Massachusetts Institute of Technology
6.042J/18.062J, Fall '02: Mathematics for Computer Science
Prof. Albert Meyer and Dr. Radhika Nagpal

Solutions to Problem Set 11-12

Problem 1. Prove that

(a) $\Pr\{A \mid B\} = \Pr\{A \mid \overline{B}\}$ iff A and B are independent.

Solution. *Proof.* ($\rightarrow$): Assuming $\Pr\{A \mid B\} = \Pr\{A \mid \overline{B}\}$,

$$\begin{aligned}
\Pr\{A\} &= \Pr\{A \mid B\}\Pr\{B\} + \Pr\{A \mid \overline{B}\}\Pr\{\overline{B}\} && \text{(Law of Total Probability)} \\
&= \Pr\{A \mid B\}\Pr\{B\} + \Pr\{A \mid B\}\Pr\{\overline{B}\} && \text{(assumption)} \\
&= \Pr\{A \mid B\}(\Pr\{B\} + \Pr\{\overline{B}\}) && \text{(distributivity)} \\
&= \Pr\{A \mid B\} \cdot 1 = \Pr\{A \mid B\} && \text{(Complement Rule),}
\end{aligned}$$

so A is independent of B by definition.

($\leftarrow$): Assuming A and B are independent, we know that A and $\overline{B}$ are also independent, so

$$\begin{aligned}
\Pr\{A \mid B\} &= \Pr\{A\} && \text{(independence of } A \text{ and } B\text{)} \\
&= \Pr\{A \mid \overline{B}\} && \text{(independence of } A \text{ and } \overline{B}\text{).}
\end{aligned}$$

For the record, here's the proof that A and $\overline{B}$ are independent:

$$\begin{aligned}
\Pr\{A \mid \overline{B}\} &= \frac{\Pr\{A \cap \overline{B}\}}{\Pr\{\overline{B}\}} && \text{(definition of conditional probability)} \\
&= \frac{\Pr\{A - B\}}{\Pr\{\overline{B}\}} && \text{(definition of set difference)} \\
&= \frac{\Pr\{A\} - \Pr\{A \cap B\}}{\Pr\{\overline{B}\}} && \text{(Difference Rule)} \\
&= \frac{\Pr\{A\} - \Pr\{A\}\Pr\{B\}}{\Pr\{\overline{B}\}} && \text{(independence of } A \text{ and } B\text{)} \\
&= \frac{\Pr\{A\}(1 - \Pr\{B\})}{\Pr\{\overline{B}\}} && \text{(distributivity)} \\
&= \frac{\Pr\{A\}\Pr\{\overline{B}\}}{\Pr\{\overline{B}\}} && \text{(Complement Rule)} \\
&= \Pr\{A\}.
\end{aligned}$$

□

■

(b) If A, B, C are mutually independent events, then A and $B \cup C$ are independent.

Solution. *Proof.* To show independence of the event A and the event $B \cup C$, it suffices to show that the probability of both events happening is equal the the product of the probabilities that each event happens. That is, we need only show that the

$$\Pr\{A \cap (B \cup C)\} = \Pr\{A\} \Pr\{B \cup C\} \tag{1}$$

But

$$\begin{aligned}
\Pr\{A \cap (B \cup C)\} &= \Pr\{(A \cap B) \cup (A \cap C)\} && \text{(distributivity of } \cap \text{ over } \cup) \\
&= \Pr\{A \cap B\} + \Pr\{A \cap C\} - \Pr\{(A \cap B) \cap (A \cap C)\} && \text{(Inclusion-exclusion)} \\
&= \Pr\{A \cap B\} + \Pr\{A \cap C\} - \Pr\{A \cap B \cap C\} && \text{(since } A \cap A = A) \\
&= \Pr\{A\} \Pr\{B\} + \Pr\{A\} \Pr\{C\} - \Pr\{A\} \Pr\{B\} \Pr\{C\} && \text{(mutual independence)} \\
&= \Pr\{A\} (\Pr\{B\} + \Pr\{C\} - \Pr\{B\} \Pr\{C\}) && \text{(distributivity of multiplication)} \\
&= \Pr\{A\} \Pr\{B \cup C\} && \text{(Inclusion-exclusion),}
\end{aligned}$$

