Разбор олимпиады ШАД

A. Май в Квадратленде

Карта одного квадратного государства разбита на $4 \times 4 = 16$ равных квадратных областей. Местный синоптик Гаврило заявляет, что в мае в Квадратленде каждый год происходят чудеса и температура на любом внутреннем участке равна среднему арифметическому температур соседних (2 области называются соседними, если у них есть общая грань). Математик Никколо узнав об этом заявил, что в мае в Квадратленде температуру на любом участке почти всегда можно восстановить, зная лишь температуру на границе государства (граничных областях). На что физик Алессио заметил, что не почти, а вообще всегда.
Разберитесь строго, правы ли Никколо и Алессио, если мы знаем, что утверждение Гаврило истинно.

Ответ: утверждение Аллесио истинно

Решение

- Карта государства в мае будет выглядеть следующим образом, значение клетки это её температура:

a_{11}	a_{12}	a_{13}	a_{14}
a_{21}	x_1	x_2	a_{24}
a_{31}	x_3	x_4	a_{34}
a_{41}	a_{42}	a_{43}	a_{44}

- По условию значения клеток будут равны

$$\begin{cases} x_1 = \dfrac{a_{12} + a_{21} + x_2 + x_3}{4} \\ x_2 = \dfrac{a_{13} + a_{24} + x_1 + x_4}{4} \\ x_3 = \dfrac{a_{31} + a_{42} + x_1 + x_4}{4} \\ x_4 = \dfrac{a_{43} + a_{34} + x_3 + x_2}{4} \end{cases} \tag{1}$$

- Система 1 эквивалентна данной:

$$\begin{cases} 4x_1 - x_2 - x_3 = a_{12} + a_{21} \\ 4x_2 - x_1 - x_4 = a_{13} + a_{24} \\ 4x_3 - x_1 - x_4 = a_{31} + a_{42} \\ 4x_4 - x_2 - x_3 = a_{43} + a_{34} \end{cases} \tag{2}$$

- Перепишем в матричном виде

$$\begin{pmatrix} 4 & -1 & -1 & 0 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} a_{12} + a_{21} \\ a_{13} + a_{24} \\ a_{31} + a_{42} \\ a_{43} + a_{34} \end{pmatrix} \tag{3}$$

- Нетрудно проверить, что матрица системы обратима, а значит мы можем всегда восстановить значения внутренних клеток по любым значениям температуры на границе, причем единственным образом.

- В качестве бонуса поймем, что матрица обратима, без вычисления определителя. Для этого заметим, что наша матрица имеет *строгое диагональное преобладание*: в каждой строке диагональный элемент по модулю строго больше суммы модулей остальных элементов. Отсюда сразу же следует обратимость!

- В самом деле, предположим, что наша матрица A необратима. Тогда существует ненулевой вектор $x = (x_1, \dots, x_n) \neq 0$, являющийся решением системы $Ax = 0$. Пусть x_j это наибольший по модулю его элемент. Рассмотрим выражение $(Ax)_j = \sum_{i=1}^{n} a_{j,i} x_i$. С одной стороны, оно равно нулю, потому что весь Ax нулевой. С другой стороны, $|a_{j,j} x_j| > \sum_{i \neq j} |a_{j,i} x_i|$, так как $|x_j| > |x_i|$ и $|a_{j,j}| > \sum_{i \neq j} |a_{j,i}|$, поэтому $(Ax)_j$ нулем быть не может. Противоречие!

B. Питерские пышки

В одной из питерских пышечных Владу показалось, что одна из пышек полностью покрыта числами и каждую секунду числа менялись определённым образом: каждое из них превращалось в среднее арифметическое ближайших. Интересная динамика решил Влад. А что будет в пределе?
Для ответа на этот вопрос Влад представил пышку как полоску клетчатой бумаги размером 4×4 клеток со следующими числами на ней:

2007	2008	2009	2010
2014	2013	2012	2011
2015	2016	2017	2018
2022	2021	2020	2019

А затем склеил левую и правую границы полоски, а также верхнюю и нижнюю. Получилась модель пышки да так, что 2007 граничит теперь с 2022 и 2010. А теперь промоделируем пышечную динамику: каждую секунду числа на пышке одновременно заменяются на средние арифметические своих 4 соседей: двух вертикальных и двух горизонтальных.
Существует ли предел значения числа в ячейке, где изначально стояло число 2018?

