

Дискретная теория вероятностей. Неофициальный конспект

Лектор: Михаил Анатольевич Лифшиц
Конспектировал Леонид Данилевич

II семестр, весна 2023 г.

Оглавление

1	Дискретная теория вероятностей	2
1.1	Основные определения и понятия	2
1.1.1	Вероятностное пространство. События	2
1.1.2	Взаимосвязь событий	3
1.2	Случайные величины	4
1.2.1	Схема Бернулли	4
1.2.2	Случайные блуждания	5
1.2.3	Про условные вероятности	6
1.3	Матожидание, дисперсия	7
1.3.1	Простейшие свойства матожидания	7
1.3.2	Неравенства, связанные с математическим ожиданием	9
1.3.3	Медиана	10
1.3.4	Дисперсия	10
1.3.5	Моменты	12
1.4	Законы больших чисел (ЗБЧ)	12
1.5	Производящие функции	14
1.5.1	Производящие функции и моменты	15
1.6	Ветвящиеся процессы	16
1.6.1	Процесс Гальтона — Ватсона	16
1.6.2	Некоторые другие виды процессов	17
1.7	Предельные теоремы Муавра — Лапласа	17
1.7.1	Локальная	17
1.7.2	Интегральная	18
1.8	Цепи Маркова	20
1.8.1	Инвариантные (стационарные) распределения	20
1.8.2	Классификация состояний в цепях Маркова	23
1.8.3	Периодичность	24
1.8.4	Связь периодов и эргодических классов	25
1.8.5	Возвратность	27
1.9	Случайное блуждание в \mathbb{Z}^1	29
1.9.1	Распределение максимума. Принцип отражения	30
1.9.2	Время пребывания на полуоси (закон арксинуса)	31
1.9.3	Задача о разорении игрока	32
1.9.4	Матожидание времени разорения	33
1.10	Случайные графы	33
1.10.1	Граф Эрдёша — Реньи	34
1.10.2	power law for degrees (степенной закон для степеней (вершин))	34
1.10.3	Дерево преимущественного присоединения	34
1.10.4	Распределение степеней вершин	35

Глава 1

Дискретная теория вероятностей

Лекция I

14 февраля 2023 г.

1.1 Основные определения и понятия

1.1.1 Вероятностное пространство. События

Рассмотрим конечное или счётное множество Ω .

Элементы множества $\omega \in \Omega$ называются *элементарными исходами*, само множество Ω называется *пространством элементарных исходов*.

Всякое подмножество $A \subset \Omega$ является *событием*.

Введём функцию $p : \Omega \rightarrow \mathbb{R}_{\geq 0}$, сопоставляющую элементарному исходу «его вероятность». Необходимым и достаточным условием является $\sum_{\omega \in \Omega} p(\omega) = 1$. Так как $p(\omega) \geq 0$, то сумма конечного или счётного числа слагаемых корректно определена. А именно, сумма счётного числа слагаемых либо расходится при любой перестановке слагаемых, либо сходится к одному и тому же числу.

Определение 1.1.1 (Вероятностное пространство). Пространство элементарных исходов Ω с заданной на нём вероятностью $p : \Omega \rightarrow \mathbb{R}_{\geq 0}$.

Определение 1.1.2 (Вероятность события). Сумма вероятностей элементарных исходов — его элементов, как множества.

Пишут $\mathbb{P} : 2^\Omega \rightarrow \mathbb{R}$; $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega)$.

Свойства (Свойства вероятностей).

- $0 \leq \mathbb{P}(A) \leq 1$.
- $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$.
- $\mathbb{P}(A) + \mathbb{P}(\bar{A}) = 1$, где $\bar{A} \stackrel{def}{=} \Omega \setminus A$.
- $\mathbb{P}\left(\bigsqcup_{j=1}^n A_j\right) = \sum_{j=1}^n \mathbb{P}(A_j)$.

Для пересекающихся событий посчитать вероятность их объединения сложнее. Используя формулу включений-исключений, можно записать

$$\begin{aligned}\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\ \mathbb{P}(A \cup B \cup C) &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) + \mathbb{P}(A \cap B \cap C)\end{aligned}$$

и так далее.

Замечание. Иногда случается так, что все элементарные исходы равновероятны. Так как сумма их вероятностей — 1, то в таком случае $|\Omega| < \infty$, и $\forall \omega \in \Omega : p(\omega) = \frac{1}{|\Omega|}$. Отсюда получаем, $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$.

1.1.2 Взаимосвязь событий

Условная вероятность

Зафиксируем некоторое событие $B \subset \Omega$, такое, что $\mathbb{P}(B) > 0$.

Определение 1.1.3 (Условная вероятность события A (при условии B)). $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$.

Об этом удобно думать, как о вероятности того, что произошло A , *при условии* того, что произошло B .



Рис. 1.1: Про условную вероятность

Красное событие довольно вероятно, что произойдёт, но при условии того, что произошло зелёное событие, вероятность красного существенно понижается.

Интуиция за этим определением следующая: все элементарные исходы, содержащиеся в B могут как произойти, так и не произойти, но все, не содержащиеся в B — точно не произошли.

Таким образом, вероятностное пространство «сузилось», ввели новую вероятностную функцию

$$\tilde{p} : \Omega \rightarrow \mathbb{R}_{\geq 0}; \quad \tilde{p} : \omega \mapsto \begin{cases} \alpha \cdot p(\omega), & \omega \in B \\ 0, & \omega \notin B \end{cases}$$

где α — коэффициент нормировки, необходимый для условия суммирования всех вероятностей в единицу. $\sum_{\omega \in B} p(\omega) = \mathbb{P}(B)$, поэтому $\alpha = \frac{1}{\mathbb{P}(B)}$.

Независимость событий

Интуитивно, независимость событий — это когда происхождение одного события не влияет на вероятность происхождения другого.

Воспользовавшись языком условной вероятности, $\mathbb{P}(A|B) = \mathbb{P}(A)$. За определение принимают формулу, полученную из этой домножением на $\mathbb{P}(B)$ — без деления.

Определение 1.1.4 (События A и B независимы). $\mathbb{P}(A) \cdot \mathbb{P}(B) = \mathbb{P}(A \cap B)$.

Замечание. Приятным бонусом формулы оказалась симметричность относительно A и B .

Можно доказать, что независимость A и B влечёт независимость \bar{A} и \bar{B} .

Независимость множества событий бывает *попарная* и *в совокупности*.

Попарная независимость — гораздо более слабое условие, оно означает лишь независимость любой пары событий. Независимость множества событий $\mathcal{A} = \{A_1, A_2, \dots\}$ в совокупности означает

$$\forall S \subset \mathcal{A} : \prod_{A \in S} \mathbb{P}(A) = \mathbb{P}\left(\bigcap_{A \in S} A\right).$$

Контрпример (Пирамидка Бернштейна). Покажем, что попарная независимость отличается от независимости в совокупности. Рассмотрим четырёхгранную пирамидку (как кубик, только четыре грани, а не шесть), у которой грани белая, синяя, красная, бело-сине-красная.

При её броске возможны 4 элементарных исхода — выпала такая-то грань. Определим вероятностное пространство на этом множестве, введя вероятности каждого исхода $1/4$.

Рассмотрим три события W, B, R — выпала грань, на которой есть белое, синее или красное соответственно. Несложно заметить, что

$$\begin{aligned}\mathbb{P}(W) &= \mathbb{P}(B) = \mathbb{P}(R) = \frac{1}{2} \\ \mathbb{P}(W \cap B) &= \mathbb{P}(B \cap R) = \mathbb{P}(W \cap R) = \frac{1}{4} \\ \mathbb{P}(W \cap B \cap R) &= \frac{1}{4}\end{aligned}$$

1.2 Случайные величины

Определение 1.2.1 (Случайная величина). Отображение $X : \Omega \rightarrow \mathbb{R}$.

Определение 1.2.2 (Независимость случайных величин X_1, \dots, X_n). $\forall r_1, \dots, r_n \in \mathbb{R}$: события $\{X = r_1\}, \dots, \{X_n = r_n\}$ независимы.

Запись $\{X = r_1\}$ является сокращением более длинной записи $\{\omega \in \Omega | X(\omega) = r_1\}$.

1.2.1 Схема Бернулли

Пусть $n \in \mathbb{N}, p \in [0, 1]$.

Введём независимые события A_1, \dots, A_n , такие, что $\mathbb{P}(A_j) = p$. Назовём их *испытаниями*, посмотрим, какие испытания завершились «успехом» (событие произошло), а какие — нет.

Пример (Схема Бернулли для $n = 2$). Обозначим $A_1 = \{\omega_1, \omega_2\}, A_2 = \{\omega_1, \omega_3\}$. Все вероятности элементарных исходов определены условием однозначно. Так, $p(\omega_1) = \mathbb{P}(A_1 \cap A_2) \stackrel{\text{независимость}}{=} \mathbb{P}(A_1)\mathbb{P}(A_2) = p \cdot p = p^2$.

Рассмотрим случайную величины $S(\omega)$ — количество успехов.

ω	A_1	A_2	$p(\omega)$	$S(\omega)$
ω_1	Успех	Успех	p^2	2
ω_2	Успех	Неудача	$p(1-p)$	1
ω_3	Неудача	Успех	$(1-p)p$	1
ω_4	Неудача	Неудача	$(1-p)^2$	0

Посчитаем для произвольного n вероятность того, что количество успехов — ровно k . Из базовой комбинаторики очевидно, что

$$\mathbb{P}(S = k) = \sum_{\omega \in \Omega: S(\omega) = k} p(\omega) = \sum_{\omega \in \Omega: S(\omega) = k} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^{n-k}$$

Для всякой случайной величины S , удовлетворяющей формуле выше, говорят, что она подчинена *биномиальному распределению* $\mathcal{B}(n, p)$.

Заметим, что $\Omega = \bigsqcup_{k=0}^n \{S = k\}$, откуда мы получаем тождество

$$\sum_{k=0}^n \mathbb{P}(S = k) = 1 \quad \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = 1$$

Теорема 1.2.1 (Пуассон). Пусть $k \in \mathbb{N}_{\geq 0}, a \in \mathbb{R}_{>0}$.

Рассмотрим последовательность схем Бернулли с параметрами (n, p_n) , где $n \cdot p_n \xrightarrow{n \rightarrow \infty} a$.

Тогда $\mathbb{P}(S_n = k) \xrightarrow{n \rightarrow \infty} e^{-a} \frac{a^k}{k!}$. Случайные величины, удовлетворяющие этой формуле, имеют *распределение Пуассона* $\mathcal{P}(a)$.

Доказательство.

$$\mathbb{P}(S_n = k) = \frac{n!}{(n-k)!k!} p_n^k (1-p_n)^{n-k} = \frac{1}{k!} \cdot \underbrace{\frac{n!}{(n-k)! \cdot n^k}}_{\rightarrow 1} \cdot \underbrace{(np_n)^k}_{\rightarrow a} \cdot \underbrace{(1-p_n)^{\frac{1}{p_n}}}_{e^{-1}} \cdot \underbrace{p_n^{n-k}}_{\rightarrow a} \rightarrow e^{-a} \frac{a^k}{k!}$$

□

Лекция II

20 февраля 2023 г.

Введём в схеме Бернулли ещё одну случайную величину T — момент первого успеха, наименьший номер первого успешного события (и формальный элемент ∞ иначе).

$T \in \{1, \dots, n, \infty\}$. (Эта запись не совсем формальна: она означает, что T , как отображение, принимает значения в данном множестве). Несложно по определению почитать

$$\mathbb{P}(T = k) = \mathbb{P}(\overline{A_1}, \dots, \overline{A_{k-1}}, A_k) = (1-p)^{k-1} \cdot p, 1 \leq k \leq n$$

Если же ни одно испытание не закончилось успехом, то $T = \infty$, $\mathbb{P}(T = \infty) = (1-p)^n$.

Рассмотрим случай $n = \infty$. Тогда событие «ни одно испытание не закончилось успехом» исключается, а сумма вероятностей остальных событий равна 1:

$$\sum_{k=1}^{\infty} \mathbb{P}(T = k) = \sum_{k=1}^{\infty} (1-p)^{k-1} p = \frac{1}{1-(1-p)} \cdot p = 1$$

Говорят, что T имеет *геометрическое распределение*.

На самом деле дискретная теория вероятностей не позволяет создать схему Бернулли со счётным (любым бесконечным) количеством испытаний (при $0 < p < 1$). Таким образом, рассматривая случай $n = \infty$, мы ведём себя неформально, в любом случае выходя за рамки дискретной теории вероятностей.

Доказательство невозможности счётной схемы Бернулли в дискретной теории вероятностей. Рассмотрим произвольный элементарный исход ω . Если ему соответствует бесконечное число успехов, то для любого m рассмотрим m успешных событий. Пусть это какие-то фиксированные A_{i_1}, \dots, A_{i_m} . Так как они произошли, то $\mathbb{P}(m) \leq p^m$, то есть на самом деле $\mathbb{P}(\omega) = 0$. (В случае бесконечного числа неудач опять же можно оценить $\forall m \in \mathbb{N} : \mathbb{P}(\omega) \leq (1-p)^m$). Но раз вероятность каждого элементарного исхода равна 0, то они не могут суммироваться в 1, противоречие. □

Это произошло из-за того, что в схеме Бернулли со счётным числом испытаний континуум возможных исходов.

Чтобы это обойти, можно рассматривать последовательность конечных схем, как в теореме Пуассона, или же просто закрыть на это глаза — в непрерывной теории вероятностей такое распределение возможно.

1.2.2 Случайные блуждания

Введём случайные величины $S_n : S_0 = 0, S_{n+1} = S_n + X_n$, где $X_n = \begin{cases} 1, & \text{с вероятностью } p \\ -1, & \text{с вероятностью } 1-p \end{cases}$, и все $\{X_n\}$ независимы.

Это та же схема Бернулли, просто успехам соответствуют движения в положительную сторону оси, и неудачам — в отрицательную.

Исследуем распределение S_n . Очевидно, возможные значения S_n — это $[-n; n]$, причём $k \equiv n \pmod{2}$.

Событие $\{S_n = k\}$ эквивалентно событию « m величин равны 1 (остальные -1)», где $k = m - (n - m) \Rightarrow m = \frac{n+k}{2}$.

Отсюда согласно схеме Бернулли получаем $\mathbb{P}(S_n = k) = \binom{n}{(n+k)/2} p^{(n+k)/2} (1-p)^{(n-k)/2}$.

В симметричном случае, при $p = 1/2$ формула упрощается, $\mathbb{P}(S_n = k) = \binom{n}{(n+k)/2} \cdot \frac{1}{2^n}$.

1.2.3 Про условные вероятности

Вероятность происхождения A при условии B : $\mathbb{P}(A|B) = \mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ (при $\mathbb{P}(B) > 0$).