which proves (1). □

■

Problem 2. There is a course—not 6.042, naturally—in which *10% of the assigned problems contain errors*. If you pick a random problem and send an email to your TA and your Lecturer asking whether the problem has an error, then *the TA's reply will be correct 80% of the time*. This 80% accuracy holds regardless of whether or not a problem has an error. Likewise, *the Lecturer's reply will be correct with only 75% accuracy*.[1]

Furthermore, the TA and Lecturers tend to be confused by different kinds of problems. The net result of this is that *the correctness of the lecturers' answer and the TA's answer are independent of each other, regardless of whether there is an error*.

For the following parts, designate events as follows:

T ::= "the TA's email says the problem has an error,"

L ::= "the Lecturer's email says the problem has an error."

E ::= "the problem has an error,"

[1]This is consistent with the view that Lecturers are chosen for their theatrical personalities, not their mastery of the material. : -)

(a) Using T, L and E translate the sentences in italics into probability notation. For example, "the correctness of the lecturers' answer and the TA's answer are independent of each other," could translate into some equations involving terms such as $\Pr\{T \mid E\}$ and $\Pr\{T \cap L \mid E\}$.

Solution. 10% of the assigned problems contain errors:

$$\Pr\{E\} = \frac{10}{100} = \frac{1}{10}. \tag{2}$$

the TA's reply will be correct 80% of the time ... regardless of whether or not a problem has an error:

$$\Pr\{T \mid E\} = \Pr\{\overline{T} \mid \overline{E}\} = \frac{80}{100} = \frac{4}{5}. \tag{3}$$

the Lecturer's reply will be correct with only 75% accuracy:

$$\Pr\{L \mid E\} = \Pr\{\overline{L} \mid \overline{E}\} = \frac{75}{100} = \frac{3}{4}. \tag{4}$$

the correctness of the lecturers' answer and the TA's answer are independent of each other, regardless of whether there is an error:

$$\Pr\{T \cap L \mid E\} = \Pr\{T \mid E\}\Pr\{L \mid E\} \tag{5}$$

$$\Pr\{T \cap L \mid \overline{E}\} = \Pr\{T \mid \overline{E}\}\Pr\{L \mid \overline{E}\}. \tag{6}$$

COMMENT FOR EDITING: Seems like a better translation than (6) would be

$$\Pr\{\overline{T} \cap \overline{L} \mid \overline{E}\} = \Pr\{\overline{T} \mid \overline{E}\}\Pr\{\overline{L} \mid \overline{E}\}.$$

I think they are equivalent, but I don't have time to check now – ARM 11/12/02 9:30PM. ■

(b) What is the probability that *both* the Lecturer and TA report an error?

Solution.

$$\begin{aligned}
\Pr\{L \cap T\} &= \Pr\{L \cap T \mid E\}\Pr\{E\} + \Pr\{L \cap T \mid \overline{E}\}\Pr\{\overline{E}\} && \text{(Law of Total Probability)}\\
&= \Pr\{L \mid E\}\Pr\{T \mid E\}\Pr\{E\} + \Pr\{L \mid \overline{E}\}\Pr\{T \mid \overline{E}\}\Pr\{\overline{E}\} && \text{(by (5) and (6))}\\
&= \frac{3}{4}\cdot\frac{4}{5}\cdot\frac{1}{10} + (1-\frac{3}{4})(1-\frac{4}{5})(1-\frac{1}{10}) = \frac{105}{1000} = 0.105. && \text{(by (4) and (3))}
\end{aligned}$$

■

(c) Is the event that "the TA says that there is an error", independent of the event that "the lecturer says that there is an error"? Prove it.

