First homework: Overture

COMP 50

Fall 2013

Due **Monday, September 9** at 1:30 PM (at the start of class)

The part of your secondary education with the most direct bearing on software design is probably high-school algebra. (Although many people, including your instructor, believe that the best indicator for a successful computer scientist is facility reading and writing one's native tongue.) With luck, your secondary education gave you some practice identifying patterns of information, capturing those patterns using simple algebraic formulas, and calculating using formulas. You probably also solved word problems. These skills correlate highly with success as a technologist. Problems 1 and 3 below are designed to reawaken those skills. Problem 2 is designed to build your intuition for measuring and estimating probabilities, which will play a major role in one of our projects.

Because some of the problems must explain domain knowledge, their descriptions are quite long. Don't be fooled; the tasks are ambitious, but short.

1. *Simple computation: comparing class schedules*. A Tufts semester includes 13 weeks of classes. The registrar rigorously adjusts the schedule so that every class gets exactly 150 minutes per week. But what if not all minutes are created equal? Please assume that students learn at top capacity only for the first 20 minutes of any class. After that, they learn at 60% capacity. I define total "learning capacity minutes" to include the first 20 minutes of every class plus 60% of the remaining minutes. So for example a 100-minute class would offer 68 "learning capacity minutes."

   With this extremely simplistic model of learning in mind, please answer these questions:

   - Alyssa P. Hacker takes a class that meets for 50 minutes three times a week. In the course of a 13-week semester, how many "learning capacity hours" of learning does she get?

     Solve this problem by writing an expression in Racket's "Beginning Student Language."

   - Ben User takes a class that meets for 75 minutes twice a week. In the course of a 13-week semester, how many "learning capacity hours" of learning does he get?

     Solve this problem by writing an expression in Racket's "Beginning Student Language."

   - How many extra classes would Ben need to accumulate the same learning as Alyssa? Or to be precise, what is the minimum number of classes that Ben needs so that his learning is at least as great as Alyssa's is.

   You will find all the Racket you need to know in Section 2 of the textbook, especially pages 5–7.

   Hint: Computer people have an abominable habit of pronouncing "at least as great" as "greater than or equal to." And they write >=.

2. *Quantifying probability*.
   One of the earliest successes in the history of computing, which played a significant role in the Allied victory in World War II, was the successful analysis ("breaking") of the German "Enigma" cryptographic machine. The code-breaking effort one of the very first real-world applications of computing devices, and at the center of it all was a mathematician named Alan Turing, who was a founder of our field. To help break the Enigma machine, Turing used the theory of probability. In COMP 50, we will use probability to solve a somewhat easier problem: writing a program to figure out what language a web page is written in. But we won't start there. Like Pascal,

Fermat, Bournoulli, Laplace, and a lot of other people who stumbled into probability, we'll start with gambling games: dice and poker.

Poker may be the quintessential American game. To win, you have to be able to reason about *odds*. "Odds" are a ratio of two probabilities; the odds of an event occurring are the probability that it occurs, divided by the probability that it doesn't occur. Odds *against* an event occurring are the other ratio (probability of not occurring divided by probability of occurring). You can complete this problem without knowing anything more about poker or odds.

Today, the word "poker" all too often means "Texas Hold 'em." But in the days of the Old West, when poker was a game of legend, "poker" meant five-card draw. The best hand in poker is called a "straight flush." [1] The chances of getting a straight flush are not good. In fact, if the dealer is fair, the odds against my being dealt a straight flush (including a royal flush) in my initial five cards are about 64,973 to 1 against. Or if you prefer, the probability of my being dealt a straight flush is about $\frac{1}{64,974}$. Although being dealt a straight flush is a rare event, I have played enough 5-card poker that I have been at a table where a straight flush was dealt in 5 cards.

Ratios like $\frac{1}{64,974}$ are hard to work with, they're hard to compare, and they don't give people much intuition. When dealing with numbers that may range from small magnitude to tens or even hundreds of billions, there is a better tool: *logarithms*. (This is the same tool used by seismologists to measure earthquakes. People who live in earthquake zones know that a magnitude 4 earthquake is something you would probably feel but wouldn't cause widespread damage, magnitude 6 or 7 is likely to cause real damage, and magnitude 9 is a major disaster.)

