

TEMPORAL WORD FREQUENCY VARIATIONS

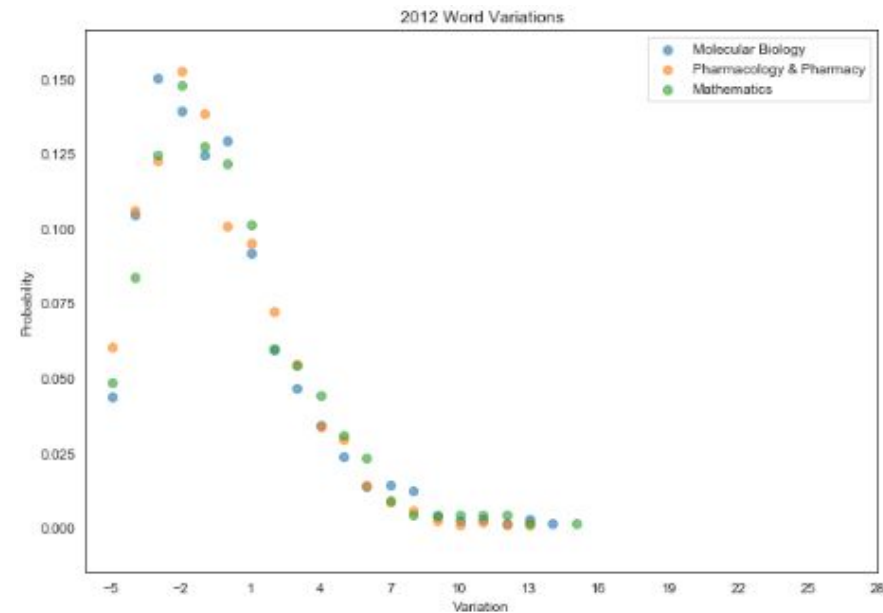
OVERVIEW

- *Goal: Study the temporal evolution of word usage in article titles*
- *We want to identify the 'big movers': those words that are increasing or decreasing much more than we would expect by chance*
- *But...*
- *What does the fluctuation of word usage look like?*

RANDOM WORD USAGE

- *For each discipline and year, consider the set of words that appear with frequency k*
- *What is the distribution of the change in word frequencies in the next year?*

