Если \mathbf{Y} не однородна по времени, мы можем определить цепь Маркова \mathbf{X} с пространством состояний $\mathbb{N} \times S$, полагая $X_t := (t, Y_t)$ и используя однородные по времени вероятности перехода $q_{(s,x),(t,y)} = \mathbf{1}_{t,s+1} p_{x,y}^s$, где $p_{x,y}^s$ — вероятности перехода \mathbf{Y} .

²Нам не обязательно, чтобы S было конечным или счётным. Но по техническим причинам (применение теоремы Колмогорова) всё же необходимо, чтобы S было разумным измеримым пространством.

Ввиду последнего замечания мы будем почти исключительно рассматривать однородные цепи Маркова в дальнейшем. Это не вполне обще, но упрощает обозначения и позволяет получить более сильные результаты.

Марковские операторы

В дальнейшем мы фиксируем вероятности перехода $(p_{x,y})$ и соответствующую цепь Маркова \mathbf{X} .

Пусть $\mathcal{B}(S)$ — пространство всех ограниченных (измеримых, если S несчётно) функций $f: S \rightarrow \mathbb{R}$. Мы определяем линейный оператор

$$P: \mathcal{B}(S) \rightarrow \mathcal{B}(S)$$

$$(Pf)(x) := \sum_{y \in S} p_{x,y} f(y) = \mathbb{E}_x[f(X_1)] = \mathbb{E}[f(X_{t+1}) \mid X_t = x], \quad t \geq 0, x \in S$$

Аналогично, для вероятностной меры $\mu \in \mathcal{P}(S)$ на S мы определяем

$$(\mu P)_x := \sum_{y \in S} \mu_y p_{y,x} = \mathbb{P}_\mu(X_1 = x)$$

Упражнение 0.5. Проверьте, что (μP) определяет вероятностную меру на S .

Упражнение 0.6. Проверьте, что P является сжатием $\sup_x (Pf)(x) \leq \sup_x f(x)$.

Примечание. Если S конечно, оператор P есть не что иное, как оператор, представляемый матрицей $(p_{x,y})$. Он действует на функции справа, а на вероятностные меры слева. Для счётного S это остаётся верным, только суммы становятся рядами, и нам нужно быть уверенными, что f или μ ограничены, чтобы определить $Pf(x)$ и $(\mu P)_x$ в этом случае.

Тогда [уравнение Чепмена-Колмогорова](#) превращается в простой факт, что вероятности перехода за несколько шагов $p_{x,y}^{(t)}$ представляют собой элементы степени P^t оператора P , поскольку действительно $p_{x,y}^{(1)} = p_{x,y}$ и $p_{x,y}^{t+1} = \sum_z p_{x,z}^t p_{z,y}$.

Цепи Маркова: Обозначения

Давайте повторим обозначения, которые мы видели до сих пор, поскольку существует несколько объектов, живущих в разных пространствах, и это может сбивать с толку.

- $(\Omega, \mathcal{F}, \mathbb{P})$ — заданное вероятностное пространство. Точная природа этого пространства не имеет значения и не канонична. Мы хотим даже избежать упоминания того, какое именно пространство Ω мы берем. Это то же отношение, которое можно иметь в геометрии, когда хотят определить многообразие как таковое, не обязательно указывая параметризацию или конкретный атлас.
- Время предполагается дискретным, обычно мы обозначаем время буквами $s, t, u \in \mathbb{N}$.
- S — пространство состояний, пространство, где цепь Маркова принимает значения. Элементы в S обозначаются x, y, z и т. д.
- $\mathbf{X}: \Omega \rightarrow S^{\mathbb{N}}$ — сама случайная цепь. Это последовательность (X_0, X_1, \dots) (измеримых) функций $X_t: \Omega \rightarrow S$.

