Теория вероятностей и линейная алгебра

Задачи

Определение: Вероятностное пространство (или множество возможных исходов) — конечное (или счётное) множество U. Вероятностное распределение — функция $\Pr : U \to [0,1]$, удовлетворяющая условию $\sum_{x \in U} \Pr[x] = 1$. $\Pr[x]$ называется вероятностью исхода $x \in U$. Событие — это какое-то подмножество исходов. Вероятность события определяется как $\Pr[A] = \sum_{x \in A} \Pr[x]$.

Задача 1. Есть 3 двери. За двумя нет ничего, за третьей — диплом победа всероса. Вы выбрали какую-то дверь, после чего из оставшихся двух вам открыли какую-то, за которой ничего нет, и предложили поменять свой выбор. Стоит ли это делать?

Задача 2. Всего есть n тем в задачах по программированию. p_i — вероятность того, что «C» всероса будет на тему i, при этом $\sum p_i = 1$ (у каждой задачи ровно одна тема). Вы решаете задачу на i-ю тему с вероятностью q_i. Какова вероятность того, что вы решите «C» в оба дня?

Задача 3. Вы становитесь победителем всероса с вероятностью p. Найдите вероятность того, что вы хоть раз станете победителем всероса за k участий.

Задача 4. Дан граф на n вершинах и m ребрах. Найдите средний размер разреза в этом графе, то есть если вершины поделить случайным образом на два множества, то сколько в среднем будет ребер между этими множествами?

Определение: Случайная величина — это функция $\alpha : U \to \mathbb{R}$. Её математическим ожиданием (матожиданием) $\mathbb{E}[\alpha]$ называется число $\mathbb{E}[\alpha] = \sum_{u \in U} \alpha(u) \cdot \Pr[u]$. Если бы все исходы были равновероятны, то это было бы среднее значение по всем исходам. А так это средневзвешенное значение.

Задача 5. Вы становитесь победителем всероса с вероятностью p. Найдите матожидание количества участий во всеросе, чтобы стать победителем.

Задача 6. Есть очень сложная задача с потестовой оценкой. Ответ в задаче — «YES» или «NO».

a) У вас есть всего одна попытка. Как наверняка получить достаточно много баллов?

b) А если есть несколько попыток, как **точно** получить достаточно много баллов?

Задача 7. Вспомните, почему алгоритм поиска медианы на основе быстрой сортировки работает за линейное время в среднем. С какой вероятностью он будет работать за квадратичное время?

(*) Как сделать так, чтобы этот алгоритм не использовал случайных чисел и всегда работал за линейное время?

Задача 8. Неравенство Маркова

Пусть α — случайная величина, **принимающая только неотрицательные значения**. Тогда для любой константы $c > 0$ выполняется неравенство

$$\Pr[\alpha \geq c \cdot E[\alpha]] \leq \frac{1}{c}$$

Насколько это может быть неверно для случайной величины, принимающей любые значения?

Задача 9. Дан цикл из 40 вершин. Каждую вершину этого цикла удалили из графа с вероятностью 1/2. Сколько в среднем будет компонент связности в получившемся графе?

Задача 10. Предложите алгоритм нахождения случайной перестановки из n элементов (то есть каждая перестановка должна выбираться с одинаковой вероятностью!), работающий за время $\mathcal{O}(n)$.

Задача 11. Есть два конверта. В одном из них в 10 раз больше денег, чем в другом. Вам дали случайный из двух конвертов, вы открыли его и увидели, что там 100 рублей. Вам предложили поменяться конвертами. Стоит ли соглашаться?

Задача 12. Чему равна вероятность того, что случайная хорда единичной окружности (отрезок, соединяющий две точки на окружности) имеет длину больше $\sqrt{3}$ (это длина стороны правильного вписанного треугольника)?

Задача 13. a) Как найти случайную точку в квадрате?

b) Как с помощью этого найти случайную точку в круге?

c) А если на сфере (не в шаре, а именно на поверхности)?

d) Вернемся к случайной точке в круге. Как выбрать ее случайно без необходимости, возможно, повторять процесс?

Задача 14. Дан массив из n чисел. Поступают запросы: «проверить, что на отрезке $[l, r]$ есть число, встречающееся хотя бы $\frac{r-l+1}{2}$ раз». Отвечать на запрос за $\mathcal{O}(\log n)$.

Задача 15. a) Представьте, что у вас есть монетка, которая выпадает орлом с вероятностью 50 процентов. Просимулируйте с ее помощью монету, которая выпадает орлом с вероятностью p. Сколько вам понадобится бросков в среднем?

b) Представьте, что у вас есть монетка, которая выпадает орлом с вероятностью p (число p вы не знаете). Просимулируйте с ее помощью монету, которая выпадает орлом с вероятностью 50 процентов. Сколько вам понадобится бросков в среднем?

Задача 16. Дано два списка из n чисел. Необходимо проверить, что произведения чисел в списках совпадают. Длинную арифметику использовать нельзя. Время $\mathcal{O}(n)$.

Задача 17. Дано дерево из $n \le 10^5$ вершин. В каждой вершине изначально записано число 0. Также даны $q \le 10^5$ запросов двух видов:

1. Дана вершина v и число d. Выбирается случайная вершина дерева r, и для всех таких вершин u, что путь от r до u проходит через v, к числу в вершине u прибавляется d.

