Study Guide and Practice for Quiz #1

- Be prepared to estimate quantities using the Fermi technique.
- Be prepared to answer basic questions in probability theory having to do with the relationship between conditional, marginal, and joint probabilities; independence versus dependence, and so on, as we discussed in class.
- In class we discussed various kinds of Monty Hall problems and similar problems like the kings and siblings. You should be prepared to answer similar questions for variants of the original problems, e.g, what if Monty has four doors, or what if the number of siblings is different from the standard problem?
- You should be able to handle problems similar to this one: A particular gene is implicated in a rather nasty form of cancer. If you have the gene, your probability of getting this cancer is 0.2, whereas if you do not have the gene your probability of getting it is 0.0002. About 1% of the population has the gene. A cheap test has been developed to test for the gene. It is 98% reliable in detecting the gene (that is, if you have the gene, the probability is 0.98 that the test will be positive). However, it also has a 5% rate of false positives. There is another test with a much smaller rate of false positives, but it is quite expensive to administer. It's proposed to use the cheap test to screen for the gene. Those that test positive would presumably take the more expensive test.

  > 1) What's the probability that a member of the general population who tests positive does not have the gene?
  > 2) What's the probability that a member of the general population who tests positive has the gene?
  > 3) What's the probability that a member of the general population who tests positive will go on to get the disease? Of these, what proportion did not have the gene?

  Assess the usefulness of using this test on the general population. Consider both the fact that people who may not know that they are at risk for the cancer might be saved as well as the fact that some people who are at much lower risk for the cancer will be unnecessarily troubled and may have to take the more expensive test.

- You should be able to handle problems similar to this one: Galaxies can be classified by the proportions of various kinds of stars and other objects in them. Suppose we have invented an automatic galaxy-classification machine. It is designed to inventory the objects in a nearby galaxy and tell us what kind of galaxy it is. The kinds of galaxies (E,S) and the objects in the galaxies that the machine can recognize and count (1, 2, 3) are shown in the table, along with the conditional probabilities that a randomly picked object will be picked in a galaxy, given the kind of galaxy it is.

	1	2	3
E	0.9	0.08	0.02
S	0.6	0.3	0.1

  Suppose that 80% of galaxies are of type E. The machine inventories 10 randomly chosen objects in a galaxy and finds that 7 of them are type 1, 2 of them are type 2, and 1 of them is type 3. What is the probability that the galaxy is of type E?
- As a guard against plagiarism, in the old days constructors of mathematical tables would often use rounding as a weapon for telling if someone had copied their table. They would round numbers ending in the digit 5 up or down randomly, e.g., if they were constructing a 5-digit table of

logarithms, they would take a number like 0.324655 when it occurred and flip a coin. If it was heads the published table would round down to 0.32465 and if tails up to 0.32466. This would have no significant effect on the legitimate users of the table, but in this way they were able to embed a random code in the table, known only to them, that they could use to detect whether someone who published a new table had actually computed his own table or was plagiarizing his table. Then if they suspected that someone was copying the table they could take them to court using as evidence the fact that the random rounding pattern was the same in their table as in the suspect one. Suppose for example that there were 1000 numbers in the table. One would expect that 100 of them would have ended in a 5 and would have been rounded up or down. Suppose that every one of those 100 numbers were rounded exactly as in the original author's table. Assuming some prior probability on the alternatives (plagiarized, not plagiarized), what is the posterior probability that the table was plagiarized?

- An urn contains a number of white balls, known to be no more than ten balls but otherwise the number is not known. You draw balls one at a time from the urn and return them to the urn, and if a ball is not marked you mark it with an "X" before returning it to the urn, afterwards shaking the urn thoroughly before drawing another ball. Each time you record whether the ball was already marked or not. You will need to put a prior probability on the number of balls in the urn and justify your choice. You draw in turn: An unmarked ball (which you mark), an unmarked ball (which you mark), a marked ball, an unmarked ball (which you mark), a marked ball. From this data, calculate the posterior probability on the number of balls in the urn
- An urn contains an unknown number N of balls, where N is not more than 10. The balls are numbered consecutively from 1 to N. You draw three balls from the urn. They are numbered 2,5 and 3. Assume a prior on N, and justify your prior. What is the posterior probability of the number N of balls in the urn. (Comment: This is similar to how the allies in World War II estimated the number of German tanks that had been produced, since the Germans made the mistake of putting serial numbers on their tanks in sequence).
- A biologist wishes to study the population of an ant's nest. There are ants in the nest, but there are also beetles that mimic ants...they look almost exactly like ants and to the ants they "smell" like ants, but microscopic examination can reveal that they are beetles and not ants. The biologist captures 20 insects and finds that 3 of them are beetles, not ants. How badly infested with beetles is this nest? Let the proportion of beetles in the nest be p. You may assume for example that p takes on the values 0.05, 0.15, 0.25,..., 0.95. You will have to assign a prior probability on the values of p, and you need to justify how you have assigned that prior probability. You may also assume that the number of ants in the nest is very large (tens of thousands). What is the distribution of posterior probability on p?
- Decision problems: You should be able to set up and solve basic decision problems, including linked decisions, using decision trees. You should be sure to label nodes with their expected values, and to cut off branches that are not desirable! Here's one for you to think about:

  You are a factory manager, and are in charge of an automatic machine that produces 24 parts every day. On 90% of the days the machine is in a "well-adjusted" state, so that 95% of the parts it produces are good and 5% of the parts are bad. However, on 10% of the days the machine is in a "maladjusted" state, producing only 70% good parts with the remaining parts bad. Good parts are worth $2,000 apiece, but bad parts are worthless. Whenever a part is produced, it is tested (testing is assumed free) so that we know whether it is good or bad.

  There's a mechanic who can adjust the machine and reliably put it into the "well-adjusted" state. The mechanic costs $150, and it takes him as long to adjust the machine as it takes to produce one part. Thus, if we routinely called him in at the beginning of the day, we would only be able to produce 23 parts, and would also have to pay him; if the machine is already in the "adjusted" state, we'll not make as much money, so there is also an opportunity cost for the possibly good part that we might have made but didn't. On the other hand, we could wait until one part is produced, test it, and then decide whether to call in the mechanic. In that case we will only be able to produce 22 more parts since it will still take the mechanic the same amount of time to adjust the machine, and again there will be an opportunity cost.

Set up a decision tree for this problem. You should seek to maximize your expected profit. Should the mechanic be called in to adjust the machine routinely every day before any parts are produced? Should the mechanic be called in to adjust the machine conditional on the whether the first part produced is good or bad?