- Для каждого t , X_t можно мыслить как случайный элемент S . Следовательно, он имеет распределение μ_t на S , определенное как $\mu_{t,x} := \mathbb{P}(X_t = x)$.
- Вероятности перехода задаются $p_{x,y} = \mathbb{P}_x(X_1 = y)$, которые можно интерпретировать как элементы марковского оператора P . Мы можем начать с цепи Маркова \mathbf{X} и определить её начальное распределение μ и вероятности перехода. Или мы можем начать с $\mu \in \mathcal{P}(S)$ и вероятностей перехода, и построить цепь Маркова \mathbf{X} .
- Мы имеем одновременно случайную траекторию на S и детерминированную траекторию³ на $\mathcal{P}(S)$
 - Случайная последовательность элементов на S : $X_0, X_1, \dots, X_t, \dots$
 - Детерминированная последовательность элементов $\mathcal{P}(S)$: $\mu_0 = \mu, \mu_1 = \mu P, \dots, \mu_t = \mu P^t$.
- В конечном счете мы можем представить цепь Маркова либо (возможно, бесконечной) матрицей с элементами $p_{x,y}$ (т.е. матрицей с неотрицательными элементами и такой, что сумма элементов в каждой строке равна 1), либо взвешенным ориентированным графом (где ребро (x, y) присутствует, если $p_{x,y} > 0$, и сумма весов стрелок, исходящих из каждой вершины, равна 1).

Пример 0.3. Оператор P , представляемый матрицей

$$P = \begin{pmatrix} 0.0 & 0.4 & 0.0 & 0.1 & 0.5 \\ 0.0 & 0.2 & 0.3 & 0.5 & 0.0 \\ 0.0 & 0.0 & 0.6 & 0.0 & 0.4 \\ 0.7 & 0.0 & 0.0 & 0.1 & 0.2 \\ 0.0 & 0.8 & 0.0 & 0.2 & 0.0 \end{pmatrix}$$

соответствует ориентированному графу

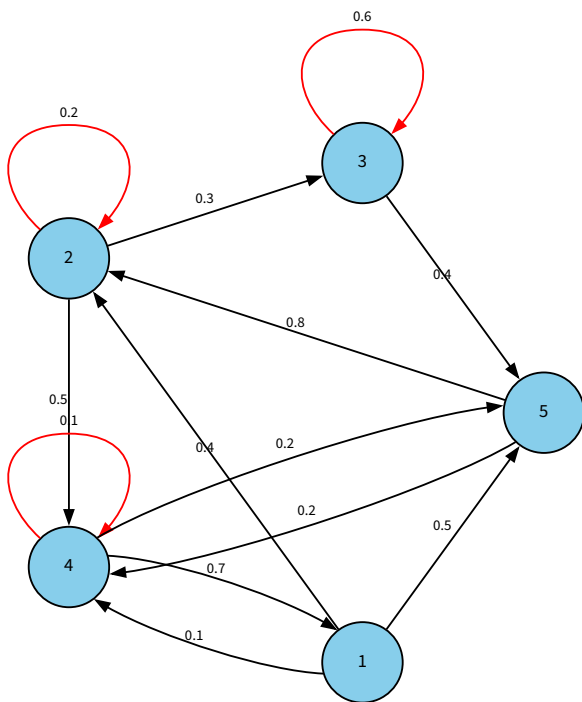


Рисунок 2

³Хотя знание \mathbf{X} в точности эквивалентно знанию X_0, X_1, \dots , знание закона распределения \mathbf{X} (вероятности на $D(S)$) содержит больше информации, чем последовательность μ_0, μ_1, \dots законов распределения X_0, X_1, \dots

Дополнительные Задачи

Упражнение 0.7. Футбольная команда играет в чемпионате, состоящем из одиннадцати игр. Их результаты сильно зависят от морального духа:

- В каждом матче они всегда имеют вероятность $1/3$ сыграть вничью.
 - Если они выиграли последний матч, они выиграют снова с вероятностью $1/2$ (и проиграют с вероятностью $1/6$ независимо от предыдущих матчей).
 - Если они проиграли последний матч, они проиграют снова с вероятностью $1/2$ (и выиграют с вероятностью $1/6$ независимо от предыдущих матчей).
 - Если последний матч был ничьей, они проиграют или выиграют с вероятностью $1/3$ (независимо от предыдущих матчей).
- a. Вычислите вероятность того, что они выиграют третий матч, если они выиграли/проиграли/сыграли вничью в первом.
- b. Вычислите вероятность того, что они выиграют последний матч, если они выиграли/проиграли/сыграли вничью в первом.

