

# Дискретная математика. Теория

Александр Сергеев

## 1 Дискретная теория вероятности

### 1.1 Введение

#### Определение

*Дискретное вероятностное пространство* – пара  $(\Omega, p)$ ,  
где  $\Omega$  – не более чем счетное множество элементарных исходов  
 $p : \Omega \rightarrow [0, 1], \sum_{\omega \in \Omega} p(\omega) = 1$

#### Определение

*Событие* –  $A \subset \Omega$

*Вероятность события* –  $P(A) = \sum_{a \in A} p(a)$

#### Определение

События  $A$  и  $B$  *независимые*, если  $P(A \cap B) = P(A)P(B)$

$P(B|A) = \frac{P(A \cap B)}{P(A)}$  – вероятность  $B$  при условии  $A$

#### Лемма

Если  $A$  и  $B$  независимы, то  $P(B|A) = P(B)$

#### Определение

Пусть  $(\Omega_1, p_1), (\Omega_2, p_2)$  – независимые ДВП

Тогда их произведение  $(\Omega = \Omega_1 \times \Omega_2, p(\omega_1 \in \Omega_1, \omega_2 \in \Omega_2) = p_1(\omega_1)p_2(\omega_2))$

#### Определение

$A_1, A_2, \dots, A_n$  – *независимы в совокупности*,

если  $\forall I \subset \{1 \dots n\} \quad P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i)$

### Теорема

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

### Формула Байеса

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}$$

### Оффтоп

Рассмотрим пример: выкинули два честных кубика

Заметим, что возможно построить две математические модели:

1. Результаты - упорядоченная пара  $\langle x, y \rangle$

$$\text{Тогда } |\Omega| = 36, p(w) = \frac{1}{36}$$

2. Результаты - неупорядоченная пара  $[x, y]$

$$\text{Тогда } |\Omega| = 21, p([x, x]) = \frac{1}{36}, p([x, y \neq x]) = \frac{1}{18}$$

Заметим, что для запросов, не содержащих информацию об упорядоченности, результат не зависит от построенной модели

## 1.2 Случайные величины

### Определение

Случайной величиной называется функция  $\xi : \Omega \rightarrow \mathbb{R}$

### Примеры случайных величин

1. Пусть кинули 10 монет. Построим случайную величину – количество выпавших орлов:  $\xi(b_1, \dots, b_n) = b_1 + \dots + b_n$
2. Пусть кинули  $n$  кубиков. Построим случайную величину – среднее значение:  $\xi(v_1, \dots, v_n) = \frac{v_1 + \dots + v_n}{n}$
3. Пусть  $n$  студентов приходят на лекцию с вероятностями  $p_1, \dots, p_n$ . Построим случайную величину - количество студентов на лекции:

$$\xi(s_1, \dots, s_n) = \sum_{i=1}^n s_i$$

Заметим, что у этой случайной величины неравномерное распределение вероятностей:  $p(s_1, \dots, s_n) = \prod_{i=1}^n \begin{cases} p_i, & s_i = 1 \\ 1 - p_i, & s_i = 0 \end{cases}$

Давайте анализировать события через их случайные величины  
Заметим, что уравнение  $\xi = 3$  задает событие  $\{w : \xi(w) = 3\}$  (аналогично и другие предикаты с  $\xi$  задают события)

### Определение

$f_\xi : \mathbb{R} \rightarrow \mathbb{R}, f_\xi(a) = P(\xi = a)$  – дискретная плотность распределения

$F_\xi : \mathbb{R} \rightarrow \mathbb{R}, f_\xi(a) = P(\xi \leq a)$  – функция распределения

### Определение

Пусть  $\xi$  – случайная величина

$E_f = \sum_{\omega \in \Omega} \xi(\omega)p(\omega)$  – математическое ожидание

$$E_\xi = \sum_{\omega \in \Omega} \xi(\omega)p(\omega) = \sum_a \sum_{\omega: \xi(\omega)=a} = \sum_a a \sum_{\omega: \xi(\omega)=a} p(\omega) = \sum_a a P(\xi = a) = \sum_a a f_\xi(a)$$

