Massachusetts Institute of Technology | Handout 30
6.042J/18.062J: Mathematics for Computer Science | April 29, 1997
Professor David R. Karger

Problem Set 11

Due: May 6, 1997.

218z NOTICE: A more compelete version will be posted on the web

Reading: Rosen 4.4-4.5 up to "independence," 4.6.

In the problems below, try to simplify your answers as much as possible. There are a lot of questions, but many of them can be answered with one equation and a brief explanation.

Problem 1. Several dice and an urn.

(a) When rolling a pair of fair dice, what is the probability of getting a sum of two on the dice? a sum of seven on the dice? a sum of eleven on the dice?

There are 6 possible outcomes for each die, and these occur independently so there are 36 equally likely outcomes. (Notice that getting the same numbers but on different dice corresponds to different outcomes). The probability of the above events is thus the number of outcomes corresponding to those events divided by 36. There is exactly 1 outcome corresponding to a sum of two on the dice, so this happens with probability $\frac{1}{36}$. There are 6 outcomes corresponding to a sum of seven on the dice, so this happens with probability $\frac{6}{36} = \frac{1}{6}$. There are 2 outcomes corresponding to a sum of eleven on the dice, so this happens with probability $\frac{2}{36} = \frac{1}{18}$.

(b) Pepys wrote Newton to ask which of three events is more likely: that a person get (i) at least one six when 6 dice are rolled, (ii) at least two sixes when 12 diced are rolled, or (iii) at least three sixes when 18 dice are rolled. What is the answer?

The answer is (i). In all cases, the number of sixes is binomially distributed. The probability of (i) is $1 - (5/6)^6$, which is about 0.665. The probability of (ii) is $1 - (5/6)^1 2 - \binom{12}{1} \cdot (1/6) \cdot (5/6)^1 1$, which is about 0.619. The probability of (iii) is $1 - (5/6)^1 8 - \binom{18}{1} \cdot (1/6) \cdot (5/6)^1 7 - \binom{18}{2} \cdot (1/6)^2 \cdot (5/6)^1 6$, which is about 0.597.

(c) A six-sided die is tossed 12 times. What is the probability of obtaining exactly 2 of each number from 1 to 6?

The number of distinct sequences in which each number appears exactly twice is the same as the number of permutations of 1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6, 6. This is

$$\frac{12!}{2!2!2!2!2!2!} = \frac{12!}{2^6} \tag{1}$$

There are 6^{12} possible outcomes in 12 tosses of a die. If we suppose that all of these have equal probabilities, then the probability of obtaining a sequence in which each face appear exactly twice is

$$\frac{\frac{12!}{2^6}}{6^12} = 3.4(10^{-3}) \qquad (2)$$

(d) An urn contains five balls numbered $1, 2, 3, 4, 5$. Suppose we draw three balls with replacement. What is the probability that all three balls are different?

There are $5^3 = 125$ possible outcomes, which we will suppose are all equally likely. The number of these outcomes for which the three draws are different is given by $5(4)(3) = 60$. Thus the probability that all three balls are different is $\frac{60}{125} = 0.48$.

Problem 2. When playing *Let's Make a Deal* in class, I decided to "speed things up" (give contestants a better chance of winning) by adding a 4th box and, after your choice, revealing 2 that don't have the prize. Draw the tree of outcomes for the "always switch" rule when 4 boxes are involved. What is the probability that you win?

The probability that you win is $\frac{3}{4}$.

Problem 3. Consider a best 3 out of 5 playoff. If team 1 has a probability 1/2 of winning each game that it plays, what is the most likely (that is, highest probability) number of games in the playoff? What if the first team's win probability is 9/10?

There are 2 (mutually exclusive) ways that the playoff can end in 3 games, each with probability $\frac{1}{8}$. So the playoff ends in 3 games with probability $2 \cdot \frac{1}{8} = \frac{1}{4}$. A playoff consisting of 4 games must involve three wins by one team and one win by the other team, with this win being any one of the first three games. (If it was the fourth, the playoff would have ended after the first three games). There are two ways of choosing which team wins, so there are 6 ways that the playoff can last exactly 4 games. Each way has probability $\frac{1}{16}$, so the probability that the playoff has exactly 4 games is $6 \cdot \frac{1}{16} = \frac{3}{8}$. The probability of the playoff having exactly 5 games is $1 - \frac{1}{4} - \frac{3}{8} = \frac{3}{8}$. Thus the most likely numbers of games are 4 and 5 (happening with equal probability).

If the first team's win probability is 9/10, then the first team can win the playoff in 3 games exactly one way, which happens with probability $(\frac{9}{10})^3$ and the second team can win the playoff in 3 games exactly one way, which happens with probability $(\frac{1}{10})^3$. So the probability that the playoff ends in exactly three games is $(\frac{9}{10})^3 + (\frac{1}{10})^3 = \frac{73}{100}$. The first team can win in exactly 4 games three different ways, each with probability $(\frac{9}{10})^3 \cdot \frac{1}{10}$. The second team can also win in exactly 4 games three different ways, but these each occur with probability $(\frac{1}{10})^3 \cdot \frac{9}{10}$. So the probability that the playoff will have exactly 4 games is $3 \cdot (\frac{9}{10})^3 \cdot \frac{1}{10} + 3 \cdot (\frac{1}{10})^3 \cdot \frac{9}{10} = \frac{1107}{5000}$. The probability that the playoff will have exactly 5 games is $1 - \frac{73}{100} - \frac{1107}{5000} = \frac{243}{5000}$. So in this case, the most likely number of games in the playoff is 3.