Решение

Матрица P с элементами $p_{x,y}$, где $x, y \in S := \{W, D, L\}$ (Победа, Ничья, Поражение), задается покомпонентно

$$P := \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/3 & 1/3 & 1/3 \\ 1/6 & 1/3 & 1/2 \end{pmatrix}$$

В силу симметрии ясно, что $p_{x,D}^{(n)} = p_{D,x}^{(n)} = 1/3$, и $a_n := p_{W,W}^{(n)} = p_{L,L}^{(n)}$ и $b_n := p_{L,W}^{(n)} = p_{W,L}^{(n)}$, причём $a_n + 1/3 + b_n = 1$. Поэтому, если мы запишем $a_n = 1/3 + c_n$ и $b_n = 1/3 - c_n$, то

$$P^n := \begin{pmatrix} 1/3 + c_n & 1/3 & 1/3 - c_n \\ 1/3 & 1/3 & 1/3 \\ 1/3 - c_n & 1/3 & 1/3 + c_n \end{pmatrix}$$

и по рекуррентности имеем $c_n = 3^{-n}/2$.

- a. Из предыдущего вычисления имеем, что искомые вероятности равны $p_{W,W}^{(2)} = 7/18$, $p_{D,W}^{(2)} = 1/3$, $p_{L,W}^{(2)} = 5/18$.
- b. Интуитивная идея о том, что все эти вероятности будут быстро сходиться к $1/3$ (как вычислено), ясна: $p_{W,W}^{(10)} \approx 0.333342$, $p_{D,W}^{(2)} = 1/3$, $p_{L,W}^{(2)} \approx 0.333325$.

Упражнение 0.8. Мои часы сошли с ума. Вместо того чтобы всё время вращаться по часовой стрелке, они ведут себя случайно. Каждый час они вращаются по часовой стрелке (то есть добавляют один час) с вероятностью $1/2$, и против часовой стрелки с вероятностью $1/2$. В полночь они показывали правильное время $x = 0$. Какова вероятность того, что этой ночью (т.е. через 24 часа) они снова покажут правильное время?

Решение

P является циркулянтной матрицей с $p_{x,x+1} = p_{x,x-1} = 1/2$ (где ± 1 понимается $(\text{mod } 12)$), и все остальные элементы равны 0. Если $\eta := e^{i2\pi/N}$, где $N = 12$, то имеем, что P имеет собственные значения $(\eta^k + \eta^{-k})/2 = \cos(2\pi k/N)$ для $k = 0, \dots, N-1$. Мы можем записать $P = VDV^{-1}$ с диагональной

матрицей с вышеупомянутыми собственными значениями и

$$V_{j,k} = e^{i2\pi jk/N} \quad V_{j,k}^{-1} = e^{-i2\pi jk/N} / N \quad (7)$$

Поскольку $P^n = VDV^{-1}$, имеем $p_{x,y}^{(n)} = \frac{1}{N} \sum_{k=0}^{N-1} \cos(2\pi k/N)^n e^{i2\pi k(x-y)/N}$. В частности

$$p_{0,0}^{(n)} = \sum_{k=0}^{N-1} \cos(2\pi k/N)^n / N$$

и при $N = 12$ и $n = 24$ имеем $p_{0,0}^{(n)} = 1486675/8388608 \approx 0.177$.

Мы также можем сказать, что часы должны сделать столько же шагов по часовой стрелке, сколько и против, $(\text{mod } N)$. Это дает

$$p_{0,0}^{(n)} = \sum_{k \in \mathbb{Z}} \binom{n}{(n+kN)/2} 2^{-n}$$

где биномиальный коэффициент интерпретируется как 0, если $n + kN/2 \notin \{0, 1, \dots, n\}$. Это, конечно, равно тому же значению, что вычислено ранее.

Упражнение 0.9. Пусть $P = (p_{x,y})_{x,y \in S}$ — матрица, где S конечно. Докажите, что следующее эквивалентно:

- P является матрицей переходных вероятностей (также называемой стохастической), т.е. $p_{x,y} \geq 0$ для всех $x, y \in S$ и $\sum_y p_{x,y} = 1$ для всех $x \in S$.
- Для каждой неотрицательной функции $f: S \rightarrow \mathbb{R}$, Pf неотрицательна. Более того, $P\mathbf{1} = \mathbf{1}$, где $\mathbf{1}$ — функция, тождественно равная 1.
- Для каждого вероятностного распределения μ на S , μP также является вероятностным распределением.

Что изменится, если S счётно, но не конечно?