We'll measure logarithms of odds using a scale invented by Alan Turing: *decibans*. The unit is named by analogy with the deci*bels* used in acoustics: a deciban is ten times the base-10 logarithm of the ratio. So 10 decibans is a ten-to-one ratio, and 20 decibans is a 100-to-one ratio. Here are some odds in decibans:

- Odds I am dealt a straight flush in five cards: 48 decibans against (or -48 decibans in favor)
- Odds of winning the jackpot in the Mass Megabucks lottery: 71 decibans against
- Odds that the New England Patriots win the next Super Bowl: 9.5 decibans against
- Odds that the Denver Broncos win the next Super Bowl: 7.8 decibans against
- Odds of a randomly chosen student earning an A in COMP 105: 3.5 decibans against
- Odds that a student who completes COMP 105 (i.e., chooses not to withdraw) earns a grade of C-minus or better: 15 decibans in favor
- Claimed threshold of human perception of differences in probability: 1 deciban

This problem has two parts:

A) Complete the following table:

Event	Odds in decibans (for or against)	Required range
		at most 1 dB
		about 5 dB
		about 10 dB
		about 15 dB
		20 or more dB

Table 1: The deciban table

Your assignment is to find *real-world* events that have probabilities which fall near 0, 5, 10, and 15 decibans, as well as a very likely (or very unlikely) event with 20 or more dB. Unlikely things that occur in the real world include terrorist attacks, lightning strikes, medical diagnoses, air disasters, automobile accidents,

---

[1] The very best of the straight flushes is usually called a "royal flush," but we'll treat the royal flush as just another straight flush.

deaths from old age, hurricanes, General Gao's at Dewick, and more. Likely things in the real world include Professor Ramsey wearing Birkenstocks to class, pizza at Dewick, failing to recognize a classmate who appears to know you, French fries arriving soggy from Tasty Gourmet, losing track of time in Tisch, and more. (Sports and games do occur in the real world, and so do games of chance, but we prefer examples from other areas.)

For each real-world event you put in the table above, please state your source of information. "Personal observation" is OK as a source information, but do tell us what you have observed and how often.

If you want to crowdsource beliefs about real-world odds, you could check out the Iowa electronic markets or other Prediction markets

If you want to compute log base 10 in Racket, you can define the following function:

```
;; log10 : number -> number
;; (log10 x) returns the base-10 logarithm of x
(define (log10 x) (/ (log x) (log 10)))
;
; tests
(check-within (log10 10)   1 0.01)
(check-within (log10 1000) 3 0.01)
```

B) Psychologists frequently measure subjective opinions using a device called a "Likert scale." Imagine you were asked to estimate the odds of a future event, such as "after the first snowstorm of 2014, Professor Ramsey will teach his next class wearing Birkenstocks." A psychologist might ask you to state your subjective opinion by choosing among the alternatives on this seven-point Likert scale:

- The event is very likely
- The event is likely
- The event is somewhat likely
- The event is equally likely and unlikely
- The event is somewhat unlikely
- The event is unlikely
- The event is very unlikely

Based on your research for part A, please **assign a range in decibans** to each point on the scale. The goal is to give you some practice in *quantifying* your subjective beliefs: in this case, your beliefs about the odds in favor of the event. Your subjective beliefs can be called **prior odds** (you'll see the term "prior probabilities" more often), and they will play a role in some of our future projects.

Hint: the rightness or wrongness of your answers to part A depends only on the quality of your research. Only one bullet in Part B has a unique right answer, but all of the bullets have cognitively rational and irrational answers. We like rationality.

3. *Devising formulas for numeric data.* This problem is similar to the "standing in a line" and "sitting by doubles" problems demonstrated in the first lecture.

Sound is communicated to the human ear by waves of compression and rarefaction that pass through the air (or another medium—we can hear underwater, for example). Within reasonable limits, the human eardrum responds to these waves by vibrating at the same frequency as the waves. These vibrations are eventually communicated to the brain, where they are perceived as sound. The number of vibrations per second is measured in Hertz, abbreviated Hz.

An expensive device called a *frequency meter* can extend the perception of sound far beyond the range of human hearing. Frequency meters are also used to measure electrical and electromagnetic signals. (Some of the early patents were held by the inventor Nikola Tesla, who also gave us our system of alternating current and who now has a car named after him.) These meters are capable, for example, of measuring the 3GHz clock signal used in your personal computer—that's three *billion* vibrations per second.

Let us suppose that Tesla wanted to test his device by measuring the frequency of sounds in a friend's home. After a period of relative silence, he measures a signal that changes in frequency every half a second. If we measure time t from the start of the sound, this is what happens:

after $t =$	Tesla measures
0.0 seconds	261.626 Hz
0.5	277.183
1.0	293.665
1.5	311.127
2.0	329.628
2.5	349.228
3.0	369.994
3.5	391.995
4.0	415.305
4.5	440.000
5.0	466.164
…	…
10.0	??
…	…
15.0	??

