Some results on random permutations

Alex Fu

2023-01-10

Let σ be a permutation chosen uniformly at random from S_n, the symmetric group under composition of permutations on n symbols $1, \dots, n$. We aim to investigate some probabilistic properties of σ. As σ is chosen *uniformly* from S_n, our investigation will be quite combinatorial: the probability that $\sigma \in S_n$ has property P is precisely the number of $\sigma \in S_n$ with property P divided by $|S_n| = n!$. We will count more "naïvely" by not using the method of generating functions.

Every permutation σ has a unique decomposition into a product of disjoint *cycles*, which permute some $k \le n$ symbols cyclically and fix the remaining $n - k$ symbols. The *order* or *period* of a permutation is the least common multiple of the lengths of its cycles (sometimes also known as orbits). The *length* of a permutation is the maximum of the lengths of its cycles.

Proposition 1 (Number of involutions).

The number of permutations of order 2 (*involutions*) in S_n is the nth telephone number

$$\sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} (2k-1)!! = n! \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{1}{(n-2k)!\, 2^k\, k!}.$$

Note that $m! = m!! \cdot (m-1)!!$, and the double factorial $(2k)!! = 2^k k!$. The combinatorial explanation: $\sigma \in S_n$ is an involution iff its orbits have order 2, so we can count the number of products of k 2-cycles for each $0 \le k \le \lfloor \frac{n}{2} \rfloor$. The term $\binom{n}{2k}$ gives the number of ways to choose $2k$ distinct symbols out of n (which also determines the other $n - 2k$ fixed points). Then,

$$(2k-1)!! = \frac{(2k)!}{2^k\, k!}$$

counts the number of permutations of the $2k$ symbols *up to* the 2^k ways to swap elements in each pair (which does not affect σ) and the $k!$ ways to rearrange k pairs (as disjoint cycles commute).

Proposition 1 generalizes quite readily to counting the number of permutations of prime order $p \ge 2$, as $\sigma \in S_n$ has order p iff its orbits have order p. In general, however, a closed form is trickier.

Proposition 2 (Number of permutations of order m).

Let π_m be the number of permutations of order m, and let ρ_m be that of order *dividing* m. Then

$$\pi_m = \sum_{d|m} \mu\left(\frac{m}{d}\right) \cdot \rho_d,$$

where $\mu(\cdot)$ is the Möbius function given by

$$\mu(r) = \begin{cases} +1 & \text{if } r \text{ is squarefree with an even number of prime factors} \\ -1 & \text{if } r \text{ is squarefree with an odd number of prime factors} \\ 0 & \text{if } r \text{ is divisible by a square.} \end{cases}$$

The key here is the *Möbius inversion formula*, the implication

$$g(m) = \sum_{d|m} f(m) \implies f(m) = \sum_{d|m} \mu\left(\frac{m}{d}\right) \cdot g(d).$$

Here, it is indeed the case that $\rho_m = \sum_{d|m} \pi_d$, the sum taken over all positive divisors d of m.

Proposition 3 (Expected number of fixed points).

The expected number of fixed points of a randomly chosen permutation $\sigma \in S_n$ is 1.

A common trick for counting fixed points: instead of summing the number of fixed points over all possible permutations $\sigma \in S_n$, we sum the number of permutations that fix each point $i = 1, \ldots, n$, which gives $n \cdot (n-1)! = n!$ total fixed points. Then $\frac{n!}{n!} = 1$ is the average number of fixed points, which is perhaps surprising because it does not depend on n. As a consequence, or generalization, of Proposition 3,

Proposition 4 (Expected number of k-cycles).

The expected number of cycles of length k (or k-*cycles*) in the decomposition of a randomly chosen $\sigma \in S_n$ is $\frac{1}{k}$. Note that a fixed point is simply a 1-cycle.

The probability that a given element i belongs to a k-cycle in σ is

$$\frac{1}{n!}\binom{n-1}{k-1}\frac{k!}{k}(n-k)! = \frac{1}{n},$$

which we find by counting the number of ways to choose and arrange $k-1$ elements distinct from i, up to the k possible "starting points" of the k-cycle, and freely permute the $n-k$ remaining elements. Or, we observe that by uniformity, i is fixed with probability $\frac{1}{n}$; otherwise, $i \to \sigma(i) \to i$ is in a 2-cycle with probability $\frac{1}{n-1} \cdot \frac{n-1}{n} = \frac{1}{n}$; otherwise, $\sigma(i) \to \sigma^2(i) \to i$ has probability $\frac{1}{n-2} \cdot \frac{n-2}{n-1} \cdot \frac{n-1}{n} = \frac{1}{n}$, etc.

