Probability

1. Let X be a random variable equal to the sum of two 6-sided die rolls.

   (a) Make a table where the rows correspond to the value of the first roll, and the columns to the value of the second. Then, the entries of the table will be the 36 possible outcomes. In each entry, write the corresponding value of X.

   Solution: Should look something like this.

	1	*2*	*3*	*4*	*5*	*6*
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

   (b) Assuming our die is fair, what is the most likely value of X?

   Solution: This is simply the value which appears most often in the table, so 7.

   (c) Now highlight (or outline, or otherwise delineate) the entire fourth row and the entire fourth column of this table. If you know that one of the rolls was a 4, what is the likelihood that $X = 7$? What about the likelihood that $X = 8$?

   Solution: As below:

	1	*2*	*3*	*4*	*5*	*6*
1	2	3	4	**5**	6	7
2	3	4	5	**6**	7	8
3	4	5	6	**7**	8	9
4	**5**	**6**	**7**	**8**	**9**	**10**
5	6	7	8	**9**	10	11
6	7	8	9	**10**	11	12

   There are 11 entries highlighted, 2 of which are 7's, 1 of which is an 8. So $P(X = 7) = \frac{2}{11}$, and $P(X = 8) = \frac{1}{11}$.

   (d) What is the probability distribution for X? You may write a formula, if you can think of one, or else simply list out the probability for each value of X.

A formula is not necessary, but if you wish one, you could express this as $P(X = n) = \frac{6-|7-n|}{36}$ for $n \in 2, 3, \ldots, 12$. Intuitively, the number of outcomes where $X = n$ is represented on the numerator. When $n = 7$, there are 6 outcomes, and for any other value of n, we will have fewer. How much fewer? Well, the number fewer turns out to be equal to the (positive) distance from n to 7, hence why we subtract $|7 - n|$.

2. Suppose I have a function p which is defined for every positive integer, such that $p(n) = (\frac{1}{2})^n$. Is p a probability distribution? If no, why not? If yes, what random event might p describe?

Solution: Yes! The two conditions are that $0 \le p(n) \le 1$ for all possible values of n, which is the case here. As for the other, infinite series are probably a little rusty, but you may recall that the value of an infinite sum is the limit of the partial sums. That is to say, $p(1) = \frac{1}{2}$, $p(1) + p(2) = \frac{3}{4}$, $p(1) + p(2) + p(3) = \frac{7}{8}$, and in general $\sum_{i=1}^{n} p(i) = \frac{2^n - 1}{2^n}$. So $\sum_{i=1}^{\infty} p(i) = \lim_{n\to\infty} \sum_{i=1}^{n} p(i) = \lim_{n\to\infty} \frac{2^n-1}{2^n} = 1$, as desired.

As for what random event this might describe - a perfect example is on the previous discussion! Namely, this probability distribution perfectly describes the scenario where we flip a coin until we get heads. There's a $\frac{1}{2}$ chance we get heads on our first flip, a $\frac{1}{4}$ chance it takes two flips, a $\frac{1}{8}$ chance it takes three, and so on.

3. One might imagine partitioning the tiles of an 8x8 chessboard into four concentric squares. Suppose that you (uniformly) randomly select a square, and then (uniformly) randomly select a tile within that square. Are all tiles equally likely to be selected? Why or why not?

Solution: We might begin with a depiction of the board in question.

4	4	4	4	4	4	4	4
4	3	3	3	3	3	3	4
4	3	2	2	2	2	3	4
4	3	2	1	1	2	3	4
4	3	2	1	1	2	3	4
4	3	2	2	2	2	3	4
4	3	3	3	3	3	3	4
4	4	4	4	4	4	4	4

It turns out that when we subdivide the tiles in this manner, not all tiles are equally likely to be chosen. For instance, the likelihood of selecting a tile in group 1 is $\frac{1}{16}$, since we have a 1 in 4 chance of choosing that group, and then a 1 in 4 chance of choosing any given tile therein. On the other hand, the outer ring has 28 tiles, giving its tiles each a $\frac{1}{4} \cdot \frac{1}{28} = \frac{1}{112}$ likelihood of being picked.

4. If you draw five cards from a standard 52-card deck, what is the likelihood of obtaining:
   (a) Four of a kind?

   Solution: Note that all hands (i.e. ignoring the order in which you draw cards) are equally likely here, thus we can either treat outcomes as hands or outcomes as draws, as long as we are consistent. When in doubt, for other similar problems, it is generally better to assume order matters; thus for the sake of promoting "good form" we will do so here.