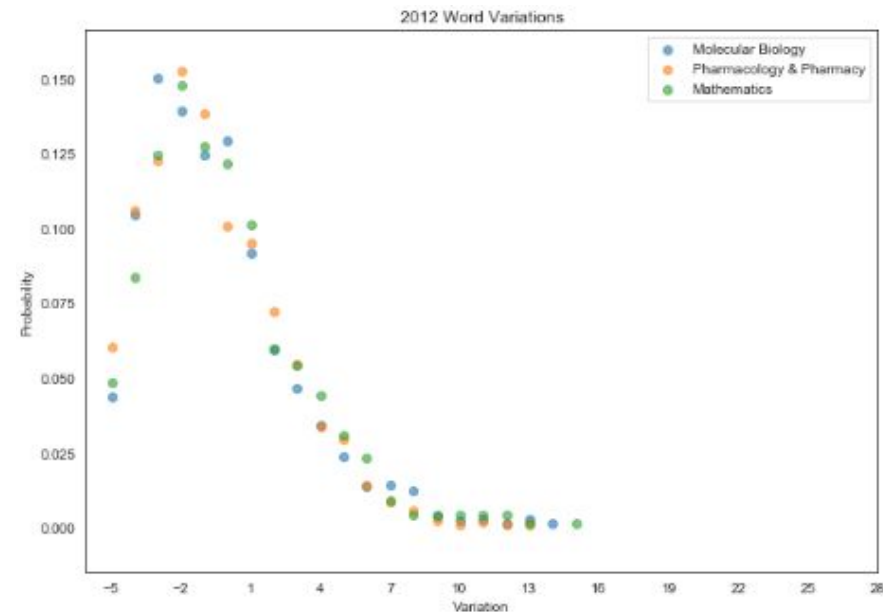
RANDOM WORD USAGE



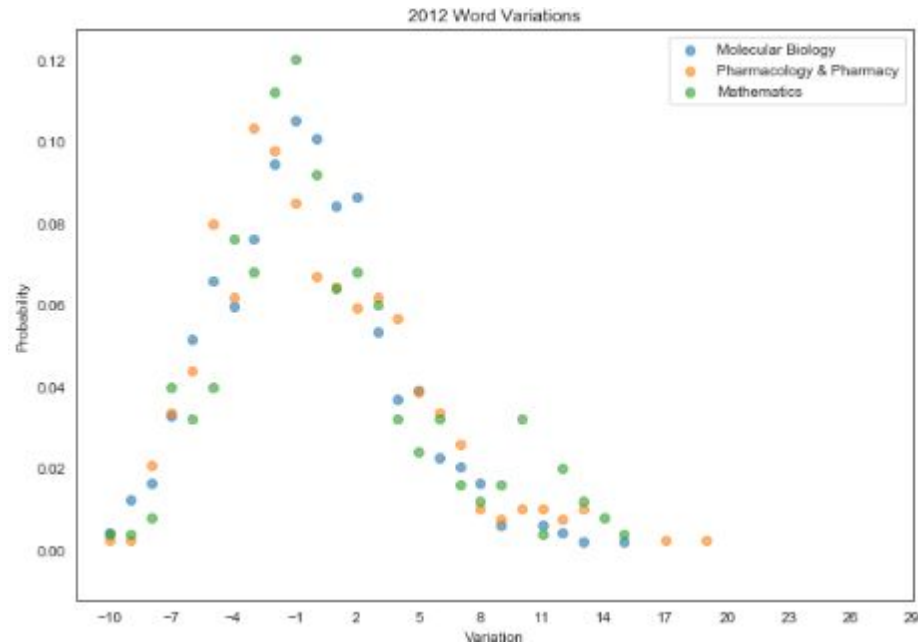
$$K = 5$$

- Take all of the words that were used $K = 5$ times in 2012
 - Molecular Biology: 1400
 - Pharmacy: 1192
 - Mathematics: 683
- Count the number of times each word was used in 2013

