Математика и её приложения

А. М. Райгородский

05.07.2025

Тема 1. Комбинаторика, гиперграфы и вероятностные неравенства

1 Комбинаторика и гиперграфы

1.1 Множества и подмножества

Пусть у нас есть множество, состоящее из чисел от 1 до n:

$$X = \{1, 2, 3, \dots, n\}$$

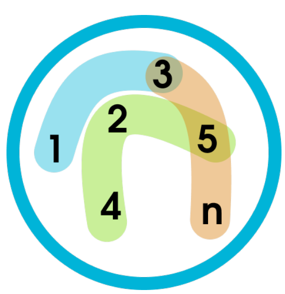
Рассмотрим **произвольные подмножества** этого множества. Обозначим их:

$$M_1, M_2, \dots, M_m \subseteq X$$

Каждое подмножество — это группа элементов из исходного множества. Например, такие подмножества:

$$M_1 = \{1, 3\}, \quad M_2 = \{2, 4, 5\}, \quad M_2 = \{3, 5, n\}, \quad \text{и т.д.}$$

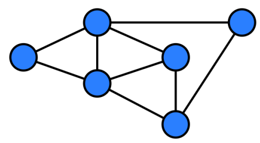
Мы можем представить элементы множества точками на плоскости, а каждое подмножество — как область, содержащую эти точки. Графически это напоминает сардельки, обводящие отдельные группы точек.



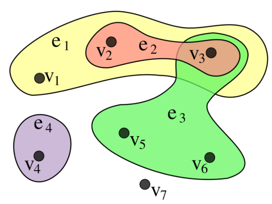
1.2 Гиперграф

Это подводит нас к понятию **гиперграфа**.

- **Граф** — структура, где вершины соединяются рёбрами по две.



- **Гиперграф** может соединять **любой набор** вершин, не обязательно только две.



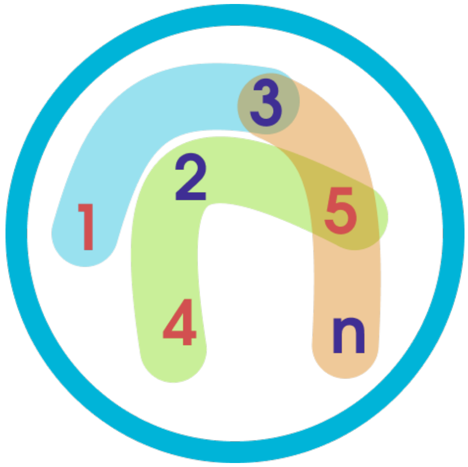
Пример:
Аудитория — множество всех студентов X.
Каждое M_i — группа студентов с общими интересами (например, физики, математики, биологи). При этом каждый студент может быть одновременно, например, и физиком, и биологом. Или математиком, информатиком и гумманитарием (не дай Бог). Как и вершины в гиперграфе могут принадлежать сразу нескольким ребрам.

2 Теорема о сбалансированной раскраске

Пусть у нас есть гиперграф $(X, \{M_1, M_2, \ldots, M_m\})$.
Мы хотим раскрасить каждую вершину (элемент множества X) **в красный или синий цвет**. То есть задать функцию:

$$f : X \to \{\text{красный}, \text{синий}\}$$

Цель: добиться, чтобы **в каждом подмножестве** M_i количества красных и синих были **почти равны**.



2.1 Формулировка теоремы

Для любых подмножеств $M_1, \dots, M_m \subseteq \{1, \dots, n\}$, существует такая раскраска элементов множества в красный и синий цвета, что **в каждом** M_i разность между числом красных и числом синих элементов по модулю не превышает:

$$\sqrt{2n \ln(2m)}$$

2.2 Интерпретация

- Представим, что каждая "сарделька"M_i содержит студентов.
- Мы красим студентов либо в синий (0), либо в красный (1).
- Эта теорема утверждает: как бы ни пересекались интересы (группы M_i), всегда можно покрасить студентов так, что **внутри каждой группы будет почти пополам** — почти одинаково красных и синих.

2.3 Пример

Пусть:

- $n = 1000$, $m = 1000$
- Тогда:

$$\sqrt{2n \ln(2m)} = \sqrt{2 \cdot 1000 \cdot \ln(2000)} \approx \sqrt{2000 \cdot 7.6} \approx \sqrt{15200} \approx 123$$

То есть если сарделек 1000 и мы поместили их в 100 групп, то разница между цветами внутри любой сардельки не больше 123.