### Определение

$D_\xi = E((\xi - E\xi)^2)$  – дисперсия

### Свойства математического ожидания

1.  $E(c\xi) = cE_\xi$
2.  $E(\xi + \eta) = E_\xi + E_\eta$  (даже для зависимых величин)
3. Для независимых  $\xi, \eta$   $E(\xi\eta) = E(\xi)E(\eta)$

### Доказательство

$$\begin{aligned} E(\xi\eta) &= \sum_a a P(\xi\omega = a) = \sum_x \sum_y xy P(\xi = x \wedge \omega = y) = \sum_x x \sum_y y P(\xi = x \wedge \omega = y) \\ &= \sum_x x \sum_y y P(\xi = x) P(\omega = y) = \sum_x x P(\xi = x) \sum_y y P(\omega = y) = E_\xi E_\omega \end{aligned}$$

4.  $E(\xi - E_\xi) = E\xi - EE\xi = E\xi - E\xi = 0$
5.  $D_\xi = E((\xi - E\xi)^2) = E(\xi^2 - 2\xi E\xi + (E\xi)^2) = E(\xi^2) - E(2\xi E\xi) + E((E\xi)^2) = E(\xi^2) - (E\xi)^2$

**Определение**

$\xi, \eta$  независимы, если  $\forall a, b$  события  $\xi = a$  и  $\eta = b$  независимы  
 Для непрерывных величин вместо  $=$  берем  $\leq$

**Пример 1**

Бросаем два кубика

$$\xi = v_1 + v_2$$

$$E_\xi = 7$$

**Пример 2**

Бросаем кубик

$$\xi = up + down$$

$$E_\xi = 7$$

**Пример 3**

$\Omega$  – перестановки  $n$  элементов

$$p(\sigma) = \frac{1}{n!}$$

$$\xi(\sigma) = |\{i : \sigma_i = i\}|$$

Утверждается, что  $E_\xi = 1$

Посчитать это через подсчет случаев сложно

Несмотря на это, мы можем посчитать матожидание

Пусть  $\xi_i = (\sigma_i = i)$

$$\xi = \xi_1 + \dots + \xi_n$$

$$E_{\xi_i} = P(\xi_i = 1) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

$$\text{Отсюда } \xi = n \frac{1}{n} = 1$$

**Свойства дисперсии**  $D(c\xi) = c^2 D(\xi)$

Дисперсия не линейна

**Теорема**

$$D(\xi + \eta) = D(\xi) + D(\eta)$$

**Доказательство**

$$\begin{aligned} D(\xi + \eta) &= E(\xi + \eta)^2 - E((\xi + \eta)^2) = E(\xi^2 + 2\xi\eta + \eta^2) - (E\xi)^2 - 2E\xi E\eta - \\ &= E\xi^2 + 2E\xi E\eta + E\eta^2 - (E\xi)^2 - 2E\xi E\eta - (E\eta)^2 = E\xi^2 - (E\xi)^2 + \\ &= E\eta^2 - (E\eta)^2 = D(\xi) + D(\eta) \end{aligned}$$

**Следствие**

$\xi_1, \dots, \xi_n$  – одинаково распределенные независимые случайные величины

$$\xi = \frac{1}{n} \sum_{i=1}^n \xi_i$$

$$E_\xi = E_{\xi_i}, D_\xi = \frac{1}{n} D_{\xi_i}$$

**Определение**

$\sigma = \sqrt{D_\xi}$  – среднеквадратичное отклонение

### 1.3 Хвостовые неравенства

**Неравенство Маркова**

Пусть  $\xi \geq 0, E_\xi > 0$

Оценим  $P(\xi \geq cE_\xi) \leq \frac{1}{c}$

**Доказательство**

$$P(\xi \geq cE_\xi) = \sum_{\substack{\omega \\ \xi(\omega) \geq cE_\xi}} p(\omega)$$