RANDOM WORD USAGE

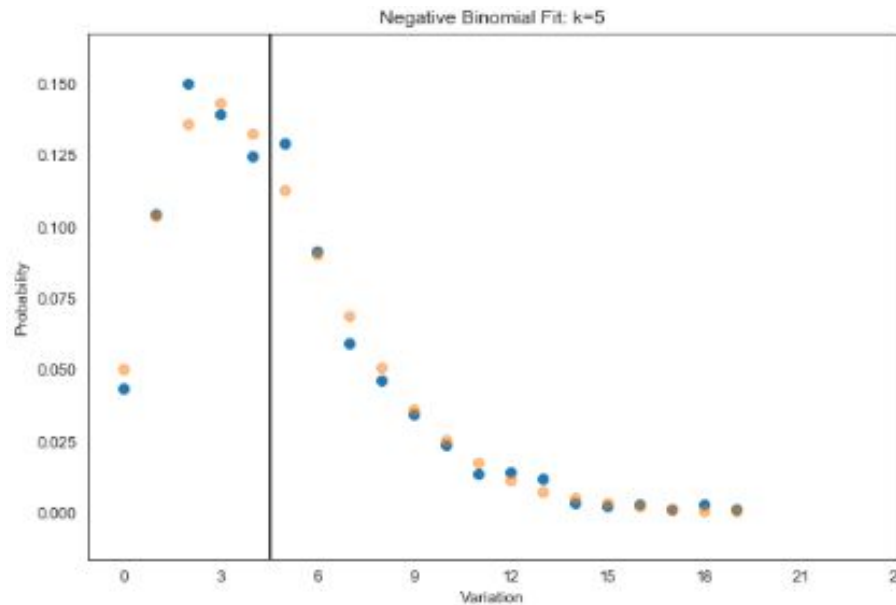


$K = 5$

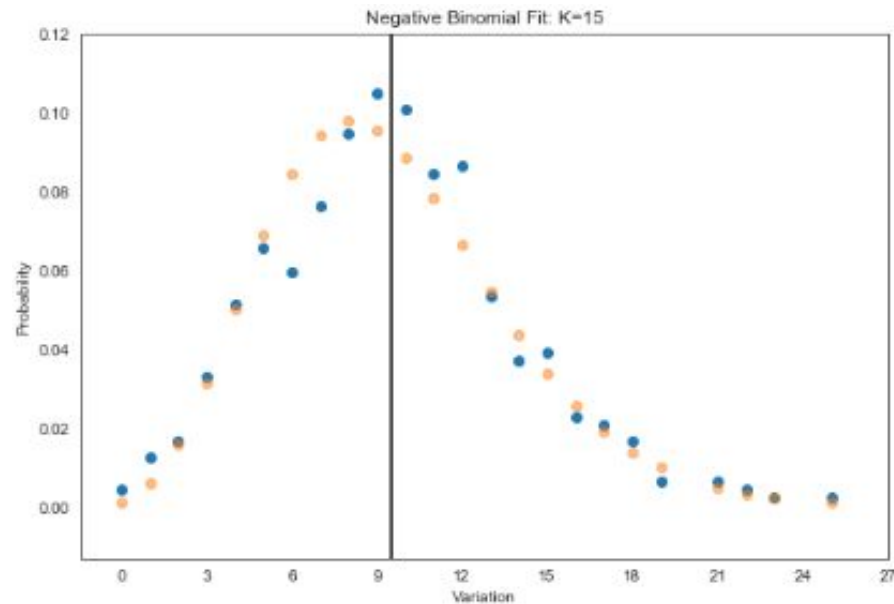


$K = 10$

NEGATIVE BINOMIAL FIT : 2012 / MOLECULAR BIOLOGY

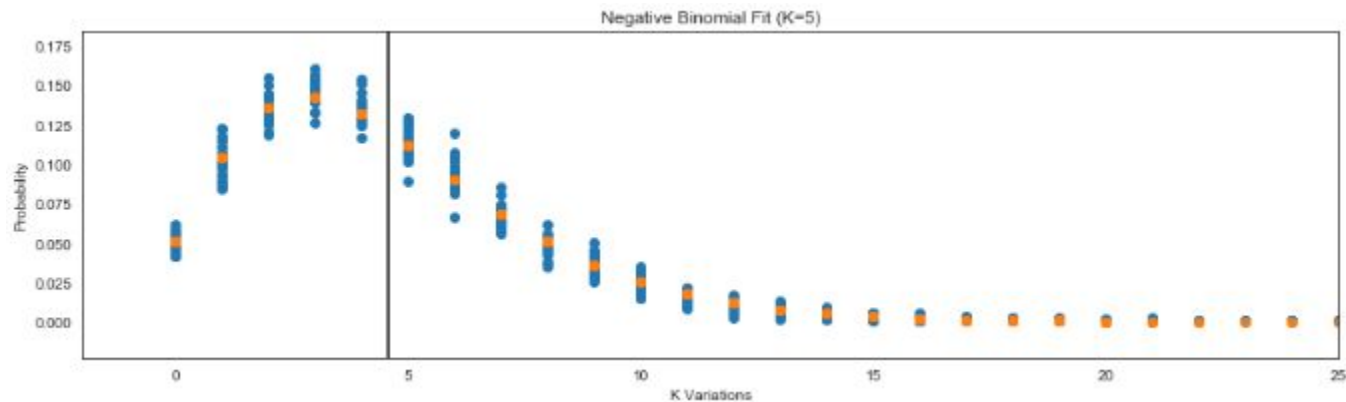


$K = 5$, SAMPLES: 1400 \rightarrow
