

Estimation based on KMV

Assume that $k=3$, the pair number is 2.

Token 1	Token 1	Token 2	Token 2	Token 3	Token 3	Token 4	Token 4
0.1	0.1	0.2	0.1	0.2	0.1	0.3	0.2
0.5	0.3	0.5	0.6	0.3	0.3	0.5	0.3
0.9	0.4	0.9	0.9	0.4	0.4	0.6	0.5

Token 1+2 Token 1+2+3 Token 1+2+3+4

0.1,1	0.1,1	0.1,1
0.2,1	0.2,2	0.2,2
0.5,2	0.3,1	0.3,2

2-occurrence,

$$2/3 * 2/0.3 = 4.45$$

Token 1+2 Token 1+2+3 Token 1+2+3+4

0.1,2	0.1,3	0.1,3
0.3,1	0.3,2	0.2,1
0.4,1	0.4,2	0.3,3

The estimation for the number of candidates are
 $4.45 * 4.45 = 19.8$