2. Дана вершина v. Найти матожидание значения в вершине v.

Все ответы необходимо вычислить по простому модулю.

Задача 18. В шляпе есть шары $n \le 20$ цветов. Про каждый цвет i известно, что в корзине есть ровно a_i шаров такого цвета. Вы берете случайный шар из шляпы, смотрите на его цвет и кладете его обратно в шляпу. Сколько в среднем надо достать шаров, чтобы увидеть все цвета?

Ответ необходимо вычислить по простому модулю.

Задача 19. Даны числа p_a, p_b и k ($k \le 1000$). Вы последовательно выписываете буквы a и b в строчку. Каждый раз вы выписываете a с вероятностью $\frac{p_a}{p_a+p_b}$ и b с вероятностью $\frac{p_b}{p_a+p_b}$. Вы останавливаетесь, когда в получившейся строке есть хотя бы k подпоследовательностей (не подстрок) ab. Каково матожидание количества букв, которые вы выпишете?

Ответ необходимо посчитать по простому модулю.

Задача 20. Кузнечик находится в точке 0. Каждый раз с вероятностью p он прыгает на одну клетку направо, а с вероятностью $1-p$ он прыгает на одну клетку налево. Найдите вероятность того, что кузнечик когда-то окажется в точке -1.

Задача 21. a) Кузнечик прыгает по прямой из n точек. Известно, что из i-й точки с вероятностью a_i он прыгает в точку с номером $i-1$ и с вероятностью $1-a_i$ прыгает в точку с номером $i+1$. Изначально он стоит в точке с индексом k. Надо за $O(n^3)$ найти матожидание числа прыжков, после которого кузнечик выпрыгнет с поля.

b) А если кузнечик прыгает по графу?

Задача 22. Есть набор из n чисел. Сколько чисел можно получить всевозможными ксорами этих чисел? А какой минимальный набор изначальных чисел можно оставить, чтобы все еще можно было получить те же самые числа ксором?

Задача 23. Вы выбираете, где купить компьютер. Вы уже выбрали, какой компьютер вы хотите. Его предлагают три продавца. Цены везде одинаковые. У одного продавца 10 отзывов, и 100 процентов из них позитивные. У другого 50 отзывов, и 96% из них позитивные. У третьего 200 отзывов, и из них 93% позитивных. Кого выбрать?

Задача 24. Есть $n \le 100$ видов монет. Про каждую монету известна вероятность p_i ($0.1 \le p_i \le 0.9$) того, что она выпадает орлом, а также c_i ($1 \le c_i \le 10^4$) — количество монет такого вида. Происходит такой процесс: мы подбрасываем все монеты, которые у нас есть, и выкидываем те, которые упали решкой. После чего мы повторяем этот процесс заново. Мы заканчиваем процесс, если у нас больше не осталось монет, либо если все оставшиеся монеты одного вида. В этом случае говорится, что этот вид монет выиграл. Для каждого вида монет определите вероятность того, что он выиграет.

Задача 25. (*) Дано подвешенное дерево. На каждой итерации удаляется какая-то случайная его вершина, а также все ее поддерево. Процесс заканчивается, когда удаляется корень. Найдите, через сколько операций в среднем этот процесс завершится.

Задача 26. (*) Обезьяну посадили за компьютер печатать «Войну и мир». Известно, что она состоит ровно из n символов, а обезьяна каждый раз нажимает на случайный из 256 символов. Через сколько нажатий в среднем обезьяна все таки сможет напечатать всю книгу подряд?

Задача 27. (*) Вам дано n случайных величин. i-я из них принимает случайное равновероятное значение на отрезке $[l_i, r_i]$. Найдите за полиномиальное время матожидание минимума этих случайных величин (ответ необходимо найти по простому модулю).

Задача 28. (*)

Пусть функция `kek` возвращает случайное вещественное число от 0 до 1. Чему равно матожидание вывода этого кода? Сколько итераций цикла в среднем произойдет?

```
double ans = 0;
while (ans < 1) {
    ans += kek();
}
cout << ans;
```

Задача 29. (*)

Дан случайный граф на n вершинах (то есть выбирается какая-то константа p_n, и каждое из $\frac{n\cdot(n-1)}{2}$ потенциальных ребер добавляется в него с вероятностью p_n). Найдите в этом графе за полиномиальное время Гамильтонов цикл. Насколько маленькой можно взять константу p_n?

Задача 30. (*)

a) Дан неориентированный граф на n вершинах. Вы хотите проверить, существует ли в этом графе путь из вершины s в вершину t. Для этого вы встаете в вершину s и запускаете random-walk на глубину K. То есть вы K раз выбираете случайного соседа текущей вершины и переходите в него. Докажите, что если взять K порядка $n^3 \log^2 n$, то такой алгоритм с очень большой вероятностью когда-то дойдет до t.

b) Докажите, что если бы граф был ориентированный, то K пришлось бы взять экспоненциально большим.

Задача 31. (*)

Вы по очереди считываете n чисел a_i ($1 \le a_i \le n$). Вам необходимо определить примерное количество различных чисел среди них (можно ошибиться, к примеру, максимум в 10 раз). При этом вам разрешено использовать только $O(\log n)$ памяти.