Last problem in problem set 1.

This is the most verbose description of a decision problem I have ever seen. Simplify it all into a simple decision tree and answer the three questions at the end.

You got to school this morning to realize that half of your class is in bed with the flu. Bad news. Time to think seriously about those flu shots? This will be our decision problem for the day. What should you factor into your decision? First, whether or not getting the shot turns out to be a good idea will depend on whether or not you will be attacked by a flu bug during the semester. In the past you have caught the flu sometimes and other times you have miraculously made it past every single bug on campus and went for a whole semester without missing a single day. We'll take this to be a random event. To start getting rigorous let's call $\theta = 0$ the event "making it past every bug on campus", and $\theta = 1$ the event "being hit by at least one bug". If $\theta = 0$ is going to happen, a flu shot is a waste of time. If $\theta = 1$ is going to happen, a flu shot is a good idea. So there is a trade off between wasting time and energy getting the shot now and and getting a bigger, but not quite certain, return.

Much of decision theory has to do with quantification of this sort of trade-offs. Let' make up some unrealistic facts to make this trade-off a little simpler to quantify. If you do not get the flu shot, and you end up running into the bugs (i.e. $\theta = 0$ does not occur,) then you will be off of work for 6 days. This is a tough flu. Otherwise, you are fine. On the other hand, if you decide to get the flu shot, you will stay at home for a day right away, due to the adverse reaction to the shot. Next we are going to make what seems like a huge assumption: we are going to attach monetary value to these outcomes, and then make our decision based only on those. Let's say that you lose a 100 dollars for each day you spend in bed sick. Another way to think of this is that the utility of one day in bed is -100, and additive in the number of days; the units are not really relevant.

The previous information can be summarized as in Table 1

Several of our classes deal with trying to understand how big is our big assumption. One question is: under which circumstances is it rational to attach a quantitative measure to the outcomes that are important in our decision, and address our trade-off based solely on that?

Having a quantitative, or even monetary, amount attached to the outcomes of the decision helps, but in most cases it is not enough to reach a unique decision. For example, if you

		Event	
		$\theta = 0$	$\theta = 1$
Decision	no shot	\$0	\$600
	shot	\$100	\$100

Table 1: Loss table for the flu shot example.

were trapped in an elevator with five coughing nurses and without heating system, you may consider $\theta = 0$ a rather remote possibility, and may be inclined to get the shot. On the other hand, if you are going to spend most of the semester on leave at the Université de Tahiti, you may think that the chance of benefiting from the shot is too small to be worth the day in bed. As you begin puzzling you realize that you need to weigh the extra \$500 loss in case of bug, against the chance that the bug will actually get you.

The expected utility approach is based on getting a quantitative measure of how likely is that $\theta = 0$, specifically a probability, and use that to compute a weighted average of the losses, or expected loss, for each decision. The Bayes decision is the one with the smallest expected loss. For example, suppose that we can estimate, based, for instance, on previous years' information, that 10% of the campus population will be hit by the bug. Imagine dividing the campus population at random in two groups, one that gets the shot and one that doesn't. Then, compute the average loss in the two groups. Approximately 10% of individuals in each group will be hit by the bug, because we are splitting at random. So the expected loss in each group is:

$$\begin{aligned} \text{Shot} \quad &: \quad \text{Expected Loss} = 0 \times 0.9 + 600 \times 0.1 = 60 \\ \text{No shot} \quad &: \quad \text{Expected Loss} = 100 \times 0.9 + 100 \times 0.1 = 100 \end{aligned}$$

The decision that minimizes the expected loss above (or equivalently maximizes expected utility) is not to take the fly shot. If the probability of $\theta = 1$ was 16.83% or greater, the best decision would be to get the shot. A big issue here is "why the average?". Several chapters in this book deal with trying to understand how big of a deal this is. Elaborating on what we said earlier, one question will be: under which circumstances is it rational (in a sense to be defined very precisely) to attach a quantitative measure of loss (or gain) to the outcomes, a probability to the unknown events, and address our trade-off based on the expected loss (or gain).

One key issue is how to determine the probability. Picturing yourself in the hypothetical clinical trial described above and gathering population data to try to choose the group you would rather belong to is on one way. But maybe you do not believe that your probability of getting a cold can be assumed to be the same as that of a randomly drawn person from that population. Maybe there are factors that might put you at higher risk, such as the number of hours of sleep you usually (don't) get, the number of colds in the last few years, the five nurses, and so forth. Maybe you want to use your own personalized probability. We'll think about this option too.

Another possibility for improving your decision making is to gather data about how likely you are to get hit by a bug. Here things get really interesting because now the decisions are not just about the flu shot, but are also about how to gather data and how to best use data to make a decision about the shot. In more technical terms, the idea is that of developing and justifying statistical procedures by looking at the use we plan to make of the results. This will keep us occupied for over one third of the book, at the end.

To be more specific about the information that you will have before deciding whether or not to get a flu shot, let's imagine that you have the opportunity to go to the hospital and do a test that helps indicating how good you are at staying free of bugs. The test has only two possible verdicts. One indicates that you are a bug-prone type and the other indicates that you are not. Sticking with the time-honored medical tradition of calling "positive" the results of tests that suggests the presence of the most devastating health conditions, we'll call positive the outcome indicating that you are bug-prone. Unfortunately, the test is not 100% accurate. Let's assume that, after some clinical experiments, it was determined that the probability that the test is positive when you really are going to be hit by a bug is 0.9 and the probability that the test is negative when you are not going to be hit is 0.8. For a perfect test, these number would be both one. In medical terminology, the first probability represents the *sensitivity* of the test, while the second one represents the *specificity* of the test.

Let us define X as the indicator variable for the event, "The test is positive". The probabilities assigned so far are:

$$\begin{aligned} Pr(\theta = 1) &= 0.1 \\ Pr(X = 1|\theta = 1) &= 0.9 \\ Pr(X = 0|\theta = 0) &= 0.8 \end{aligned}$$

- A natural question is how to use the results of the test to make a better decision. The test seems reliable enough that we may want to proceed with the shot if the test is positive and forego the shot otherwise. Is this right?
- Another natural question is whether the test is worthwhile. Is it?
- If the test had no side effect, what would be the expected value of the information provided by it? How would it compare to that of a test with no side effects and perfect sensitivity and specificity?