$$E_\xi = \sum_{\omega} p(\omega) \xi(\omega) = \sum_{\substack{\omega \\ \xi(\omega) \geq cE_\xi}} p(\omega) \xi(\omega) + \sum_{\substack{\omega \\ \xi(\omega) < cE_\xi}} p(\omega) \xi(\omega) \geq \sum_{\substack{\omega \\ \xi(\omega) \geq cE_\xi}} p(\omega) \xi(\omega) \geq$$

$$cE_\xi \sum_{\substack{\omega \\ \xi(\omega) \geq cE_\xi}} p(\omega) = cE_\xi P(\xi \geq cE_\xi)$$

$$1 \geq cP(\xi \geq cE_\xi)$$

$$P(\xi \geq cE_\xi) \leq \frac{1}{c}$$

**Неравенство Чебышева**

$P(|\xi - E_\xi| \geq c\sqrt{D_\xi}) \leq \frac{1}{c^2}$  – относительная форма неравенства Чебышева

**Доказательство**

Возьмем  $\eta = (\xi - E_\xi)^2$

**Неравенство Чебышева (ver. 2)**

$$c := \frac{a}{\sqrt{D_\xi}}$$

$P(|\xi - E_\xi| \geq a) \leq \frac{D_\xi}{a^2}$  – абсолютная форма неравенства Чебышева

**Пример**

Возьмем честную монету

$$E_\xi = \frac{1}{2}$$

$$D_\xi = \frac{1}{4}$$

$$D_\xi = E_\xi - (E_\xi)^2$$

$$P(|\xi - E_\xi| \geq \frac{1}{2}) \leq 1$$

$$P(|\xi - E_\xi| \geq 1) \leq \frac{1}{4} \text{ (на самом деле 0)}$$

Видим, что оценка сверху неточная

### Пример 2

$\xi_1, \dots, \xi_n$  — одинаково распределенные независимые случайные величины

$$\xi = \frac{1}{n} \sum_{i=1}^n \xi_i$$

$$P(|\xi - E_\xi| \geq \varepsilon) \leq \frac{D_\xi}{\varepsilon^2}$$

$$P(|\xi - E_\xi| \geq \varepsilon) \leq \frac{D_{\xi_i}}{n\varepsilon^2}$$

Пусть мы хотим не попадать в  $\varepsilon$ -окрестность с вероятностью не более  $\delta$  (вероятность промаха)

$$P(|\xi - E_\xi| < \varepsilon) > 1 - \delta$$

$$P(|\xi - E_\xi| \geq \varepsilon) \leq \delta$$

$$\text{Тогда } \frac{D_{\xi_i}}{n\varepsilon^2} \leq \delta$$

$$n \geq \frac{D_{\xi_i}}{\varepsilon^2 \delta} \sim \frac{1}{\varepsilon^2 \delta}$$

### Граница Чернова для монеты Бернулли

$$P(\xi \geq (1 + \delta)np) \leq e^{-\frac{\delta^2}{2+\delta}np}$$

$$P(\xi \leq (1 - \delta)np) \leq e^{-\frac{\delta^2}{2}np}$$

### Доказательство

Доказательства не будет, жди теорвер

## 1.4 Введение в информатику

### Определение

*информация* = -неопределенность (по Шеннону)

Рассмотрим модель случайного источника

Пусть есть вероятностное пространство  $\Omega$  и распределение  $p$

Получая событие  $\omega$ , мы получаем информацию о том, что оно произошло

Определим, сколько информации мы получаем

Заметим, что оно не зависит от  $\Omega$

Пусть  $H(p_1, p_2, \dots)$  – количество информации в зависимости от вероятностей событий

$H$  удовлетворяет следующим свойствам:

1. Для любого числа  $n$   $H(p_1, \dots, p_n)$  – непрерывно  
(т.к. при малом изменении вероятностей количество информации мало изменяется)

2.  $H(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}) = h(n)$   
 $h(n) \uparrow$  – (т.к. чем больше вариантов, тем больше информации)

