И. ВЫСОЦКИЙ,
г. Москва

Задача 11

ЗАДАЧИ С УЛИЦЫ

Задача о циклах

Мы уже знакомы со случайными перестановками. Они представляют собой поистине неисчерпаемый источник вдохновения. Решая задачу о совпадениях (№ 1, 2023), мы выяснили, сколько в среднем совпадений (неподвижных точек) в случайной перестановке. Более того, мы нашли распределение случайной величины «*число неподвижных точек*», то есть вероятности того, что совпадений не будет, будет ровно 1, ровно 2 и т.д. При этом мы использовали сюжет с рассыпавшимися ключами, которые завуч Светлана Николаевна в спешке развешивает как попало.

Решая задачу об узком шоссе и о театралах (№ 2–3, 2023), мы научились разбивать перестановку на *вереницы*[1], каждую из которых возглавляет *лидер*. Вереницы могут следовать по убыванию лидера (первая машина в каждой веренице движется медленнее, чем все машины перед нею) или по возрастанию лидера (зрителю не придется вставать, только если он пришел позже всех тех, кто сидит дальше).

Снова используем сюжет о перепутанных ключах для введения в другую вероятностную ситуацию, в которую автор попал после переезда в новый дом пару лет назад. Дом был настолько новый, что в подъезде (115 квартир) еще не было почтовых ящиков, и все счета и рекламы кучей сваливались на стол около каморки консьержки, что не приводило в восторг ни ее, ни жильцов.

И вот настал радостный день: двое брутальных молчаливых рабочих притащили несколько блоков почтовых ящиков, быстро повесили их на отведенное место и так же быстро удалились, столкнувшись в дверях с нами: мы с Патриком возвращались с утренней прогулки. Так случилось, что я застал заключительный акт этой пьесы. Я еще не знал, что это пьеса, я думал — просто почтовые ящики, и обратился к консьержке с вопросом, а где, мол, ключи. Давясь от смеха, та сказала, что ключи в ящиках, но ящики заперты. Мало того, оказалось, что рабочие добросовестно покидали в каждый ящик ровно по одной связке с парой ключей, но... в случайном порядке. Да и кто их разберет, эти ключи?! — они все одинаковые, хотя на каждой паре висела пластиковая бирка с номером.

Итак, перед нами оказалась случайная перестановка длины 115, в которой предстояло разыскать элемент № 156. Я спросил Патрика, правильно ли я помню, что оптимального алгоритма поиска в случайной перестановке не существует и при любой стратегии средняя длительность поиска не меньше, чем половина длины перестановки. Патрик удрученно кивнул и поджал хвост. То есть об ускорении поиска речь не шла, можно было только надеяться на везение, думать о том, как не запутаться, и еще о том, как облегчить жизнь соседям, неизбежно последующим за мной. А подходящий для этих целей алгоритм существует. Нужно следовать циклам перестановки.

[1] Напомним себе, что слова «вереница» и «лидер» не являются общепринятыми терминами и использованы только в связи с сюжетом, где по шоссе едет много автомобилей. По понятной причине мы не можем использовать слово «группа».

Есть дополнительные материалы на сайте raum.math.ru.

Позаимствовав у консьержки скрепку и изогнув ее наподобие интеграла, я подцепил связку в ящике № 156 и извлек ее на белый свет. Это оказалась связка[2] № 98. Открыв ящик № 98, я достал из него ключи № 112, которыми открыл ящик № 112, и нашел там, скажем, связку № 195. И так далее — до тех пор, пока в очередном ящике я не обнаружил свои ключи, № 156. Я уже не помню, сколько именно ящиков мне пришлось открыть. Чем больше ящиков я открыл по дороге, тем большее количество соседей я избавил от поисков.

Ясно, что я «прошел» по циклу, начав с ящика № 156 и закончив этим же ящиком, при этом каждый следующий шаг мне подсказывал номер связки, лежащей в очередном открытом ящике. Цикл мог иметь длину 1, если бы случайным образом в ящике № 156 оказались ключи № 156. Цикл мог быть длинным: максимальная возможная длина цикла равна 115. Так случилось бы, если, например, в каждом ящике лежали ключи от следующего ящика, а в последнем — от первого.