Ответ: Предела не существует

Решение

- У этой задачи есть несколько решений. По-видимому, самое естественное решение такое: раскрасим пышку в шахматном порядке и заметим, что каждый ход сумма чисел одного цвета переходит в сумму чисел другого цвета, и наоборот. Получается, что через два хода сумма чисел одного цвета не меняется. Обратим внимание, что сумма чисел у двух цветов разная. После этого аккуратно докажем, что с течением времени все значения одного цвета на четных секундах будут стремиться к своему среднему арифметическому, откуда будет следовать, что предела нет: на четных секундах значение в каждой клетке будет близко к одному среднему, а на нечетных — к другому.

- Мы же пойдем немного другим путем! Заметим, что описанное в задаче преобразование (обозначим его $\mathcal{A}$) является линейным отображением на пространстве матриц размера 4×4. Это отображение очень симметрично устроено, поэтому у него можно довольно быстро найти кучу собственных векторов. Например, если v_1 это матрица, заполненная единицами, то $\mathcal{A}(v_1)$ будет равно v_1, что означает, что v_1 это собственный вектор для собственного числа 1.

- Аналогично, нетрудно проверить, что следующие матрицы (назовем их v_2, v_3, v_4 и v_5 соответственно) будут собственными векторами нашего отображения:

$$\begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 \end{pmatrix},$$

В самом деле, $\mathcal{A}(v_2) = -1 \cdot v_2$, $\mathcal{A}(v_3) = -\frac{1}{2}v_3$, $\mathcal{A}(v_4) = 0$, $\mathcal{A}(v_5) = \frac{1}{2}v_5$.

- Матрицы $v_1, \ldots, v_5$ выбраны не случайно: оказывается, наша исходная матрица, назовем ее P, является их линейной комбинацией:

$$P = \begin{pmatrix} 2007 & 2008 & 2009 & 2010 \\ 2014 & 2013 & 2012 & 2011 \\ 2015 & 2016 & 2017 & 2018 \\ 2020 & 2021 & 2020 & 2019 \end{pmatrix} = 2014.5v_1 - \frac{1}{2}v_2 - v_3 - 2v_4 - 4v_5.$$

- Представив исходную матрицу в виде линейной комбинации собственных векторов, мы получаем возможность очень просто выражать результат многократного применения преобразования $\mathcal{A}$: при каждом применении собственные вектора будут умножаться на соответствующие собственные числа:

$$\mathcal{A}^n(P) = 2014.5 \cdot 1^n \cdot v_1 - \frac{1}{2} \cdot (-1)^n \cdot v_2 - (-\frac{1}{2})^n \cdot v_3 - 2 \cdot 0^n \cdot v_4 - 4 \cdot 0.5^n \cdot v_5.$$

- С ростом n последние три слагаемых будут стремиться к нулю (слагаемое с v_4 станет нулем уже при $n = 1$), а вот первые два — нет. При четных n получаем $\mathcal{A}^n(P) \to 2014.5v_1 - 0.5 \cdot v_2$, а при нечетных — $\mathcal{A}^n(P) \to 2014.5v_1 + 0.5 \cdot v_2$. Так как пределы при четных и нечетных n не совпадают, то общего предела нет ни в одной клетке исходной матрицы.

- Напоследок ответим на два вопроса. Во-первых, как по имеющимся векторам $v_1, \ldots, v_5$ определить, с какими коэффициентами они будут участвовать в разложении P? Для ответа на этот вопрос заметим, что все вектора являются попарно ортогональными относительно скалярного произведения $\langle A, B\rangle = \sum_{i,j} A_{i,j}B_{i,j}$. В этом случае коэффициент λ_i перед собственным вектором v_i определяется по формуле

$$\lambda_i = \frac{\langle P, v_i\rangle}{\langle v_i, v_i\rangle}.$$

- Во-вторых, можно ли обойтись без поиска всех собственных векторов, участвующих в разложении P? Да, можно. В данной задаче достаточно найти собственные вектора v_1 и v_2, соответствующие собственным числам 1 и -1, после чего найти λ_1 и λ_2 по формуле выше, посчитать матрицу $R = P - \lambda_1 v_1 - \lambda_2 v_2$ и посмотреть, что с ней будет происходить, если мы будем последовательно применять к ней преобразование $\mathcal{A}$. Довольно быстро станет понятно, что она стремительно уменьшается, поэтому не будет влиять на разницу между $\lambda_1 v_1 - \lambda_2 v_2$ и $\lambda_1 v_1 + \lambda_2 v_2$ на четных и нечетных секундах.