3. Аддитивность

Пусть  $\Omega \subset \mathbb{R}^2$  – множество пар

$$A_a = \{(a, *) \in \Omega\}$$

$$P(A_a) = p_a$$

$$P(\{(a, b)\} | A_a) = q_{a,b}$$

$$p(\{(a, b)\}) = p_a q_{a,b}$$

$$\text{Тогда } H(p_1 q_{1,1}, \dots, p_1 q_{1,k_1}, \dots, p_n q_{n,1}, \dots, p_n q_{n,k_n}) = H(p_1, \dots, p_n) + \sum_i p_i H(q_{i,1}, \dots, q_{i,k_i})$$

Рассмотрим случай с равными вероятностями

$$h(nm) = h(n) + \sum_{i=1}^n \frac{1}{n} h(m) = h(n) + h(m)$$

**Лемма**

$$h(n) = c \log_2(n)$$

Традиционно  $c = h(2)$  – бит

**Доказательство**

$$h(2^k) = k h(2) = ck$$

Рассмотрим  $n^t, n, t \in \mathbb{N}$

Пусть  $2^k \leq n^t < 2^{k+1}$

Тогда  $h(2^k) \leq h(n^t) \leq h(2^{k+1})$   
 $ck \leq th(n) \leq c(k+1)$   
 $\frac{ck}{t} \leq h(n) \leq \frac{c(k+1)}{t}$   
 $k \leq t \log_2(n) \leq k+1$   
 $\frac{ck}{t} \leq c \log_2(n) \leq \frac{c(k+1)}{t}$   
 $|h(n) - c \log_2(n)| \leq \frac{c}{t}$  — при всех  $t$   
 Отсюда  $h(n) = c \log_2(n)$

"Разберемся" с  $H$   
 Начнем с  $p_i, q_i \in \mathbb{Q}$   
 Пусть  $p_i = \frac{a_i}{b}$   
 $k_i = a_i, q_{ij} = \frac{1}{a_i}$

Отсюда  $H(\frac{1}{b}, \dots, \frac{1}{b}) = H(p_1, \dots, p_n) + \sum_{i=1}^n p_i H(\frac{1}{a_i}, \dots, \frac{1}{a_i}) = H(p_1, \dots, p_n) +$

$$\sum_{i=1}^n c \log_2(a_i)$$

$$H(p_1, \dots, p_n) = -c(\sum_{i=1}^n p_i \log_2(a_i) - \log_2(b)) = -c(\sum_{i=1}^n p_i \log_2(a_i) - \sum_{i=1}^n p_i \log_2(b)) =$$

$$-c \sum_{i=1}^n p_i (\log_2(a_i) - \log_2(b)) = -c \sum_{i=1}^n p_i \log_2(\frac{a_i}{b}) = -c \sum_{i=1}^n p_i \log_2(p_i)$$

Из непрерывности формула верна для всех  $p_i \in \mathbb{R}$

Выберем  $c = 1$  бит

Тогда  $H = - \sum_{i=1}^n p_i \log_2(p_i)$  бит

Или  $H = \sum_{i=1}^n p_i \log_2(\frac{1}{p_i})$  бит — энтропия

Флешбеш: арифметическое кодирование использует в среднем  $H$  бит на каждый символ

Отсюда арифметическое кодирование — оптимальное кодирование для данных, которые можно аппроксимировать случайным источником

Ограничение в  $H$  бит на символ называют *энтропийным барьером*

Энтропийный барьер можно преодолеть лишь учетом закономерностей



в последовательности символов

Энтропия Шеннона хорошо описывает случайные последовательности и плохо описывает "регулярные" строки (строчки, имеющие закономерности)

Для измерения информации в более сложных объектах используется *Колмогоровская сложность*

Колмогоровская сложность зависит от *декодера* и равна количеству информации, необходимому для кодирования объекта

$K_A(s) \leq K_B(s) + C_{A,B}$ , где  $A, B$  – декодеры,  $C$  – константа

$K(s) \leq H(s) + C$

## 1.5 Цепи Маркова

### Определение

*Марковская цепь* – взвешенный ориентированный граф с неотрицательными весами и суммарным весом исходящих ребер, равным 1

