

Теорвер

Тимофей Равнушкин

Август 2025

Содержание

1	Матожидание	1
2	Теорема Кэли о числе деревьев через мартингалы	2
3	Теорема об остановке	3
4	Задача о разорении игрока	3
5	Слайм и бисквиты	4
6	Матожидание времени первого вхождения заданной подстроки в случайную строку	5

1 Матожидание

Определение 1.1

Пусть у нас есть множество Ω , для некоторых подмножеств которого мы можем говорить "вероятность" этого подмножества (т. е. для некоторого $\mathcal{F} \subset 2^\Omega$ определена $P : \mathcal{F} \rightarrow [0, 1]$). **Случайной величиной** на этом множестве называется такая функция $X : \Omega \rightarrow \mathcal{U}$, что $\mathcal{U}$ - конечное подмножество $\mathbb{R}$ и для всех $u \in \mathcal{U}$ $P^{-1}(u) \in \mathcal{F}$.

Матожиданием случайной величины X называется $E[X] = \sum_{x \in \mathcal{U}} P(X = x) \cdot x$.

Предложение 1.2

$$\begin{aligned} E[X + Y] &= \sum_x \sum_y (x + y) P(X = x, Y = y) \\ &= \sum_x x \sum_y P(X = x, Y = y) + \sum_y y \sum_x P(X = x, Y = y) \\ &= \sum_x x P(X = x) + \sum_y y P(Y = y) \\ &= E[X] + E[Y] \\ E[aX] &= aE[X] \end{aligned}$$

То есть матожидание линейно даже для зависимых случайных величин. Например, матожидание количества локальных максимумов в случайной перестановке из n элементов — $\frac{n-2}{3}$, потому что для каждой тройки подряд идущих элементов максимум в ней находится в её середине с вероятностью $\frac{1}{3}$. Если просуммировать это по всем тройкам, получим $\frac{n-2}{3}$.

Задача 1. Выбран случайный простой граф на n вершинах с $2n$ ребрами. Хотим найти матожидание количества треугольников в нём.

Для каждой тройки вершин вероятность, что все три необходимых для превращения её в треугольник ребра окажутся в выбранном множестве, равна $\frac{\binom{M-3}{2n-3}}{\binom{M}{2n}}$, где $M = \binom{n}{2}$. При этом $\frac{\binom{M-3}{2n-3}}{\binom{M}{2n}} = \frac{2n(2n-1)(2n-2)}{M(M-1)(M-2)}$. По линейности матожидания, $E[\#\Delta] = \binom{n}{3} \frac{2n(2n-1)(2n-2)}{M(M-1)(M-2)}$.

Определение 1.3

Матожиданием случайной величины X при условии события A (подмножества Ω , для которого мы умеем считать вероятность) называется значение $E[X \mid A] = \frac{E[X \cdot \mathbf{1}_A]}{P(A)} =$

$\sum_{w \in A} X(w) \frac{P(w)}{P(A)}$ где $\mathbf{1}_A$ - индикатор подмножества A и w - минимальные измеримые множества (нетрудно заметить, что на всех элементах w случ. величина принимает одно и то же значение).

Матожиданием случайной величины X при условии случайной величины Y называется случайная величина, заданная $E[X \mid Y](\omega) = E[X \mid Y = Y(\omega)] = E[X \mid Y^{-1}(\omega)]$. Аналогично определяется условное матожидание относительно многих переменных.

Втупую проверяется, что условное матожидание также линейно.

Определение 1.4

Случайным процессом называется последовательность случайных величин $X_0, X_1, \dots$ на одном и том же вероятностном пространстве.

Случайный процесс называется **мартингалом**, если для любого n выполнено $E[X_{n+1} \mid X_n, X_{n-1}, \dots, X_0] = X_n$.

Случайный процесс называется **марковской цепью**, если для любого n и любых $x_0, x_1, \dots, x_{n+1}$ выполнено $P(X_{n+1} = x_{n+1} \mid X_n = x_n, \dots, X_0 = x_0) = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$.

Простейшим примером мартингала является равновероятное случайное блуждание, когда мы с шансом 0.5 прибавляем или вычитаем 1 к текущему значению.

