Probability of Individual Items in a Randomly Shuffled List

Echeveria A. Poliszuk

September 2024

1 Premise

Let $c_1, c_2, c_3, ..., c_n$ be rational numbers between 0 and 1. The following process is applied to them:

1. Randomly shuffle the numbers.
2. Select the first number.
3. Generate a random rational number between 0 and 1.
4. If the random number is less than the selected number, the selected number is the output. If not, select the next number and go to step 3.
5. If no more numbers can be selected from the list, no item is selected.

Given the above process, for all possible random shuffles, what is the percentage (expressed as a number between 0 and 1) that you will get each number from the list?

2 Process

I rely on the wealth of symmetry this problem has. Let's focus on a single value, m. Let all other values be represented as $x_1, x_2, x_3, ..., x_{n-1}$. for all j, let $d_j = 1 - x_j$.

Lets start by analyzing how the probability of getting m is affected by its index in the random shuffling.

If m is first, the chance of getting m is just m itself. This happens $(n-1)!$ times, given all possible random shuffles.

If m is second, there is a number before it in the list. That means that the chance of getting m is $m(1 - x_1)$ or $m \cdot d_1$. Any x_j can be in front of m in the list, meaning across all random permutations every other value has an effect on the probability of m. The amount of times a specific value appears in front of m across all random shuffles is $(n-2)!$. So the probability of getting m across all shuffles where it is in the second index of the list is $(n-2)! \cdot m \cdot (d_1 + d_2 + d_3 + ... + d_{n-1})$.

For further indexes, I'm going to need to use elementary symmetric polynomials. Let $e_k(S)$ be the elementary symmetric polynomial of degree k of S. We also let $D = \{d_1, d_2, d_3, ..., d_{n-1}\}$. We will let $e_0(...) = 1$. We can now rewrite the previous calculated probability as $(n-2)! \cdot m \cdot e_1(D)$

For m in index 3, there are two unique values in front of m at any time. The probability for m is therefore $d_1 \cdot d_2 \cdot m$. The amount of times a specific set of 2 values (accounting for multiplicative commutativity) appears in front of m is $(n-3)!$. Accounting for multiplicative commutativity, there are 2! unique orderings of the same values in front of m. We can find the probability of getting m when it is in the 3rd index across all possible random orderings is $2! \cdot (n-3)! \cdot m \cdot e_2(D)$.

This finding can be generalized. for an index i, the probability of getting m in that index across all possible random orderings is $(i-1)! \cdot (n-i)! \cdot m \cdot e_{i-1}(D)$. We can sum that up across all indexes:

$$\sum_{j=1}^{n} (j-1)! \cdot (n-j)! \cdot m \cdot e_{j-1}(D)$$

We can then factor m out, getting:

$$m \cdot \sum_{j=1}^{n} (j-1)! \cdot (n-j)! \cdot e_{j-1}(D)$$

Now, we average it across all possible random orderings:

$$\frac{m}{n!} \cdot \sum_{j=1}^{n} (j-1)! \cdot (n-j)! \cdot e_{j-1}(D)$$

This should solve for the probability of any value in the list, given an input of m being that specific value, and D being the set of all values that aren't m run through the function $1-x$.

3 Conclusion

In conclusion, the probability of getting any c_k can be obtained by constructing $D = \{i \neq k | 1 - c_i\}$ and evaluating:

$$\frac{c_k}{n!} \sum_{j=1}^{n} (j-1)!(n-j)! \cdot e_{j-1}(D)$$

What is used in the code is a modified version, closer to:

$$c_k \sum_{j=0}^{n-1} \frac{1}{j+1} \binom{n}{j+1}^{-1} e_j(D)$$