Сначала докажем, что таким образом я обязательно нашел бы свои ключи. Это почти очевидно: в каждом ящике лежит связка, и если бы в первых открытых так и не оказалось моей, я продолжал бы открывать ящики до бесконечности. Но ящиков конечное число. Противоречие. Итак, алгоритм результативен.

Любая перестановка может быть представлена в виде объединения нескольких *циклов*. Рассмотрим пример. Перестановка

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 5 & 7 & 6 & 3 & 1 & 2 & 4 & 8 \end{pmatrix} \qquad (1)$$

разбивается на циклы

$$\begin{pmatrix} 1 & 5 \\ 5 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 7 & 4 & 3 & 6 \\ 7 & 4 & 3 & 6 & 2 \end{pmatrix} \text{ и } \begin{pmatrix} 8 \\ 8 \end{pmatrix}.$$

Можно взять циклы в скобки и записать одной строкой:

[15] [2 7 4 3 6] [8].

Мы используем квадратные скобки, чтобы не путать с обычными скобками, которые используем в записи самой перестановки. Циклы можно менять местами, а каждый цикл длины 2 или больше можно записать разными способами, начиная с разных чисел. Например, эту же перестановку можно записать как

[6 2 7 4 3] [8] [51].

Не думаю, что требуется давать определение цикла, поскольку это интуитивно ясный объект. На всякий случай: цикл можно понимать как часть перестановки, граф которой является циклом. Граф перестановки (1) показан на рисунке 1. Видно, что она распадается на три цикла.

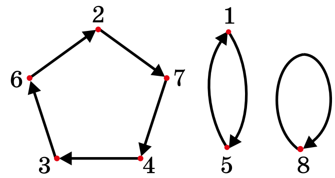


Рис. 1

Поставим и решим несколько задач. Начнем с двух наиболее естественных и несложных.

Задача 1. Какова вероятность того, что данный элемент случайной перестановки длины n принадлежит циклу длины m?

Обратите внимание: мы говорим не о случайно выбранном цикле, а о цикле, содержащем случайно выбранный элемент, а это не одно и то же! В истории с ключами — это вероятность того, что в ходе поисков придется открыть ровно m ящиков.

Решение. Поставим вопрос о распределении вероятностей случайной величины L *«длина цикла, содержащего элемент a»*.

Очевидно, событие $L = 1$ состоит в том, что в этой перестановке число a является совпадением, неподвижной точкой, то есть переходит в себя. Можно записать $a \to a$. Число a может перейти в любой из n равновозможных элементов. Поэтому вероятность того, что оно перейдет именно в себя, равна $\frac{1}{n}$.

Событие $L = 2$ имеет вид $a \to b \to a$, где b — любое число, но не a. Первый переход имеет вероятность $\frac{n-1}{n}$, и при этом условии вероятность второго равна $\frac{1}{n-1}$. Поэтому

$$\mathrm{P}(L=2)=\frac{n-1}{n}\cdot\frac{1}{n-1}=\frac{1}{n}.$$

Уже ясно, что

$$\mathrm{P}(L=3)=\frac{n-1}{n}\cdot\frac{n-2}{n-1}\cdot\frac{1}{n-2}=\frac{1}{n}$$

и т.д.: каким бы ни было число l, вероятность неизменна для $l = 1, 2, \ldots, n$:

$$\mathrm{P}(L=l)=\frac{1}{n}.$$

Задача 2. Какова средняя длина цикла, содержащего конкретный элемент?

Иными словами, каково среднее время поиска ключа?

Решение. Интуитивно ясно: чтобы найти нужный ключ в куче ключей, при средненьком везении придется перебрать половину кучи. Проверим интуицию (мы помним, что при поиске

[2] Может, и не 98. Не помню, но это неважно.

среднего чутье обычно не подводит). Учитывая результат задачи 1, получаем:

$$\mathrm{E}L = 1\cdot\frac{1}{n} + 2\cdot\frac{1}{n} + \ldots + n\cdot\frac{1}{n} =$$

$$= \frac{1}{n}\cdot(1+2+\ldots+n) = \frac{n(n+1)}{2n} = \frac{n+1}{2}.$$

Интуиция почти не подвела.

Если бы я действовал совершенно тупо, залезая скрепкой по очереди во все ящики, то никакого ожидаемого проигрыша или выигрыша не было бы: математическое ожидание числа открытых ящиков все равно было бы $\frac{n+1}{2}$. Следование циклу имеет два неоспоримых преимущества. Во-первых, скрепка нужна только один раз, во-вторых, по мере открывания ящиков уменьшается число беспорядков, так как открытые ящики снабжаются своими собственными ключами.