By the linearity of expectation, the expected number of elements in a k-cycle is $n \cdot \frac{1}{n} = 1$, and this is k times the expected number of k-cycles, which then must be $\frac{1}{k}$.

As a result, by applying the linearity of expectation to the total number of cycles, we find that on average, the number of cycles in $\sigma \in S_n$ is approximately logarithmic in n:

Proposition 5 (Expected number of cycles in a random permutation)**.**

The expected number of cycles in the decomposition of a random permutation σ is the nth harmonic number $H_n = \sum_{i=1}^{n} \frac{1}{i}$. Note that 1-cycles are counted as well, and $\ln n \leq H_n \leq \ln n + 1$.

And, the total length of all cycles of all permutations is $n! \cdot n$. Dividing by $n! \cdot H_n$, the total number of cycles across all permutations, we have the following result:

Proposition 6 (Expected length of cycle in a random permutation)**.**

The expected length of a cycle in a random permutation $\sigma \in S_n$ is $\frac{n}{H_n}$.

We remark that Proposition 6 gives a surprising solution to a *prisoner's problem*. $n = 100$ prisoners are tasked with each finding their own names, which have been placed uniformly at random in 100 envelopes. However, each prisoner is only allowed to look inside 50 envelopes, and they cannot communicate in any way with one another. This seems like an impossible task: the naïve strategy has a $(\frac{1}{2})^{100}$ probability of success. But, by labelling the names $1, \ldots, 100$, we can describe the problem using a random permutation $\sigma \in S_{100}$. The strategy: prisoner i looks inside envelope i, then $\sigma(i)$, $\sigma^2(i)$, and so on until they find i. In this case, the prisoners fail if there is a cycle in σ of length $\geq \frac{n}{2}$, which has probability

$$\sum_{k=\lceil n/2 \rceil}^{n} \frac{1}{k} = H_n - H_{\lceil n/2 \rceil} < \ln 2$$

per Proposition 4, since σ can contain at most one cycle of length $k \geq \frac{n}{2}$, so the probability is precisely the expected value. This strategy guarantees a probability of success of at least 30%!

Proposition 6 finds the expected length of a cycle chosen uniformly at random among all cycles. Instead, if the probability of a cycle being chosen were weighted by its length, we find the following related result.

Proposition 7 (Expected length of cycle containing a given point in a random permutation)**.**

Let i be a point in $1, \ldots, n$, without loss of generality $i = 1$. Then the expected length of the cycle containing i in a random permutation $\sigma \in S_n$ is $\frac{n+1}{2}$.

This follows from the probability of i being in a k-cycle in σ equalling $\frac{1}{n}$ for all $k = 1, \ldots, n$, and

$$\sum_{k=1}^{n} \frac{1}{n} \cdot k = \frac{1}{n} \cdot \frac{n(n+1)}{2} = \frac{n+1}{2}.$$

There is much more to be studied about the statistics of random permutations, including cycle type or *cycle shape*, which uniquely determines conjugacy classes in S_n. However, in the interest of time, we will leave you with two interesting but miscellaneous results.

Proposition 8 (Number of derangements).

The number of permutations with no fixed points (*derangements*) is

$$n! \sum_{k=0}^{n} \frac{(-1)^k}{k!} \sim \frac{n!}{e}.$$

A simple proof is given by the principle of inclusion-exclusion. The number of permutations with at least one fixed point in $1, \ldots, n$ is

$$\sum_{i=1}^{n} (n-1)! - \sum_{i<j} (n-2)! + \cdots + (-1)^{n-1} = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} (n-k)! = n! \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!}.$$

Subtracting this count from $n!$ gives the total number of derangements.

Proposition 9 (Number of permutations of given cycle shape).

The number of permutations in S_n with cycle shape $1^{m_1} 2^{m_2} 3^{m_3} \cdots$ is

$$\prod_{k=1}^{n} k^{m_k} m_k!.$$

Here, m_k gives the multiplicity or number of k-cycles. The number of possible cycle shapes in S_n is the number of partitions $p(n)$ of n, i.e. the number of ways to write n as a sum of positive integers.

$m_k!$ counts the number of ways to permute the m_k many k-cycles, and k^{m_k} counts the number of ways to internally rearrange each k-cycle on top of that. We do not need to consider additional rearrangements as disjoint cycles commute.

Results on random permutations are not merely useful for the study of symmetric groups, but also in the analysis of (possibly randomized) sorting algorithms, which we leave to another day.

■