Problem 4. Suppose you are dealt a random poker hand of 5 cards from a deck of 52 cards. Compute the probability that you are dealt

(a) Four of a kind (four cards with the same value)

There are $C(52,5) = \frac{52!}{5!47!}$ ways to choose 5 cards. There are 13 ways to choose four of a kind and 48 ways to choose the 5^{th} card. The probability of four of a kind is $\frac{13 \cdot 48}{\frac{52!}{5!47!}}$.

(b) A flush (5 cards from the same suit) There is 4 ways to choose a suit and $C(13,5)$ ways to choose 5 cards from the same suit. The probability of a flush is $\frac{4 \cdot C(13,5)}{\frac{52!}{5!47!}} = \frac{4\frac{13!}{5!(13-5)!}}{\frac{52!}{5!47!}}$.

(c) Two pair (two pairs, where each pair is two cards with the same value, but the two pairs have different values and the remaining card does not have either value)

There are 13 of a kind and $C(4,2)$ ways to choose 2 cards out of 4 of a kind. There are 12 of a kind and $C(4,2)$ ways to choose 2 cards out of 4 of a kind and 44 ways to choose the 5^{th} card. The probability is $\frac{13 \cdot C(4,2) \cdot 12 \cdot C(4,2) \cdot 44}{\frac{52!}{5!47!}}$.

Problem 5. A batch of 50 items contains 10 defective items. Suppose 5 of these items are selected at random (without replacement) and tested.

(a) What is the probability that exactly 4 of the items tested are defective?

$$\frac{\binom{10}{4}\binom{40}{1}}{\binom{50}{5}} \tag{3}$$

(b) What is the most likely number of defective items in the sample?

$$\begin{aligned}
\Pr[0 \text{ defects}] &= \frac{\binom{10}{0}\binom{40}{5}}{\binom{50}{5}} \\
\Pr[1 \text{ defect}] &= \frac{\binom{10}{1}\binom{40}{4}}{\binom{50}{5}} \\
\Pr[2 \text{ defects}] &= \frac{\binom{10}{2}\binom{40}{3}}{\binom{50}{5}} \\
\Pr[3 \text{ defects}] &= \frac{\binom{10}{3}\binom{40}{2}}{\binom{50}{5}}
\end{aligned}$$

$$\begin{aligned}
\Pr[4 \text{ defects}] &= \frac{\binom{10}{4}\binom{40}{1}}{\binom{50}{5}} \\
\Pr[5 \text{ defects}] &= \frac{\binom{10}{5}\binom{40}{0}}{\binom{50}{5}}
\end{aligned}$$

Thus, it is most likely that that there exists 1 defective item in the sample.

Problem 6. Suppose that I randomly permute the letters in MISSISSIPPI. What is the probability that all the S's end up next to each other?

There are $\binom{11}{1,4,4,2}$ permutations of the letters (all equally likely). We can count the number of permutations with all the S's next to each other by replacing them with a single new letter. So there are $\binom{8}{1,4,1,2}$ of them. Thus the probability that all the S's end up next to each other is $\frac{\binom{8}{1,4,1,2}}{\binom{11}{1,4,4,2}} = \frac{4}{165}$.

Problem 7. Prove that if $A_1, \ldots, A_n$ is a partition of the sample space, S, then

$$\Pr[B] = \sum_{i=1}^{n} \Pr[B \mid A_i] \cdot \Pr[A_i]$$

Let S be the sample space partitioned by $A_1, \ldots, A_n$.

$$\begin{aligned}
\Pr[B] &= \Pr[B \cap S] \\
&= \Pr[B \cap (A_1 \cup A_2 \cup \ldots \cup A_n)] \qquad \text{since } A_1, \ldots, A_n \text{ partions } S \\
&= \Pr[(A_1 \cap B) \cup \cdots \cup (A_n \cap B)] \\
&= \Pr[A_1 \cap B] + \Pr[A_2 \cap B] + \cdots + \Pr[A_n \cap B] \\
&= \Pr[B \mid A_1] \cdot \Pr[A_1] + \ldots + \Pr[B \mid A_n] \cdot \Pr[A_n] \qquad \text{by conditional probability} \\
&= \sum_{i=1}^{n} \Pr[B \mid A_i] \cdot \Pr[A_i]
\end{aligned}$$

Problem 8. A witness to a crime says that the criminal had a beard. We know from experience that if a criminal has a beard, witnesses will say so with probability 0.8, and if the criminal does not have a beard, witnesses will say that he does with probability 0.12. If 15% of the population has beards, what is the probability that a random criminal has a beard if the witness says he does?