Стандартная форма перестановки. Связь с вереницами и лемма Фóты

Мы видели, что записать перестановку с помощью циклов можно разными способами: циклы можно менять местами, а внутри цикла можно двигать начало отсчета. Нужно навести порядок. Условимся в каждом цикле первым записывать наименьший элемент. А циклы упорядочим по убыванию первого элемента. Тогда перестановка

(5 7 6 3 1 2 4 8)

(используем однострочную запись) запишется определенным образом:

[8] [2 7 4 3 6] [1 5].

Такое представление перестановки называется *стандартным представлением* или *стандартной формой.*[3]

В каждом цикле наименьший элемент поставлен первым и выделен красным шрифтом, а циклы расположены по убыванию этого наименьшего элемента. Ничего не напоминает? Мы получили что-то, очень похожее на вереницы из задачи про узкое шоссе. Не просто похожее, а конкретно — вереницы. Только не в перестановке (5 7 6 3 1 2 4 8), а в перестановке (8 2 7 4 3 6 1 5).

Интересно: удаление квадратных скобок в стандартной форме превращает одну перестановку в другую. А удастся ли восстановить исходную? Очевидно, да. Нужно разбить эту перестановку на вереницы, взять каждую в квадратные скобки и считать, что теперь вереницы — это циклы.

Таким образом, получается преобразование на множестве перестановок длины n; опускание скобок в стандартной форме дает однострочную запись другой перестановки, при этом преобразование само по себе однозначно и однозначно обратимо, а значит, взаимно однозначно. Это утверждение называют *леммой Фóты* (Dominique Foata), а Ричард Стэнли (Richard P. Stanley) даже называет это преобразование *фундаментальной биекцией* на множестве перестановок.

Что это нам дает? Это значит, что неважно, что изучать, — циклы или вереницы в случайных перестановках. Они по сути одно и то же, а значит, у них одинаковые комбинаторные и вероятностные свойства. Теперь если нужно выяснить что-то про циклы, но удобнее рассуждать о вереницах, будем говорить о вереницах. Если какое-то утверждение про вереницы легче изучать на циклах, так и будем делать.

Более подробное изучение верениц (или циклов)

В связи с вереницами (и циклами, поскольку обнаружилось, что это одно и то же) возникает много самых разных задач. Мы поставим несколько, но решим не все, а иногда и не полностью, предложив недоделанное в качестве упражнений. Во всех задачах речь идет о случайной перестановке длины n. Я постарался расположить задачи в некоторой более или менее разумной последовательности, а именно в той, в какой мне удобнее эти задачи решать. Буду рад, если хотя бы некоторые из этих задач вызовут у вас интерес.

Задача 3. Сколько в среднем циклов в случайной перестановке длины n?

В терминах ключей это вопрос о том, скольким жильцам-*первопроходцам* придется повторять трюк со скрепкой, если все они будут следовать описанному алгоритму поиска по циклу (остальным не придется, поскольку они воспользуются плодами трудов первопроходцев).

Задача 4. Какова вероятность того, что в случайной перестановке длины n ровно k циклов?

Это равносильно вопросу о вероятности того, что трюк со скрепкой придется повторять ровно k раз, чтобы разыскать все ключи.

Задача 5. Сколько в среднем в случайной перестановке длины n циклов заданной длины m?

Это то же самое, что спросить, в скольких вереницах на шоссе ровно m автомобилей.

Задача 6. Какова вероятность того, что k-я по счету вереница в случайной перестановке длиной n имеет длину ровно m?

Эту задачу мы поставили, но не решили в статье об узком шоссе.

