CSE 312: Foundations of Computing II
Additional Midterm Review Problems (solutions)

Counting and Basic Probability (including Pigeonhole, Inclusion-Exclusion, etc)

1. (This problem is similar to a problem from HW1, except that "=" has been replaced by "≤".)

   Consider the following inequality: $a_1+a_2+a_3+a_4+a_5+a_6 \leq 70$. A solution to this inequality over the nonnegative integers is a choice of a nonnegative integer for each of the 6 variables $a_1, a_2, a_3, a_4, a_5, a_6$ that satisfies the inequality. To be different, two solutions have to differ on the value assigned to some a_i. How many different solutions are there to the inequality?

   This is equivalent to asking how many different solutions are there to the equation $a_1 + a_2 + a_3 + a_4 + a_5 + a_6 + a_7 = 70$. The answer is $\binom{76}{6}$, using the method from HW1.

2. Given any set of 18 integers, show that one may always choose two of them so that their difference is divisible by 17.

   By the pigeonhole principle, two of them, say x and y, must have the same remainder when divided by 17. That means $x \equiv y \pmod{17}$, which in turn means 17 divides $x - y$.

3. The space shuttle has 6 O-rings: these were involved in the Challenger disaster. When the space shuttle is launched, each O-ring has a probability of failure of 0.0137, independent of whether other O-rings fail.

   (a) What is the probability that, during 23 launches, no O-ring will fail, but that at least one O-ring will fail during the 24th launch?

   The probability that no O-ring fails on a single launch is $(1 - 0.0137)^6 \approx 0.921$. The probability that this happen for 23 launches and doesn't happen on the 24th launch is $0.921^{23}(1 - 0.921) \approx 0.0118$.

   (b) What is the probability that no O-ring fails during 24 launches?

   $0.921^{24} \approx 0.137$

4. Suppose you record the birthdays of a large group of people, one at a time, until you have found a person whose birthday matches your own birthday. What is the probability that it takes exactly 20 people for this to occur? Assume that there are 365 possible birthdays and each one is equally probable for a randomly chosen person.

   This is a geometric distribution:

$$\left(\frac{364}{365}\right)^{19} \frac{1}{365} \approx 0.0026$$

5. Two fair 6-sided dice are thrown n times in succession. Compute the probability that double 6 (i.e., 6 on each die) appears at least once in the n throws.

The probability that double 6 does not occur is $\left(\frac{35}{36}\right)^n$, so the probability that it occurs at least once is $1 - \left(\frac{35}{36}\right)^n$.

How large need n be to make this probability at least 1/2?

Solving $1 - \left(\frac{35}{36}\right)^n \geq \frac{1}{2}$ gives the solution $n \geq 24.6$. Since n must be an integer, $n \geq 25$.

6. For $n > 1$, let $a_1, ..., a_n \in [0, 1]$. Show that there exist numbers $x_1, ..., x_n \in \{-1, 0, 1\}$ not all zero such that $\left|\sum_{i=1}^n a_i x_i\right| \leq \dfrac{n}{2^n - 2}$.

First realize that, by the pigeonhole principle, for an interval of length ℓ, if you put k points in them, there must be two such that the distance between them is $\leq \dfrac{\ell}{k-1}$. To see this, split the interval $[0, \ell]$ into $k-1$ equal length subintervals: $\left[0, \dfrac{\ell}{k-1}\right], \left[\dfrac{\ell}{k-1}, \dfrac{2\ell}{k-1}\right], ..., \left[\dfrac{(k-2)\ell}{k-1}, 1\right]$, and by the pigeonhole principle, two of the k points must belong to the same interval and hence have distance between them $\leq \dfrac{\ell}{k-1}$. For each of the $2^n - 1$ nonempty subsets $I \subseteq [n]$, consider the sum $S_I = \sum_{i \in I} a_i$. Notice that $0 \leq S_I \leq n$ since each $0 \leq a_i \leq 1$. Thus, there exist nonempty subsets I and J of $[n]$ with $I \neq J$ such that $|S_I - S_J| \leq \dfrac{n}{2^n - 2}$, by applying the previous result with $\ell = n, k = 2^n - 1$. Note that $|S_I - S_J| = \left|\sum_{i \in I} a_i - \sum_{j \in J} a_j\right|$. For each $k = 1, ..., n$, define $x_k = 1$ if $a_k \in I \setminus J$, $x_k = 0$ if $a_k \in I \cap J$, and $x_k = -1$ if $a_k \in J \setminus I$. Then $|S_I - S_J| = \left|\sum_{i \in I} a_i - \sum_{j \in J} a_j\right| = \left|\sum_{i=1}^n a_i x_i\right| \leq \dfrac{n}{2^n - 2}$.