Let A be the event that the witness says the criminal has a beard and B be the event that the criminal actually does have a beard. The given information is that $\Pr[A \mid B] = 0.8, \Pr[A \mid \bar{B}] = 0.12$, and $\Pr[B] = 0.15$. So $\Pr[A \cap B] = \Pr[A \mid B] \cdot Pr[B] = 0.8 \cdot 0.15 = 0.12$ and $\Pr[A \cap \bar{B}] = \Pr[A \mid \bar{B}] \cdot Pr[\bar{B}] = 0.12 \cdot (1 - 0.15) = 0.102$. Then $\Pr[A] = Pr[A \cap B] + Pr[A \cap \bar{B}] = 0.12 + 0.102 = 0.222$. We are asked to find $Pr[B \mid A]$, which is given by $Pr[B \mid A] = \frac{\Pr[B \cap A]}{\Pr[A]} = \frac{0.12}{0.222} = \frac{20}{37}$.

Optional Problem 9. Suppose you are playing the *Let's Make a Deal* game, but you happen to know that the prize is going to be behind doors 1, 2, or 3 with probabilities $p_1 \geq p_2 \geq p_3$, respectively. How should you play? What are your chances of winning? Can you prove your strategy is optimal?

Suppose your strategy is to pick door i and stick with it regardless of what door the host opens. Then you win iff the prize was behind door i, which happens with probability p_i.

Suppose your strategy is to pick door i and switch to door k if door j is opened but stay with door i if door k is opened. If the prize is behind door k, door j will be opened, you will switch, and you will win. This happens with probability p_k. If the prize is behind door j, door k will be opened, you will stay with door i, and you will lose. If the prize is behind door i, half the time door j will be opened (in which case you switch and lose) and half the time door k will be opened (in which case you stay and win). Thus the conditional probability of your winning given that the prize is behind door i is 0.5. The probability that the prize is behind door i and you win is $0.5 \cdot p_i$. So with this strategy, you win with probability $p_k + 0.5 \cdot p_i$.

Suppose your strategy is to pick door i and then switch (regardless of whether this means switching to door j or to door k). Then you win iff the prize is behind door j or door k (by the same analysis as in the original problem). This happens with probability $p_j + p_k$.

This accounts for all possible strategies. The best probability of winning is $p_1 + p_2$, which is for the strategy of picking door 3 first and then always switching. Note that this may not be the unique optimal strategy in the special case where the p_i are not all distinct.

Optional Problem 10. Let f be a function $f \colon \{0,1\}^n \to \{0,1\}$. We think of f as an electoral system in which if people vote $X_1, \ldots, X_n$, $f(X_1, \ldots, X_n)$ wins the elections.

A natural requirement on f is monotonicity:

$$\forall i. X_i \leq Y_i \Rightarrow f(X_1, \ldots, X_n) \leq f(Y_1, \ldots, Y_n),$$

i.e. if some people change their vote from V to $1 - V$, this can only change the result of the elections from V to $1 - V$.

We measure the quality of an electoral system f by the probability that a randomly selected person R can change the outcome of the elections by changing his vote, when all other people vote independently and uniformly at random.

What is the the "best" electoral system, i.e., which monotone function f maximizes the probability that

$$f(X_1, \ldots, X_R, \ldots, X_n) \neq f(X_1, \ldots, 1 - X_R, \ldots, X_n)$$

when $X_1, \ldots, X_n$ are independent uniform binary random variables and R is an independent uniform random variable with values in $\{1, \ldots, n\}$?

Let g be a particular optimal monotone function.

Lemma 0.1 *For any input with more ones than zeroes, g is 1.*

Proof. We use induction on the number k of zeroes in the input.

Base Case: $k = 0$. Then the input is all ones. This input has to make g be 1 or else g will always be 0 (since g is monotone), which clearly is not optimal.

Inductive Step: Consider an input with exactly $k + 1$ zeroes (with $2(k + 1) < n$ so it has more than $k + 1$ ones). Suppose for sake of contradiction that g is 0 on this input. By the IH, every input obtained from this one by changing a zero to a one is an input for which g is 1. This means that we can change the value of g to 1 on this input (without changing any other values of g) and still have a monotone function. By monotonicity, every input obtained from this one by changing a one to a zero is an input for which g is 0. Thus changing the value of g to a 1 on this particular input loses $2 \cdot (k + 1)$ variable flips that were counted in the probability in question but gains $2 \cdot (n - (k + 1))$ variable flips toward the probability in question. This strictly increases the probability, contradicting the fact that g maximized it. ∎

One can give a similar proof that g is 0 on any input with more zeroes than ones. This completely defines g if n is odd. If n is even, this defines g except on inputs with an equal number of zeroes and ones. An analysis similar to that used in proving the lemma shows that it doesn't matter what the value of g is on any of these inputs. So all functions that are 0 when there are more zeroes in the input than ones and are 1 when there are more ones in the input than zeroes are optimal.