[3] Сергей Китаев и Мартин Айнер называют эту запись стандартной формой перестановки. Термин «стандартное представление» использует Ричард Стэнли, но он и Айнер предпочитают симметричную запись: первыми в циклах записаны наибольшие элементы, а циклы расположены по возрастанию. Для нашего примера получилось бы [5 1] [7 4 3 6 2] [8]. Миклош Бона называет эту же запись канонической цикловой записью. В общем, единая терминология не установилась. А потому как решим, так и будет.

	n\k	1	2	3	4	5	6	7	8	9	10	11	12
4	1	1,00000											
5	2	0,50000	0,50000										
6	3	0,33333	=Числа Стирлинга 1ГО/ФАКТР(Вероятности!$B6)										
7	4	0,25000	0,45833	0,25000	0,04167								
8	5	0,20000	0,41667	0,29167	0,08333	0,00833							
9	6	0,16667	0,38056	0,31250	0,11806	0,02083	0,00139						
10	7	0,14286	0,35000	0,32222	0,14583	0,03472	0,00417	0,00020					
11	8	0,12500	0,32411	0,32569	0,16788	0,04861	0,00799	0,00069	0,00002				
12	9	0,11111	0,30198	0,32552	0,18542	0,06186	0,01250	0,00150	0,00010	0,00000			
13	10	0,10000	0,28290	0,32316	0,19943	0,07422	0,01744	0,00260	0,00024	0,00001	0,00000		
14	11	0,09091	0,26627	0,31950	0,21068	0,08560	0,02260	0,00395	0,00045	0,00003	0,00000	0,00000	
15	12	0,08333	0,25166	0,31507	0,21974	0,09602	0,02785	0,00551	0,00075	0,00007	0,00000	0,00000	0,00000

Рис. 2. Вероятности $p_{n,k} = \dfrac{c(n;k)}{n!}$

Задача 7. Какова вероятность того, что j-й по счету элемент случайной перестановки длины n принадлежит k-й по счету веренице?

Задача 8. Какова средняя длина k-й по счету вереницы в случайной перестановке длины n?

Или: какова средняя длина k-го по счету цикла в стандартной форме перестановки.

Задача 9. Будем называть цикл *длинным*, если его длина больше половины длины перестановки. Какова вероятность того, что в случайной перестановке длины n есть длинный цикл?

Ну или будем искать вероятность того, что на шоссе есть длинная вереница, которая содержит больше половины всех машин.

Задача 10. Какова вероятность того, что наперед заданный элемент случайной перестановки длины n попадет в длинный цикл?

Насколько велика была вероятность того, что нам с Патриком пришлось бы открыть больше половины ящиков, прежде чем мы нашли свои ключи?

Уже известные постоянным читателям предварительные сведения

Все нужные объекты и факты, которые нам потребуются дальше, есть в статьях «Задача об узком шоссе» (№ 2, 2023) и «Задача коллекционера» (№ 9, 2022). Чтобы не заставлять читателя листать подшивку журнала туда-сюда, повторим основное.

Число Стирлинга 1-го рода $c(n;\ k)$ мы нашли как количество перестановок длины n, содержащих ровно k вереницы (рис. 2). Поэтому вероятность того, что случайная перестановка содержит ровно k верениц, равна

$$p_{n,k} = \frac{c(n;k)}{n!}. \tag{2}$$

Теперь мы знаем: в силу леммы Фоты вереницы можно отождествить с циклами некоторой другой перестановки. Поэтому определим числа Стирлинга иначе, как это обычно и делают.

Определение. Числом Стирлинга 1-го рода $c(n;\ k)$ называется количество перестановок длины n, содержащих ровно k циклов.

Напомним равенство, с помощью которого построена таблица на рисунке 3:

$$c(n;\ k) = c(n-1;\ k-1) + (n-1)c(n-1;\ k) \tag{3}$$

при $n \geq 2$ и $k \geq 1$, где для начала рекурсии [4] следует положить, если $k > n$:

$$c(1;\ 0) = c(n;\ 0),\ c(1;\ 1) = 1,\ c(n;\ k) = 0.$$

Числа первого столбца равны факториалам:

$$c(n;\ 1) = (n-1)!, \tag{4}$$

поскольку каждое такое число равно числу случайных перестановок с единственной вереницей, то есть с числом 1 в начале (рис. 3). Таких перестановок ровно $(n-1)!$.