Table 2: First experiment with the frequency meter

Guess a formula for calculating what frequency Tesla measures at time t.

Check the formula for the first five table entries. If it doesn't work, guess some more.

Once the formula works for the first five entries, use it and a calculator (or DrRacket's Interactions Window) to fill in the two boxes with ?? in the table above.

Please submit both your **formula** and your **two table entries**.

Five minutes later, Tesla returns, and this is what he measures:

after $t =$	300.0	300.5	301.0	301.5	302.0	…
Tesla measures	261.626	293.665	329.628	349.228	391.995	…

Table 3: Second experiment with the frequency meter

Karma Problem: at the five-minute mark, what is happening?

Hint: look at the words in the following sentence, and think about an alternative meaning for one of the words:

The frequency meter is a scientific instrument.

4

What to submit

Submit your answers on *printer paper*. Please write on *one side* and *put your name on every page*. We will be running the submissions through a scanner, so *please do not use notebook paper*.

1. Submit two expressions in Racket's "Beginning Student Language:" one for Alyssa and one for Ben.

   State the number of extra classes Ben needs (as a whole number).

2. For part A, submit a deciban table with five entries. Each entry should list the event, the odds in decibans, and whether the odds are for or against.

   For part A, submit five sources (or personal observations) telling us where you got your numbers from.

   For part B, submit a table with seven entries; each entry should give the likelihood in English and your subjective opinion of the *odds* in decibans.

3. For part C, submit a *formula* and submit *two frequencies*: the ones you expect to be observed at times 10.0 and 15.0, respectively. (The formula *need not be written using Racket*.)

How your work will be evaluated

	Exemplary	Satisfactory	Must Improve
Computation	• For Alyssa, and Ben, solution to problem 1 contains correct expressions in Racket's *Beginning Student Language*. • For Alyssa, and Ben, problem 1 contains expressions that are *almost* correct, but they compute minutes (1482 and 1378) instead of hours. • Solution to problem 1 states correctly how many extra classes Ben would needs to accumulate the same learning as Alyssa.	• Solution to problem 1 contains well-formed expressions in Racket's *Beginning Student Language*, but they don't compute what was asked for Alyssa and Ben. • Solution to problem 1 contains expressions or formulas that compute what was asked for Alyssa and Ben, but they are not well-formed expressions in Racket's *Beginning Student Language*. • Solution to problem 1 states almost correctly how many extra classes Ben would need: it's off by one.	• Solutions for Ben and Alyssa are not well-formed expressions in Racket's *Beginning Student Language*, and they don't compute the right answers, either. • Solution to how many classes Ben would need is off by more than one.

	Exemplary	**Satisfactory**	**Must Improve**
Probability	• In problem 2A, course staff can easily verify that the odds in decibans are stated correctly. (Staff will check cited sources or will look at personal observations stated.) Answers state correctly whether odds are 'for' or 'against'. • Any personal observations used in problem 2A are *quantified* (that is, stated with numbers). • In problem 2A the student has mistakenly used log of *probability* instead of log of *odds*. Minor deduction. • In problem 2B, the entries in the table reflect a rational interpretation of all the words. • In problem 2B the student has mistakenly used log of *probability* instead of log of *odds*. (Easily spotted because "equally likely and unlikely" will be 3.01dB.) Minor deduction.	• Odds in problem 2A are not easily verifiable, but staff believe them to be plausible. • Answers mix up 'for' and 'against'. • In 2B, the entries in the table reflect a rational interpretation of words like "very" and "somewhere", but the interpretation of the words "likely" and "unlikely" is open to challenge. • *Or*, the entries in the table would reflect a rational interpretation, but the interpretation is exactly backwards.	• One or more entries in the deciban table is shown to be wrong; that is, the decibans have not been calculated correctly. • One or more entries in the deciban table is dramatically unbelievable, and the value of the entry is not supported by either a cited source or by *quantitative* personal observations. • The entries in the table reflect an irrational interpretation of the words "very" and "somewhat."
Formulas	• In problem 3, a formula is given that is accurate to 0.01Hz, and all the numbers that appear are whole numbers (integers) • *Or*, a formula is given that is accurate to 0.1Hz. • *Or*, the formula uses a numeric power law, and the answer for $t = 10$ is 830 ± 10Hz and the answer for $t = 15$ is 1490 ± 30Hz.	• Answers for times 10.0 and 15.0 are accurate to 1Hz, but a formula is given which is not as accurate.	• Answers for times 10.0 and 15.0 are accurate to within 10Hz, regardless of the accuracy of the formula given. • Answers are not accurate to within 10Hz, or no formula is given (**serious fault**).