Solution.

$$\begin{aligned}
\Pr\{T\} &= \Pr\{T \mid E\}\Pr\{E\} + \Pr\{T \mid \overline{E}\}\Pr\{\overline{E}\} && \text{(Law of Total Probability)}\\
&= \Pr\{T \mid E\}\Pr\{E\} + (1 - \Pr\{\overline{T} \mid \overline{E}\})(1 - \Pr\{E\}) && \text{(Complement Rule)}\\
&= \Pr\{T \mid E\}\Pr\{E\} + (1 - \Pr\{T \mid E\})(1 - \Pr\{E\}) && \text{(by (3))}\\
&= \frac{4}{5}\frac{1}{10} + (1 - \frac{4}{5})(1 - \frac{1}{10}) = \frac{13}{50}, && \text{(by (3) and (2))}
\end{aligned}$$

A similar argument proves that

$$\Pr\{L\} = \Pr\{L \cap E\} + \Pr\{L \cap \overline{E}\} = \frac{3}{4}\frac{1}{10} + (1 - \frac{3}{4})(1 - \frac{1}{10}) = \frac{3}{10}. \tag{7}$$

Finally,

$$\Pr\{L\}\Pr\{T\} = \frac{3}{10} \cdot \frac{13}{50} = .078 > 0.105 = \Pr\{L \cap T\},$$

so L and T are not independent.

There is a simple intuitive argument supporting this conclusion. Suppose you receive an email from your TA that says the problem has an error. Then the odds that the Lecturer's email will now say that the problem has an error are higher than before the email. That is, $\Pr\{L \mid E\} > \Pr\{L\}$. The reason is that

- Because the TA is usually right, when the TA says that the problem has an error, the likelihood that there really is an error is increased.
- But the lecturer is also usually right, so increasing the likelihood of there *being* an error also increases the likelihood that the lecturer will *report* an error.

■

Problem 3. In the board game Monopoly, the number of squares that a player advances in a single turn is determined by up to three rolls of a pair of dice as follows.

- Initially, the player rolls two dice, sums the results, and advances that many squares.
- If the same number comes up on both dice—this is called *rolling a double*—then the player rolls the dice a second time, sums the results, and advances that many additional squares.
- If the player rolled on double on the second roll, then the player rolls the dice a third time, sums the results, and advances that many additional squares.
- However, if third roll was a double, then the player "goes to jail". We will treat this as though he had to reset his position to where he was just before the first roll, so the total number of squares advanced by the player in this case is zero.

Assume that the dice are fair and six-sided and that rolls are mutually independent.

We want to calculate the expected number of squares that a player advances in a single turn. Here is one approach to the problem. Lets assume that the pair of dice are *always* rolled three times, with the player simply ignoring results of any roll which was preceded by a non-double roll.

Let the random variable R_i be the sum of the two dice on the ith roll, and let I_i be an indicator for the event that the ith roll was a double.

(a) Let the random variable R, be the number of squares advanced by the player. Express R is terms of R_i and I_i.

Solution. We can express the random variable R, the number of squares advanced by the player, as follows:

$$R = R_1 + I_1 R_2 + I_1 I_2 R_3 - I_1 I_2 I_3 (R_1 + R_2 + R_3)$$

■

(b) Calculate $\mathrm{E}[R]$. You will almost surely use some assumptions about independence. Be sure to indicate where you do so.

Solution. Since the dice rolls are independent, we know that R_i's and I_j's at different rolls are independent. For example, knowing the value of the first roll has no effect on the probability that the second or third roll is a double. We can take expectations of both sides and reason as follows:

$$\begin{aligned}
\mathrm{E}[R] &= \mathrm{E}[R_1 + I_1 R_2 + I_1 I_2 R_3 - I_1 I_2 I_3 (R_1 + R_2 + R_3)] && \text{definition of R} \\
&= \mathrm{E}[R_1 + I_1 R_2 + I_1 I_2 R_3 - I_1 I_2 I_3 R_1 - I_1 I_2 I_3 R_2 - I_1 I_2 I_3 R_3] && \text{distributivity} \\
&= \mathrm{E}[R_1] + \mathrm{E}[I_1 R_2] + \mathrm{E}[I_1 I_2 R_3] && \text{linearity of expectation} \\
&\quad - \mathrm{E}[I_1 I_2 I_3 R_1] - \mathrm{E}[I_1 I_2 I_3 R_2] - \mathrm{E}[I_1 I_2 I_3 R_3] \\
&= \mathrm{E}[R_1] + \mathrm{E}[I_1] \cdot \mathrm{E}[R_2] + \mathrm{E}[I_1] \cdot \mathrm{E}[I_2] \cdot \mathrm{E}[R_3] \\
&\quad - \mathrm{E}[I_2] \cdot \mathrm{E}[I_3] \cdot \mathrm{E}[I_1 R_1] - \mathrm{E}[I_1] \cdot \mathrm{E}[I_3] \cdot \mathrm{E}[I_2 R_2] - \mathrm{E}[I_1] \cdot \mathrm{E}[I_2] \cdot \mathrm{E}[I_3 R_3]
\end{aligned}$$

The last step follows because mutual independence of the rolls implies that each the following sets of random variables are independent: $\{I_1, R_2\}$, $\{I_1, I_2, R_3\}$, $\{I_1, I_2, R_3\}$, $\{I_2, I_3, (I_1 R_1)\}$, $\{I_1, I_3, (I_2 R_2)\}$, and $\{I_1, I_2, (I_3 R_3)\}$.

Note that the expectation of R_i is 7, and the expectation of I_i is $1/6$. Finally, note that $\mathrm{E}[I_i R_i] = 7/6$ by direct calculation. Substituting into the formula above gives:

$$\mathrm{E}[R] \quad = \quad \frac{595}{72}.$$

■

Problem 4. Suppose successive digits from zero to nine are generated independently until the four digit sequence **9999** appears. What is the expected number of digits generated? (*Hint:* Parse the sequence into consecutive "tries to get 9999", where a "try" is either a single digit other than 9, or a 9 followed by a single digit other than 9, etc. For example, the sequence 92349982999769999 parses into eight tries: /92/3/4/998/2/9997/6/9999/.)

Solution. Solution: 11110.

By Wald's Theorem, the expected number is the mean time to a successful try, namely 10^4, times the expected length of a try. Notice that a try is of length four iff it begins with three nines, so the probability of a length four try is $1/10^3$. That means the expected length of a try is

$$1 \cdot \frac{9}{10} + 2 \cdot \frac{1}{10}\frac{9}{10} + 3 \cdot (\frac{1}{10})^2 \frac{9}{10} + 4 \cdot (\frac{1}{10})^3 = (900 + 180 + 27 + 4)/1000 = 1111/1000$$

Thus on average, we will have to read 11110 digits in order to see "9999". ■

Problem 5. Suppose that I choose a permutation of the numbers $1, 2, \ldots, n$ uniformly at random. What is the expected number of entries that are greater than all preceding entries? For example, in the permutation 4, 2, 1, 5, 3, the numbers 4 and 5 are greater than all preceding entries (Hint: What's the probability that the first entry is greater than all the preceding entries? What about the second one?)

Solution. We can solve this problem in a similar manner to the cool stickers problem from class. Let i_k be an indicator for the event that the k-th entry is greater than the preceding $k-1$ entries. The I_k are not independent, however independence is not required for using linearity of expectation

$$\begin{aligned}
\mathrm{E}\left[\#\text{ new maxima}\right] &= \sum_{k=1}^{n} \mathrm{E}(i_k) \\
&= \sum_{k=1}^{n} 1 \cdot \Pr(i_k = 1) + 0 \cdot \Pr(i_k = 0) \\
&= \sum_{k=1}^{n} \Pr(i_k = 1)
\end{aligned}$$

All that remains is to compute $\Pr(i_k = 1)$, which is the probability that the k-th entry is greater than all preceding entries. Consider a k-permutation selected from n elements. One of the elements in the permutation must be the largest, since all the elements are distinct. However by symmetry, the largest number is equally likely to be in the first or second or third ... k position. Therefore, $\Pr(i_k = 1) = \frac{1}{k}$.