Гармоническое число H_n — сумма величин, обратных первым n натуральным числам:

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}.$$

Последовательность гармонических чисел возрастает, но медленно — как логарифм:

$$H_n = \ln n + \gamma_n, \text{ где } \lim_{n\to\infty} \gamma_n = \gamma = 0{,}577\ldots$$

[4] Часто для общности добавляют число $c(0;\ 0) = 1$.

	n\k	1	2	3	4	5	6	7	8	9	10	11	12
4	1	1											
5	2	1	1										
6	3	2	3	1									
7	4	6	11	6	1								
8	5	24	50	35	10	1							
9	6	120	274	225	85	15	1						
10	7	720	1764	1624	735	175	21	1					
11	8	5040	13068	13132	6769	1960	322	28	1				
12	9	40320	109584	118124	67284	22449	4536	546	36	1			
13	10	362880	1026576	1172700	723680	269325	63273	9450	870	45	1		
14	11	3628800	10628640	12753576	8409500	3416930	902055	157773	18150	1320	55	1	
15	12	39916800	120543840	150917976	105258076	45995730	13339535	2637558	357423	32670	1925	66	1

Рис. 3. Числа Стирлинга 1-го рода до $c(12;\ 12)$

Число γ называется константой Эйлера–Маскерони. При больших n для большинства практических нужд можно считать, что

$$H_n \approx \ln n + 0{,}577... \qquad (5)$$

Гармоническое число 2-го порядка $H_{n,\,2}$ — сумма величин, обратных первым n натуральным квадратам:

$$H_{n,\,2} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \ldots + \frac{1}{n^2}.$$

Про гармонические числа 2-го порядка известно, что

$$H_{n,\,2} \to \frac{\pi^2}{6} \text{ при } n \to \infty,$$

причем сходимость довольно быстрая.

В статье про задачу об узком шоссе мы нашли математическое ожидание и дисперсию числа верениц X в случайной перестановке:

$$\mathrm{E}X = H_n \text{ и } \mathrm{D}X = H_n - H_{n,\,2}. \qquad (6)$$

Решение задач 3–10

3. Столько же, сколько в среднем верениц, то есть H_n (см. равенство (6)).

Пример. В перестановке порядка 115 среднее число циклов равно

$$H_{115} \approx \ln 115 + 0{,}577 \approx 5{,}32.$$

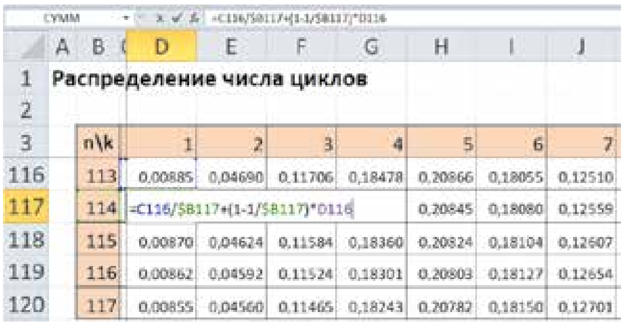
4. Такова же, какова вероятность того, что в случайной перестановке ровно k вереница (см. (2)):

$$p_{n,\,k} = \frac{c(n;\,k)}{n!}.$$

Из (3) следует рекуррентное соотношение для вероятностей:

$$p_{n,\,k} = \frac{1}{n} p_{n-1,\,k-1} + \left(1 - \frac{1}{n}\right) p_{n-1,\,k}. \qquad (7)$$

Пример. Вероятность того, что в тот день в перепутанных ключах от 115 почтовых ящиков образовалось ровно 5 циклов, примерно равна 0,20824 (рис. 4):


СУММ =C116/$B117+(1-1/$B117)*D116

Распределение числа циклов

n\k	1	2	3	4	5	6	7
113	0,00885	0,04690	0,11706	0,18478	0,20866	0,18055	0,12510
114	=C116/$B117+(1-1/$B117)*D116				0,20845	0,18080	0,12559
115	0,00870	0,04624	0,11584	0,18360	0,20824	0,18104	0,12607
116	0,00862	0,04592	0,11524	0,18301	0,20803	0,18127	0,12654
117	0,00855	0,04560	0,11465	0,18243	0,20782	0,18150	0,12701

Рис. 4. Продолжение таблицы на рисунке 2.

Вероятности посчитаны по (7)

5. Мы знаем, что в перестановке в среднем одно совпадение, то есть один цикл длины 1 («Задача о совпадениях», № 1, 2023). А сколько циклов другой длины?