Применение условных вероятностей:

- Вычисление вероятностей вложенных событий. Пусть $A_1 \supset A_2 \supset \dots A_n$.

$$\mathbb{P}(A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \dots \cdot \mathbb{P}(A_n|A_{n-1})$$

Доказательство. По индукции.

$$\mathbb{P}(A_n) \underset{\text{события вложены}}{=} \mathbb{P}(A_n \cap A_{n-1}) = \mathbb{P}(A_{n-1}) \cdot \mathbb{P}(A_n|A_{n-1})$$

□

- Формула полной вероятности Пусть вероятностное пространство Ω разбито на конечное (или счётное) число дизъюнктивных событий H_1, \dots, H_n .



Рис. 1.2: Разбиение вероятностного пространства

Рассмотрим произвольное событие $A \subset \Omega$.

$$\mathbb{P}(A) = \sum_{i=1}^n \mathbb{P}(A \cap H_i) = \sum_{i=1}^n \mathbb{P}(A|H_i) \mathbb{P}(H_i)$$

- Формула Байеса. Пусть вероятностное пространство Ω разбито на конечное (или счётное) число дизъюнктивных событий H_1, \dots, H_n . Теперь мы хотим узнать вероятность H_i для некоего i при условии наступления события A .

Запишем

$$\mathbb{P}(H_i|A) \underset{\text{по определению}}{=} \frac{\mathbb{P}(H_i \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|H_i) \mathbb{P}(H_i)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A|H_i) \mathbb{P}(H_i)}{\sum_{j=1}^n \mathbb{P}(A|H_j) \mathbb{P}(H_j)}$$

1.3 Матожидание, дисперсия

Говорят, что X и Y *одинаково распределены*, если $\forall r \in \mathbb{R} : \mathbb{P}(X = r) = \mathbb{P}(Y = r)$. Например, в схеме Бернулли из 6 испытаний случайные величины «количество успехов на первых 2 испытаниях» и «количество успехов на последних 2 испытаниях» одинаково распределены.

Если X и Y определены на одном и том же вероятностном пространстве, то можно определить арифметические действия (сумму, произведение...) случайных величин, как соответствующие арифметические действия над отображениями поточечно.

Определение 1.3.1 (Математическое ожидание случайной величины X). Обозначается

$$\mathbb{E}X \stackrel{def}{=} \sum_{\omega \in \Omega} X(\omega)p(\omega)$$

Математическое ожидание довольно неплохо описывает случайную величину одним числом: (задача 1.3.1).

После приведения подобных членов, можно записать $\mathbb{E}(X) = \sum_r \mathbb{P}(X = r)r$. Если Ω конечно, то сумма считается; если же Ω — бесконечное вероятностное пространство, то матожидание может быть не определено, как сумма бесконечного ряда (тем не менее, сумма всегда существует, если X всегда принимает неотрицательные значения). Чтобы было удобно оперировать с матожиданиями, будем считать, что матожидание определено, если и только если ряд сходится **абсолютно**.

Чтобы исследовать существование $\mathbb{E}X$, введём функции положительной и отрицательной частей числа:

$$x_+ \stackrel{def}{=} \max\{x, 0\}$$
$$x_- \stackrel{def}{=} \max\{-x, 0\}$$

Несложно видеть, что равенство $x = x_+ - x_-$ выполнено всегда.

Посчитав матожидание положительной и отрицательной частей X , $\mathbb{E}(X_+)$ и $\mathbb{E}(X_-)$, можно утверждать, что $\mathbb{E}(X)$ существует, если и только если хотя бы одно из $\mathbb{E}(X_+)$ и $\mathbb{E}(X_-)$ конечно. Если же $\mathbb{E}(X_+) = \mathbb{E}(X_-) = +\infty$, то $\mathbb{E}(X)$ не определено. (Если ровно одно из $\mathbb{E}(X_+)$ или $\mathbb{E}(X_-)$ бесконечно, то $\mathbb{E}X$ тоже можно мыслить, как бесконечность того или иного знака)

1.3.1 Простейшие свойства матожидания

- $X \geq 0 \Rightarrow \mathbb{E}(X) \geq 0$.
- $\forall c \in \mathbb{R} : \mathbb{E}(cX) = c\mathbb{E}(X)$.
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$.

Доказательство.

$$\begin{aligned} \mathbb{E}(X) + \mathbb{E}(Y) &= \sum_{r_1} r_1 \mathbb{P}(X = r_1) + \sum_{r_2} r_2 \mathbb{P}(Y = r_2) = \\ &= \sum_{r_1} r_1 \sum_{r_2} \mathbb{P}(X = r_1 \wedge Y = r_2) + \sum_{r_2} r_2 \sum_{r_1} \mathbb{P}(X = r_1 \wedge Y = r_2) = \\ &= \sum_{r_1, r_2} (r_1 + r_2) \mathbb{P}(X = r_1 \wedge Y = r_2) \end{aligned} \quad \square$$

Здесь важно заметить, что X и Y лишь должны быть определены на одном вероятностном событии; они не обязаны быть, например, независимы.

- $X \geq Y \Rightarrow \mathbb{E}(X) \geq \mathbb{E}(Y)$. Для доказательства можно записать $Y = X + (Y - X)$. Тогда $\mathbb{E}Y = \mathbb{E}X + \mathbb{E}(Y - X)$.

Примеры (Матожидания случайных величин).

- X имеет распределение Бернулли с параметром p , записываемое $\mathcal{B}(p)$. Это по определению значит

$$\begin{cases} \mathbb{P}(X = 1) = p \\ \mathbb{P}(X = 0) = 1 - p \end{cases}$$

В таком случае $\mathbb{E}X = 1 \cdot \mathbb{P}(X = 1) + 0 \cdot \mathbb{P}(X = 0) = p$.

- Пусть S имеет распределение $\mathcal{B}(n, p)$ — число успехов в схеме Бернулли.

$$\mathbb{P}(S = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Матожидание S можно посчитать по определению: $\mathbb{E}S = \sum_{k=0}^n k \binom{n}{k} p^k (1 - p)^{n-k}$.

Но это неоправданно сложно. Для упрощения работы запишем $S = \mathbb{1}_1 + \dots + \mathbb{1}_n$, где $\mathbb{1}_1, \dots, \mathbb{1}_n$ — индикаторы событий A_1, \dots, A_n соответственно. По определению $\mathbb{1}_i = \begin{cases} 1, & A_i \text{ успешно} \\ 0, & A_i \text{ неуспешно} \end{cases}$

Каждый индикатор по отдельности имеет распределение Бернулли с параметром p , таким образом,

$$\mathbb{E}S = \sum_{i=1}^n \mathbb{E}X_i = n \cdot p$$

- Пусть X имеет распределение Пуассона $\mathcal{P}(a)$:

$$\mathbb{P}(X = k) = e^{-a} \frac{a^k}{k!}$$

Матожидание такой случайной величины равно

$$\sum_{k=0}^{\infty} k e^{-a} \frac{a^k}{k!} = e^{-a} \cdot a \sum_{k=1}^{\infty} \frac{a^{k-1}}{(k-1)!} = e^{-a} a e^a = a$$

Оказывается, параметр a в Пуассоновском распределении — матожидание данной случайной величины.

Лекция III

27 февраля 2023 г.

Известно, что $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$. Верно ли, что $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$?

Выберем в качестве X величину, распределённую по закону $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$.

В качестве Y возьмём эту же случайную величину: $Y = X$.

Тогда замечаем, что $\mathbb{E}X = 0, \mathbb{E}Y = 0, \mathbb{E}XY = \mathbb{E}X^2 = 1$, равенство не выполняется. «Увы, так устроен мир»

К счастью, можно наложить дополнительные условия, а именно, о **независимости** случайных величин X и Y .

В таком случае формула выполняется:

$$\mathbb{P}(X = r_1, Y = r_2) \underset{\text{определение независимости}}{=} \mathbb{P}(X = r_1) \cdot \mathbb{P}(Y = r_2)$$

откуда

$$\sum_{r_1, r_2} r_1 r_2 \mathbb{P}(X = r_1, Y = r_2) = \sum_{r_1, r_2} r_1 r_2 \mathbb{P}(X = r_1) \cdot \mathbb{P}(Y = r_2) = \left(\sum_{r_1} r_1 \mathbb{P}(X = r_1) \right) \left(\sum_{r_2} r_2 \mathbb{P}(Y = r_2) \right) = \mathbb{E}X \cdot \mathbb{E}Y$$

Конечно, можно доказать по индукции формулу для любого конечного числа сомножителей:

$$\mathbb{E}(X_1 \cdot \dots \cdot X_n) = \prod_{i=1}^n \mathbb{E}X_i$$

для независимых событий X_1, \dots, X_n .

Рассмотрим следующую задачу, показывающее, что матожидание — число, наилучшим образом приближает случайную величину:

Задача 1.3.1. Дана случайная величина $X : \mathbb{E}X^2 < \infty$. Надо найти число r , минимизирующее $\mathbb{E}((X - r)^2)$.

Значит, надо минимизировать $\mathbb{E}(X^2 - 2rX + r^2) = \mathbb{E}(X^2) - 2r\mathbb{E}(X) + r^2$. Это квадратный трёхчлен по r , минимум достигается при $r = \mathbb{E}(X)$.

1.3.2 Неравенства, связанные с математическим ожиданием

Пусть $f : \mathbb{R} \rightarrow \mathbb{R}$ — неубывающая неотрицательная функция.

Факт 1.3.1. $\forall X$ — случайная величина и $\forall r \in \mathbb{R}$ имеет место неравенство:

$$\mathbb{P}(X \geq r) \leq \frac{\mathbb{E}f(X)}{f(r)}$$

Доказательство. Рассмотрим вторую функцию $g(x) = \begin{cases} 0, & x < r \\ f(r), & x \geq r \end{cases}$. Несложно проверить, что $g(x) \leq f(x)$. Отсюда $g(X) \leq f(X)$ ($f(X)$ — композиция двух функций), и, как следствие, $\mathbb{E}(g(X)) \leq \mathbb{E}(f(X))$. Но несложно видеть, что $\mathbb{E}(g(X)) = 0 \cdot \mathbb{P}(X < r) + f(r) \cdot \mathbb{P}(X \geq r) = f(r) \cdot \mathbb{P}(X \geq r)$, и неравенство выполнено. \square

- **Следствие 1.3.1** (Экспоненциальное неравенство Чебышёва). Рассмотрим $f(x) = e^{\lambda x}$, где $\lambda > 0$.

$$\text{Тогда } \mathbb{P}(X \geq r) \leq \frac{\mathbb{E}(e^{\lambda X})}{e^{\lambda r}}.$$

Более того, здесь возможна более сильная форма — оптимизация по λ :

$$\mathbb{P}(X \geq r) \leq \inf_{\lambda > 0} \frac{\mathbb{E}(e^{\lambda X})}{e^{\lambda r}}$$

- **Следствие 1.3.2** (Неравенство Маркова). $\forall r > 0 : \mathbb{P}(|X| \geq r) \leq \frac{\mathbb{E}(|X|)}{r}$.

Доказательство. Применим неравенство (факт 1.3.1) для $f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases}$ и случайной величины $|X|$. Получим

$$\mathbb{P}(|X| \geq r) \leq \frac{\mathbb{E}f(|X|)}{r} = \frac{\mathbb{E}|X|}{r}$$

что и требовалось доказать. \square

- **Следствие 1.3.3.** $\mathbb{P}(|X| \geq r) \leq \frac{\mathbb{E}(X^2)}{r^2}$

Доказательство. Следует из предыдущего применением $\mathbb{P}(|X| \geq r) \iff \mathbb{P}(X^2 \geq r^2)$. \square

Замечание. Несмотря на то, что это практически то же, что и выше, в мире случайных величин нам будет удобно оценивать не случайную величину, а её квадрат.

- **Следствие 1.3.4** (Вероятностное неравенство Йенсена). Пусть X — случайная величина с конечным матожиданием, а $\phi : \mathbb{R} \rightarrow \mathbb{R}$ выпукла вниз (как x^2).

Тогда $\mathbb{E}(\phi(X)) \geq \phi(\mathbb{E}X)$. (картинка, где X принимает два значения).

Доказательство. Пусть X принимает конечное число значений. Тогда

$$\mathbb{E}(\phi(X)) = \sum_{\omega \in \Omega} p(\omega) \phi(X(\omega)) \geq_{\text{по неравенству Йенсена}} \phi \left(\sum_{\omega \in \Omega} p(\omega) X(\omega) \right) = \phi(\mathbb{E}X)$$

Если X принимает счётное число значений, то можно устроить предельный переход. □

1.3.3 Медиана

Ещё одно число, которым можно характеризовать случайную величину — *медиана*.

Определение 1.3.2 (Медиана случайной величины X). Такое число m , что $\mathbb{P}(X \geq m) \geq \frac{1}{2}$ и $\mathbb{P}(X \leq m) \geq \frac{1}{2}$.

1. Можно доказать, что медиана (в отличие от матожидания) всегда существует.
2. Медиана необязательно единственна. Так, в случае случайной величины X , распределённой по закону $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$ медианой является любое число $m \in [-1, 1]$.
3. Пусть X — случайная величина, такая, что $\mathbb{P}(X = -1) = \mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{3}$. Единственная медиана — это 0, причём $\mathbb{P}(X \geq 0) = \frac{2}{3}$, и $\mathbb{P}(X \leq 0) = \frac{2}{3}$ тоже.
4. На самом деле, медиана — плохая метрика, которой никто не пользуется. Так, только матожидание линейно: медиана суммы вообще не выражается через медианы слагаемых.
5. *Интересный факт.* Если в задаче (задача 1.3.1) заменить $\mathbb{E}((X - r)^2)$ на $\mathbb{E}(|X - r|)$, то минимизирующим r окажется не матожидание, но медиана.

1.3.4 Дисперсия

«Слово дисперсия знакомо тем, кто имеет дело с садоводством. Садоводы используют так называемую дисперсионную краску»

Вообще говоря, дисперсия описывает «меру разброса» данной случайной величины.

Пусть X — случайная величина, такая, что $\mathbb{E}(X^2) < \infty$.

Определение 1.3.3 (Дисперсия X). $\mathbb{D}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$.

Докажем эквивалентность двух определений:

Доказательство.

$$\mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X^2 - 2X\mathbb{E}X + (\mathbb{E}X)^2) = \mathbb{E}(X^2) - 2\mathbb{E}X \cdot \mathbb{E}X + (\mathbb{E}X)^2 = \mathbb{E}(X^2) - (\mathbb{E}X)^2 \quad \square$$

Замечание. В англоязычных текстах дисперсию обозначают $\text{Var}(X)$ — от слова Variance.

1. $\mathbb{D}(X) \geq 0$, как матожидание неотрицательной величины.
2. У константы нет дисперсии: $\mathbb{D}(C) = 0$
3. Из определения очевидно $\mathbb{D}(X + C) = \mathbb{D}X$.
4. Из определения очевидно $\mathbb{D}(C \cdot X) = C^2 \cdot \mathbb{D}(X)$. В частности, $\mathbb{D}(-X) = \mathbb{D}(X)$.
5. Аддитивность: для **независимых** случайных величин X, Y : $\mathbb{D}(X + Y) = \mathbb{D}(X) + \mathbb{D}(Y)$.

