Shakespear's constraint

kl

8 February 2018

Suppose you were uniformly distributing 28 'templates' at random among 2140 poetic lines. Assuming the choice of a template for any given line is independent of what has gone before. What is the probability that the observed frequencies would reveal at least one example of every one of the 28 templates?

Combinatorics

It is certain that each tempalte is unique. If each poetic line is also unique, the total enumeration of uniformly distributing $k = 28$ templates among $n = 2140$ lines is k^n. Given the restriction that each template appears in at least one line, the total enumeration is $S(n, k) \times n!$, where

$$S(n, k) = \frac{1}{k!} \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} j^n$$

is the Stirling number of the second kind. Hence, the probability that each of the 28 templates appears at least once in the observed frequencies is

$$p = \frac{S(n, k) \times n!}{k^n}. \tag{1}$$

Both top and bottom of the above fraction are really large integers, so we calculate log of the probability first. This becomes

$$\begin{aligned} \log p &= \log S(n, k) + \log n! - n \log k & (2) \\ &= 7063.028 + 67.88974 - 7130.918 & (3) \\ &= -0.00026 & (4) \end{aligned}$$

The final probability $p = \exp(-0.00026) = 0.99974$.

This can also be justified experimentally by the following. The following code randomly samples an integer from the range $[1, 28]$. And it samples in totoal 2140 times. The it marks how many times there is at least one occurence from each number. The experiment is repeated 10000 times and the result shows there is never a chance that you see 0 occurence of any one of the 28 integers.

```
cnt = 0
for (i in 1:10000) {
  k = 28
  n = 2140
  x = sample(k, n, replace = TRUE)
  l = length(unique(x))
  if (l == k) cnt = cnt + 1
}
cnt
```

```
## [1] 10000
```