7. The probability that a customer pays with cash is 40%, independent of other customers. Find the probability that the 12th customer to arrive at the cashier is the 8th one that pays with cash.

$$\binom{11}{7}(0.4)^7(0.6)^4(0.4) \approx 0.028$$

Conditional Probability and Bayes Rule

8. You are trying to diagnose the probability that a patient with a positive blood sugar test result has diabetes, even though she is in a low risk group. The probability of a woman in this group having diabetes is 0.8%. 90% of women with diabetes will test positive in the blood sugar test. 7% of women without diabetes will test positive in the blood sugar test. Your patient tests positive in the blood sugar test. What is the probability that she has diabetes?

Let D be the event that she has diabetes and $+$ be the event of a positive test.

$$\mathbb{P}(D \mid +) = \frac{\mathbb{P}(+ \mid D)\mathbb{P}(D)}{\mathbb{P}(+ \mid D)\mathbb{P}(D) + \mathbb{P}(+ \mid \overline{D})\mathbb{P}(\overline{D})} = \frac{0.9 \times 0.008}{0.9 \times 0.008 + 0.07 \times 0.992} \approx 0.09$$

Notice that the posterior probability 0.09 of diabetes is approximately 10 times as great as the prior probability 0.008 of diabetes, but still small.

9. A very long multiple choice exam has 4 choices for each question. Charlie has studied enough so that he knows the correct answer for 1/2 of the questions; for an additional 1/4 of the questions he can eliminate one choice and chooses randomly and uniformly among the other three, and for the remaining 1/4 of the questions he chooses randomly and uniformly among all four answers.

   As the teacher, you want to determine how many answers the student actually knows. For a randomly chosen question, if Charlie answers it correctly, what is the probability he knew the answer?

   Let C be the event that Charlie has the correct answer and K be the event that Charlie knew the answer. Then

$$\begin{aligned} \mathbb{P}(C) &= \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot\frac{1}{3} + \frac{1}{4}\cdot\frac{1}{4} = \frac{31}{48} \\ \mathbb{P}(K \mid C) &= \frac{\mathbb{P}(C \mid K)\mathbb{P}(K)}{\mathbb{P}(C)} = \frac{1\cdot\frac{1}{2}}{31/48} = \frac{24}{31} \end{aligned}$$

Random Variables and Expectation

10. Let X be the outcome of rolling a fair 6-sided die once. Let Y be the sum of the outcomes of rolling the same die n times independently.

    (a) Compute $\mathbb{E}[X]$.

$$\mathbb{E}[X] = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2}$$

    (b) Compute $\mathrm{Var}(X)$ and the standard deviation σ of X.

$$\begin{aligned} \mathbb{E}[X^2] &= \frac{1}{6}(1+4+9+16+25+36) = \frac{91}{6} \\ \mathrm{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \\ \sigma &= \sqrt{\frac{35}{12}} \approx 1.7 \end{aligned}$$

    (c) Compute $\mathbb{E}[Y]$.

    For $1 \le i \le n$, let X_i be the outcome of the i-th die roll.

$$\mathbb{E}[Y] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n}\mathbb{E}[X_i] = \sum_{i=1}^{n}\frac{7}{2} = \frac{7}{2}n$$

(d) Compute Var(Y).

Because $X_1, X_2, \ldots, X_n$ are independent,

$$\text{Var}(Y) = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}(X_i) = \sum_{i=1}^{n} \frac{35}{12} = \frac{35}{12}n$$

11. Suppose we have r independent random variables $X_1, \ldots, X_r$ that each represent the number of coins flipped up to and including the first head, where $\mathbb{P}(\text{head}) = p$. Recall that each X_i has probability mass function,

$$p_{X_i}(k) = \mathbb{P}(X_i = k) = (1-p)^{k-1} p$$

(a) What do you think $\mathbb{E}[X_i]$ should be (without calculations) if $p = \frac{1}{2}$? If $p = \frac{1}{3}$? In the general case? (Proof in lecture soon.)

They should be 2, 3, and $\frac{1}{p}$ in general.

(b) Suppose we define $X = X_1 + \ldots + X_r$. What does X represent, in English words? (Hint: think of performing each "trial" one after the other.)

The number of coins flipped up to and including the r^{th} head.

(c) What is Ω_X? Find $p_X(k)$, the probability mass function for X.

$$\Omega_X = \{r, r+1, \ldots\}$$

$$p_X(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1, \ldots$$

(d) Find $\mathbb{E}[X]$ using linearity of expectation.

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{r} X_i\right] = \sum_{i=1}^{r} \mathbb{E}[X_i] = \sum_{i=1}^{r} \frac{1}{p} = \frac{r}{p}$$

12. Let the random variable X be the number of heads in n independent flips of a fair coin.

(a) What is the probability mass function of X?

$$p_X(i) = \binom{n}{i} 2^{-n} \text{ for } i \in \{0, 1, \ldots, n\}$$

(b) Find $\mathbb{E}[X]$ directly by applying the definition of expectation to the result from part (a).