Доказательство.

$$\begin{aligned}\mathbb{D}(X+Y) &= \mathbb{E}(X+Y)^2 - (\mathbb{E}(X+Y))^2 = \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}X + \mathbb{E}Y)^2 = \\ &= (\mathbb{E}X^2 - (\mathbb{E}X)^2) + (\mathbb{E}Y^2 - (\mathbb{E}Y)^2) + \underbrace{(2\mathbb{E}(XY) - 2\mathbb{E}X \cdot \mathbb{E}Y)}_{0 \text{ из-за независимости}} = \mathbb{D}(X) + \mathbb{D}(Y) \quad \square\end{aligned}$$

6. Определение дисперсии без вычитания матожидания: пусть X, X' независимы и одинаково распределены.

$$\text{Тогда } \mathbb{D}X = \frac{1}{2}\mathbb{E}(X - X')^2.$$

Доказательство.

$$\frac{1}{2}\mathbb{E}(X - X')^2 = \frac{1}{2}\mathbb{E}(X^2 + X'^2 - 2XX') = \frac{1}{2}(\mathbb{E}X^2 + \mathbb{E}X'^2 - 2(\mathbb{E}X \cdot \mathbb{E}X')) = \mathbb{E}X^2 - (\mathbb{E}X)^2 \quad \square$$

7. «Элементарное, но нетривиальное свойство».

Пусть $f: \mathbb{R} \rightarrow \mathbb{R}$ — 1-липшицева функция, то есть $|f(x) - f(y)| \leq |x - y|$.

Тогда для любой случайной величины $X: \mathbb{D}(f(X)) \leq \mathbb{D}(X)$.

Доказательство. Воспользоваться свойством $\mathbb{D}X = \frac{1}{2}\mathbb{E}(X - X')^2$, а также тем, что $(X - X')^2 \geq (f(X) - f(X'))^2$ (поточечно). □

8. **Факт 1.3.2** (Неравенство Чебышёва). Пусть X — случайная величина, такая, что $\mathbb{D}(X) < \infty$.

$$\text{Тогда } \mathbb{P}(|X - \mathbb{E}X| \geq r) \leq \frac{\mathbb{D}X}{r^2}.$$

Доказательство.

$$\mathbb{P}(|X - \mathbb{E}X| \geq r) \leq \frac{\mathbb{E}(X - \mathbb{E}X)^2}{r^2} = \frac{\mathbb{D}X}{r^2} \quad \square$$

(следствие 1.3.3)

Замечание (О единицах измерения). Если случайная величина принимает значения некоей размерности (рубли, очки, километры), то матожидание имеет ту же размерность, а дисперсия — размерности квадрата измеряемой величины. Чтобы избавиться от такого неудобства, вводят *среднеквадратическое отклонение*.

Определение 1.3.4 (Среднеквадратическое отклонение случайной величины X). $\sigma(X) \stackrel{\text{def}}{=} \sqrt{\mathbb{D}(X)}$.

Пример. Пусть X имеет распределение Пуассона $\mathcal{P}(a)$.

По формуле $\mathbb{D}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$ получаем, что для вычисления дисперсии надо получить $\mathbb{E}(X^2)$ (нам уже известно, что $(\mathbb{E}X)^2 = a^2$).

Необыкновенным образом получаем, что легче посчитать $\mathbb{E}(X(X-1)) = \mathbb{E}(X^2) - \mathbb{E}X$.

$$\mathbb{E}(X(X-1)) = \sum_{k=2}^{\infty} k(k-1) \frac{a^k}{k!} = a^2 \sum_{k=2}^{\infty} e^{-a} \frac{a^{k-2}}{(k-2)!} = a^2$$

Отсюда $\mathbb{E}(X^2) = \mathbb{E}(X(X-1)) + \mathbb{E}(X) = a^2 + a$, и, наконец, $\mathbb{D}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = a + a^2 - a^2 = a$.

Лекция IV

6 марта 2023 г.

Пусть X, Y — случайные величины.

Определение 1.3.5 (Ковариация X и Y).

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$$

Про ковариацию говорят, что это *мера линейной зависимости* X и Y .

Ковариация билинейна (линейна по обоим аргументам) и симметрична.

Определение 1.3.6 (X и Y некоррелированы). Ковариация X и Y равна 0, т. е. $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

В частности, независимые величины с конечным матожиданием модуля некоррелированы. Из ковариации следует, что для некоррелированных случайных величин $X, Y : \mathbb{D}(X + Y) = \mathbb{D}(X) + \mathbb{D}(Y)$.

Когда говорят про некоррелированность случайных величин, то имеют в виду попарную некоррелированность.

1.3.5 Моменты

Для $k \in \mathbb{N}$ определяют k -й момент случайной величины X , он по определению равен $\mathbb{E}(X^k)$. Для $k = 1$ это матожидание.

Также определяют k -й центральный момент случайной величины X , он по определению равен $\mathbb{E}(X - \mathbb{E}X)^k$. Для $k = 2$ это дисперсия.

Если в определении звучит слово *абсолютный*, то матожидание берётся от модуля аргумента (k -й абсолютный момент, k -й абсолютный центральный момент).

k -й момент однороден — при домножении случайной величины на c он домножается на c^k или $|c|^k$. Для чётных k абсолютные моменты совпадают с обычными.

1.4 Законы больших чисел (ЗБЧ)

Если сложить много случайных величин, то в сумме получится что-то близкое к сумме их матожиданий.

Теорема 1.4.1 (Закон больших чисел Чебышёва). Пусть X_1, X_2, \dots, X_n — некоррелированные случайные величины, такие, что $\mathbb{E}X_i^2 < \infty$. Запишем это как $\exists \sigma \in \mathbb{R} : \sup_i \mathbb{D}X_i \leq \sigma^2$.

Тогда

$$\forall \varepsilon > 0 : \quad \mathbb{P} \left(\left| \frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n \mathbb{E}X_i}{n} \right| > \varepsilon \right) \xrightarrow{n \rightarrow \infty} 0$$

Доказательство.

$$\mathbb{P} \left(\left| \frac{\sum_{i=1}^n X_i}{n} - \frac{\sum_{i=1}^n \mathbb{E}X_i}{n} \right| > \varepsilon \right) = \mathbb{P} \left(\left| \sum_{i=1}^n (X_i - \mathbb{E}X_i) \right| > n\varepsilon \right)$$

Согласно неравенству Чебышёва (факт 1.3.2), это оценивается следующим образом:

$$\mathbb{P} \left(\left| \sum_{i=1}^n (X_i - \mathbb{E}X_i) \right| > n\varepsilon \right) \leq \frac{\mathbb{D} \left(\sum_{i=1}^n X_i \right)}{(n\varepsilon)^2} \leq \frac{n\sigma^2}{(n\varepsilon)^2} \xrightarrow{n \rightarrow \infty} 0$$

□

Следствие 1.4.1. Пусть X_1, \dots, X_n — независимые одинаково распределённые случайные величины.

Если $\mathbb{E}X_i^2 < \infty, \mathbb{E}X_i = a$, то

$$\forall \varepsilon > 0 : \quad \mathbb{P} \left(\left| \frac{\sum_{i=1}^n X_i}{n} - a \right| \geq \varepsilon \right) \xrightarrow{n \rightarrow \infty} 0$$

Замечание. В заключении следствия ничего не говорится про второй момент величин X_j , и, на самом деле, следствие как теорема верно и без оценки $\mathbb{E}X_i^2$ в посылке. Это мы докажем через пару лет совсем не тривиальной математикой.

Следствие 1.4.2 (Закон больших чисел Бернулли). Пусть S_n — число успехов в схеме Бернулли с параметрами n, p . Тогда

$$\forall \varepsilon > 0 : \quad \mathbb{P} \left(\left| \frac{S_n}{n} - p \right| > \varepsilon \right) \xrightarrow{n \rightarrow \infty} 0$$

На самом деле, в 1613 году Бернулли доказал закон больших чисел, названный позднее в честь него, используя довольно сложные вычисления.

Лишь только в 1870 году Чебышёв доказал общий закон больших чисел и следствие из него.

Докажем полученными средствами теорему из матанализа, не использующую в своей формулировке ничего случайного.

Теорема 1.4.2 (Вейерштрасс). Пусть $f : [0, 1] \rightarrow \mathbb{R}$ — непрерывная функция. Тогда

$$\exists \{P_n\}_{n=1}^{\infty} : \max_{t \in [0, 1]} |f(t) - P_n(t)| \xrightarrow{n \rightarrow \infty} 0$$

Доказательство.

Лемма 1.4.1 (О математических ожиданиях). Пусть $\{Z_n\}_{n=1}^{\infty}$ — последовательность случайных величин, такая, что $\exists a \in \mathbb{R}$:

$$\forall \varepsilon > 0 : \mathbb{P}(|Z_n - a| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0$$

Пусть дана функция f , заданная в окрестности точки a , непрерывная в a и ограниченная неким числом $M \in \mathbb{R}$.

Тогда $\mathbb{E}(f(Z_n)) \xrightarrow{n \rightarrow \infty} f(a)$.

Доказательство леммы.

$$\begin{aligned} |\mathbb{E}f(Z_n) - f(a)| &= |\mathbb{E}(f(Z_n) - f(a))| && \leq \\ &&& \text{например, по неравенству Йенсена для модуля} \\ &\leq \mathbb{E}|f(Z_n) - f(a)| \leq \mathbb{E} \left[\underbrace{|f(Z_n) - f(a)| \cdot \chi_{\{|Z_n - a| \geq \varepsilon\}}}_{\leq 2M \cdot \mathbb{P}(|Z_n - a| \geq \varepsilon)} + \underbrace{|f(Z_n) - f(a)| \cdot \chi_{\{|Z_n - a| < \varepsilon\}}}_{\leq w(f, a, \varepsilon)} \right] \end{aligned}$$

где $w(f, a, \varepsilon) = \sup_{|s-a| < \varepsilon} |f(s) - f(a)|$. В силу непрерывности f это сходится к 0 при $\varepsilon \rightarrow 0$.

Устремив ε к нулю, получаем, что $|\mathbb{E}f(Z_n) - f(a)| < \delta(\varepsilon)$, где $\delta(\varepsilon) \xrightarrow{\varepsilon \rightarrow 0} 0, \delta > 0$.

Левая часть не зависит от ε , получается $|\mathbb{E}f(Z_n) - f(a)| = 0$. □

Рассмотрим последовательность случайных величин S_n — число успехов в схеме Бернулли с параметрами (n, p) , где p — фиксированное число из $[0, 1]$.

Согласно закону больших чисел Бернулли $\mathbb{P}(|\frac{S_n}{n} - p| > \varepsilon) \xrightarrow{n \rightarrow \infty} 0$.

Применим лемму для p и f : $\mathbb{E}f(\frac{S_n}{n}) \xrightarrow{n \rightarrow \infty} f(p)$. Подставим определение математического ожидания, отсюда

$$\sum_{k=0}^n \mathbb{P}(S_n = k) f\left(\frac{k}{n}\right) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} f\left(\frac{k}{n}\right) \xrightarrow{n \rightarrow \infty} f(p)$$

Осталось сказать, что сходимость к $f(p)$ равномерна при всех $p \in [0, 1]$. Для этого улучшим оценку: заметим, что из леммы на самом деле следует, что

$$\left| \mathbb{E}f\left(\frac{S_n}{n}\right) - f(p) \right| \leq 2M \cdot \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq \varepsilon\right) + w(f, p, \varepsilon)$$

Первое слагаемое оценивается сверху в виде

$$2M \cdot \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq \varepsilon\right) = 2M \cdot \mathbb{P}(|S_n - np| \leq n\varepsilon) \leq 2M \frac{np(1-p)}{(n\varepsilon^2)} \leq \frac{2M}{n\varepsilon^2}$$

Чтобы показать, что $|\mathbb{E}f(\frac{S_n}{n}) - f(p)| \leq \delta$, выберем $\varepsilon > 0$ такой, что $\forall p \in [0, 1] : w(f, p, \varepsilon) < \frac{\delta}{2}$ (это можно сделать, так как согласно теореме Кантора непрерывная на отрезке функция равномерно непрерывна), затем выберем настолько большое n , что $\frac{2M}{n\varepsilon^2} \leq \frac{\delta}{2}$. \square

1.5 Производящие функции

Пусть X — случайная величина, принимающая целые неотрицательные значения.

Определение 1.5.1 (Производящая функция величины X). Степенной ряд

$$\phi_X(z) = \mathbb{E}(z^X) = \sum_{k=0}^{\infty} \mathbb{P}(X = k) z^k$$

Так как $\sum_{k=0}^{\infty} \mathbb{P}(X = k) = 1$, то ряд сходится при $|z| \leq 1$.

При рассмотрении производящих функций мы будем брать аргументы $z \in [0, 1]$.

Заметим, что $\phi_X(0) = \mathbb{P}(X = 0)$, $\phi_X(1) = 1$, а сама функция неубывает и выпукла вниз (как x^2). Это следует из того, что $\phi_X(z)$ — линейная комбинация стандартных мономов, каждый из которых неубывает и выпуклый вниз.

Если X и Y независимы, то $\phi_{X+Y}(z) = \phi_X(z) \phi_Y(z)$.

Доказательство.

$$\phi_{X+Y}(z) = \mathbb{E}(z^{X+Y}) = \mathbb{E}(z^X \cdot z^Y) \underset{X \text{ и } Y \text{ независимы}}{=} \mathbb{E}(z^X) \cdot \mathbb{E}(z^Y) = \phi_X(z) \cdot \phi_Y(z)$$

\square

Лекция V

13 марта 2023 г.

Обобщим данную формулу.

1. Пусть X_1, \dots, X_n независимы. Тогда $\phi_{S_n}(z) = \prod_{j=1}^n \phi_{X_j}(z)$, где $S_n := \sum_{j=1}^n X_j$ — тоже случайная величина.

2. В частности, если X_1, \dots, X_n независимы и одинаково распределены, то $\phi_{S_n}(z) = \phi_{X_1}(z)^n$.
3. Пусть X_1, \dots, X_n, \dots — независимы (независимо любое конечное подмножество) и одинаково распределены. Пусть $N \in \mathbb{N}_0$ — случайная величина (формальнее, $N : \Omega \rightarrow \mathbb{N}_0$, где Ω — вероятностное пространство), не зависящая от всех X -ов.

Положим $S := \sum_{i=1}^N X_i$.

Тогда $\phi_S(z) = \phi_N(\phi_{X_1}(z))$.

Замечание. Предыдущий пункт — частный случай данного. В самом деле, для неслучайной величины N , всегда равной n , производящая функция равна z^n .