Упражнение 0.10. Проверьте, что если вероятности перехода $p_{x,y} = m_y$ не зависят от x , то независимо от начального распределения X_0 , последовательность $X_1, X_2 \dots$ является i.i.d. (независимой и одинаково распределённой).

Найдите пример (для конкретного начального условия), где (X_0, X_1, \dots) являются i.i.d., несмотря на то, что $p_{x,y}$ зависит от x .

i Абстракция

Стандартная конструкция цепи Маркова

Попробуем найти базовые ингредиенты для общего определения цепей Маркова с теоретико-мерной точки зрения. Основным моментом заключается в том, что мы можем взять в качестве пространства состояний S общее измеримое пространство, хотя мы будем брать его Польским пространством, наделенным его Борелевской σ -алгеброй, чтобы избежать любых проблем с **регулярностью**. Тогда вероятности перехода будут измеримыми отображениями $S \ni x \mapsto p(x, \cdot) \in \mathcal{P}(S)$. Другими словами, когда цепь Маркова находится в точке $x \in S$, она перейдет в множество $A \subset S$ с вероятностью $p(x, A)$. Здесь *измеримость* означает, что отображение $x \mapsto p(x, A)$ измеримо как отображение из S в \mathbb{R} . Все конструкции и результаты, представленные здесь, обобщаются на этот случай.

Более сложный слой абстракции

Можно попытаться вписать цепи Маркова в более общую структуру, где X_t живет в разном пространстве для каждого t , а множество Θ моментов времени — просто частично упорядоченное множество. Существует не так много актуальных применений Θ за пределами подмножеств \mathbb{Z} и \mathbb{R} , заметным исключением являются компактные подмножества в \mathbb{C} , упорядоченные по включению.

Определение 0.4 (Измеримое расслоение). $(E, \Theta, \pi, (\mathcal{G}_t)_{t \in \Theta})$ является **измеримым расслоением**, если:

- E — непустое множество, называемое **тотальным пространством**.
- Θ — непустое множество, называемое **базой**.
- $\pi: E \rightarrow \Theta$ — сюръективное отображение.
- \mathcal{G}_t — σ -алгебра на $\pi^{-1}(\{t\})$, для $t \in \Theta$.

Сечение — это отображение $\mathbf{X}: \Theta \rightarrow E$, которое является правым обратным к π , а именно $\pi(X_t) = t$ для $t \in \Theta$. Сечение можно эквивалентно рассматривать как коллекцию $\mathbf{X} = (X_t)_{t \in \Theta}$ с $X_t \in \pi^{-1}(\{t\})$. Пространство сечений обозначается $D(E)$.

Определение 0.5 (Стохастическое сечение). Пусть (Ω, \mathcal{F}) — **измеримое пространство**, а $(E, \Theta, \pi, (\mathcal{G}_t)_{t \in \Theta})$ — измеримое расслоение. Отображение $\mathbf{X}: \Omega \rightarrow D(E)$ называется **стохастическим сечением**, если отображение $\Omega \ni \omega \mapsto X_t(\omega) \in \pi^{-1}(\{t\})$ является $\mathcal{F}/\mathcal{G}_t$ -измеримым.

Определение 0.6 (Марковский процесс). Пусть $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \Theta}, \mathbb{P})$ — **фильтрованное вероятностное пространство**, а $(E, \Theta, \pi, (\mathcal{G}_t)_{t \in \Theta})$ — измеримое расслоение. Стохастическое сечение \mathbf{X} называется **Марковским процессом**, если

$$\mathbb{P}(X_t \in A \mid \mathcal{F}_s) = \mathbb{P}(X_t \in A \mid X_s) \quad \forall s \leq t, A \in \mathcal{G}_t$$

Если мы определим $S_t := \pi^{-1}(\{t\})$, то вероятности перехода являются отображениями $p^{(s,t)}: S_s \rightarrow \mathcal{P}(S_t)$, задаваемыми⁴

$$p^{(s,t)}(x, A) := \mathbb{P}(X_t \in A \mid X_s = x)$$

⁴Мы предполагаем минимальную регулярность пространства S_t , чтобы гарантировать, что $\mathbb{P}(X_t \in A \mid X_s)$ действительно является измеримой функцией от X_s , так что мы можем записать $\mathbb{P}(X_t \in A \mid X_s = x)$ однозначно.