2 Теорема Кэли о числе деревьев через мартингалы

Сейчас мы будем самым смешным способом доказывать теорему Кэли о числе деревьев (изначально оно из статьи с arXiv, но перевод я украл у Александра Бабина).

Положим, что у нас есть n ящиков и n ключей от них, при чем ключи для ящиков $2, \dots, n$ мы раскладываем равновероятно и независимо по этим ящикам. А ключ 1 забираем к себе. За один ход мы открыть ящик с помощью ключа и забрать оттуда все ключи. Расклад ключей считается успешным, если возможно открыть все ящики. Рассмотрим величину X_k — математическое ожидание количества ключей, которое мы найдем в ящике, после k шагов алгоритма. Оказывается, что $\{X_k\}$ — мартингал. Действительно, пусть на r -м шаге у нас было не открыто m ящиков и в них лежало k ключей, тогда математическое ожидание ключей в любом из неоткрытых ящиков равно $\frac{k}{m}$. Если же мы откроем любой из доступных ящиков, то математическое ожидание число ключей в других ящиках равно $k - \frac{k}{m}$, а в каждом из них оно будет равно, соответственно:

$$\frac{k - \frac{k}{m}}{m - 1} = \frac{k \frac{m-1}{m}}{m - 1} = \frac{k}{m}$$

Следовательно, X_n - мартингал.

У условного матожидания есть одно полезное свойство – $E[E[X | Y]] = E[X]$:

$$\begin{aligned} E[E[X | Y]] &= \sum_y E[X | Y = y] \cdot P(Y = y) \\ &= \sum_y \left(\frac{\sum_{w \in Y^{-1}(y)} X(w)P(w)}{P(Y = y)} \right) P(Y = y) \\ &= \sum_{w \in \Omega} X(w)P(w) = E[X] \end{aligned}$$

В приложении к мартингалам, из этого можно получить $E[X_n] = E[E[X_n | X_{n-1}, \dots, X_0]] = E[X_{n-1}] = \dots = E[X_0]$, то есть матожидание мартингала в произвольный момент времени равно матожиданию в начальный.

Если мы завершили неудачей, то $X_{n-1} = 1$, потому что все ключи от неоткрытых ящиков остались в них. Иначе $X_{n-1} = 0$, поскольку в какой-то момент мы получили ключи от всех ящиков. Тогда $P(\text{неудача}) = E[X_{n-1}] = E[X_0] = \frac{n-1}{n} \implies P(\text{успех}) = \frac{1}{n}$.

Остаётся заметить, что успешные расстановки соответствуют деревьям с корнем 1 (ребра идут из ящика в ящики, ключи от которых в нём лежат). При этом всего изначальных расстановок n^{n-1} , из них хороших $P(\text{успех})$, следовательно, деревьев на n вершинах всего n^{n-2} .

3 Теорема об остановке

Определение 3.1

Случайная величина $T : \Omega \rightarrow \mathcal{U}$, где $\mathcal{U} \subset \mathbb{Z}_+ \cup \{+\infty\}$ называется **марковским моментом** для процесса $X_0, X_1, \dots$, если мы можем для данного i проверить, что $T = i$, зная только первые i шагов процесса. Марковский момент называется **моментом остановки**, если $P(T = +\infty) = 0$.

Теорема 3.2: (об остановке)

X_n является мартингалом $\implies$ для любого ограниченного момента остановки $\tau \leq N$ выполнено $E[X_\tau] = E[X_0]$.

Доказательство. Введём $Y_n = X_{\min(\tau, n)}$. Очевидно, что Y_n тоже является мартингалом, потому что для неостановленной части процесса наследует это свойство от X_n , а для остановленной остаётся константой. Тогда $E[Y_n]$ константно по n , но для $n \geq N$ верно $Y_n = X_\tau$ и при этом $Y_0 = X_0$. Получается, что

$$E[X_\tau] = E[Y_N] = E[Y_0] = E[X_0] \quad \square$$

С помощью некоторых трюков со сходимостями можно показать, что если $|X_{\min(\tau, n)}| \leq c$ для какой-то константы c , то теорема об остановке верна даже для неограниченных моментов остановки.