 $N = 3.79$, $P = 0.45$

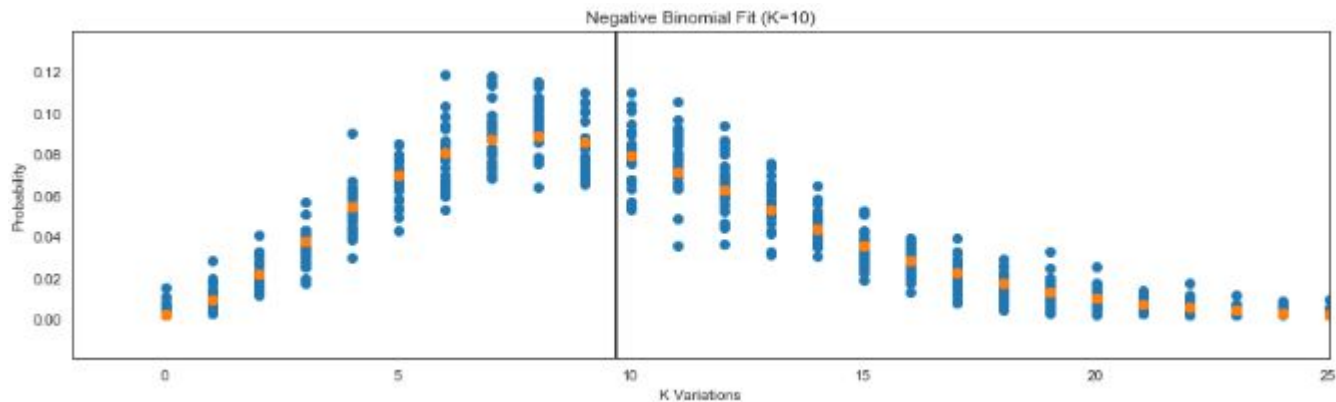


$K = 10$, SAMPLES: 486
 $\rightarrow N = 10.42$, $P = 0.52$

INCREASING SAMPLE SIZE (1990 - 2016)...

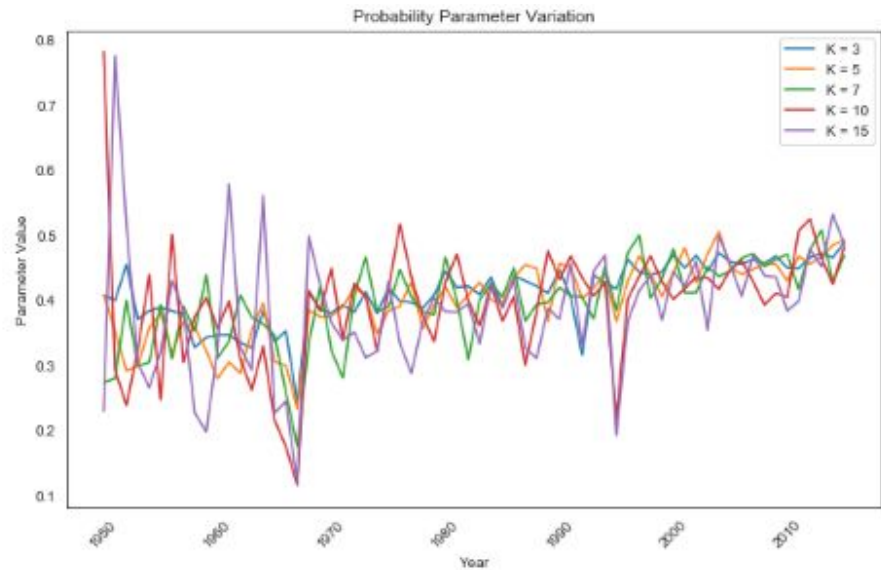
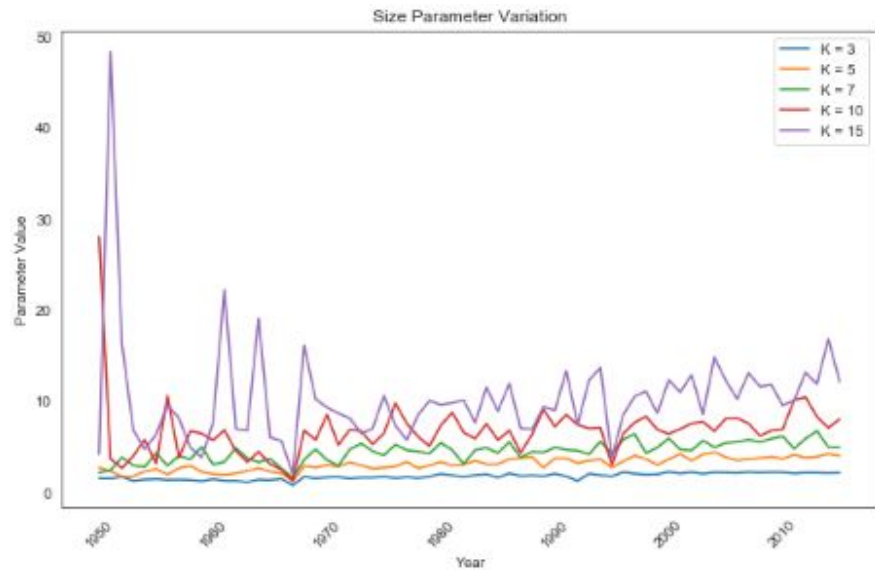