C. Дороги под Минском

В минской области дороги между городами устроены максимально просто: любые 2 города соединяет трасса (на трассе может быть и больше городов), а любые две трассы пересекаются ровно в одном городе, а вне городов не пересекаются.
Мы знаем, что всего в минской области n трасс. Какое в минской области может быть число городов?
Напишите в ответ минимальное и максимальное возможные значения числа городов для $n = 2071$

Ответ: 2071 и 2071

Решение

- Пусть m — число городов. Заведём матрицу $A_{m\times n} = \{A_{ij}\}$, где A_{ij} будет равно 1, если через i-ый город проходит j-ая трасса и 0 иначе.

- Рассмотрим матрицу AA^T размера $m \times m$ и матрицу A^TA размера $n \times n$:

$$AA^T = \begin{pmatrix} a_1 & 1 & \ldots & 1 \\ 1 & a_2 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & a_m \end{pmatrix} \qquad A^TA = \begin{pmatrix} b_1 & 1 & \ldots & 1 \\ 1 & b_2 & \ldots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \ldots & b_n \end{pmatrix}$$

- $(AA^T)_{ij}$ = *количество общих трасс для городов i и j*, причем в случае $i = j$ это количество трасс, проходящих через данный город. По условию через два различных города проходит ровно одна трасса, поэтому все недиагональные элементы матрицы равны 1. Диагональные элементы удовлетворяют неравенству $a_i \geq 2$, потому что иначе через какой-то город проходит ровно одна трасса, и значит на этой трассе находятся все остальные города, что невозможно.

- $(A^T A)_{ij}$ = *количество общих городов для трасс i и j*, причем в случае $i = j$ это количество городов на данной трассе. По условию две трассы пересекаются ровно в одном городе, поэтому все недиагональные элементы матрицы равны 1, а диагональные удовлетворяют неравенству $b_i \geq 2$.

- Заметим, что обе матрицы положительно определены: каждая из них является суммой матрицы из всех единиц и диагональной матрицы, у которой на диагонали стоят числа не меньшие 1. Матрица из всех единиц является неотрицательно определенной, а упомянутая диагональная матрица положительно определенной. Сумма неотрицательно определенной матрицы и положительной определенной является положительно определенной.

- Как известно, положительно определенные матрицы невырождены, но такое может быть только если $n = m$, так как ранг матриц $A^T A$ и AA^T равен $\min\{m, n\}$ и он должен совпадать с размером обеих матриц.

- Окончательный ответ: $m = 2071$.

Разбор Олимпиады ШАД

D. Цу-е-фа!

После пар 10 студентов ШАДа из Еката решили отдохнуть и поиграть в плойку. Но вот беда: в этот вечер она сломалась.
Тогда они вспомнили старинную игру под названием «Камень, ножницы, бумага».
Правила игры от «YandexGPT 2.0»

1. Все игроки должны встать или сесть друг напротив друга.
2. Одновременно с другими игроками нужно показать одну из фигур: «камень» (сжатый кулак), «ножницы» (указательный и средний пальцы) или «бумага» (открытая ладонь).
3. Камень побеждает ножницы, так как он их ломает. Ножницы побеждают бумагу, так как они её режут. Бумага побеждает камень, так как она его заворачивает.
4. Если игроки показали одинаковые фигуры или встретились фигуры всех 3 типов, то объявляется ничья и можно играть ещё раз
5. Если все игроки показали фигуры лишь 2 типов, проигравший выбирается случайным образом из показавших более слабую фигуру.