Из теоремы об остановке следует, что если мы играем в казино, в котором есть ограничение на максимальную ставку (или у нас ограниченное количество денег) матожидание нашего выигрыша 0, как бы мы ни играли. (((

4 Задача о разорении игрока

Игрок играет в казино. У него изначально x денег, он хочет в итоге получить $b > x$, но если его деньги достигнут уровня $a < x$, он разоряется и перестаёт играть. Каждый раз игрок ставит 1 единицу денег, результаты розыгрышей независимы и одинаково распределены. Найти вероятность разорения игрока.

Пусть ξ_n - независимые о.р.с.в., причём $P(\xi_n = 1) = p, P(\xi_n = -1) = 1 - p = q$. Обозначим $S_0 = x, S_n = x + \xi_1 + \dots + \xi_n$. Положим $\tau = \min\{n : S_n \in \{a, b\}\}$ - момент окончания игры. Считаем $p \in (0, 1)$, иначе тривиально. Очевидно, что τ является моментом остановки для S_n .

Случай $p = q = \frac{1}{2}$. Заметим, что в этом случае $E[\xi_n] = 0$, следовательно, S_n является мартингалом. Для всех n выполнено $S_{\min(n, \tau)} \in [a, b]$, т. е. можно применять расширенную теорему об остановке. Получается, что

$$E[S_\tau] = E[S_0] = x.$$

С другой стороны, S_τ принимает всего два значения:

$$E[S_\tau] = a \cdot P(S_\tau = a) + b \cdot P(S_\tau = b).$$

С учётом того, что $P(S_\tau = b) = 1 - P(S_\tau = a)$ видим, что

$$P(S_\tau = a) = \frac{b - x}{b - a}$$

Случай $p \neq q$. Рассмотрим $X_n = (q/p)^{S_n}$ и проверим, что это мартингал:

$$E\left[\left(\frac{q}{p}\right)^{\xi_n}\right] = \frac{q}{p} \cdot p + \frac{p}{q} \cdot q = q + p = 1$$

Снова проверим возможность применения расширенной теоремы об остановке: для всех n выполнено, что $S_{\min(\tau, n)} \in [a, b]$, отсюда следует, что $|X_{\min(\tau, n)}|$ тоже ограничено. Получается, что

$$E[X_\tau] = E[X_0] = \left(\frac{q}{p}\right)^x.$$

При этом

$$E[X_\tau] = \left(\frac{q}{p}\right)^a \cdot P(S_\tau = a) + \left(\frac{q}{p}\right)^b \cdot P(S_\tau = b)$$

Поэтому

$$P(S_\tau = a) = \frac{(q/p)^b - (q/p)^a}{(q/p)^b - (q/p)^a}.$$

5 Слайм и бисквиты

В начале игры у i -го игрока есть a_i бисквитов. Каждую секунду Слайм выберет один бисквит равновероятно среди $a_1 + a_2 + \dots + a_n$ бисквитов, и владелец этого бисквита передает его случайному игроку, равновероятно среди $n - 1$ игроков кроме себя. Игра закончится, когда у одного игрока будут все бисквиты.

Как хозяин вечеринки, Слайм хочет узнать математическое ожидание времени, когда игра закончится, чтобы успеть провести следующее мероприятие.

Эта задача показывает основной метод применения мартингалов в спортивном программировании. Если мы найдём такой потенциал F_n , матожидание изменения которого на каждом шаге нашего процесса равно -1 , можно будет рассмотреть процесс $X_n = F_n + n$. Очевидно, что в таком случае X_n - мартингал, и если τ - момент остановки для него, то $E[X_\tau] = E[X_0]$, из чего следует $E[F_\tau] + E[\tau] = E[F_0]$, то есть $E[\tau] = E[F_0 - F_\tau]$. Если при этом F_τ - константа, то $E[\tau] = F_0 - F_\tau$.

Часто потенциалы удобно вводить, как сумму каких-то функций от размеров подмножеств, играющих роль в нужной задаче. Например, если в данной задаче мы найдём такую f , что $E[\sum_i f(a_i)]$ уменьшается на 1 каждый шаг, то ответ будет равен $\sum f(a_i) - ((n - 1)f(0) + f(m))$.

