1 Random permutations

What are the properties of randomly generated permutations of N objects?

Such permutations are a special case of random graphs, which have applications in many fields including coding theory and puzzles like this one:

> Example 1. A team of $N = 100$ contestants must choose a strategy for the following game.
>
> The members of the team are numbered from 1 to N.
>
> When the game begins, no further communications between team members are permitted. The game host creates a board with N little doors, numbered, on their fronts, from 1 to N. Behind the doors, he also writes the numbers from 1 to N, in a random permutation. (So behind door number 1, we might find the hidden number is 42; all permutations are equally probable.) Each hidden number is written either in red ink or in blue ink, chosen independently and randomly.
>
> Each contestant is allowed to open up to $N/2$ doors, looking at the number hidden behind each. (He may choose his sequence of doors in any manner, for example, in a way that depends on what he sees behind the doors he has opened, or on his own supply of random numbers.) Then all the contestants must guess the colour of *their* hidden number. For example, contestant number 42 must guess the colour of the hidden number 42, wherever it is located.
>
> The team wins a big prize if *all* the contestants guess their hidden numbers' colours correctly.
>
> Find a strategy that gives the team a substantial probability of winning.

The permutation defines a graph of N vertices and N directed edges. Every vertex has one directed edge leaving it, and one directed edge arriving at it. This graph is called a *cycle graph*, because it consists of one or more directed cycles.

1.1 One-cycles and N-cycles

Call the permutation π. If the permutation maps a particular vertex v_1 to itself, *i.e.*, $\pi(v_1) = v_1$, then we'll describe that little part of the graph as a one-cycle.

Ex.1 Show that the probability that a particular vertex n is in a one-cycle is

$$\frac{1}{N}.$$

If the graph contains only one long cycle, linking all N nodes, we'll say that the graph is an N-cycle.

Ex.2 Show that the probability that the graph contains one big N-cycle is

$$P(\text{number of cycles}=1) = \frac{1}{N}.$$

Solution: Start from a vertex v_1, and follow the permutation. The probability that there is one long N-cycle is the probability that: $v_2 \equiv \pi(v_1)$ is not v_1; and $v_3 \equiv \pi(v_2)$ is not v_1; and $v_4 \equiv \pi(v_3)$ is not v_1; and $v_5 \equiv \pi(v_4)$ is not v_1; $\cdots$ and $v_{N-1} \equiv \pi(v_{N-2})$ is not v_1. This probability is

$$P(C=1) = \frac{N-1}{N}\,\frac{N-2}{N-1}\,\frac{N-3}{N-2}\,\frac{N-4}{N-3}\cdots\frac{1}{2} = \frac{1}{N}.$$

1.2 M-cycles

We've found that the probability that a given vertex is in a 1-cycle is $1/N$, and the probability that it is in an N-cycle is $1/N$. What about other lengths of cycles?

Ex.3 What is the probability that a given vertex v_1 is in an M-cycle?

Solution: We require that: $v_2 \equiv \pi(v_1)$ is not v_1; and $v_3 \equiv \pi(v_2)$ is not v_1; ... and $v_{M-1} \equiv \pi(v_{M-2})$ is not v_1; and $v_M \equiv \pi(v_{M-1})$ *is* v_1. So

$$P(v_1 \text{ is in an } M\text{-cycle}) = \frac{N-1}{N}\frac{N-2}{N-1}\frac{N-3}{N-2}\cdots\frac{N-M}{N-M+1}\frac{1}{N-M} = \frac{1}{N}.$$

That's a nice simple result: the probability that v_1 is in an M-cycle is $1/N$, independent of M.

Ex.4 What is the expected length of the cycle in which vertex v_1 finds itself?

Solution: The vertex is in a cycle of length drawn from $\{1, 2, 3, \ldots N\}$ with equal probability; the average of these numbers is $(N+1)/2$. There is a probability of $1/2$ that v_1's cycle has length between 1 and $N/2$.

1.3 Probability of a large cycle

For this section, let M be a cycle length greater than $N/2$. We'll call such a cycle a large cycle.

We know that the probability that there is a cycle of the maximum length $M = N$ is $1/N$; and that the probability that a particular vertex v_1 is in an M-cycle is $1/N$. What is the probability that there exists *an* M-cycle? (Remember that the M-cycle doesn't necessarily contain vertex v_1, if M is smaller than N.)

Call the set of vertices in the M-cycle S. For any given permutation, there can be at most one such set, since we decided a moment ago that M was constrained to be bigger than $N/2$. Now, by the sum rule, the probability that there exists one M-cycle can be decomposed by summing over all possible subsets S of size M:

$$\begin{aligned} P(\text{an } M\text{-cycle exists}) &= \sum_S P(S \text{ contains an } M\text{-cycle}) && (1)\\ &= \sum_S \frac{M-1}{N}\frac{M-2}{N-1}\frac{M-3}{N-2}\cdots\frac{1}{N-M+2}\frac{1}{N-M+1} && (2)\end{aligned}$$

The number of terms in the sum, the number of subsets, is $\binom{N}{M}$. So

$$\begin{aligned} P(\text{an } M\text{-cycle exists}) &= \frac{N}{M}\frac{N-1}{M-1}\frac{N-2}{M-2}\cdots\frac{N-M+1}{1}\times \\ &\quad \frac{M-1}{N}\frac{M-2}{N-1}\frac{M-3}{N-2}\cdots\frac{1}{N-M+2}\frac{1}{N-M+1} \\ &= \frac{1}{M}. && (3)\end{aligned}$$

Ex.5 What is the probability that there is a cycle with length larger than $N/2$?