$K=5$,
SAMPLES: 30574
→ $N=3.68$, $P = 0.44$

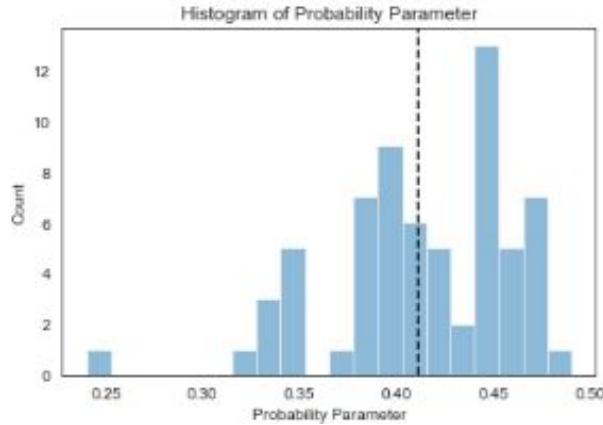


$K=10$,
SAMPLES: 9886
→ $N=6.97$, $P = 0.41$

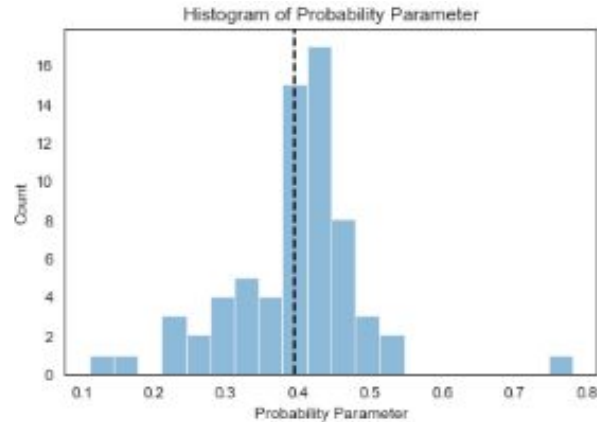
PARAMETER VARIATIONS



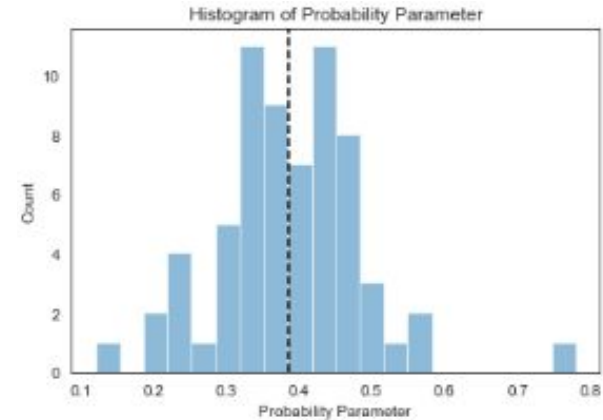
DISTRIBUTION OF PROBABILITY PARAMETER (1950 - 2016) ...



$K = 5,$
 $N = 3.09, P = 0.41$



$K = 10,$
 $N = 6.77, P = 0.39$



$K = 15,$
 $N = 10.30, P = 0.38$