

Exercise: Add-one smoothing for language models

Consider fitting a multinomial distribution to a bag of words with the following statistics:

Token	1	2	3	4	5	6	7	8	9	10
Word	mary	lamb	little	big	fleece	white	black	snow	rain	unk
Count	2	4	4	0	1	1	0	1	0	3

If we use a Dirichlet prior with $\alpha_j = 1$, what is the posterior predictive distribution for each of the 10 words in the vocabulary? What are the most likely next set of words?