Доказательство.

$$\phi_S(z) = \sum_{k=0}^{\infty} \mathbb{P}(S = k) z^k = \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \mathbb{P}(S = k, N = n) z^k = \mathbb{P}(S_n = k, N = n) z^k =$$

Воспользуемся независимостью, продолжив равенство

$$\begin{aligned} &= \sum_{k=0}^{\infty} \sum_{n=0}^{\infty} \mathbb{P}(S_n = k) \mathbb{P}(N = n) z^k = \sum_{n=0}^{\infty} \mathbb{P}(N = n) \underbrace{\sum_{k=0}^{\infty} \mathbb{P}(S_n = k) z^k}_{\phi_{S_n}(z)} = \\ &= \sum_{n=0}^{\infty} \mathbb{P}(N = n) \cdot \phi_{X_1}(z)^n = \phi_S(\phi_{X_1}(z)) \end{aligned}$$

□

1.5.1 Производящие функции и моменты

Предложение 1.5.1. $\phi_X^{(k)}(1) = \mathbb{E}(X(X-1) \cdot \dots \cdot (X-k+1))$.

В частности, для $k = 1$: $\phi_X'(1) = \mathbb{E}X$; для $k = 2$: $\phi_X''(1) = \mathbb{E}(X(X-1)) = \mathbb{E}X^2 - \mathbb{E}X$.

Доказательство. Докажем для $k = 1$.

Формально продифференцировав ряд, получаем $\phi_X'(z) = \left(\sum_{k=0}^{\infty} \mathbb{P}(X = k) z^k \right)' = \sum_{k=1}^{\infty} \mathbb{P}(X = k) k \cdot z^{k-1}$.

При подстановке $z = 1$ действительно получается $\mathbb{E}X$, но надо обосновывать, почему производная ряда в граничной точке круга сходимости равна сумме производных слагаемых ряда. □

Другой вариант доказательства. Данный вариант тяжелее в смысле выкладок, но легче — в смысле теорем, на которые опирается доказательство.

Рассмотрим $z \in (0, 1)$, близкое к единице.

$$\frac{\phi_X(1) - \phi_X(z)}{1 - z} = \frac{1 - \phi_X(z)}{1 - z} = \frac{1 - \sum_{k=0}^{\infty} \mathbb{P}(X = k) z^k}{1 - z} = \sum_{k=0}^{\infty} \mathbb{P}(X = k) \frac{1 - z^k}{1 - z}$$

По теореме Коши найдутся точки $\tilde{z}_k \in (z, 1)$, такие, что $\frac{1 - z^k}{1 - z} = k \tilde{z}_k^{k-1}$.

Отсюда получаем оценку $\frac{\phi_X(1) - \phi_X(z)}{1 - z} \leq \sum_{k=0}^{\infty} \mathbb{P}(X = k) \cdot k$ (пользуемся тем, что все $\tilde{z}_k \leq 1$). В пределе

$$\phi_X'(X) \leq \sum_{k=0}^{\infty} \mathbb{P}(X = k) \cdot k.$$

Чтобы получить оценку с другой стороны, заменим сумму на конечную, совершив предельный переход, получим $\phi'_X(X) \geq \sum_{k=0}^K \mathbb{P}(X = k) \cdot k \cdot \tilde{z}^{k-1}$. Устремив z к единице, получаем оценку $\phi'_X(X) \geq \sum_{k=0}^K \mathbb{P}(X = k) \cdot k$, затем можно перейти к предельному переходу по $K \rightarrow \infty$. \square

Замечание. Производная бесконечна \iff матожидание бесконечно.

1.6 Ветвящиеся процессы

1.6.1 Процесс Гальтона — Ватсона

График: в момент времени $t = 0$ есть частица (человек, электрон), которая в каждый момент времени порождает случайное число потомков.

Получается, если можно так выразиться, дерево. Будем считать, что числа потомков у каждой частицы — независимые одинаково распределённые случайные величины.

Гальтон и Ватсон интересовались генеалогией знатных родов, но потом внезапно оказалось, что процесс прекрасно описывает ядерные реакции.

Определение 1.6.1 (Процесс Галтона — Ватсона). Пусть $(X_{n,j})_{n \geq 0, j \geq 1}$ — независимые одинаково распределённые случайные величины.

Последовательность случайных величин определяется формулой $M_0 = 1$, $M_{n+1} = \sum_{j=1}^{M_n} X_{n,j}$ и называется *ветвящимся процессом*.

Согласно рекурсивной формуле, M_n не зависит от $X_{n,1}, X_{n,2}, \dots$.

Значит, $\phi_{M_{n+1}}(z) = \phi_{M_n}(\phi_X(z))$, где ϕ_X — производящая функция любой из величин $X_{n,j}$.

Получаем $\phi_{M_0}(z) = z$, $\phi_{M_1}(z) = \phi_X(z)$, $\phi_{M_2}(z) = \phi_X(\phi_X(z))$. Вообще, $\phi_{M_n}(z) = \phi_X^{\circ n}(z)$.

Задача о выживании и вырождении ветвящегося процесса

Определим вероятность того, что на n -м шаге процесс не выжил $q_n = \mathbb{P}(M_n = 0)$.

Очевидно, $q_{n+1} \geq q_n$, так как если процесс выродился, то так потом и будет, но он может выродиться на $n + 1$ -м шаге впервые.

Так как $q_n \leq 1$, то последовательность $\{q_n\}_{n \in \mathbb{N}}$ имеет предел q .

Говорят, что *процесс вырождается*, если $q = 1$.

Нарисуем график $\phi_X(z)$ при $z \in [0, 1]$.

Предложение 1.6.1. q — наименьший корень уравнения $\phi_X(z) = z$.

Доказательство. Рассмотрим M — множество корней уравнения. $1 \in M$, M замкнуто — прообраз нуля некоторого непрерывного отображения.

Отсюда следует, что в M существует наименьший элемент z_* .

Так как $q_n = \mathbb{P}(M_n = 0) = \phi_{M_n}(0) = \phi_X^{\circ n}(0)$, то $q_{n+1} = \phi_X^{\circ n+1}(0) = \phi_X(q_n)$.

Запишем $0 \leq z_*$, откуда $\phi_X(0) \leq \phi_X(z_*) = z_*$. Так можно применять много раз, получаем $\forall n \in \mathbb{N} : \phi_X^{\circ n}(0) \leq z_*$.

Перейдя к пределу у $q_{n+1} = \phi_X(q_n)$ получаем $q = \phi_X(q)$.

Используя $q \leq z_*$ и $\phi_X(q) = q$, получаем $q = z_*$. \square

Обозначим $m = \mathbb{E}X$ — среднее число потомков частицы.

Теорема 1.6.1. Процесс M_n не вырождается \iff либо $m > 1$, либо $X = 1$ всегда, то есть X — величина неслучайная.

Доказательство.

- Рассмотрим $m > 1$. При z , близком к единице, $\phi_X(z) = 1 - m(1 - z) + o(1 - z)$, что при z достаточно близких к 1 меньше z .

Таким образом, нашлась точка $z : \phi_X(z) < z$. С другой стороны, $\phi_X(0) \geq 0$, значит, существует корень уравнения $\phi_X(z) = z$, строго меньший единицы. Отсюда следует, что процесс не вырождается.

- Рассмотрим $m < 1$. Функция $\phi_X(z)$ выпукла вниз, поэтому $\forall z \in [0, 1] : \phi_X(z) \geq 1 + m(z - 1)$.

Таким образом, единственный корень уравнения $\phi_X(z) = z$ — $z = 1$.

- Рассмотрим $m = 1$. Касательная прямая к $\phi_X(z)$ проходит по диагонали $y = z$.

Рассмотрим наименьший корень уравнения $\phi_X(z) = z$. Есть варианты:

1. Касание единицы происходит только в самом конце: $\phi_X(z) > z$ для $z < 1$. Это случай вырождения процесса.
2. $\forall z \in [0, 1] : \phi_X(z) = z$. Процесс не вырождается, $X_{n,j} = 1$ всегда.
3. Остался один случай, которого не бывает. Для некоего $a \in (0, 1)$, совпадение $\phi_X(z) = z$ происходит только при $z \in [a, 1]$.

На самом деле, с производящими функциями такое невозможно: если $\phi_X(z) = z$ в окрестности 1, то $\phi_X''(1) = 0$.

Но $\phi_X''(1) = \mathbb{E}(X(X - 1)) = \mathbb{E}X^2 - \mathbb{E}X$, а мы знаем, что $\mathbb{E}X = 1$. Получается, $\mathbb{E}X^2 = 1$, и дисперсия этой величины нулевая: $\mathbb{D}X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = 0$. Таким образом, X — величина неслучайная.

□

1.6.2 Некоторые другие виды процессов

Процессы Беллмана — Харриса

Отличие от процессов Гальтона — Ватсона состоит в том, что каждый субъект живёт случайное время. В конце своего жизненного времени частица распадается на случайное количество частиц.

Многотиповые процессы

Распределение числа потомков зависит от типа данной частицы: синяя частица порождает либо два синие, либо две красные, а красная — одну жёлтую, и, возможно, одну зелёную.

Процессы с иммиграцией

На каждом поколении число частиц меняется каким-то фиксированным образом — частицы «прибывают откуда-то снаружи».

1.7 Предельные теоремы Муавра — Лапласа

1.7.1 Локальная

Запишем число успехов в схеме Бернулли $\mathcal{B}(n, p)$. Зафиксируем p и изучим $\mathbb{P}(S_n = k)$ для «типичных» значений k при $n \rightarrow \infty$.

Вспомним, что $\mathbb{E}S_n = np$, $\mathbb{D}S_n = np(1-p)$ для любого $n \in \mathbb{N}$.

Так как дисперсия — это квадрат «типичного отклонения», то для некой константы C величина S_n должна часто отклоняться от своего матожидания не больше, чем на $C\sqrt{n}$.

Определение 1.7.1 (Последовательности $A_{n,k}$ и $B_{n,k}$ равномерно эквивалентны при $n \rightarrow \infty$ на некоторой области $k \in C_n$).

$$\max_{k \in C_n} \left| \frac{A_{n,k}}{B_{n,k}} - 1 \right| \xrightarrow{n \rightarrow \infty} 0$$

Теорема 1.7.1 (Локальная предельная теорема Муавра — Лапласа). Локальность означает, что в рассмотрении находится фиксированное k .

Пусть последовательность ε_n стремится к нулю. Утверждается, что

$$\mathbb{P}(S_n = k) \sim \frac{1}{\sqrt{2\pi np(1-p)}} \cdot \exp \left\{ -\frac{(k-np)^2}{2p(1-p)n} \right\}$$

равномерно по области $\{k \in \mathbb{N} \mid |k-np| \leq \varepsilon_n \cdot n^{2/3}\}$. «Название теоремы — историческое недоразумение. Теорему Муавра — Лапласа доказал Муавр, а Лаплас — лишь включил её в свой учебник. Впрочем, к распространению этой теоремы он всё-таки имел какое-то отношение»

Доказательство. Запишем

$$\begin{aligned} \mathbb{P}(S_n = k) &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{(n-k)!k!} p^k (1-p)^k \\ \frac{n!}{(n-k)!k!} p^k (1-p)^k &\sim \frac{(n/e)^n \sqrt{2\pi n} \cdot p^k (1-p)^{n-k}}{((n-k)/e)^{n-k} \sqrt{2\pi(n-k)} \cdot (k/e)^k \sqrt{2\pi k}} \sim \frac{n^n p^k (1-p)^{n-k}}{\sqrt{2\pi np(1-p)} \cdot (n-k)^{n-k} k^k} \end{aligned}$$

Преобразовав ещё чуть-чуть выражение, получаем

$$\frac{1}{\sqrt{2\pi np(1-p)}} \cdot \frac{n^n p^k (1-p)^{n-k}}{(n-k)^{n-k} k^k} = \frac{1}{\sqrt{2\pi np(1-p)}} \cdot \left(\frac{np}{k}\right)^k \cdot \left(\frac{n(1-p)}{n-k}\right)^{n-k}$$

Определим новую переменную v таким образом: $k = np + v$. В таком случае $\left(\frac{k}{np}\right)^k = \left(\frac{np+v}{np}\right)^{np+v} = \left(1 + \frac{v}{np}\right)^{np+v} = \exp \left(\log \left(1 + \frac{v}{np}\right) (np+v) \right)$. Разложим \log в ряд с точностью до второго члена: $\exp \left(\left(\frac{v}{np} - \frac{v^2}{2(np)^2} + \mathcal{O} \left(\frac{v^3}{(np)^3} \right) \right) (np+v) \right) = \exp \left(v - \frac{v^2}{2np} + \frac{v^2}{np} - \frac{v^3}{2(np)^2} + \mathcal{O} \left(\frac{v^3}{(np)^2} \right) \right) = \exp \left(v + \frac{v^2}{2np} + o(1) \right)$. Слагаемое под \mathcal{O} стремится к нулю, так как $|v| \leq \varepsilon_n \cdot n^{2/3}$ по условию на рассматриваемую область k .

Таким образом, $\left(\frac{np}{k}\right)^k = \exp \left(-(k-np) - \frac{(k-np)^2}{2np} + o(1) \right)$. Аналогично (подставив $p \rightsquigarrow (1-p)$; $k \rightsquigarrow (n-k)$; $v \rightsquigarrow -v$) получаем $\left(\frac{n(1-p)}{n-k}\right)^{n-k} = \exp \left(-(np-k) - \frac{(np-k)^2}{2n(1-p)} + o(1) \right)$

Перемножив, получаем

$$\left(\frac{np}{k}\right)^k \cdot \left(\frac{n(1-p)}{n-k}\right)^{n-k} = \exp \left\{ -\frac{(k-np)^2}{2n} \left(\frac{1}{p} + \frac{1}{1-p} \right) + o(1) \right\} = \exp \left\{ -\frac{(np-k)^2}{2np(1-p)} \right\} + o(1)$$

□

1.7.2 Интегральная

Что можно сказать о вероятности попадания числа успехов в определённый интервал?

Теорема 1.7.2 (Интегральная теорема Муавра — Лапласа). Пусть $a < b$.

$$\mathbb{P}\left(S_n \in \left[np + a\sqrt{p(1-p)n}; np + b\sqrt{p(1-p)n}\right]\right) \sim \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$$

Вероятность переписывается в виде $\mathbb{P}\left(a \leq \frac{S_n - np}{\sqrt{p(1-p)n}} \leq b\right) = \mathbb{P}\left(a \leq \frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{D}S_n}} \leq b\right)$

Доказательство.

$$\mathbb{P}\left(S_n \in \left[np + a\sqrt{p(1-p)n}; np + b\sqrt{p(1-p)n}\right]\right) = \sum_{k \in \left[np + a\sqrt{p(1-p)n}; np + b\sqrt{p(1-p)n}\right]} \mathbb{P}(S_n = k)$$

Так как $k - np \sim \mathcal{O}(\sqrt{n})$, то все последующие оценки равномерны по k .

$$\sum_k \mathbb{P}(S_n = k) \sim \sum_k \frac{1}{\sqrt{2np(1-p)}} \exp\left\{-\frac{(k - np)^2}{2p(1-p)n}\right\}$$

Слагаемые в сумме можно заменить на эквивалентные, так как оценка равномерна. Заменим сумму интегралом: для начала покажем $\frac{1}{\sqrt{2np(1-p)}} \exp\left\{-\frac{(k - np)^2}{2p(1-p)n}\right\} \sim \int_k^{k+1} \frac{1}{\sqrt{2np(1-p)}} \exp\left\{-\frac{(x - np)^2}{2p(1-p)n}\right\} dx$.