Пусть число циклов длины m равно C_m. Это случайная величина, и нужно найти ее математическое ожидание. Обозначим общее число элементов во всех циклах длины m буквой Y. Тогда

$$mC_m = Y.$$

Мы уже знаем, что вероятность события «*один конкретный элемент перестановки принадлежит циклу длины m*» равна $\frac{1}{n}$ и не зависит от m (задача 1). Значит,

$$\mathrm{E}Y = \frac{1}{n} \cdot n = 1,$$

то есть в среднем в перестановке один элемент принадлежит циклу длины m. Перейдем к математическим ожиданиям в равенстве $mC_m = Y$:

$$\mathrm{E}(mC_m) = \mathrm{E}Y = 1,$$

откуда

$$\mathrm{E}C_m = \frac{1}{m}.$$

Это согласуется с результатом задачи 3: в перестановке в среднем один цикл длины 1, в среднем полцикла длины 2, треть цикла длины 3 и т.п. Всего в среднем циклов

$$1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} = H_n,$$

что нам и так уже известно.

Пример. Предположим, что мои соседи, обнаружив, что в их ящике чужой ключ, действовали тем же способом — шли по циклу. Сколько из них нашли бы свой ключ за три попытки? В среднем... $\frac{1}{3}$ соседа, что бы это ни значило.

6. Решим задачу только для случая $k = 1$, то есть найдем вероятность того, что первая вереница имеет длину m. Воспользуемся схематической диаграммой верениц в перестановке.

Рассмотрим сначала случай $m < n$, когда в перестановке есть еще хотя бы одна вереница (рис. 5):

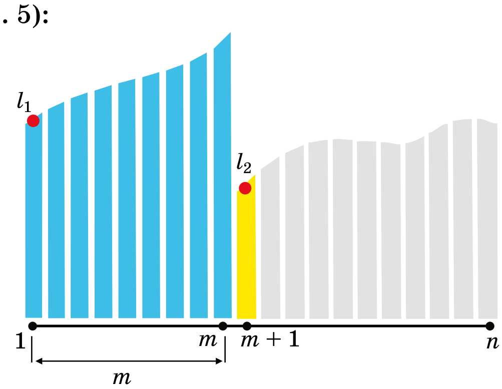


Рис. 5

Тогда лидер второй вереницы l_2 — наименьшее из первых $(m + 1)$ чисел, и это число должно стоять на позиции $(m + 1)$. Вероятность этого $\frac{1}{m+1}$. При этом условии первые m чисел должны образовывать случайную перестановку, в которой ровно одна вереница. Вероятность этого

$$\frac{c(m;1)}{m!}=\frac{(m-1)!}{m!}=\frac{1}{m}.$$

Здесь использовано равенство (4), но можно проще: это вероятность того, что самое малое из первых m чисел окажется ровно на первой позиции.

Таким образом, вероятность того, что первая вереница в этом случае имеет длину m, равна $\frac{1}{m(m+1)}$.

Теперь случай $m = n$. Вереница единственная, и потому этот случай проще. Нужно, только чтобы наименьшее число стояло первым. Вероятность этого $\frac{1}{m}$.

Таким образом, можно записать распределение случайной величины L_1 *«длина 1-й вереницы»*:

$$L_1 \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & \ldots & m & \ldots & n-1 & n \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{12} & \frac{1}{20} & \frac{1}{30} & \ldots & \frac{1}{m(m+1)} & \ldots & \frac{1}{(n-1)n} & \frac{1}{n} \end{pmatrix}. \quad (8)$$

Не правда ли, интересно, что если первая вереница не «уперлась» в конец перестановки, то вероятность того, что она имеет длину m, не зависит от длины самой перестановки. Например, если спросить, какова вероятность того, что первая вереница имеет длину 7, то ответ будет $\frac{1}{56}$ и для перестановки длины 8, и для перестановки длины 2023.

Для второй и последующих вереницы зависимость от n уже есть. Предлагаю решить задачу в общем случае самостоятельно (см. задачу 5* в конце статьи), модифицировав приведенное рассуждение. Советую воспользоваться диаграммой, как на рисунке 5.

Пример. Если на узком шоссе $n = 100$ автомобилей, то за первым тянутся еще три машины ($L_1 = 4$) с вероятностью $\frac{1}{4\cdot 5}=0{,}05$.

7. Задача кажется сложной только на первый взгляд. Достаточно понять, что вопрос сводится к тому, какова вероятность того, что первые j элементов сами по себе образуют случайную перестановку, в которой ровно k вереницы. Из равенства получаем:

$$p_{j,k}=\frac{c(j;k)}{j!}.$$

Здесь тоже отсутствует зависимость от длины перестановки.