3 Вероятностные неравенства

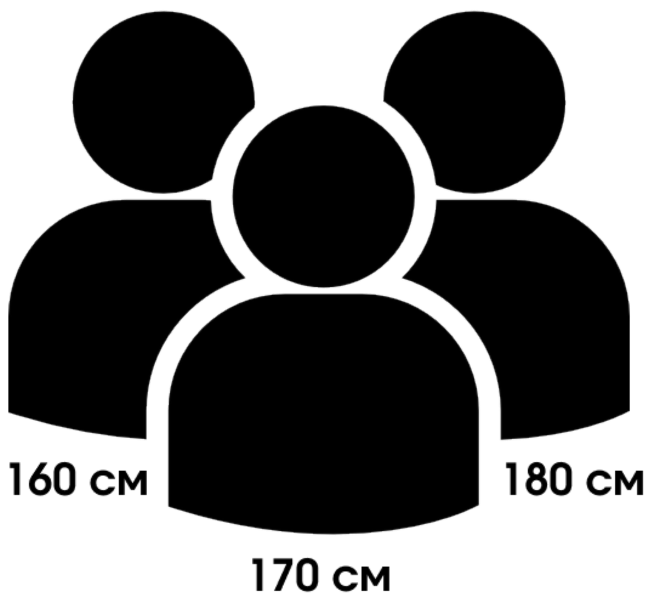
Чтобы доказать такую теорему, нужны инструменты теории вероятностей. Начнём с базовых понятий.

3.1 Матожидание (ожидаемое значение)

Пусть X — случайная величина. Тогда её **математическое ожидание** — это среднее значение, которое мы ожидаем от X, если будем многократно повторять эксперимент.

$$\mathbb{E}[X] = \sum_k y_k \cdot P(X = y_k)$$

Пример:
Вы выбираете случайного человека в аудитории и измеряете его рост.



- Пусть возможные значения роста: $y_1 = 160$, $y_2 = 170$, $y_3 = 180$
- Соответствующие вероятности: (будем считать, у каждого студента разная вероятность вызова к доске, т.к. у преподавателя могут быть любимчики, некоторые студенты имеют плохую успеваемость и т.д.) $P(X = y_1) = 0.2$, $P(X = y_2) = 0.5$, $P(X = y_3) = 0.3$

Тогда:

$$\mathbb{E}[X] = 160 \cdot 0.2 + 170 \cdot 0.5 + 180 \cdot 0.3 = 32 + 85 + 54 = 171$$

3.2 Дисперсия

Дисперсия показывает, насколько разбросаны значения случайной величины вокруг среднего.

$$D(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

3.3 Неравенство Маркова

Пусть $X \geq 0$, тогда:

$$P(X > a) \leq \frac{\mathbb{E}[X]}{a}$$

Доказательство:

$$E[X] = \sum_x x \cdot P(X = x) \quad \text{по всем возможным } x$$

Разобьём сумму на две части: где $x \geq a$ и где $x < a$

$$E[X] \geq \sum_{x \geq a} x \cdot P(X = x) \geq a \cdot \sum_{x \geq a} P(X = x) = a \cdot P(X \geq a)$$

Отсюда:

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Пример: Если ожидаемое значение роста равно 171 см, то вероятность того, что рост > 200 см, не превышает $\frac{171}{200} \approx 0.855$.

3.4 Неравенство Чебышёва

$$P(|X - \mathbb{E}[X]| > a) \leq \frac{D(X)}{a^2}$$

Доказательство через неравенство Маркова:
Рассмотрим $Y = (X - E[X])^2$

$$P(|X - E[X]| \geq a) = P(Y \geq a^2)$$

Применим неравенство Маркова к Y:

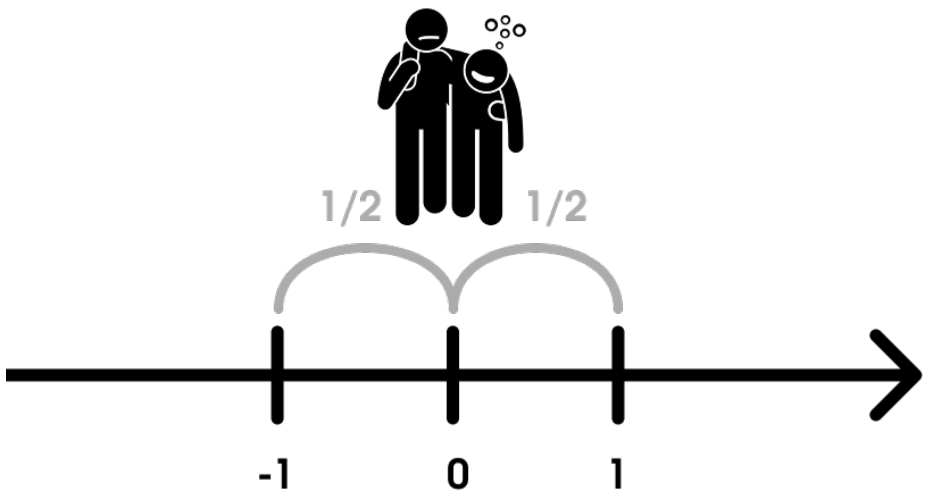
$$P(Y \geq a^2) \leq \frac{E[Y]}{a^2} = \frac{D[X]}{a^2}$$

Пример: Если дисперсия роста равна 100, то вероятность отклонения от среднего более чем на 20 см:

$$\leq \frac{100}{400} = 0.25$$

4 Случайное блуждание и пьяница

Представим, что по прямой линии (на оси чисел) из точки 0 (из кабака) выходит **пьяный человек**:



- Каждый шаг: либо $+1$, либо -1, с вероятностью $\frac{1}{2}$
- После n шагов его положение:

$$Y_n = X_1 + X_2 + \cdots + X_n$$

где каждый $X_i \in \{+1, -1\}$.

4.1 Матожидание пьяницы

$$\mathbb{E}[Y_n] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] = 0$$

Пьяница **в среднем** остаётся рядом с кабаком.

4.2 Дисперсия положения

$$\mathrm{Var}(Y_n) = \mathrm{Var}(X_1 + \cdots + X_n) = n$$

4.3 Оценим вероятность ухода далеко

Используем неравенство Чебышёва:

$$P(|Y_n| > a) \le \frac{n}{a^2}$$

Пьяница делает 10^6 шагов. Какова вероятность, что он уйдёт дальше, чем на 10^4 шагов? То есть казалось бы, он сделал целых миллион шагов и уйти ему надо на какое-то относильно малое расстояние — 10000. Однако подставим в формулу:

$$P(|Y_n| > 10^4) \leq \frac{10^6}{(10^4)^2} = \frac{10^6}{10^8} = \frac{1}{100}$$

Однако это все еще остается большой вероятностью по сравнению с более точным подсчетом. Давайте к нему перейдем

4.4 Центральная предельная теорема (ЦПТ)

При больших n поведение суммы Y_n приближается к нормальному распределению:

$$P(Y_n > a) \approx e^{-\frac{a^2}{2n}}$$

Если $n = 10^6$, $a = 10^4$, то:

$$P(Y_n > 10^4) \approx e^{-50}$$

Стоим сказать, что это ужасно маленькое число, которое меньше даже компьютерного нуля.

5 Неравенство Азума–Хефдинга

Мощное неравенство концентрации:

$$P(Y_n > a) \leq \exp\left(-\frac{a^2}{2n}\right)$$

Доказательство с помощью Маркова:

$$\begin{aligned} P(Y_n > a) &= P(e^{\lambda Y_n} > e^{\lambda a}) \\ &\leq \frac{\mathbb{E}[e^{\lambda Y_n}]}{e^{\lambda a}} \end{aligned}$$

Используем независимость X_i:

$$\mathbb{E}[e^{\lambda Y_n}] = \prod_{i=1}^{n} \mathbb{E}[e^{\lambda X_i}] = \left(\frac{e^{\lambda} + e^{-\lambda}}{2}\right)^n = (\cosh(\lambda))^n$$

Используем неравенство:

$$\cosh(\lambda) \leq e^{\lambda^2/2} \Rightarrow (\cosh(\lambda))^n \leq e^{n\lambda^2/2}$$

Тогда:

$$P(Y_n > a) \leq \exp\left(\frac{n\lambda^2}{2} - \lambda a\right)$$

Берём оптимальное $\lambda = \frac{a}{n}$, получаем:

$$P(Y_n > a) \leq \exp\left(-\frac{a^2}{2n}\right)$$