Покажем корректность этой эквивалентности, заменив $x = k + \theta$. $\frac{1}{\sqrt{2np(1-p)}} \exp\left\{-\frac{(x - np)^2}{2p(1-p)n}\right\} = \frac{1}{\sqrt{2np(1-p)}} \exp\left\{-\frac{(k - np)^2 + 2(k - np)\theta + \theta^2}{2p(1-p)n}\right\}$ Так как $(k - np)\theta = \mathcal{O}(\sqrt{n})$, то этими слагаемыми действительно можно пренебречь — знаменатель порядка n , эти слагаемые — $o(1)$.

$$\sum_k \mathbb{P}(S_n = k) \sim \int_{np + a\sqrt{np(1-p)}}^{np + b\sqrt{np(1-p)}} \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(x - np)^2}{2np(1-p)}\right) dx + o(1)$$

Сделаем замену переменной: $u = \frac{x - np}{\sqrt{np(1-p)}}$. Тогда $du = \frac{dx}{\sqrt{np(1-p)}}$.

Интеграл упрощается до $\frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{u^2}{2}\right) du$ □

Следствие 1.7.1.

$$\mathbb{P}\left(S_n \leq np + b\sqrt{np(1-p)}\right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{x^2}{2}} dx$$

Доказательство.

- Докажем, что $\lim_{n \rightarrow \infty} \mathbb{P}(S_n < np) = \frac{1}{2}$. Для этого покажем $\forall \varepsilon > 0 : \left| \lim_{n \rightarrow \infty} \mathbb{P}(S_n < np) - \frac{1}{2} \right| \leq \varepsilon$.

Воспользуемся тем, что $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1$. Значит, найдётся $M > 0$:

$$\frac{1}{\sqrt{2\pi}} \int_{-M}^0 e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_0^{+M} e^{-\frac{x^2}{2}} dx \geq \frac{1}{2} - \varepsilon$$

Отсюда

$$\begin{aligned} \mathbb{P}\left(np - M\sqrt{np(1-p)} \leq S_n < np\right) &\leq \mathbb{P}(S_n < np) \leq 1 - \mathbb{P}\left(np \leq S_n \leq np + M\sqrt{np(1-p)}\right) \\ &\downarrow n \rightarrow \infty \\ \int_{-M}^0 e^{-\frac{x^2}{2}} dx &\leq \lim_{n \rightarrow \infty} \mathbb{P}(S_n < np) \leq 1 - \int_0^M e^{-\frac{x^2}{2}} dx \end{aligned}$$

- Теперь осталось посчитать $\mathbb{P}\left(S_n \leq np + b\sqrt{np(1-p)}\right)$. Без потери общности $b \geq 0$, тогда

$$\begin{aligned} \mathbb{P}\left(S_n \leq np + b\sqrt{np(1-p)}\right) &= \mathbb{P}(S_n < np) + \mathbb{P}\left(np \leq S_n \leq np + b\sqrt{np(1-p)}\right) \xrightarrow{n \rightarrow \infty} \\ &\rightarrow \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^b e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^b e^{-\frac{x^2}{2}} dx \end{aligned}$$

□

Интересный факт. Интеграл в правой части описывает нормальное распределение, он не берётся.

Теорема Леви «выросла» из интегральной теоремы Муавра — Лапласа.

Интересный факт (Теорема Леви). Пусть X_1, \dots, X_n, \dots — независимо распределённые случайные величины, $S_n := X_1 + \dots + X_n$. Предположим, что $\mathbb{E}X_j^2 < \infty$ для любого j .

Тогда для $\forall a < b$: $\mathbb{P}\left(a \leq \frac{S_n - \mathbb{E}S_n}{\sqrt{\mathbb{D}S_n}} \leq b\right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx$.

Лекция VI

22 марта 2023 г.

1.8 Цепи Маркова

Лекция пропущена.

Лекция VII

27 марта 2023 г.

Было: \mathcal{X} — множество состояний. $X_0, X_1, \dots \in \mathcal{X}$. $\pi_n(x) = \mathbb{P}(X_n = x), x \in \mathcal{X}$. Вероятность перехода $p(x \rightarrow y) = \mathbb{P}(X_{n+1} = y | X_n = x)$. π_0, p определяют состояние цепи. $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) = \pi_0 p(x_0 \rightarrow x_1) \cdot \dots \cdot \pi_n = \pi_0 \cdot p^n$.

1.8.1 Инвариантные (стационарные) распределения

Определение 1.8.1 (Распределение на множестве \mathcal{X}). Такое отображение $\pi : \mathcal{X} \rightarrow [0, 1]$, что $\sum_{x \in \mathcal{X}} \pi(x) = 1$.

Определение 1.8.2 (Инвариантное распределение). Такое распределение π , что $\pi \cdot p = \pi$.

$$\forall y \in \mathcal{X} : \pi(y) = \sum_{x \in \mathcal{X}} \pi(x) p(x \rightarrow y).$$

Если π_0 инвариантно, то $\forall n \geq 0 : \pi_n = \pi \cdot p^n = \pi_0$. Следует из ассоциативности умножения матриц.

Примеры.

- «Хороший пример»: блуждание по конечному неориентированному графу.

Обозначим за E общее число рёбер, $\deg x$ — число рёбер, инцидентных x . Очевидно. $\sum_{x \in \mathcal{X}} \deg x = 2E$.

Рассмотрим цепь Маркова, где $\forall y : p(x \rightarrow y) = \begin{cases} \frac{1}{\deg x}, & \exists(x, y) \\ 0, & \nexists(x, y) \end{cases}$.

Выберем распределение $\pi(x) = \frac{\deg(x)}{2E}$. Покажем, что оно инвариантно:

$$\frac{\deg(y)}{2E} = \pi(y) = \sum_{x \in \mathcal{X}} \pi(x) p(x \rightarrow y) = \sum_{x \in \mathcal{X}, \exists(x, y)} \frac{\deg(x)}{2E} \cdot \frac{1}{\deg x} = \sum_{x \in \mathcal{X}, \exists(x, y)} \frac{1}{2E} = \frac{\deg y}{2E}$$

- «Плохой пример»: случайное блуждание на множестве целых чисел \mathbb{Z} . Вероятности перехода $p(n \rightarrow n+1) = p(n \rightarrow n-1) = \frac{1}{2}$.

Граф бесконечный. и это всё разрушает. Поищем инвариантное распределение. Пусть это π .

Тогда $\pi(y) = \frac{1}{2}(\pi(y-1) + \pi(y+1))$. Отсюда можно выразить $\forall y \in \mathbb{Z} : \pi(y) = \pi(0) + ky$, где k — некая константа. Несложно видеть, что во всех трёх случаях ($k < 0, k > 0, k = 0$) π не является распределением.

Таким образом, для случайного блуждания на \mathbb{Z} нет инвариантного распределения.

Теорема 1.8.1 (Марков). Пусть \mathcal{X} — конечная цепь, причём вероятность любого перехода ненулевая: $\delta := \min_{x, y \in \mathcal{X}} p(x \rightarrow y) > 0$.

Тогда $\exists \pi$ — такое распределение, что

$$\forall x, y \in \mathcal{X}, n \in \mathbb{N} : |p^n(x \rightarrow y) - \pi(y)| \leq (1 - \delta)^n \quad (1.1)$$

При этом π — единственное инвариантное распределение цепи. Любое начальное распределение π_0 влечёт $\pi_n \xrightarrow{n \rightarrow \infty} \pi$.

Доказательство. В предположении истинности (1.1) получаем

$$\pi_n(y) = (\pi_0 p^n)(y) = \sum_{x \in \mathcal{X}} \pi_0(x) p^n(x \rightarrow y) \xrightarrow{n \rightarrow \infty} \underbrace{\left(\sum_{x \in \mathcal{X}} \pi_0(x) \right)}_1 \pi(y) = \pi(y)$$

Предположим, что $\tilde{\pi}$ — произвольное инвариантное распределение. Рассмотрим цепь для $\pi_0 = \tilde{\pi}$. С одной стороны, в таком случае $\forall n \in \mathbb{N} : \pi_n = \tilde{\pi}$. С другой стороны, $\pi_n \xrightarrow{n \rightarrow \infty} \pi$. Значит, $\tilde{\pi} = \pi$. Таким образом, все инвариантные распределения совпадают с π .

Теперь докажем что-то. Запишем в координатном виде $p^{n+1} = p^n \cdot p$.

$$\begin{aligned} p^{n+1}(x \rightarrow y) &= \sum_{z \in \mathcal{X}} p^n(x \rightarrow z) p(z \rightarrow y) \\ &\downarrow n \rightarrow \infty \\ \pi(y) &= \sum_{z \in \mathcal{X}} \pi(z) p(z \rightarrow y) \end{aligned}$$

Интересно, что мы доказали?

Покажем, что π — распределение, то есть сумма $\sum_{x \in \mathcal{X}} \pi(x) = 1$. Для любого фиксированного $x \in \mathcal{X}$

$$1 = \sum_{y \in \mathcal{X}} p^n(x \rightarrow y) \xrightarrow{n \rightarrow \infty} \sum_{y \in \mathcal{X}} \pi(y)$$

Осталось доказать (1.1). Зафиксируем $y \in \mathcal{X}$. Рассмотрим последовательности $m_n = \min_{x \in \mathcal{X}} p^n(x \rightarrow y)$ и $M_n = \max_{x \in \mathcal{X}} p^n(x \rightarrow y)$.

m_n неубывает, M_n невозрастает:

$$p^{n+1}(x \rightarrow y) = \sum_{z \in \mathcal{X}} p(x \rightarrow z) p^n(z \rightarrow y)$$

Так как $p^n(z \rightarrow y) \in [m_n, M_n]$, то $p^{n+1}(x \rightarrow y)$, как барицентрическая комбинация таких вероятностей, тоже лежит в $[m_n, M_n]$. Отсюда действительно m_n неубывает, M_n невозрастает.

Ещё докажем их сближение: $(M_{n+1} - m_{n+1}) \leq (1 - \delta)(M_n - m_n)$: Выберем такие x_1, x_2 , что максимум и минимум достигаются: $M_{n+1} = p^{n+1}(x_1 \rightarrow y)$, $m_{n+1} = p^{n+1}(x_2 \rightarrow y)$.

$$M_{n+1} - m_{n+1} = p^{n+1}(x_1 \rightarrow y) - p^{n+1}(x_2 \rightarrow y) = \sum_{z \in \mathcal{X}} [p(x_1 \rightarrow z) - p(x_2 \rightarrow z)] p^n(z \rightarrow y)$$

Оценим эту сумму следующим образом:

$$\sum_{z \in \mathcal{X}} [p(x_1 \rightarrow z) - p(x_2 \rightarrow z)] p^n(z \rightarrow y) \leq \sum_{z \in \mathcal{X}} [p(x_1 \rightarrow z) - p(x_2 \rightarrow z)]_+ M_n - \sum_{z \in \mathcal{X}} [p(x_1 \rightarrow z) - p(x_2 \rightarrow z)]_- m_n$$

Покажем равенство

$$\sum_{z \in \mathcal{X}} [p(x_1 \rightarrow z) - p(x_2 \rightarrow z)]_+ = \sum_{z \in \mathcal{X}} [p(x_1 \rightarrow z) - p(x_2 \rightarrow z)]_-$$

Это верно, так как

$$\begin{aligned} & \sum_{z \in \mathcal{X}} [p(x_1 \rightarrow z) - p(x_2 \rightarrow z)]_+ - \sum_{z \in \mathcal{X}} [p(x_1 \rightarrow z) - p(x_2 \rightarrow z)]_- = \\ & \sum_{z \in \mathcal{X}} (p(x_1 \rightarrow z) - p(x_2 \rightarrow z)) = \sum_{z \in \mathcal{X}} p(x_1 \rightarrow z) - \sum_{z \in \mathcal{X}} p(x_2 \rightarrow z) = 1 - 1 = 0 \end{aligned}$$

Таким образом

$$M_{n+1} - m_{n+1} \leq \sum_{z \in \mathcal{X}} [p(x_1 \rightarrow z) - p(x_2 \rightarrow z)]_+ (M_n - m_n)$$

Если все слагаемые $[p(x_1 \rightarrow z) - p(x_2 \rightarrow z)]_+$ равны нулю, то доказывать нечего. Иначе найдётся положительное слагаемое $p(x_1 \rightarrow z) - p(x_2 \rightarrow z) > 0$. Согласно определению $\delta : p(x_1 \rightarrow z) - p(x_2 \rightarrow z) \leq p(x_1 \rightarrow z) - \delta$.

Доказали сближение $(M_{n+1} - m_{n+1}) \leq (1 - \delta)(M_n - m_n)$.

Таким образом, m_n неубывает, M_n невозрастает, $M_n - m_n \xrightarrow{n \rightarrow \infty} 0$. Назначим за $\pi(y)$ общий предел последовательностей m_n и M_n . Так как $|p^n(x \rightarrow y) - \pi(y)| \leq M_n - m_n \leq (1 - \delta)^n$, то (1.1) доказана. \square

Примеры (Теорема Маркова здесь неприменима).

- «Бесконечно плохой пример»: случайное блуждание на квадрате из четырёх вершин. Вероятность перехода в диагонально противоположную вершину равна 0, вероятности $p^n(x \rightarrow y)$ не сходятся — они периодически меняются с $\frac{1}{2}$ до 0.
- Случайное блуждание по пятиугольнику из пяти вершин. Есть рёбра с вероятностью перехода 0, напрямую теорема неприменима. Но здесь за четыре шага можно попасть в любую вершину: $\forall x, y : p^4(x \rightarrow y) > 0$.

Факт 1.8.1. Пусть цепь Маркова такова, что для некоторого $m \in \mathbb{N} : \forall x, y \in \mathcal{X} : p^m(x \rightarrow y) > 0$. Тогда $\exists!$ инвариантное распределение $\pi : p^n(x \rightarrow y) \xrightarrow{n \rightarrow \infty} \pi(y)$.

Доказательство. Доказательство Маркова применимо к прореженной цепи X_0, X_m, \dots с матрицей перехода p^m . Согласно ему, $p^{mn}(x \rightarrow y) \xrightarrow{n \rightarrow \infty} \pi(y)$.

$$p^k(x \rightarrow y) = p^{mn+l}(x \rightarrow y) = \sum_{z \in \mathcal{X}} p^l(x \rightarrow z) p^{mn}(z \rightarrow y) \xrightarrow{n \rightarrow \infty} \underbrace{\left(\sum_{z \in \mathcal{X}} p^l(x \rightarrow z) \right)}_1 \pi(y) = \pi(y)$$

□

Лекция VIII

3 апреля 2023 г.