Пример. 15-я машина из 100 идущих по шоссе окажется в третьей веренице с вероятностью

$$p_{15,3}=\frac{c(15;3)}{15!}=0{,}2998\ldots$$

(проверьте с помощью приложенной таблицы Excel).

8. Пойдем проторенным путем: пусть индикатор I_j равен 1, если наступает событие A_j *«j-й элемент принадлежит k-й веренице»*, и равен в противном случае. Разумно считать, что $j \ge k$. Математическое ожидание каждого индикатора равно (см. задачу 7)

$$\mathrm{E}I_j=\mathrm{P}\left(A_j\right)=p_{j,k}=\frac{c(j;k)}{j!}.$$

Случайная величина L_k *«длина k-й вереницы»* равна сумме всех индикаторов по всем элементам, которые теоретически могут попасть в k-ю вереницу:

$$L_k = I_k + I_{k+1} + \ldots + I_j + \ldots + I_n.$$

Перейдем к математическим ожиданиям:

$$\mathrm{E}L_k=\frac{c(k;k)}{k!}+\frac{c(k+1;k)}{(k+1)!}+\ldots+\frac{c(n;k)}{n!}. \quad (9)$$

В частности, средняя длина первой вереницы равна H_n:

$$\mathrm{E}L_1=\frac{c(1;1)}{1!}+\frac{c(2;1)}{2!}+\frac{c(3;1)}{3!}+\ldots+\frac{c(n;1)}{n!}=$$

$$=\frac{0!}{1!}+\frac{1!}{2!}+\frac{2!}{3!}+\ldots+\frac{(n-1)!}{n!}=1+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n}=H_n.$$

В качестве совсем простого упражнения получите это равенство из распределения .

Казалось бы, все. Но не совсем. Сумму (9) можно упростить:

$$\mathrm{E}L_k=\frac{c(n+1;k+1)}{n!}. \quad (10)$$

Упрощение это не имеет отношения к теории вероятностей, поэтому предоставим его читателю в качестве несложного упражнения на математическую индукцию (задача 2 в конце статьи).

Пример. Если на шоссе 100 автомобилей, то ожидаемая численность третьей вереницы равна

$$\frac{c(101;4)}{100!}=101\cdot p_{101,4}=101\cdot 0{,}19232\approx 19{,}424.$$

9. Пусть C_m, как и прежде, — число циклов длины m. Мы знаем (задача 5), что

$$\mathrm{E}C_m=\frac{1}{m}.$$

Но если цикл длинный, то есть $m>\frac{n}{2}$, то случайная величина C_m принимает только два значения: 0 (если цикла длиной m нет) или 1 (если есть). В перестановке может быть только один длинный цикл. Пусть $b=\left[\frac{n}{2}\right]+1$ — наименьшее целое[5], превосходящее $\frac{n}{2}$. Общее число длинных циклов Y равно сумме чисел C_m:

$$Y = C_b + C_{b+1} + \ldots + C_n,$$

[5] Квадратными скобками здесь обозначена целая часть числа.

при этом все слагаемые равны нулю, кроме, быть может, одного, которое равно 1, поскольку, повторим, двух длинных циклов быть не может.

Значит, число EY равно вероятности того, что $Y = 1$, то есть вероятности нужного события «*есть длинный цикл*». Перейдем к математическим ожиданиям:

$$P(Y=1)=EY=EC_b+EC_{b+1}+\ldots+EC_n=$$
$$=\frac{1}{b}+\frac{1}{b+1}+\ldots+\frac{1}{n}=H_n-H_{b-1}=H_n-H_{\left[\frac{n}{2}\right]}.$$

Это и есть ответ, только неясно, много это или мало. Чтобы выяснить это, воспользуемся приближением:

$$P(Y=1)\approx\ln n+\gamma-\ln\left[\frac{n}{2}\right]-\gamma\approx\ln n-\ln\left(\frac{n}{2}\right)=\ln 2=0{,}693\ldots$$

Таким образом, в случайной перестановке, скорее всего, есть длинный цикл. Перестановки, состоящие только из коротких циклов (или вереениц), встречаются примерно вдвое реже, чем перестановки с длинным циклом.

10. В перестановке 115 ключей от почтовых ящиков с большой вероятностью мог оказаться длинный цикл, и если бы мы с Патриком «попали в него», начав с нашего ящика № 156, то нам пришлось бы открыть не меньше 58 ящиков. Высока ли вероятность такой неприятности?

