Why would "leave-one-out" work?

20 word by author1

abc abc ab c d d
abc cd d d
abd ab ab ab ab
cd d e cd e

20 word by author2

abc abc ab c d d
abe cb e f
acf fb ef aff abef
cdc db ge f s

Suppose we keep sampling and get 10 more words. Which author is likely to "write" more new words?

Now, suppose we leave "e" out…

$$p_{ml}("e"|author1) = \frac{1}{19}$$

$$p_{ml}("e"|author2) = \frac{0}{19}$$

μ doesn't have to be big

$$p_{smooth}("e"|author1) = \frac{20}{20+\mu}\frac{1}{19} + \frac{\mu}{20+\mu}p("e"|REF)$$

$$p_{smooth}("e"|author2) = \frac{20}{20+\mu}\frac{0}{19} + \frac{\mu}{20+\mu}p("e"|REF)$$

μ must be big! more smoothing

The amount of smoothing is closely related to the underlying vocabulary size