1.8.2 Классификация состояний в цепях Маркова

Рассмотрим для примера цепь Маркова на таком графе:



Существенные и несущественные состояния

Определение 1.8.3 (Состояние y достижимо из x). Существует такая последовательность состояний x_0, \dots, x_m , такая, что

$$x_0 = x; x_m = y; \quad p(x_i \rightarrow x_{i+1}) > 0$$

Обозначается $x \cdots \rightarrow y$

Определение 1.8.4 (Существенное состояние x). $\forall y$ такого, достижимого из x , можно вернуться:

$$(x \cdots \rightarrow y) \Rightarrow (y \cdots \rightarrow x)$$

Пример. A — единственное несущественное состояние в графе в начала раздела.

Факт 1.8.2. Из существенного состояния можно перейти только в существенное.

Доказательство. От противного: $\exists z \in \mathcal{X} : (y \cdots \rightarrow z) \wedge \neg(z \cdots \rightarrow y)$. Тогда в частности $\neg(z \cdots \rightarrow x)$, но $x \cdots \rightarrow z$. Противоречие. □

Факт 1.8.3. В конечной цепи Маркова всегда найдётся хотя бы одно существенное состояние.

Доказательство. Рассмотрим цепочку состояний. Если $x_0 \in \mathcal{X}$ (произвольный элемент) — существенное состояние, то доказывать нечего. Иначе выберем x_1 как такое состояние, что $x_0 \cdots \rightarrow x_1$, но не наоборот.

Так дальше продолжим цепочку: $x_n \cdots \rightarrow x_{n+1}$. От противного: пусть она стала бесконечной, никакие состояния в ней не оказались существенными. Если в какой-то момент окажется, что $x_i = x_j$, то значит мы нашли цикл $x_i \cdots \rightarrow \dots \rightarrow x_j$, и получили противоречие. □

Контрпример. В бесконечной цепи $p(n \rightarrow n+1) = p(n \rightarrow n+2) = \frac{1}{2}$ существенных состояний нет. Но, она, конечно, бесконечная.

На множестве существенных состояний можно ввести отношение эквивалентности:

$$x \sim y \iff x \cdots \rightarrow y \vee x = y$$

Симметричность: по определению того, что x — существенное состояние: $(x \cdots \rightarrow y) \Rightarrow (y \cdots \rightarrow x)$. Транзитивность и рефлексивность очевидны из определения.

Следствие 1.8.1. *Множество существенных состояний распадается на классы достижимых — эргодические классы.*

Факт 1.8.4. *Каждый эргодический класс замкнут: из любого эргодического класса нельзя выйти.*

Доказательство. Из всякого x из данного эргодического класса можно попасть только в существенные y , которые по определению эквивалентны x . \square

Пример. В графе выше эти классы — треугольник BCD и четырёхугольник $EFGH$.

Определение 1.8.5 (Неприводимая цепь Маркова). В данной цепи нет замкнутых множеств кроме всего пространства \mathcal{X} .

Рассмотрим произвольное состояние $x \in \mathcal{X}$. По определению, множество точек, достижимых из x (обозначим его T_x), замкнуто.

В неприводимой цепи $\forall x \in \mathcal{X} : T_x = \mathcal{X}$, значит, эквивалентным определением неприводимой цепи является то, что из любого состояния можно добраться до любого другого.

В частности, в неприводимой цепи все состояния — существенны, образуют один эргодический класс.

1.8.3 Периодичность

Рассмотрим произвольное состояние $x \in \mathcal{X}$, обозначим $I_x := \{k \in \mathbb{N} \mid p^k(x \rightarrow x) > 0\}$. Будем считать, что I_x непустое.

Определение 1.8.6 (Период состояния x). $d(x) = \gcd(I_x)$.

Замечание. Для произвольного x : I_x — полугруппа по сложению.

Факт 1.8.5. *Существует конечное подмножество $I'_x \subset I_x$, такое, что $\gcd(I_x) = \gcd(I'_x)$.*

Доказательство. Положим $d_M := \gcd(I_x \cap [1, M])$. С ростом M последовательность множеств увеличивается по включению, d_M убывает.

Так как d_M — натуральные числа, то последовательность стабилизируется: $\exists M_0 : \forall M > M_0 : \gcd(I_x \cap [1, M]) = d_{M_0}$. Очевидно, в таком случае $I_x \cap [1, M]$ — искомое подмножество. \square

Факт 1.8.6. $\exists k_0 \in \mathbb{N}$:

$$\{k \cdot d(x) \mid k \in \mathbb{N}, k \geq k_0\} \subset I_x \subset \{k \cdot d(x) \mid k \in \mathbb{N}\}$$

Доказательство. Правое включение очевидно верно независимо от k_0 .

1. Найдём конечное множество $I'_x \subset I_x$, такое, что $\gcd(I'_x) = d(x)$.
2. Найдём линейную комбинацию элементов I'_x , такую, что $d(x) = \sum_j v_j \lambda_j, v_j \in I'_x, \lambda_j \in \mathbb{Z}$.
3. Выберем $b := \sum_j v_j |\lambda_j|$. $b \in I_x$, как линейная комбинация его элементов с неотрицательными коэффициентами $|\lambda_j|$.
Значит, b представимо в виде $b = \beta \cdot d(x)$.

4. Заметим, что $(\beta + 1)d(x) = \sum_j v_j \cdot (\lambda_j + |\lambda_j|)$, что опять-таки линейная комбинация с неотрицательными коэффициентами, лежит в I_x .
5. Рассмотрим достаточно большое $k \in \mathbb{N}$. Разделив на β с остатком, получаем $k = r\beta + v$, где $0 \leq v < \beta$.

$$k = r\beta + v(\beta + 1) - v\beta = (r - v)\beta + v(\beta + 1)$$

Для $r - v \geq 0$, например, для $k \geq \beta^2$: $k \cdot d(x) \in I_x$, как линейная комбинация $\beta \cdot d(x)$ и $(\beta + 1) \cdot d(x)$.

Таким образом, $k_0 = \beta^2$ подходит.

□

Следствие 1.8.2. В частности, $\exists k \in \mathbb{N} : kd \in I_x \wedge (k + 1) \cdot d(x) \in I_x$ (например, $k = \beta$).

Факт 1.8.7. Если два состояния сообщаются: $x \cdots \rightarrow y$ и $y \cdots \rightarrow x$, то $d(x) = d(y)$.

Доказательство. Пусть $p^a(x \rightarrow y) > 0$. Воспользуемся (следствие 1.8.2) применительно к y : есть два цикла, содержащих y , длин $k \cdot d(y)$ и $(k + 1) \cdot d(y)$.

Тогда $a + k \cdot d(y) \in I_x$ и $a + (k + 1) \cdot d(y) \in I_x$ тоже. Отсюда сразу получаем $d(x) \mid d(y) = (a + (k + 1) \cdot d(y) - a - k \cdot d(y))$. Аналогично $d(y) \mid d(x)$, значит они равны. □

1.8.4 Связь периодов и эргодических классов

Для произвольного эргодического класса $C \subset \mathcal{X}$: $x, y \in C \Rightarrow d(x) = d(y)$.

Доказательство. x и y сообщаются, так как они в одном эргодическом классе. □

Циклические подклассы

Рассмотрим один эргодический класс, например, $C = \{E, F, G, H\}$. Заметим, что для $C_0 = \{E, G\}$ и $C_1 = \{F, H\}$: из одного класса на следующем шаге можно попасть только в другой.

Пусть C — эргодический класс с периодом d . Тогда существует разбиение $C = C_0 \sqcup C_1 \sqcup \cdots \sqcup C_{d-1}$, такое, что вероятность перехода из C_i в $C_{(i+1) \pmod d}$ равна 1.

Иными словами, $\forall x \in C_i : p(x \rightarrow y) > 0 \Rightarrow y \in C_{(i+1) \pmod d}$. Это называется *разбиением на циклические подклассы*.

Доказательство. Выберем произвольное $x_0 \in C$. Для всякого $y \in C$ найдём такое $l(y) : p^{l(y)}(x_0 \rightarrow y) > 0$.

Положим $j(y) = l(y) \pmod d$ ($0 \leq j(y) < d$).

Определим $\forall j = 0..d - 1 : C_j := \{y \in C \mid j(y) = j\}$. Ясно, что $\bigcup_j C_j = C$.

Заметим, что если $p(y \rightarrow z) > 0$, то $p^{l(y)}(x_0 \rightarrow y) > 0 \Rightarrow p^{l(y)+1}(x_0 \rightarrow z) > 0$. Значит, действительно, $p(x \rightarrow y) > 0 \Rightarrow y \in C_{(i+1) \pmod d}$.

Осталось показать, что C_j не пересекаются.

Пойдём от противного: пусть $y \in C_{j_1} \cap C_{j_2}$. Тогда $\exists l_1, l_2 \in \mathbb{N} : p^{l_1}(x_0 \rightarrow y) > 0, p^{l_2}(x_0 \rightarrow y) > 0$. Так как x_0 и y в одном эргодическом классе, то для некоторого $b \in \mathbb{N} : p^b(y \rightarrow x_0)$. значит, $l_1 + b \in I_{x_0}$ и $l_2 + b \in I_{x_0}$. Значит, они оба делятся на d , их разность делится на d , значит, $l_1 \equiv l_2 \pmod d$. □

Лекция IX

10 апреля 2023 г.

Теорема 1.8.2 (Марков). Самая общая формулировка, которая у нас покамест встречалась, звучит так:

Если для конечной цепи \mathcal{X} существует $m \in \mathbb{N} : \forall x, y \in \mathcal{X} : p^m(x \rightarrow y) > 0$, то

$\exists \pi$ — распределение, такое, что $p^n(x \rightarrow y) \xrightarrow{n \rightarrow \infty} \pi$, а ещё $\pi \cdot p = \pi$ и $\forall \pi_0 : \pi_n(y) \xrightarrow{n \rightarrow \infty} \pi$.

На этой лекции мы рассмотрим ещё две теоремы, далее обобщающие теорему Маркова.

Теорема 1.8.3 (Марков, для апериодических цепей). Пусть \mathcal{X} конечно и состоит из единственного эргодического класса с периодом 1.

Утверждается, что тогда верно утверждение предыдущей теоремы (истинна посылка).

Доказательство. Пусть x — произвольное состояние. Тогда, согласно предыдущей лекции, существует достаточно большое $K(x) : \forall k \geq K(x) : p^k(x \rightarrow x) > 0$.

По определению эргодического класса, $\exists a(x, y) : p^{a(x, y)}(x \rightarrow y) > 0$. Тогда

$$\forall k \geq K(x) + a(x, y) : p^k(x \rightarrow y) > p^{k-a(x, y)}(x \rightarrow x) \cdot p^{a(x, y)}(x \rightarrow y) > 0$$

Так как $\max_{x, y} (K(x) + a(x, y))$ — конечное число, то $m := \max_{x, y} (K(x) + a(x, y))$ подойдёт. □

Замечание. Рассмотрим цепь, в которой есть один эргодический класс \mathcal{C} и много несущественных состояний, из которых достигим данный класс.

Формально, под условие теоремы эта цепь не подходит. Тем не менее, доказательство работает и здесь.

Упражнение. Если \mathcal{X} конечно, и содержит единственный эргодический класс \mathcal{C} , причём его период — 1, то утверждение теоремы Маркова тоже верно (правда, посылка в записанной форме не истинна), причём предельное распределение сосредоточено на эргодическом классе: $\sum_{y \in \mathcal{C}} \pi(y) = 1$.

Теорема 1.8.4 (Марков, для периодических цепей). Пусть \mathcal{X} конечно и состоит из единственного эргодического класса с периодом $d > 1$.

Для краткости записи обозначим $i \oplus j := (i + j \pmod{d})$.

Тогда, как уже доказано, $\mathcal{X} = \mathcal{C}_0 \sqcup \dots \sqcup \mathcal{C}_{d-1}$, таких, что

$$p(x \rightarrow y) > 0 \Rightarrow \exists j \in [0, d) : x \in \mathcal{C}_j, y \in \mathcal{C}_{j \oplus 1}$$

Утверждается, что $\exists \{\pi_j\}_{j=0}^{d-1}$ — система распределений, такая, что $\forall j : \pi_j$ сосредоточено на \mathcal{C}_j , и

$$\forall x \in \mathcal{C}_i, y \in \mathcal{C}_{i \oplus j} : \lim_{n \rightarrow \infty} p^{nd+j}(x \rightarrow y) = \pi_{i \oplus j}(y)$$

Кроме того, условие инвариантности заменяется на условие $\pi_j \cdot p = \pi_{j \oplus 1}$.

Доказательство. Зафиксируем подкласс \mathcal{C}_i и рассмотрим на нём марковскую цепь с переходной вероятностью $q := p^d$. Заметим, что тогда \mathcal{C}_i — эргодический класс в новой цепи, причём его период — 1. В самом деле,

$$\forall x \in \mathcal{C}_i : \exists K : \forall k \geq K : p^{kd}(x \rightarrow x) > 0 \\ q^k(x \rightarrow x) > 0$$

Таким образом, период новой цепи равен 1, откуда получаем, что к новой цепи применима предыдущая теорема.

А именно, существует распределение π_i на \mathcal{C}_i :

$$\forall x, y \in \mathcal{C}_i : q^n(x \rightarrow y) \xrightarrow{n \rightarrow \infty} \pi_i(x \rightarrow y) \iff p^{nd}(x \rightarrow y) \xrightarrow{n \rightarrow \infty} \pi_i(x \rightarrow y)$$

Теперь рассмотрим два подкласса \mathcal{C}_i и $\mathcal{C}_{i \oplus j}$ и произвольные $x \in \mathcal{C}_i, y \in \mathcal{C}_{i \oplus j}$.

$$p^{nd+j}(x \rightarrow y) = p^{nd}p^j(x \rightarrow y) = \sum_{z \in \mathcal{C}_{i \oplus j}} p^j(x \rightarrow z) \cdot p^{nd}(z \rightarrow y)$$

Так как $p^{nd}(z \rightarrow y) \xrightarrow{n \rightarrow \infty} \pi_{i \oplus j}(y)$, то $p^{nd+j}(x \rightarrow y) \xrightarrow{n \rightarrow \infty} \left(\sum_{z \in \mathcal{C}_{i \oplus j}} p^j(x \rightarrow z) \right) \cdot \pi_{i \oplus j}(y) = \pi_{i \oplus j}(y)$.

Осталось доказать, что $\pi_j \cdot p = \pi_{j \oplus 1}$. Положим $y \in \mathcal{C}_{j+1}$, запишем

$$\pi_{j \oplus 1}(y) = \sum_{x \in \mathcal{C}_j} \pi_j(x) \cdot p(x \rightarrow y)$$

Для этого вспомним, что $\forall x_0 \in \mathcal{C}_j : \pi_j(x) = \lim_{n \rightarrow \infty} p^{nd}(x_0 \rightarrow x)$. Тогда

$$\pi_{j \oplus 1}(y) = \lim_{n \rightarrow \infty} \sum_{x \in \mathcal{C}_j} p^{nd}(x_0 \rightarrow x) \cdot p(x \rightarrow y) = \lim_{n \rightarrow \infty} p^{nd+1}(x \rightarrow y) \stackrel{\text{предыдущее утверждение для } j=1}{=} \pi_{j \oplus 1}(y)$$

□

1.8.5 Возвратность

Пусть \mathcal{X} — быть может бесконечное пространство состояний.