С одной стороны, интуиция говорит, что раз вероятность присутствия длинного цикла большая, да и сам цикл длинный, то вероятность попасть в него должна быть велика.

С другой стороны, средняя длина цикла равна $\frac{n+1}{2}$ (задача 2), что близко к $\frac{n}{2}$. Медиана случайной величины «*длина цикла*», скорее всего, близка к ее математическому ожиданию. Поэтому вероятности попасть в длинный цикл и в короткий должны быть примерно одинаковы. И это тоже говорит интуиция.

Какая из интуиций более права, чем другая?

Вероятность наличия цикла длины $m\ge b=\left[\frac{n}{2}\right]+1$ равна $\frac{1}{m}$ (задачи 5 и 9), а если он есть, то вероятность случайно попасть в него равна $\frac{m}{n}$, поэтому вероятность того, что наперед заданный элемент принадлежит длинному циклу, равна

$$\underbrace{\frac{1}{b}\cdot\frac{b}{n}+\frac{1}{b+1}\cdot\frac{b+1}{n}+\ldots+\frac{n}{n}\cdot\frac{1}{n}}_{n-b+1\text{ слагаемых}}=\frac{1}{n}\cdot(n-b+1)=$$
$$=\frac{1}{n}\cdot\left(n-\left[\frac{n}{2}\right]\right)=1-\frac{1}{n}\cdot\left[\frac{n}{2}\right].$$

Если n четно, то это в точности $\frac{1}{2}$, а если n нечетно, то получается $\frac{1}{2}+\frac{1}{2n}$, то есть чуть больше $\frac{1}{2}$ при больших n. В случае $n = 115$ это будет $\frac{58}{115}\approx 0{,}504$, что меньше 0,5 лишь на 0,003.

Второе рассуждение оказалось правдоподобным. Кстати, первое тоже правильное — вероятность 0,5 попасть в длинный цикл действительно большая, и это как раз потому, что он длинный.

А чему равна условная вероятность попасть в длинный цикл, если уже известно, что он есть? Эта дополнительная задачка несложная, и мы выносим ее в задачи для самостоятельного решения.

Как обычно, все использованные электронные таблицы размещены на сайте ptlab.mccme.ru на странице «Задачи с улицы».

Задачи для самостоятельного решения

1. Получите из равенства (3) равенство (7), то есть покажите, что верно равенство

$$p_{n,k}=\frac{1}{n}p_{n-1,k-1}+\left(1-\frac{1}{n}\right)p_{n-1,k}$$

для вероятностей $p_{n,\,k}$ событий «*перестановка длины n содержит ровно k циклов*» при $n\ge 2$, $1\le k\le n$ и начальных условиях $p_{1,0}=0$, $p_{1,1}=1$.

2. Получите из равенства (9) равенство (10), то есть покажите, что

$$\frac{c(k;k)}{k!}+\frac{c(k+1;k)}{(k+1)!}+\ldots+\frac{c(n;k)}{n!}=\frac{c(n+1;k+1)}{n!}.$$

3. Известно, что в обширной (n велико) случайной перестановке есть длинный цикл. Какова при этом условии вероятность того, что случайный элемент перестановки принадлежит именно этому циклу?

4. Получите среднюю длину первой вереницы $EL_1 = H_n$ из распределения (8).

5*. Решите задачу 6 полностью: докажите, что вероятность события «*k-я по счету вереница имеет длину ровно* $m > 0$» равна

$$P(L_k=m)=\sum_{j=k-1}^{n-m-1}\frac{c(j;k-1)}{j!(j+m)(j+m+1)}+\frac{c(n-m;k-1)}{n(n-m)!}.$$

Ответы к задачам из статьи «Задача о саженцах и овербукинге»

1. Не меньше 3. **2.** 104 билета. Вероятность 2 лишних пассажиров при этом 0,001. **3.** Вероятность того, что RX единиц будет недостаточно, приближается к 0,5 с ростом X, а вот вероятность того, что хватит $MX = R^2X$ единиц, достигает 0,999 уже при $X = 24$. При $X \ge 100$ и «утруске» 30% с вероятностью 0,999 хватит $1{,}65X$ единиц. Коэффициент 2,04 — расточительство! Теория вероятностей значительно экономит средства.