   If we are to construct a hand with four of a kind, we would first choose which value gets drawn. There are 13 options here. Then, having ruled out 4 cards in doing so, we now have 48 options for the last card. Factor in the 5! different orderings in which we can draw any given hand, and there are $13 \cdot 4 \cdot 5!$ ways to draw a four-of-a-kind. On the other hand, the total number of draws we could have made is $52 \cdot 51 \cdot 50 \cdot 49 \cdot 48 = P(52,5)$. Thus, our probability is $\frac{13 \cdot 4 \cdot 5!}{P(52,5)}$, or about 0.024%.

   (b) A full house?

   Recall that a full house is a three-of-a-kind plus a pair. Selecting one value to be the triplet and one to be the pair gives $13 \cdot 12$ possibilities. Then, we must choose which of the five cards we draw will come from the pair; here, there are $C(5,2)$ options. Lastly, we must choose the particular ordering in which we draw our triplet and pair, giving $P(4,3)$ and $P(4,2)$ possibilities, respectively. So the probability is the unwieldy $\frac{13 \cdot 12 \cdot C(5,2) \cdot P(4,3) \cdot P(4,2)}{P(52,5)}$, or approximately 0.14%. (It is perhaps worth noting that all answers here will have the same denominator, since the number of total 5-card hands you can draw never changes.)

   (c) At least a pair?

   This is best accomplished with complementary counting. In order to fully avoid getting a pair, we would have to draw five cards of different values. All 52 cards are fine to draw on our first draw, but we cannot draw any cards of the same value on the second, lest we create a pair. Thus, there are only 48 possibilities on the second draw, and by similar logic 44 on the third, 40 on the fourth, and 36 on the fifth. So the probability is $1 - \frac{52 \cdot 48 \cdot 44 \cdot 40 \cdot 36}{P(52,5)}$, or approximately 49.29%.

(d) Only a pair?

First, decide which value gives us our pair: here, there are 13 possibilities. Next, which two of our five draws will be the pair? This gives us an additional $C(5,2)$ possibilities. Isolating the pair cards now gives $P(4,2)$ possible suit orderings for our pair. Lastly, we want to avoid making a pair with our remaining three cards. Thus, the first card of these only has 48 possibilities (to avoid making a triplet with our pair), the second has 44, and the third has 40. So our likelihood is $\frac{13\cdot C(5,2)\cdot P(4,2)\cdot 48\cdot 44\cdot 40}{P(52,5)}$, or aproximately 42.25%. Shockingly close to the above, which I suppose hammers home just how unlikely you are to get anything other than a pair (approximately 7%).

(e) Both a pair and a flush?

This is impossible: in a flush, all cards are of the same suit, but in a pair, two cards have the same value. But there are no cards with both the same value and the same suit. So the probability is simply 0%.

(Recall that, if all outcomes are equally likely, then the likelihood of an event E is $P(E) = \frac{|E|}{|S|}$, the number of outcomes in E over the total number of outcomes.)

5. Suppose that I (uniformly) randomly choose a number from 0 to 10, and then (uniformly) randomly choose a number from 0 to that number. I write this second number on a slip of paper and hand it to you. What number do you think you are most likely to see? What is the probability distribution of the value I write down?

These problems roughly got harder as they went along, so perhaps I ought to have marked this optional. The insight, though, comes from problems 1 and 3: we would like to do something similar to problem 1, and make perhaps a table, but just as in problem 3, not all tiles would be equally likely. This poses a challenge. There are a couple workarounds, but maybe the easiest is just to draw the table as follows:

<table>
<tr><td colspan="60">0</td></tr>
<tr><td colspan="30">0</td><td colspan="30">1</td></tr>
<tr><td colspan="20">0</td><td colspan="20">1</td><td colspan="20">2</td></tr>
<tr><td colspan="15">0</td><td colspan="15">1</td><td colspan="15">2</td><td colspan="15">3</td></tr>
<tr><td colspan="12">0</td><td colspan="12">1</td><td colspan="12">2</td><td colspan="12">3</td><td colspan="12">4</td></tr>
</table>

$\vdots$

This continues down for some 11 rows that I'm too lazy to draw. In essence, this is the table we desire, but rescaled so that area correlates with likelihood: each row corresponds to a different choice of first number, and each column within that row correlates to a different choice of second number. To find the likelihood of a given outcome, we simply add up the area that all the corresponding entries take up. So 0 takes up all of the first row, half the second, a third of the third, and so on, for a total of $1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{11}$ rows. Each row is $\frac{1}{11}$ the area of the table so $P(0) = \frac{1}{11}(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{11})$. By similar logic, $P(1) = \frac{1}{11}(\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{11})$, and in general, $P(n) = \frac{1}{11}\sum_{k=n+1}^{11}\frac{1}{k}$ for $n \in \{0, 1, \ldots, 10\}$.