Выберем $x_0 \in \mathcal{X}$, обозначим за f_i вероятность вернуться в \mathcal{X} на i -м шаге:

$$f_i(x_0) := \mathbb{P}((x_1 \neq x_0) \wedge \dots \wedge (x_{i-1} \neq x_0) \wedge (x_i = x_0))$$

Так как события несовместны, то $\sum_{i=1}^{\infty} f_i(x_0) \leq 1$.

Определение 1.8.7 ($x_0 \in \mathcal{X}$ — возвратное состояние). Такое состояние, для которого $\sum_{i=1}^{\infty} f_i(x_0) = 1$.

При этом говорят, что x_0 — *положительно возвратно*, если $\sum_{i=1}^{\infty} i \cdot f_i(x_0) < \infty$, то есть матожидание времени возврата конечно. Иначе x_0 называется *нуль-возвратным*.

Теорема 1.8.5 (Критерий возвратности). $x \in \mathcal{X}$ возвратно $\iff \sum_{n=1}^{\infty} p^n(x \rightarrow x) = \infty$.

Доказательство. Запишем двумя способами вероятность события пройти цикл из x в x .

$$p^n(x \rightarrow x) = \sum_{i=1}^n f_i(x) \cdot p^{n-i}(x \rightarrow x)$$

Введём производящие функции $\mathcal{F}(z) = \sum_{i=1}^{\infty} f_i(x)z^i$ и $\mathcal{P}(z) = \sum_{n=0}^{\infty} p^n(x \rightarrow x)z^n$, действующие на $z \in [0, 1)$. Для них

$$\mathcal{P}(z) = 1 + \mathcal{F}(z)\mathcal{P}(z)$$

Таким образом,

$$1 - \frac{1}{\mathcal{P}(z)} = \mathcal{F}(z)$$

Перейдём к пределу при $z \rightarrow 1$. Равенство обратится в

$$1 - \frac{1}{\sum_{n=1}^{\infty} p^n(x \rightarrow x)} = \sum_{i=1}^{\infty} f_i(x)$$

□

Факт 1.8.8. Если x и y сообщаются, то они либо оба возвратны, либо оба — невозвратны.

Доказательство. $\exists a, b \in \mathbb{N} : p^a(x \rightarrow y) > 0$ и $p^b(y \rightarrow x) > 0$. Тогда запишем

$$p^{n+a+b}(x \rightarrow x) \geq p^a(x \rightarrow y)p^n(y \rightarrow y)p^b(y \rightarrow x)$$

Отсюда видим, что ряды $\sum_{n=1}^{\infty} p^n(x \rightarrow x)$ и $\sum_{n=1}^{\infty} p^n(y \rightarrow y)$ сходятся (или нет) одновременно. \square

Следствие 1.8.3. Если в цепи все состояния сообщаются, то они все одновременно либо возвратны, либо нет.

Пример (Самый знаменитый пример). **Простое симметричное случайное блуждание на \mathbb{Z}^d .**

Пусть мы находимся в произвольной точке пространства $\mathbb{Z}^d \ni (x_1 \dots x_d)$. На каждом шагу меняется произвольная координата с вероятностью $\frac{1}{2d}$ — на ± 1 .

Все точки сообщаются, значит, все они возвратны или невозвратны одновременно.

Лекция X

17 апреля 2023 г.

Теорема 1.8.6 (Пойа). Симметричное случайное блуждание на целочисленной решётке \mathbb{Z}^d возвратно $\iff d \leq 2$.

Доказательство. Будем пользоваться не определением возвратности, а критерием — про сходимость ряда. Идея состоит в том, что $p^n(x \rightarrow x) \asymp n^{-d/2}$. Этот ряд сходится при $d \geq 3$.

$p^{2n+1}(0 \rightarrow 0) = 0$, поэтому для проверки расходимости ряда будем рассматривать чётные индексы.

$d = 1$.

$$p_1^n(0 \rightarrow 0) = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{n!n!2^{2n}} \sim \frac{(2n/e)^n \sqrt{2\pi 2n}}{(n/e)^n (n/e)^n \cdot \sqrt{2\pi n} \sqrt{2\pi n} \cdot 2^{2n}} = \frac{1}{\sqrt{\pi n}}$$

Так как ряд расходится, то блуждание возвратно.

$d = 2$. Представим себе блуждание по плоскости x, y и рассмотрим замену переменных: $\begin{cases} u = x + y \\ v = x - y \end{cases}$.

Теперь обе координаты (u, v) независимы:

$$\begin{cases} x \rightsquigarrow x+1 & u \rightsquigarrow u+1, v \rightsquigarrow v+1 & 1/4 \\ x \rightsquigarrow x-1 & u \rightsquigarrow u-1, v \rightsquigarrow v-1 & 1/4 \\ y \rightsquigarrow y+1 & u \rightsquigarrow u+1, v \rightsquigarrow v-1 & 1/4 \\ y \rightsquigarrow y-1 & u \rightsquigarrow u-1, v \rightsquigarrow v+1 & 1/4 \end{cases}$$

Таким образом, случайные блуждания по заменённым координатам независимы, откуда:

$$p_2^{2n}(0 \xrightarrow{(xy)} 0) = p_2^{2n}(0 \xrightarrow{(uv)} 0) = p_1^{2n}(0 \xrightarrow{u} 0) \cdot p_1^{2n}(0 \xrightarrow{v} 0) = \frac{1}{\pi n}$$

Ряд расходится, блуждание возвратно.

$d = 3$. Введём M_1, M_2, M_3 — число шагов вдоль осей 1, 2, 3 — случайные величины, такие, что $M_1 + M_2 + M_3 = 2n$. Также введём событие

$$A_{m_1, m_2, m_3} = \{M_1 = 2m_1, M_2 = 2m_2, M_3 = 2m_3\}$$

Запишем $p_3^{2n}(0 \rightarrow 0) = \sum_{m_1, m_2, m_3} p_3^{2n}(0 \rightarrow 0 \wedge A_{m_1, m_2, m_3})$ — формулу полной вероятности.

Здесь есть плохие слагаемые, в которых одно из m_1, m_2, m_3 слишком мало.

$$\mathbb{E}M_1 = \mathbb{E}M_2 = \mathbb{E}M_3 = \frac{2n}{3}; \quad \mathbb{D}M_1 = \mathbb{D}M_2 = \mathbb{D}M_3 \sim \text{const} \cdot n$$

Согласно неравенству Чебышёва

$$\mathbb{P}\left(M_1 \leq \frac{n}{3}\right) = \mathbb{P}\left(M_1 - \mathbb{E}M_1 \leq -\frac{n}{3}\right) \leq \mathbb{P}\left(|M_1 - \mathbb{E}M_1| \geq \frac{n}{3}\right) \leq \frac{\mathbb{D}M_1^2}{(n/3)^2} = \frac{\text{const}}{n}$$

Эта оценка слишком слабая, она расходится и не помогает доказать сходимости.

Воспользуемся лучше экспоненциальным неравенством Чебышёва:

$$\begin{aligned} \mathbb{P}\left(M_1 \leq \frac{n}{3}\right) &= \mathbb{P}\left(-M_1 \geq -\frac{n}{3}\right) \leq \frac{\mathbb{E}(e^{-M_1})}{e^{-n/3}} = \\ &= \mathbb{E}(e^{-M_1}) \cdot e^{n/3} = \left(\frac{2}{3} + \frac{1}{3} \cdot e^{-1}\right)^{2n} \cdot e^{n/3} = \left(\left(\frac{2+e^{-1}}{3}\right)^2 e^{1/3}\right)^n \approx 0.87^n \end{aligned}$$

Теперь

$$\sum_{m_1, m_2, m_3} p_3^{2n}(0 \rightarrow 0 \wedge A_{m_1, m_2, m_3}) \leq \sum_{\substack{m_1, m_2, m_3 \\ m_1, m_2 \text{ или } m_3 \text{ меньше } n/3}} \mathbb{P}(A_{m_1, m_2, m_3}) + \sum_{\substack{m_1, m_2, m_3 \\ \text{иначе}}} \mathbb{P}(0 \rightarrow 0 \wedge A_{m_1, m_2, m_3})$$

Первая сумма сходится: оценивается суммой $\mathbb{P}(m_1 \leq \frac{n}{3}) + \mathbb{P}(m_2 \leq \frac{n}{3}) + \mathbb{P}(m_3 \leq \frac{n}{3})$, где каждое слагаемое оценено выше.

Вторая сумма оценивается из формулы полной вероятности: $p_3^{2n}(0 \rightarrow 0 \wedge A_{m_1, m_2, m_3}) = p_3^{2n}(0 \rightarrow 0 | A_{m_1, m_2, m_3}) \cdot \mathbb{P}(A_{m_1, m_2, m_3})$. Дальше $p_3^{2n}(0 \rightarrow 0 | A_{m_1, m_2, m_3})$ раскладывается на три множителя по каждой координате:

$$p_3^{2n}(0 \rightarrow 0 | A_{m_1, m_2, m_3}) = p_1^{m_1}(0 \rightarrow 0) p_2^{m_2}(0 \rightarrow 0) p_1^{m_3}(0 \rightarrow 0) \leq \frac{\text{const}}{\sqrt{m_1}} \cdot \frac{\text{const}}{\sqrt{m_2}} \cdot \frac{\text{const}}{\sqrt{m_3}} \leq \frac{\text{const}}{n^{3/2}}$$

Таким образом, $\sum_{\substack{m_1, m_2, m_3, \text{ одно меньше } n/3 \\ A_{m_1, m_2, m_3}}} p_3^{2n}(0 \rightarrow 0 | A_{m_1, m_2, m_3}) \cdot \mathbb{P}(A_{m_1, m_2, m_3}) \leq \frac{\text{const}}{n^{3/2}}$ — события A_{m_1, m_2, m_3} не пересекаются. Итак, ряд сходится, блуждание не возвратно.

$d > 3$. Доказывается аналогично $d = 3$.

□

1.9 Случайное блуждание в \mathbb{Z}^1

Случайное блуждание на \mathbb{Z} можно воспринимать либо как сумму независимых случайных величин X_j , распределённых по закону $X_{i,j} = \begin{cases} +1, & \text{с вероятностью } p \\ -1, & \text{с вероятностью } q \end{cases}$ (и $S_n = X_1 + \dots + X_n$), или как марковскую цепь

$$\begin{aligned} \mathbb{P}(S_{n+1} = s+1 | S_n = s) &= p \\ \mathbb{P}(S_{n+1} = s-1 | S_n = s) &= q \end{aligned}$$

Исследуем некоторые параметры данного случайного блуждания.

Обозначим за R_n количество шагов вправо среди первых n шагов. Это величина с биномиальным распределением $\mathcal{B}(n, p)$. $S_n = R_n - (n - R_n) = 2R_n - n$, откуда вероятность $\mathbb{P}(S_n \geq m)$ переписывается в виде $\mathbb{P}(2R_n - n \geq m) = \mathbb{P}(R_n \geq \frac{n+m}{2})$.

Интегральная теорема Муавра — Лапласа говорит, что $\mathbb{P}\left(R_n \geq np + b\sqrt{np(1-p)}\right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_b^\infty e^{-\frac{x^2}{2}} dx$.

В частности, для $p = 1/2$ получаем $\mathbb{P}(S_n \geq b\sqrt{n}) = \mathbb{P}(R_n \geq \frac{n}{2} + \frac{1}{2}b\sqrt{n}) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_b^\infty e^{-\frac{x^2}{2}} dx$.

1.9.2 Время пребывания на полуоси (закон арксинуса)

Рассмотрим симметричное блуждание с $p = q = \frac{1}{2}$. Изобразим на своеобразном графике точки (k, S_k) , соединив последовательные отрезками.



Назовём временем, проводимым на положительной оси $T_n = \sum_{k=1}^n \mathbb{1}_{\{s_k \geq 0, s_{k-1} \geq 0\}}$. Пусть $a, b \in (0, 1)$, найдём, чему пропорциональна вероятность $\mathbb{P}(a \leq \frac{T_n}{n} \leq b)$.

Будем рассматривать чётные n , то есть обозначим их $2n$. Интересно заметить, что T_{2n} всегда чётно: точки $S_k = 0$ появляются всегда при чётных k , и между соседними точками либо всё время — пребывание на положительной полуоси, либо всё время — пребывание на отрицательной полуоси.

Будем использовать без доказательства факт $\mathbb{P}(T_{2n} = k) = \mathbb{P}(S_{2k} = 0) \cdot \mathbb{P}(S_{2(n-k)} = 0)$. (доказательство можно найти в учебнике Ширяева «Вероятность», глава 1, параграф 10).

Таким образом, мы можем выразить $(T_{2n} = k)$ с помощью простых методов:

$$\begin{aligned}\mathbb{P}(S_{2k} = 0) &= 2^{-2k} \binom{2k}{k} = \frac{1}{\sqrt{\pi k}} (1 + o(1)) \\ \mathbb{P}(S_{2(n-k)} = 0) &= \frac{1}{\sqrt{\pi(n-k)}} (1 + o(1))\end{aligned}$$

Теперь можно записать

$$\mathbb{P}\left(a < \frac{T_n}{n} < b\right) = \sum_{a < \frac{k}{n} < b} \frac{1}{\sqrt{\pi k}} \cdot \frac{1}{\sqrt{\pi(n-k)}} (1 + o(1)) = \frac{1}{\pi} \cdot \sum_{a < \frac{k}{n} < b} \frac{1}{\sqrt{\frac{k}{n}}} \cdot \frac{1}{\sqrt{1 - \frac{k}{n}}} \cdot \frac{1}{n} (1 + o(1))$$

Заметим, что теперь под суммой стоит сумма Римана — Дарбу, можем записать свойство интеграла Римана

$$\frac{1}{\pi} \cdot \sum_{a < \frac{k}{n} < b} \frac{1}{\sqrt{\frac{k}{n}}} \cdot \frac{1}{\sqrt{1 - \frac{k}{n}}} \cdot \frac{1}{n} (1 + o(1)) \xrightarrow{n \rightarrow \infty} \frac{1}{\pi} \int_a^b \frac{du}{\sqrt{u(1-u)}} = I(b) - I(a)$$

где в качестве I подойдёт любая первообразная. Любопытно, что здесь есть две разные естественно выглядящие первообразные

$$\begin{aligned}I_1(x) &= \frac{1}{\pi} \arcsin(2x - 1) \\ I_2(x) &= \frac{2}{\pi} \arcsin(\sqrt{x})\end{aligned}$$

Это можно видеть из тождества $\arcsin(2x - 1) + \frac{\pi}{2} = 2 \arcsin(\sqrt{x})$ при $x \in [0, 1]$. (Проверяется взятием косинуса от обеих частей)

График $\frac{1}{\sqrt{u(1-u)}}$ выглядит, как U-образная кривая, с концами, уходящими в бесконечность, поэтому распределение сосредоточено около границ.