Найдите математическое ожидание числа раундов до выявления победителя. Напомним, что студентов всего 10

Ответ: $\sum_{n=2}^{10} \frac{3^{n-1}}{2^n-2}$

Решение

- Для фиксированного количества участников $n \geq 2$ найдем вероятность p_n того, что будет проигравший. Количество различных вариантов того, что выберут участники, равно 3^n. Количество вариантов, когда среди выбранных фигур встретятся в точности две, равно $3 \cdot (2^n - 2)$: тремя способами можно выбрать две различные фигуры из трех; всего есть 2^n вариантов выбора участников из этих двух фигур; из этих вариантов нам не подходит два — когда все выбрали одну фигуру и когда все выбрали другую. Таким образом вероятность того, что в данном раунде будет проигравший, равна

$$p_n = \frac{2^n - 2}{3^{n-1}} \tag{1}$$

- Заметим, что для каждого n от 10 до 2, количество ходов, пока кто-то не проиграет, это геометрически распределённая случайная величина с параметром p_n, её матожидание равно

$$\frac{1}{p_n} = \frac{3^{n-1}}{2^n - 2} \tag{2}$$

Матожидание числа шагов, которое потребуется, чтобы число игроков сократилось с 10 до одного, равно сумме таких матожиданий по n от 10 до 2:

$$\sum_{n=2}^{10} \frac{3^{n-1}}{2^n - 2} \tag{3}$$

E. Хобби

В новосибирском ШАДе есть 20 клубов по интересам, начиная с любителей сидра и заканчивая поклонниками методов рационального мышления. Разумеется, далеко не все в них состоят, а кто-то может состоять и в нескольких. Алла, новый куратор ШАДа, решила разобраться, какие именно студенты входят в эти клубы. Но сами студенты не готовы делиться этой закрытой информацией. Они лишь готовы отвечать ей на вопросы вида «Какие именно студенты входят в объединение и пересечение двух данных клубов?».
В ответе необходимо написать минимальное число вопросов для идентефикации участников всех клубов

Ответ: 20

Решение

- Для того, чтобы доказать, что ответ в этой задаче 20, нам нужно сделать два шага: объяснить, почему за 20 вопросов всегда удастся идентифицировать участников, и объяснить, почему меньше чем за двадцать в общем случае нельзя.

- Способ узнать за 20: возьмём три множества A, B, C и попарно зададим три вопроса, узнав шесть множеств: $A \cup B$, $A \cap B$, $A \cup C$, $A \cap C$, $B \cup C$, $B \cap C$. Эти множества мы можем при желании объединять, пересекать, вычитать и так далее. Так, например, мы можем определить множество $A \cup B \cup C = (A \cup B) \cup (A \cup C)$. По этим множествам мы можем определить множество A:
$$A = (A \cup B \cup C) \setminus (B \cup C) \cup (A \cap C) \cup (A \cap B) \tag{4}$$
Аналогичным образом определим и множества B и C.

- Остальные 17 вопросов используем, чтобы найти множества вида $A \cup S_i$, $A \cap S_i$. Зная A и эти множества, мы можем найти каждое S_i по формуле:
$$S_i = (A \cup S_i) \setminus (A \setminus (A \cap S_i)). \tag{5}$$

- Покажем, почему за 19 вопросов, вообще говоря, нельзя определить все клубы. Для этого построим граф, в котором вершинами являются клубы, а ребрами соединены вершины, про которые был задан вопрос. Таким образом получим граф на 20 вершинах с не более чем 19 ребрами.

- Заметим, что в графе, в котором ребер меньше чем вершин обязательно будет существовать компонента связности, являющаяся деревом: если в каждой компоненте есть цикл, то ребер в ней будет по крайней мере столько же сколько вершин, и значит общее число ребер будет не меньше общего числа вершин.

- Если какое-то из деревьев состоит из одной вершины, то про этот клуб мы не задали ни одного вопроса, и поэтому конечно же его определить не удастся. Если какое-то из деревьев имеет более одной вершины, то давайте раскрасим его вершины в два цвета, черный и белый, так, чтобы соседние вершины имели разный цвет. После этого рассмотрим следующие две группы студентов: те, кто входит во все черные клубы и больше ни в какой другой, и те, кто входит во все белые клубы и больше ни в какой другой.