Пронумеруем состояния (вершины)

Пусть  $b = (b_1 \ b_2 \ \dots \ b_n)$  – матрица состояния  $B$ , где  $b_i$  – вероятность находиться в  $i$ -ом состоянии ( $P(B = i)$ )

Пусть  $c = (c_1 \ c_2 \ \dots \ c_n)$  – матрица состояния  $C$

Рассмотрим матрицу переходов  $P = (p_{ij})_{n \times n}$ , где  $p_{ij}$  – вероятность перейти из  $i$  в  $j$

Найдем зависимость между  $b$  и  $c$

$$c_i = P(C = i) = \sum_{j=1}^n P(C = i | B = j) P(B = j) = \sum_{j=1}^n p_{ji} b_j$$

Отсюда  $c = b \cdot P$

Тогда распределение вероятностей на  $i$ -ом шаге  $b^i = b^0 P^i$ , где  $b^0$  – начальное состояние

Рассмотрим цепь Маркова как граф

Вершина в цепи Маркова называется *состоянием*

*Поглощающее (существенное) состояние* – состояние с кольцевым ребром веса 1

Цепь Маркова называется *поглощающей*, если из любого состояния можно попасть в поглощающее

Пример непоглощающей цепи: цепь с циклом длины 2 и более, где все ребра веса 1

*Эргодический класс* – компонента сильной связности графа Марковских цепей

*Компонента сильной связности* – максимальное по включению множество вершин, где из каждой можно дойти до каждой (класс эквивалентности для отношения достижимости)

Эргодический класс называется *поглощающим*, если из него нет исходящих переходов

Цепь Маркова можно представить как граф эргодических классов (но оценить веса ребер не всегда просто)

Цепь Маркова называется *эргодической*, если она содержит ровно 1 эргодический класс

Эргодический класс называется *периодическим с циклом  $d$* , если любая длина цикла в этом классе делится на  $d > 1$

### Теорема о классификации Марковских цепей

Любая Марковская цепь содержит поглощающие эргодические классы  
Марковская цепь с вероятностью 1 рано или поздно оказывается в состоянии из поглощающего эргодическим классом

Для непериодического поглощающего эргодического класса в случае попадания в него существует предельное распределение вероятностей  $b$  :  
 $b = bP$

Для любого распределения  $b^0$   $b^0 P^n \rightarrow b$

Для цепей Маркова существуют две независимые задачи: задача поглощения – задача определения, в какой поглощающий эргодический класс мы попадем, и задача стационарного распределения внутри поглощающего эргодического класса

Займемся задачей поглощения (т.е. определим, в какой эргодический класс мы попадем)

В ходе решения этой задачи поглощающие эргодический классы можно заменить на одно поглощающее состояние

Занумеруем состояния так, чтобы сначала шли непоглощающие, а потом поглощающие

Пусть  $1 \dots m$  – непоглощающие состояния,  $m + 1 \dots n$  – поглощающие

Тогда  $P = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}$ , где

$$Q = P[1 \dots m][1 \dots m]$$

$$R = P[1 \dots m][m+1 \dots n]$$

$0 = P[m+1 \dots n][1 \dots m]$  – нулевая матрица

$I = P[m+1 \dots n][m+1 \dots n]$  – единичная матрица

Возьмем матрицу состояния  $b = (b_1 \ b_2 \ \dots \ b_n)$

Пусть  $a = (b_1 \ b_2 \ \dots \ b_m)$

$$a^n = a^0 Q^n$$

**Лемма**

$$Q^n \rightarrow 0$$

**Доказательство**

Пусть  $L$  – максимальная длина кратчайшего пути от  $i$  до поглощающей

Найдем  $X = Q^L$

$$x_{ij} = \sum_{k_1, k_2, \dots, k_{L-1}} q_{ik_1} q_{k_1 k_2} \dots q_{k_{L-1} j}$$

$$\sum_{j \text{ -- не погл.}} x_{ij} = \sum_{k_1, k_2, \dots, k_{L-1}, j} q_{ik_1} q_{k_1 k_2} \dots q_{k_{L-1} j} = \delta_i < 1 \text{ -- т.к. это вероятность}$$