Если рассмотреть случайную величину Z с распределением $\mathbb{P}(Z \in [a, b]) = \frac{1}{\sqrt{\pi}} \int_a^b \frac{du}{\sqrt{u(1-u)}}$, то окажется, что она с очень большой вероятностью распределена близко к краю:

$$\mathbb{P}(Z \leq 0.024) \approx 0.1 \quad \mathbb{P}(Z \leq 0.006) \approx 0.05$$

1.9.3 Задача о разорении игрока

Пусть у I игрока есть $|A|$ монет (мы будем считать $A < 0$), у II игрока — B монет, и пусть они играют в азартную игру. У I игрока вероятность выигрыша p , у II игрока — $q = 1 - p$. По выигрышу проигравший платит одну монету другому, игра заканчивается, когда один из них разорится.

Исследуем эту модель. Заметим, что это на самом деле тоже случайное блуждание, заканчивающееся когда S_k выходит из интервала $[A, B]$:



Положим $\beta_k(x)$ — вероятность выйти на B раньше, чем на A не более чем за k шагов, исходя из точки x . Эти величины мы можем рассматривать в дискретной теории вероятностей, так как бесконечных траекторий несчётное количество.

Заметим, что $\beta_k(x)$ монотонно возрастает по k , но, очевидно, $\beta_k(x)$ ограничена. Значит, имеется предел, который мы и хотим вычислить.

Запишем своеобразную рекурренту на β : с вероятностью p первый шаг — в положительном направлении, с вероятностью q — в отрицательном

$$\beta_k(x) = p\beta_{k-1}(x+1) + q\beta_{k-1}(x-1)$$

Перейдя к пределу по k получаем

$$\begin{aligned}\beta(x) &= p\beta(x+1) + q\beta(x-1) \\ p\beta(x) + q\beta(x) &= p\beta(x+1) + q\beta(x-1) \\ q(\beta(x) - \beta(x-1)) &= p(\beta(x+1) - \beta(x)) \\ \beta(x+1) - \beta(x) &= \frac{q}{p}(\beta(x) - \beta(x-1))\end{aligned}$$

Таким образом, последовательные разности $\beta(x+1) - \beta(x)$ образуют геометрическую прогрессию. Воспользовавшись начальными условиями $\begin{cases} \beta(B) = 1 \\ \beta(A) = 0 \end{cases}$ можно получить точную формулу. В

частности, для $p = q = \frac{1}{2}$ получается неожиданно простая формула

$$\beta(0) = \frac{|A|}{B + |A|}$$

Если $p \neq q$, то можно решить систему из $B + |A| + 1$ линейных уравнений, результатом будет

$$\beta(x) = \frac{(q/p)^x - (1/p)^A}{(q/p)^B - (1/p)^A}$$

Замечание. Случайное блуждание не может бесконечное время болтаться внутри ограниченного отрезка. Вероятность того, что рано или поздно кто-то выиграет стремится к единице. Доказательство остаётся читателю в качестве упражнения.

1.9.4 Матожидание времени разорения

Задача прежняя — есть два игрока с капиталами $|A|$, B , p, q — вероятности их выигрышей соответственно.

Обозначим $T(x)$ — время разорения одного из игроков, если блуждание началось в точке x . Чему равно $\mathbb{E}T(x)$?

Как и в предыдущей задаче, ограничим игру конечным числом ходов: $T_k(x) = \begin{cases} T(x), & T(x) \leq k \\ k, & T(x) \geq k \end{cases}$.

Используемая выше $T(x)$ — величина, которую мы не можем рассматривать в дискретной теории вероятностей. Чтобы этого избежать, рассмотрим величины $T_k(x)$ и найдём $\lim_{k \rightarrow \infty} \mathbb{E}T_k(x)$.

Обозначим $m_k(x) := \mathbb{E}T_k(x)$. $m_k(x)$ тоже монотонно возрастет по k . Более того, у него есть предел — вероятность того, что $T(x) > n$ экспоненциально убывает, но выкладок, обосновывающих это, нет.

$$m_k(x) = \begin{cases} pm_{k-1}(x+1) + qm_{k-1}(x-1) + 1, & x \in (A, B) \\ 0, & x = A \vee x = B \end{cases}$$

Преобразуем первое равенство, перейдя в нём к пределу по $k \rightarrow \infty$.

$$\begin{aligned} pm(x) + qm(x) &= pm(x+1) + qm(x-1) + 1 \\ p(m(x+1) - m(x)) &= q(m(x) - m(x-1)) - 1 \\ m(x+1) - m(x) &= \frac{q}{p}(m(x) - m(x-1)) - \frac{1}{p} \end{aligned}$$

Это опять же решаемая система, но для экономии времени лекции приведём лишь решение для $p = q = \frac{1}{2}$:

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

Решением является многочлен второй степени с корнями в A и B . $m(x) = K(x-A)(x-B)$. Подгоняя K так, чтобы выполнялось уравнение $m(x+1) - m(x) = (m(x) - m(x-1)) - 2$ понимаем, что $K = -1$.

$$m(x) = (B-x)(x-A); \text{ в частности, } m(0) = |A| \cdot B$$

Если $p \neq q$, то ответ чуть более противный:

$$m(0) = \frac{B-A}{p-q} \cdot \frac{1 - (q/p)^A}{(q/p)^B - (q/p)^A} + \frac{A}{p-q}$$

1.10 Случайные графы

В нашей жизни есть огромное множество графов: графов друзей социальных сетей, граф аэропортов и авиалиний, граф совместных научных публикаций и граф цитирований. . .

small world — маленькость мира, диаметры реальных графов (длина пути — количество рёбер) очень малы. Так, в графе совместных публикаций научного мира диаметр порядка 10.

Графы бывают статические и динамические — во времени меняются последние.

Типичная статическая модель: граф Эрдёша — Реньи на n вершинах $G(n, p)$, в котором каждое из $\binom{n}{2}$ рёбер проведено с вероятностью p .

Самая знаменитая динамическая модель: модель преимущественного присоединения. Начнём с какого-то простого графа, на каждом шаге добавляем вершину и одно ребро из неё, ведущее к какой-нибудь из существующих вершин, причём вероятность пропорциональна степени вершины.

Лекция XII

22 мая 2023 г.

1.10.1 Граф Эрдёша — Реньи

Рассмотрим множество из n вершин, каждое из $\binom{n}{2}$ рёбер проведено с вероятностью p независимо от других — случайный граф $G(n, p)$.

Рассмотрим последовательность p_n и изучим поведение $G(n, p)$ при $n \rightarrow \infty$.

Интересный факт (Условие связности).

- Если $\lim_{n \rightarrow \infty} \frac{p_n}{\log n/n} > 1$, то $\mathbb{P}(G(n, p_n) \text{ связан}) \xrightarrow{n \rightarrow \infty} 1$.
- Если $\overline{\lim}_{n \rightarrow \infty} \frac{p_n}{\log n/n} < 1$, то $\mathbb{P}(G(n, p_n) \text{ связан}) \xrightarrow{n \rightarrow \infty} 0$.

Обозначим за M_n размер максимальной компоненты связности в $G(n, p_n)$.

Интересный факт (О гигантской компоненте).

- Если $\lim_{n \rightarrow \infty} \frac{p_n}{1/n} = \gamma > 1$, то $\exists a(\gamma) : \mathbb{P}(M_n > a(\gamma) \cdot n) \xrightarrow{n \rightarrow \infty} 1$.
- Если $\overline{\lim}_{n \rightarrow \infty} \frac{p_n}{1/n} = \gamma < 1$, то $\exists b(\gamma) : \mathbb{P}(M_n \leq b(\gamma) \cdot \log n) \xrightarrow{n \rightarrow \infty} 1$.

1.10.2 power law for degrees (степенной закон для степеней (вершин))

Рассмотрим большой граф из n вершин; обозначим за $V_n^{(d)}$ количество вершин степени d .

Оказывается, часто имеет место приближение $V_n^{(d)} \approx (ad^{-\alpha})n$, где $\alpha \in (2, 5)$ — для разных графов предлагались разные значения. a и α — константы, зависящие от типа графа, но не зависящие от d , иначе было бы совсем неинтересно. Тем не менее, α меняется не очень сильно, а a находится из уравнения $V_n^{(0)} + V_n^{(1)} + \dots + V_n^{(n)} = n$.

1.10.3 Дерево преимущественного присоединения

Рассмотрим в качестве начального состояния граф K_2 , состоящий из двух вершин и одного ребра.

На k -м шаге в граф добавляется вершина с номером $k + 2$, и из неё добавляется ровно одно случайное ребро, причём оно проведено к вершине $i \in [1, k + 1]$ с вероятностью, пропорциональной $\deg(i)$, где \deg — степень в графе на первых $k + 1$ вершинах.

После шага n в графе $n + 2$ вершины, $n + 1$ ребро, несложно видеть, что граф связан и является деревом.

Поведение степеней вершин

Обозначим за $X_n^{(m)}$ степень вершины m после шага n .

«Кто не успел, тот опоздал»

Рассмотрим $m = 1$. $X_0^{(1)} = 1$ — после 0-го шага величина пока неслучайная. Запишем уравнения на развитие случайной переменной $X_n^{(1)}$.

$$\begin{aligned}\mathbb{P}\left(X_{n+1}^{(1)} - X_n^{(1)} = 1 \mid X_n^{(1)} = k\right) &= \frac{k}{2(n+1)} \\ \mathbb{P}\left(X_{n+1}^{(1)} - X_n^{(1)} = 0 \mid X_n^{(1)} = k\right) &= 1 - \frac{k}{2(n+1)}\end{aligned}$$

Посчитаем от величины $X_n^{(1)}$ только её матожидание.

$$\mathbb{E}\left(X_{n+1}^{(1)}\right) - \mathbb{E}\left(X_n^{(1)}\right) = \mathbb{E}\left(X_{n+1}^{(1)} - X_n^{(1)}\right) = \mathbb{P}\left(X_{n+1}^{(1)} - X_n^{(1)} = 1\right) = \sum_{k=1}^{\infty} \mathbb{P}\left(X_n^{(1)} = k\right) \frac{k}{2(n+1)}$$

В этом месте чудесным образом появляется матожидание, получаем рекурренту на матожидание

$$\mathbb{E}\left(X_{n+1}^{(1)}\right) - \mathbb{E}\left(X_n^{(1)}\right) = \mathbb{E}\left(X_n^{(1)} \cdot \frac{1}{2(n+1)}\right)$$

откуда $\mathbb{E}X_{n+1}^{(1)} = \mathbb{E}X_n^{(1)} \left(1 + \frac{1}{2(n+1)}\right) = \mathbb{E}X_n^{(1)} \cdot \frac{2n+3}{2(n+1)} = \frac{(2n+3)!!}{(2n+2)!!}$. Используя формулу Стирлинга, получаем

$$(2n)!! = 2^n n! \sim 2^n \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \quad (2n+1)!! = \frac{(2n+1)!}{(2n)!}$$

откуда

$$\mathbb{E}X_n^{(1)} = \frac{(2n+1)!}{((2n)!!)^2} \sim \frac{\sqrt{2\pi(2n+1)} \left(\frac{2n+1}{e}\right)^{2n+1}}{2^{2n} (2\pi n) \left(\frac{n}{e}\right)^{2n}} \sim \frac{1}{\sqrt{\pi}} \frac{\sqrt{2 \cdot 2n}}{2n} \underbrace{\left(\frac{2n+1}{2n}\right)^{2n}}_{\rightarrow e} \frac{2n+1}{e} \sim \frac{2}{\sqrt{\pi}} \sqrt{n}$$

Заметим, что в графе $2(n+1)$ рёбер всего, поэтому в среднем степень вершины порядка 2. Таким образом, видим, что степень первой вершины сильно больше средней степени.

Очевидно, $\mathbb{E}X_n^{(2)} = \mathbb{E}X_n^{(1)}$. Можно написать формулу для произвольной вершины, она доказывается примерно так же.

$$\mathbb{E}X_n^{(l+1)} \sim \frac{2}{\sqrt{\pi}} \frac{(2l-2)!!}{(2l-1)!!} \sqrt{n}, \text{ где можно записать } \frac{(2l-2)!!}{(2l-1)!!} \sim l^{-1/2}$$

1.10.4 Распределение степеней вершин

Пусть $V_n^{(d)}$ — количество вершин степени d после шага n .

Рассмотрим $d = 1$. После 0 шагов $V_0^{(d)} = 2$ — величина ещё неслучайная. Опять же, выпишем условные вероятности. Заметим, что $V_{n+1}^{(1)} - V_n^{(1)}$ — всегда либо 0, либо 1 (добавляется вершина степени 1, но, быть может, одна из вершин степени 1 станет вершиной степени 2).

$$\begin{aligned}\mathbb{P}\left(V_{n+1}^{(1)} - V_n^{(1)} = 0 \mid V_n = k\right) &= \frac{k}{2(n+1)} \\ \mathbb{P}\left(V_{n+1}^{(1)} - V_n^{(1)} = 1 \mid V_n = k\right) &= 1 - \frac{k}{2(n+1)}\end{aligned}$$

Аналогично подсчёту $\mathbb{E}X_n^{(1)}$ получаем

$$\mathbb{E}V_{n+1}^{(1)} - \mathbb{E}V_n^{(1)} = \mathbb{E}\left(V_{n+1}^{(1)} - V_n^{(1)}\right) = \mathbb{P}\left(V_{n+1}^{(1)} - V_n^{(1)} = 1\right) = \sum_{k=1}^{\infty} \left(1 - \frac{k}{2(n+1)}\right) \mathbb{P}\left(V_n^{(1)} = k\right)$$

Суммируя вероятности $\mathbb{P}(V_n^{(1)} = k)$ получаем 1; во второй половине правой части формулы опять получается матожидание. Значит,

$$\mathbb{E}V_{n+1}^{(1)} - \mathbb{E}V_n^{(1)} = 1 - \frac{\mathbb{E}V_n^{(1)}}{2(n+1)}$$

Чтобы решить эту рекурренту, предположим, что $\mathbb{E}V_n^{(1)} \sim \alpha n$ для некоего $\alpha \in \mathbb{R}$. По-хорошему, это надо обосновать, но давайте опустим.

Тогда решая уравнение $\alpha = 1 - \frac{\alpha}{2}$, получаем $\alpha = \frac{2}{3}$.

$$\mathbb{E}V_n^{(1)} \underset{n \rightarrow \infty}{\sim} \frac{2}{3}n$$

В общем случае получится формула

$$\mathbb{E}V_n^{(d)} \underset{n \rightarrow \infty}{\sim} \frac{4}{d(d+1)(d+2)}n \underset{d \rightarrow \infty}{\sim} \frac{4}{d^3}n$$