Substituting this into the equation above gives:

$$
\begin{aligned}
\mathrm{E}\left[\#\text{ new maxima}\right] &= \sum_{k=1}^{n} \Pr(i_k = 1) \\
&= \sum_{k=1}^{n} \frac{1}{k} \\
&= H_n
\end{aligned}
$$

■

Problem 6. There are n MIT students who are taking 6.042 and 6.003 this term. To make it easier on themselves, the professors in charge of these classes have decided to randomly permute their class lists and then assign students grades based on their rank in the permutation[2]. Assume all permutations are equally likely and that the ranking in each class is independent of the other.

(a) What is the expected number of students that have a higher rank in 6.042 than 6.003?

Solution. Let X be a random variable whose value is the number of students whose ranks are higher in 6.042. Let X_i be a random variable whose value is 1 if student i has higher rank in 6.042 than 6.003 and 0 otherwise. By the Law of Total Probability, the probability that student i has higher rank in 6.042 is

$$
\Pr\{X_i = 1\} = \sum_{1 \le r \le n} \Pr\{\text{6.033 rank } < r \mid \text{6.042 rank = r}\} \tag{8}
$$

Since there are n possible rankings and $r-1$ rankings lower than r, the probability that student i has higher rank in 6.042 than in in 6.003 is $(r-1)/n$. The probability that student i has rank r in 6.042 is $1/n$. Thus, 8 becomes

$$
\Pr\{X_i = 1\} = \sum_{1 \le r \le n} \frac{1}{n}\frac{r-1}{n} = \frac{n-1}{2n}.
$$

Since $X = \sum_{1 \le i \le n} X_i$, we use linearity of expectation to compute that the expected number of students with a higher rank on 6.042 is rank on 6.042 is

$$
\mathrm{E}[X] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n] = \frac{n-1}{2}.
$$

Another way to solve this problem is to observe that students with higher rank in one class have lower rank in the other class, so by symmetry, the expected number of higher- and lower-ranked students in either class is the same. The expected number of students with the same rank equals 1 by linearity of expectation, since probability student ranked the same in both schools is $1/n$. These observations now yield same answer as above. ■

[2] . . . just as many students have long suspected : -)

(b) What is the expected number of students that have a ranking at least k higher in 6.042 than in 6.003?

Solution. This part can be done in the same way as the previous one. Let X be a random variable whose value is the number of students who are ranked at least k higher in 6.042 thank 6.003. Let X_i be the random variable whose value is 1 if student i is ranked at least k higher in 6.042 than 6.003 and 0 otherwise. By the Law of Total Probability, the probability that student i has at least k higher rank in 6.042 is

$$\Pr\{X_i = 1\} \quad = \sum_{k+1 \leq r \leq n} \Pr\{6.033 \text{ rank } \leq r-k \mid 6.042 \text{ rank = r}\} \tag{9}$$

COMMENT FOR EDITING: Why are we summing from $k+1$ to n rather than 1 to $n-k$? Does "higher rank" mean higher number or lower number?

(Notice that the $\Pr\{X_i = 1 \mid i \leq k\} = 0$.) Since there are n possible rankings and $r-k$ rankings no higher than $r-k$, the probability that student i has k higher rank in 6.042 than in in 6.003 is $(r-k)/n$. The probability that student i has rank r in 6.042 is $1/n$. Thus, 9 becomes

$$\Pr\{X_i = 1\} = \sum_{k+1 \leq r \leq n} \frac{1}{n}\frac{r-k}{n} = \frac{(n-k)(n-k+1)}{2n^2}.$$

Since $X = \sum_{1 \leq r \leq n} X_i$, we use linearity of expectation to compute that the expected number of students with a k higher rank on 6.042 is

$$\mathrm{E}[X] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n] = \frac{(n-k)(n-k+1)}{2n}.$$

■

Problem 7. **(a)** Suppose we flip a fair coin until two heads in a row come up. What is the expected number, F, of flips we perform?

