Random Sequences and Permutations

Luca Ballan

Oct/2018

Problem: How many distinct elements are in a random sequence with k elements? What is the average number of distinct permutations of a random sequence with k elements? What is the probability distribution of the multiplicity of each distinct element in a random sequence?

Multiplicity in a Random Sequence

An ordered sequence of k elements is built by randomly selecting k elements from a set of n values with possible repetitions. The probability that this sequence contains k_1 repetitions of an element, k_2 repetitions of a different element, ..., and k_s repetitions of another different element s, such that $\sum k_i = k$, is

$$P\left(\overline{K} = (k_1, \ldots, k_s)\right) = \mathbf{C}\left(n, s\right)\left(\frac{1}{n}\right)^k \frac{k!}{k_1! \ldots k_s!}$$

Proof: We first need to select s distinct elements from the pool of n values, $\mathbf{C}\left(n, s\right)$. Only these elements will be used to build the sequence. The probability of building a sequence with $(k_1, \ldots, k_s)$ distinct elements is

$$\mathbf{C}\left(n, s\right)\left(\frac{1}{n}\right)^k$$

The probability for each sequence is $(1/n)^k$ because there is only one possible sequence that given s distinct elements can be form with a $(k_1, \ldots, k_s)$ repetition scheme. By including al the possible permutations of the $(k_1, \ldots, k_s)$ distinct elements, the final probability increases to

$$\mathbf{C}\left(n, s\right)\left(\frac{1}{n}\right)^k \frac{k!}{k_1! \ldots k_s!}$$

Distinct Elements and Average Number of Permutations

An ordered sequence of k elements is built by randomly selecting elements from a set of n values with possible repetitions. The number $\overline{S}$ of distinct elements in the sequence has probability distribution

$$P\left(\overline{S} = s\right) = \sum_{\substack{k_1, \ldots, k_s \geqslant 1 \\ \sum k_i = k}} P\left(\overline{K} = (k_1, \ldots, k_s)\right)$$

Let $\overline{\Phi}\left(k, n\right)$ denotes the number of possible permutations of such a sequence. The average number of $\overline{\Phi}\left(k, n\right)$ is

$$E\left[\overline{\Phi}\left(k, n\right)\right] = \sum_{s=1}^{k} E\left[\overline{\Phi}\left(k, n\right) \mid \overline{S} = s\right] P\left(\overline{S} = s\right)$$

where $\overline{\Phi}(k,n) \mid \overline{S}=s$ is the number of possible permutations of a sequence with s distinct elements. Note that, a sequence can have only 1 to k distinct elements.
If the sequence has s distinct elements, the average number of possible permutations of this sequence is

$$E\left[\overline{\Phi}(k,n) \mid \overline{S}=s\right] = \sum_{\substack{k_1,\ldots,k_s \geqslant 1 \\ \sum k_i = k}} \frac{k!}{k_1! \ldots k_s!} P\left(\overline{K}=(k_1,\ldots,k_s) \mid \overline{S}=s\right)$$

Proof: The number of permutations of a sequence with a $(k_1,\ldots,k_s)$ repetition scheme is

$$\frac{k!}{k_1! \ldots k_s!}$$

while $P\left(\overline{K}=(k_1,\ldots,k_s) \mid \overline{S}=s\right)$ is the probability of such a sequence. Note that

$$P\left(\overline{K}=(k_1,\ldots,k_s) \mid \overline{S}=s\right) = \frac{P\left(\overline{K}=(k_1,\ldots,k_s) \wedge \overline{S}=s\right)}{P\left(\overline{S}=s\right)}$$

and $P\left(\overline{K}=(k_1,\ldots,k_s) \wedge \overline{S}=s\right) = P\left(\overline{K}=(k_1,\ldots,k_s)\right)$ because $(k_1,\ldots,k_s)$ contains all the information about the number of distinct elements $\overline{S}$.
As an example, when $k=4$, the probability of having $\overline{S}$ distinct elements in the array is

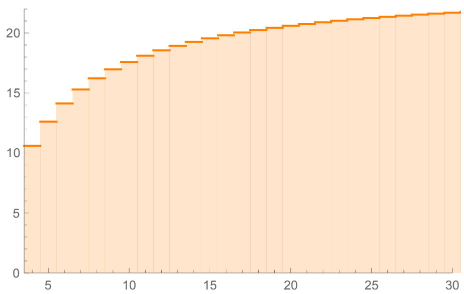
$$\begin{aligned} P\left(\overline{S}=1\right) &= \frac{1}{n^3} \\ P\left(\overline{S}=2\right) &= \frac{7(n-1)}{n^3} \\ P\left(\overline{S}=3\right) &= \frac{6(n-1)(n-2)}{n^3} \\ P\left(\overline{S}=4\right) &= \frac{(n-1)(n-2)(n-3)}{n^3} \end{aligned}$$

which sums to 1. The average number of possible permutations of a sequence with s distinct element is

$$\begin{aligned} E\left[\overline{\Phi}(1,k,n)\right] &= 1 \\ E\left[\overline{\Phi}(2,k,n)\right] &= 4.85 \\ E\left[\overline{\Phi}(3,k,n)\right] &= 12 \\ E\left[\overline{\Phi}(4,k,n)\right] &= 24 \end{aligned}$$

Note that the case $\overline{S}=2$ is fractional because the possible multiplicities of two distinct elements in a sequence of 4 are $(1,3)$, $(3,1)$, and $(2,2)$. The permutations for these cases are $4!/(3!1!)$, $4!/(1!3!)$, and $4!/(2!2!)$. The expected value of the number of permutations is the weighted average of these single permutations.
The average number of possible permutations of a generic sequence with $k=4$ elements is therefore

$$E\left[\overline{\Phi}(k,n)\right] = \frac{-33+82n-72n^2+24n^3}{n^3}$$



$E\left[\overline{\Phi}(k=4,n)\right]$ with varying n

From the figure, it is visible that, when the number of total elements n approaches the number of selected elements k, the average number of total permutations is far from 24, i.e., the permutation of k distinct elements. This is due to the fact that, even when $n = k$, the probability of choosing 4 distinct values is very low, precisely in this case 9%.

$$\begin{aligned}
P\left(\overline{S} = 1\right) &= 1\% &\longrightarrow\quad E\left[\overline{\Phi}\left(1, k, n\right)\right] &= 1 \\
P\left(\overline{S} = 2\right) &= 32\% &\longrightarrow\quad E\left[\overline{\Phi}\left(2, k, n\right)\right] &= 4.85 \\
P\left(\overline{S} = 3\right) &= 56\% &\longrightarrow\quad E\left[\overline{\Phi}\left(3, k, n\right)\right] &= 12 \\
P\left(\overline{S} = 4\right) &= 9\% &\longrightarrow\quad E\left[\overline{\Phi}\left(4, k, n\right)\right] &= 24
\end{aligned}$$

From the figure, it is also visible that $E\left[\overline{\Phi}\left(k, n\right)\right]$ increases with n, and tend to 24, i.e., to $E\left[\overline{\Phi}\left(k, k, n\right)\right]$, when $n \longrightarrow \infty$. This is due to the fact that the probability of selecting k distinct elements is equal to 1 when the pool is very large $n \longrightarrow \infty$. Therefore the number of total permutations is equal to the one of k distinct elements. In other words, given a large pool of elements (n), an ordered sequence of k elements built by randomly selecting elements from this pool has on average $k!$ possible permutations.