Solution:

$$P(\text{large cycle}) = \sum_{M=N/2+1}^{N}\frac{1}{M} \simeq \int_{N/2}^{N}\frac{1}{x}\,\mathrm{d}x = \ln 2 \simeq 0.69. \quad (4)$$

So the probability that the graph contains *no* large cycles (where 'large' means 'having length bigger than $N/2$' is 0.31.

How does this relate to the puzzle? Hopefully you can complete that step now.

In case you don't want to the solution to the puzzle yet, let's discuss random permutations a little more.

2 Number of cycles, and their expected lengths

We've found the expected length of the cycle containing one particular vertex, and the probability that there are no large cycles. What do we expect the *number* of M-cycles to be? And what do we expect the *total number* of cycles to be? What's the probability distribution of that total number?

We already know the probability that the number of cycles is 1.

2.1 Probability that the number of cycles is 1

It's $1/N$.

2.2 Probability that the number of cycles is N

Ex.6 Show that the probability that all N vertices are in one-cycles is

$$P(\text{number of cycles} = N) = \frac{1}{N!}.$$

[Hint: it's the probability that the permutation is the identity mapping.]

2.3 Probability that the number of cycles is C

We know the answers for $C = 1$ and $C = N$.

Ex.7 What happens in between?

I can't see an easy way to answer this question.

However, we can establish what the mean number of cycles is.

2.4 Number of M-cycles

Ex.8 What is the expected number of M-cycles?

Solution: The probability that v_1 is on an M-cycle is $1/N$. If we ask all N vertices to say 'eek' if they are on an M-cycle, then the expected number of 'eeks' we will hear is $N \times 1/N = 1$. The eeks will come along in clumps of size M, each clump corresponding to a single M-cycle. So the expected number of M-cycles is

$$\mathcal{E}(M) = 1/M.$$

(We saw this expression before, when we asked for the *probability* that there is an M-cycle, for M greater than $N/2$.)

2.5 Number of cycles

Ex.9 What is the expected number of cycles?

Solution: The expected number is exactly

$$\sum_{M=1}^{N} \mathcal{E}(M) = \sum_{M=1}^{N} 1/M \simeq \ln N.$$

2.6 Average cycle length

We already computed the average length of the cycle containing a given vertex, and found it was $(N+1)/2$. But we can define another average. (See the busstop paradox and the rolling sixes examples in my textbook for further examples of these two sorts of average.)

Ex.10 What is the average length of a cycle, selecting uniformly randomly from all cycles?

Solution: The total length of N edges is shared between C cycles. So, averaging uniformly over all cycles in the ensemble, the average length is

$$N/\langle C \rangle = \frac{N}{\ln N}.$$

3 Solution to puzzle

Here are some strategies the team might adopt.

Strategy 1 Everyone opens 50 random doors. If you don't find your own hidden number, you guess its colour at random.

The probability that a given person finds their number is 1/2; the probability that they say the right colour is thus 3/4. The doors opened are independent, so the probability of team success is

$$(3/4)^{100} \simeq 3 \times 10^{-13}.$$

Strategy 2 Everyone opens doors 1–50. If you don't find your own hidden number, you guess its colour at random. The probability that a given person finds their number is 1/2; the probability that they say the right colour is 3/4. However, is it at all likely that the team will all guess correctly? Exactly half of them are guaranteed to learn their own number's colour – the 50 of them whose numbers lie behind doors 1–50. The other 50 will certainly guess. The probability of team success is

$$(1/2)^{50} \simeq 9 \times 10^{-16}.$$

That's worse than the random strategy, strategy 1! It would be nice if there were a chance that all the people would learn their own number's colour.

Strategy 3 The nth person opens doors $(n+1)$–$(n+50)$, modulo N. If you don't find your own hidden number, you guess its colour at random. What's the probability of team success? I haven't worked it out. I expect it's similar to strategy 1.

Strategy 4 How about using the permutation to determine which doors are opened? The nth person opens door n, revealing $v_2 = \pi(n)$; he then opens the door number v_2, revealing $v_3 = \pi(v_3)$; and so on. Now, if vertex n is on a 'short' cycle (of length M equal to $N/2$ or less) then he will, at the Mth step, open a door that reveals his number n. Recall from section 1.2 that there is a chance of 1/2 that he is on a cycle of length between 1 and $N/2$. So there is a 1/2 chance that he will find his own number. If his vertex is on a 'long' cycle, then he will, sadly, never reach his own number; and neither will any of the other people on the same long cycle.

This is the secret of success of strategy 4: the shared source of randomness, π, causes failures of the participants to be positively correlated. And if they are more likely to simultaneously fail; they must also be more likely to simultaneously succeed!

The team succeeds if the permutation contains no 'long' cycle. This occurs with probability 0.31.

This is an addition to *Information Theory, Inference, and Learning Algorithms* (Cambridge Univ. Press, 2003), which is available online from

`http://www.aims.ac.za/~mackay/`