Hint: prove and use the identity $i\binom{n}{i} = n\binom{n-1}{i-1}$.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{n} i p_X(i) \\
&= \sum_{i=0}^{n} i \binom{n}{i} 2^{-n} \\
&= \sum_{i=1}^{n} i \binom{n}{i} 2^{-n} \\
&= 2^{-n} \sum_{i=1}^{n} i \cdot \frac{n!}{i!(n-i)!} \\
&= 2^{-n} \sum_{i=1}^{n} \frac{n!}{(i-1)!(n-i)!} \\
&= 2^{-n} \sum_{i=1}^{n} n \cdot \frac{(n-1)!}{(i-1)!((n-1)-(i-1))!} \\
&= n \cdot 2^{-n} \sum_{i=1}^{n} \binom{n-1}{i-1} \\
&= n \cdot 2^{-n} \sum_{j=0}^{n-1} \binom{n-1}{j} \\
&= n \cdot 2^{-n} \cdot 2^{n-1} \\
&= n/2
\end{aligned}$$

13. At a reception, n people give their hats to a hat-check person. When they leave, the hat-check person gives each of them a hat chosen at random from the hats that remain. What is the expected number of people who get their own hats back? (This is closely related to, but much simpler than, the challenge problem from the worksheet from quiz section #2. Notice that the hats returned to two people are not independent events: if a certain hat is returned to one person, it cannot also be returned to the other person.)

For $1 \leq i \leq n$, let

$$\begin{aligned}
X_i &= \begin{cases} 1, & \text{if } i\text{-th person gets own hat back} \\ 0, & \text{otherwise} \end{cases} \\
\mathbb{E}[X] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i] \\
&= \sum_{i=1}^{n} (1 \cdot \mathbb{P}(X_i = 1) + 0 \cdot \mathbb{P}(X_i = 0)) \\
&= \sum_{i=1}^{n} \mathbb{P}(i\text{-th person gets own hat back}) \\
&= \sum_{i=1}^{n} \frac{1}{n} = 1
\end{aligned}$$

14. (This exercise is the same as #1 on the section worksheet, but with an ordinary 6-sided die rather than a 3-sided die.) Let the random variable X be the sum of two independent rolls of a fair 6-sided die.

(a) What is the probability mass function of X?

$$p(i) = \begin{cases} (i-1)/36 \text{ for } i \in \{2,3,4,5,6,7\} \\ (13-i)/36 \text{ for } i \in \{8,9,10,11,12\} \end{cases}$$

(b) Find $\mathbb{E}[X]$ directly by applying the definition of expectation to the result from part (a).

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{i=2}^{12} i p_X(i) \\
&= \sum_{i=2}^{7} i(i-1)/36 + \sum_{i=8}^{12} i(13-i)/36 \\
&= 7 \cdot 6/36 + \sum_{i=2}^{6} (i + 14 - i)(i-1)/36 \\
&= 7/6 + \sum_{j=1}^{5} 14j/36 \\
&= 7/6 + (14/36) \sum_{j=1}^{5} j \\
&= 7/6 + 15 \cdot 14/36 \\
&= 7/6 + 35/6 \\
&= 7
\end{aligned}
$$

(c) Find $\mathbb{E}[X]$ again using linearity of expectation.

Let Y be the result of the first roll and Z be the result of the second.

$$\mathbb{E}[X] = \mathbb{E}[Y + Z] = \mathbb{E}[Y] + \mathbb{E}[Z] = 2\mathbb{E}[Y] = 2 \cdot \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 2 \cdot \frac{7}{2} = 7$$

(d) Check that your answers to parts (b) and (c) are the same. Which way of computing the expectation was simpler, (a)+(b), or (c)?

15. You roll three fair dice, each with a different numbers of faces: die 1 has six faces (numbered 1 … 6), die 2 has eight faces (numbered 1 … 8), and die 3 has twelve faces (numbered 1 … 12). Let the random variable X be the sum of the three values rolled. What is $\mathbb{E}[X]$?

Let D_1, D_2, D_3 be the values of die 1, die 2, and die 3, respectively. $\mathbb{E}[D_1] = 3.5$, $\mathbb{E}[D_2] = 4.5$, and $\mathbb{E}[D_3] = 6.5$. Therefore, $\mathbb{E}[X] = \mathbb{E}[D_1 + D_2 + D_3] = \mathbb{E}[D_1] + \mathbb{E}[D_2] + \mathbb{E}[D_3] = 3.5 + 4.5 + 6.5 = 14.5$.