Applying Naive Bayes to WSD

- P(c) is the prior probability of that sense
  - Counting in a labeled training set.
- P(w|c) conditional probability of a word given a particular sense
  - P(w|c) = count(w,c)/count(c)
- We get both of these from a tagged corpus like SemCor

- Can also generalize to look at other features besides words.
  - Then it would be P(f|c)
    - Conditional probability of a feature given a sense

Dan Jurafsky

$$\hat{P}(c) = \frac{N_c}{N}$$

$$\hat{P}(w \mid c) = \frac{count(w,c)+1}{count(c)+|V|}$$

	Doc	Words	Class
Training	1	fish smoked fish	f
	2	fish line	f
	3	fish haul smoked	f
	4	guitar jazz line	g
Test	5	line guitar jazz jazz	?

V = {fish, smoked, line, haul, guitar, jazz}

Priors:

$P(f) = \frac{3}{4}$

$P(g) = \frac{1}{4}$

Choosing a class:

$P(f \mid d5) \propto 3/4 * 2/14 * (1/14)^2 * 1/14$

≈ 0.00003

$P(g \mid d5) \propto 1/4 * 2/9 * (2/9)^2 * 2/9$

≈ 0.0006

Conditional Probabilities:

$P(line \mid f) = (1+1) / (8+6) = 2/14$

$P(guitar \mid f) = (0+1) / (8+6) = 1/14$

$P(jazz \mid f) = (0+1) / (8+6) = 1/14$

$P(line \mid g) = (1+1) / (3+6) = 2/9$

$P(guitar \mid g) = (1+1) / (3+6) = 2/9$

$P(jazz \mid g) = (1+1) / (3+6) = 2/9$

25