пройти от  $i$  до непоглощающего состояния (если бы до любого состояния, то было бы 1)

$$\text{Отсюда } \max_{i=1 \dots m} \sum_{j \text{ -- не погл.}} x_{ij} = \max \delta_i = \delta < 1$$

Тогда  $Q^n = Q^L Q^{n-L}$

Пусть  $\max Q^{n-L} = v_{n-L}$

$$Q^n_{ij} = (Q^L Q^{n-L})_{ij} = \sum_k Q^L_{ik} Q^{n-L}_{kj} \leq \sum_k Q^L_{ik} v_{n-L} \leq \delta v_{n-L}$$

$$v_n \leq \delta^{\lfloor \frac{n}{L} \rfloor} \rightarrow 0$$

Тогда  $Q^n \rightarrow 0$

**Теорема о поглощении**

Поглощающая Марковская цепь переходит в состояние поглощения с вероятностью 1

**Доказательство**

Следует из леммы

Научимся определять, где же мы поглотимся

Для этого найдем мат. ожидание времени до поглощения

$b^0$  – начальное распределение

$T$  – случайная величина – число шагов до поглощения

$T = \sum_{i=1}^m T_i$ , где  $T_i$  – число посещений  $i$ -ого состояния

$T_i = \sum_{j=0}^{\infty} T_{ij}$ , где  $T_{ij} = \begin{cases} 1, & \text{если на } j\text{-ом шаге мы в состоянии } i \\ 0, & \text{иначе} \end{cases}$

**Лемма**

$$\sum_{j=0}^{\infty} Q^j = (I - Q)^{-1}$$

**Доказательство**

$$(I - Q)(I + Q + Q^2 + \dots + Q^n) = I + Q + Q^2 + \dots + Q^n - Q - Q^2 - \dots - Q^{n+1} = I - Q^{n+1} \rightarrow I$$

**Определение**

$N = (I - Q)^{-1}$  – фундаментальная матрица поглощения Марковской цепи

$$\begin{aligned} ET &= \sum_{i=1}^m ET_i = \sum_{i=1}^m \sum_{j=0}^{\infty} ET_{ij} = \sum_{i=1}^m \sum_{j=0}^{\infty} P(\text{цепь в состоянии } i \text{ на шаге } j) = \\ &= \sum_{i=1}^m \sum_{j=0}^{\infty} (a^0 Q^j)_i = \sum_{i=1}^m \left( \sum_{j=0}^{\infty} a^0 Q^j \right)_i = \sum_{i=1}^m \left( a^0 \sum_{j=0}^{\infty} Q^j \right)_i = \sum_{i=1}^m (a^0 N)_i = a^0 N \mathbb{1} \end{aligned}$$

Заметим, что  $a^0 N = (ET_1 \quad ET_2 \quad \dots \quad ET_m)$

$$\begin{aligned} P(\text{погл. в } j) &= \sum_{i=1}^m P(\text{погл. в } j \text{ из } i) P(\text{быть в } j) = \sum_{t=0}^{\infty} \sum_{i=1}^m P(\text{погл. в } j \text{ из } i) P(\text{быть в } i \text{ на шаге } t) = \\ &= \sum_{t=0}^{\infty} \sum_{i=1}^m R_{i,j-m} P(\text{быть в } i \text{ на шаге } t) = \sum_{i=1}^m R_{i,j-m} \sum_{t=0}^{\infty} P(\text{быть в } i \text{ на шаге } t) = \\ &= \sum_{i=1}^m (a^0 N)_i R_{i,j-m} = (a^0 N R)_{j-m} \end{aligned}$$

$$\text{Отсюда } A = \begin{pmatrix} P(\text{погл. в } m+1) & P(\text{погл. в } m+2) & \dots & P(\text{погл. в } n) \end{pmatrix} = a^0 N R$$