Hint: Let F_{T} be the expected number of further flips until two heads comes up, given that the previous flip was T, and likewise let F_{H} be the expected number of further flips until two heads comes up given that the previous flip was H. Argue that F_{T} will equal 1 plus the average of F_{T} and F_{H}.

Solution. $F = F_{\mathrm{T}}$, since the initial condition has no chance of achieving two heads on the first flip and is thus the same as if the previous flip had been tails.

To calculate the expected number of flips, we take 1 for the first flip and then consider the two equally-likely outcomes: the first flip is tails, or the first flip is heads. The remaining flips we can expect are simply F_{T} and F_{H}, respectively. This gives us a total expectation of:

$$F_{\mathrm{T}} = 1 + \frac{1}{2}F_{\mathrm{T}} + \frac{1}{2}F_{\mathrm{H}}$$

If the previous flip was heads, we get a similar calculation. The biggest difference is that if the first flip is heads, we have no more flips beyond the current one (we are done).

$$F_{\mathrm{H}} = 1 + \frac{1}{2}F_{\mathrm{T}} + \frac{1}{2}(0)$$

Solving for F_{T} (by substitution) yields $F = 6$.

This can also be solved using a single equation formed by examining that cases for the first and second flips, but it is not as clean.

■

(b) Suppose we flip a fair coin until a head followed by a tail come up. What is the expected number, G, of flips we perform?

Solution. Clearly $G = G_{\mathrm{T}}$. As above

$$\begin{aligned} G_{\mathrm{T}} &= 1 + \frac{G_{\mathrm{T}} + G_{\mathrm{H}}}{2}, \\ G_{\mathrm{H}} &= 1 + \frac{0 + G_{\mathrm{H}}}{2}. \end{aligned}$$

So $G = G_{\mathrm{T}} = 4$. ■

(c) Suppose we now play a game: flip a fair coin until either HH or HT first occurs. You win if HT comes up first, lose if HH comes up first. What odds should you offer an opponent to make this a fair game?

Solution. Even money odds. Although the expected time for HH is larger than for HT, the game of waiting for one or the other to come up first is fair!

To prove this, let W be the probability of winning. Clearly,

$$W = (1/2)\Pr\{W \mid \text{first roll is } T\} + (1/2)\Pr\{W \mid \text{first roll is } H\} = (1/2)W + (1/2)(1/2)$$

so $W = 1/2$.

By the way, an even more dramatic contrast comes up if we consider longer sequences, where the sequence with the longer waiting time may be *more* likely to appear first. For example, the expected waiting time for HHH is 14, while the expected waiting time for TTHH is 16, as can be shown by reasoning about conditional expectations as in the previous problem parts. However, the probability that HHH appears before TTHH is only 1/3, which can be shown by reasoning about conditional probabilities as above (see also, Friday Week 10, Class Problem 3).

This surprising result reflects the unexpected `:-)` technical properties of expectation. A simple intuitive explanation of this phenomenon is that although one pattern may usually occur before a second one, it may be that when the second pattern does appear earlier, the time for the first pattern to appear later is very large, so that overall, the expected time for the more likely first pattern is much larger.

A simpler, contrived example crystallizes this idea: consider an experiment which begins by flipping a coin biased 3 to 1 in favor of Tails. If a Head comes up, then continue flipping a two-Headed coin forever. If a Tail comes up, then flip a two-Tailed coin once, and then continue flipping a two-Headed coin forever.

So an occurrence of TT is 3 times as likely to appear before HH than vice-versa, because TT appears before HH precisely when the first flip comes up Tails. But HH is guaranteed to appear in at most 4 flips in either case, so its expectation is obviously at most 4 (actually it's $(1/4)2 + (3/4)4 = 3.5$).■

On the other hand, when the first flip comes up Heads, TT *never* appears. Since this happens with probability $1/4$, the expected time for TT to appear is infinite!