

Homework 2

Due: 19 Sep 2006

REMEMBER:

- Include work in symbolic form (e.g. $p(X = \text{foo} | Y = 3)$). I should be able to tell where every number came from.
- Sanity-check your answers. If your answer seems crazy but you can't find the error, at least make it clear you know there's a problem.

Problem 2.1

Let's say a dice bag has in it one ten-sided and four four-sided dice (all fair). You pull one die out at random, and roll it.

- a. What is the probability that the die is ten-sided *and* you roll a 7?
- b. What is the probability that you roll a 4, if *given that* you pick a four-sider?
- c. What is the overall probability that you roll a 3?

Problem 2.2 ($\times 2$)

Let's say a bag (with fair dice) has in it one four-sided die, two six-sided dice, four twelve-sided dice, and one twenty-sided die. Both the d6 are traditional dice with pips (dots), one red and one white. The others are gamer dice with numbers on them; two of the d12 are red, and all the rest (1d4, 1d20, 2d12) are blue.

Someone pulls one die out at random, and rolls it.

- a. What is the probability that the roll is a 9?
- b. What is the probability that the roll shows at least three pips?
- c. What is the probability that the roll is a red 3?

- d. If the person reports that they in fact rolled a 3, calculate the probability that the selected die was red.
- e. If the person reports that they in fact rolled a 3, calculate the probability that the selected die was six-sided.

Show your work.

Problem 2.3 (×2)

Write a program to calculate and report bigram statistics over a text given to it on standard input, just as described in yesterday's lab, with one additional requirement: the percentages in both the single-word ("unigram") statistics and in the bigram statistics should be printed to six decimal places. (The printf conversion for this is "%.6f".) I'm less picky about the whitespace, though it should look generally as described in the lab handout.

When you handin your code (as `hwk2`), include a text file named `README` if you need to tell me anything important about your code.

Fair warning: part of my grading process here will be

- a. Compile all the source files (.java, .C, whatever),
- b. Type `a.out < lab2-vicar.txt > output` or `java BigramStats < lab2-vicar.txt > output` or the equivalent, and
- c. Strip whitespace from `output` and diff against my own output.

Handins that actually *break* on steps 1 or 2 will not be received favourably. Handins that merely have some errors in step 3 are way less problematic.