Для удовлетворения условию изменения матожидания, у нас есть уравнение

$$\begin{aligned}\sum_i f(a_i) &= \sum_i \frac{a_i}{m} \left((f(a_i - 1) + 1) + \left(\sum_{j \neq i} \frac{1}{n-1} f(a_j + 1) + \frac{n-2}{n-1} f(a_j) \right) \right) \\ &= \sum_i \left(\frac{a_i}{m} (f(a_i - 1) + 1) + \frac{(m - a_i)}{m(n-1)} f(a_i + 1) + \frac{(m - a_i)(n-2)}{m(n-1)} f(a_i) \right) \\ f(a) &= \frac{a}{m} (f(a-1) + 1) + \frac{m-a}{m(n-1)} f(a+1) + \frac{(m-a)(n-2)}{m(n-1)} f(a)\end{aligned}$$

Получается, что можно вычислить $f(a)$ для всех возможных a_i с помощью ДП, и дальше найти ответ по указанной выше формуле.

6 Матожидание времени первого вхождения заданной подстроки в случайную строку

Пусть нам задан алфавит $1, 2, \dots, C$, функция $p : \{1, 2, \dots, C\} \rightarrow [0, 1]$ такая, что $\sum p(i) = 1$, и строка s из n символов нашего алфавита. Изначально у нас есть пустая строка t . Каждую секунду к t добавляется случайный символ из алфавита i с вероятностью $p(i)$. Хотим найти матожидание прошедшего времени на момент первого вхождения s в t .

Построим на s автомат префикс-функции. Теперь наша случайная генерация строки - случайное блуждание по этому автомату, переход по каждому символу выбирается с соответствующей ему вероятностью.

Обозначим за $h(i)$ матожидание времени, необходимого для попадания в вершину, соответствующую концу строки, из вершины i . В такой терминологии ответ равен $h(0)$. Тогда $h(n) = 0, h(k) = \sum_{a=1, \dots, C} p(a)h(\delta(k, a))$, где $\delta(k, a)$ - переход из k по символу a . Мы получили какую-то СЛАУ на $h(i)$, которую можно в общем случае решить за $O(n^3)$ (если решение существует) методом Гаусса. Однако, в данном конкретном случае можно найти формульное решение.

Обозначим за $\pi(k)$ суффиксную ссылку из вершины k (длину наибольшего собственного суффикса $s[1..k]$, который также является префиксом s ; $\pi(0)$ полагается равным 0). Введём $d(k) := h(\pi(k)) - h(k), k \geq 1$. Тогда $h(0) = \sum_{i=0}^k d(b_i)$, где $b_0 = n$ и $b_i = \pi(b_{i-1}), i \geq 1$ и $b_{k+1} = 0$ (b 'шки - вершины, посещённые по суффикссылкам из последней вершины). Осталось найти $d(i)$.

Автомат Ахо-Корасик характеризуется

$$\delta(k, a) = \begin{cases} k+1, & a = s_{k+1} \\ \delta(\pi(k), a), & a \neq s_{k+1} \end{cases}$$

Тогда

$$\begin{aligned}h(k) &= \sum_{a \neq s_{k+1}} p(a)h(\delta(\pi(k), a)) + p(s_{k+1})h(k+1) \\ h(\pi(k)) &= \sum_a p(a)h(\delta(\pi(k), a)) \\ d(k) &= p(s_{k+1})h(\delta(\pi(k), s_{k+1})) - p(s_{k+1})h(k+1) \\ &= p(s_{k+1})(h(\delta(\pi(k), s_{k+1})) - h(k+1))\end{aligned}$$

Но при этом можно заметить, что $\pi(\delta(k, s_{k+1})) = \delta(\pi(k), s_{k+1})$. Получается, что $d(k) = p(s_{k+1})d(k+1)$. Если $d(1) = \frac{1}{p(s_1)}$, то $d(k) = \frac{1}{p(s_1)p(s_2)\dots p(s_k)}$. Итого, ответ равен

$$\sum_{k | s[1..k] = s[(n-k+1)..n]} \frac{1}{p(s_1)p(s_2)\dots p(s_k)}$$