- Заметим, что по заданным вопросам нам не удастся определить, какие студенты входят в первую группу, а какие во вторую: студенты обеих групп не лежат ни в одном известном пересечении (никакие два черных или белых клуба не соединены ребром), но лежат во всех известных объединениях внутри дерева (любое ребро дерева соединяет белый и черный клуб). Раз не можем определить указанные две группы, то не сможем определить и участников черных и белых клубов по отдельности.

F. Приходите завтра

Захотелось Вале при подготовке к олимпиаде в ШАД почитать Кормена на бумаге. Да вот беда! Как не придёт она в библиотеку, там всегда говорят: «Мы закрыты, приходите завтра!». Валя знает, что библиотека точно не работает, начиная с какого-то фиксированного времени. И более того, если в ней нет посетителей, она может и раньше закрыться.
Решила Валя оценить официальное время закрытия следующим образом:

- Валя предположила, что официальное время закрытия (T) распределено равномерно на отрезке $[0, 1]$. (Тут она конечно для удобства отнормировала время.)
- 5 раз она приходила в разные дни и фиксировала время, когда библиотека была закрыта: $(s_1, \ldots, s_5)$.

В ответе требуется написать математическое ожидание T при наблюдаемых Валей событиях: $s_1 = 0.3, s_2 = 0.6$, $s_3 = 0.4, s_4 = 0.25, s_5 = 0.5$.
Мы считаем, что времена самовольного закрытия в разные дни независимы, а время самовольного закрытия распределено равномерно на отрезке $[0, T]$.

Ответ: Ниже

Решение

- Пусть T это официальное время закрытия, F_i это фактическое время закрытия в i-й день, а s_i — время в i-й день, когда пришла Валя. Известно, что

    1. $T \sim U([0, 1])$,

    2. F_i независимы при условии, что нам известно T, и имеют распределение $U([0, T])$,

    3. $F_i < s_i$ для всех i.

- Искомая величина это
$$\mathbb{E}\left(T \mid F_1 < s_1, \ldots, F_5 < s_5\right).$$

- Для того, чтобы посчитать это матожидание, посчитаем условную плотность T по формуле Байеса с точностью до множителя (посчитаем только числитель):
$$p\left(T \mid F_1 < s_1, \ldots, F_5 < s_5\right) \propto p\left(F_1 < s_1, \ldots, F_5 < s_5 \mid T\right) p(T) = p(T) \prod_{i=1}^{5} p\left(F_i < s_i \mid T\right),$$
где $\propto$ это значок пропорциональности, а последнее равенство получается из независимости событий $F_i < s_i$ при условии, что нам известно T.

- Заметим, что
$$p\left(F_i < s_i \mid T\right) = \begin{cases} 1, & T < s_i, \\ \frac{s_i}{T}, & s_i < T \end{cases}.$$

- Получаем, что
$$f(T) := p\left(T \mid F_i < s_i, \forall i\right) \propto p(T) \prod_i p\left(F_i < s_i \mid T\right) = \begin{cases} 1, & 0 \le T \le 0.25, \\ \frac{0.25}{T}, & 0.25 < T \le 0.3, \\ \frac{0.25 \cdot 0.3}{T^2}, & 0.3 < T \le 0.4, \\ \frac{0.25 \cdot 0.3 \cdot 0.4}{T^3}, & 0.4 < T \le 0.5, \\ \frac{0.25 \cdot 0.3 \cdot 0.4 \cdot 0.5}{T^4}, & 0.5 < T \le 0.6, \\ \frac{0.25 \cdot 0.3 \cdot 0.4 \cdot 0.5 \cdot 0.6}{T^5}, & 0.6 < T \le 1, \\ 0, & T \notin [0, 1]. \end{cases}$$

- Чтобы получить настоящую условную плотность, осталось разделить $f(T)$ на ее интеграл по T:
$$p(t \mid F_i < s_i, \forall i) = \frac{f(t)}{\int_0^1 f(t) dt}.$$

  В ответе имеем
$$\mathbb{E}(T \mid F_i < s_i, \forall i) = \int_0^1 t\, p(t | F_i < s_i, \forall i)\, dt = \frac{\int_0^1 t f(t) dt}{\int_0^1 f(t) dt}.$$