Homework 1

Due Tue, Oct 10

1. Random variables, sample space and events
Consider the random experiment of flipping a coin 4 times.

(a) Define the appropriate random variables.
(b) List all the outcomes in the sample space Ω. How many are there?
(c) Let A be the event that the first flip is a Heads. List all the outcomes in A. How many are there?
(d) Let B be the event that the third flip is a Heads. List all the outcomes in B. How many are there?
(e) Let C be the event that the first flip and the third flip are both Heads. List all the outcomes in C. How many are there?
(f) Let D be the event that the first flip or the third flip is a Heads. List all the outcomes in D. How many are there?
(g) Are the events A and B disjoint? Express the event C in terms of A and B. Express the event D in terms of A and B.
(h) Suppose now the coin is flipped $n \geq 3$ times instead of 4 flips. Define the random variables, compute $|\Omega|, |A|, |B|, |C|, |D|.$

2. More probability models
Suppose you have two coins, one is biased with a probability of $p$ coming up Heads, and one is biased with a probability of $q$ coming up Heads. Answer the questions below, but you don’t need to provide justifications.

(a) Suppose $p = 1$ and $q = 0$.
   (a) You pick one of the two coins randomly and flip it. You repeat this process $n$ times, each time randomly picking one of the two coins and then flipping it. Let $X_i$ be the result of the $i$th flip for $i = 1, \ldots, n$. Describe the sample space and give a reasonable probability assignment to model the situation.
   (b) Now you pick one of the two coins randomly, but flip the same coin $n$ times. Again let $X_i$ be the result of the $i$th flip. Identify the sample space for this experiment together with a reasonable probability assignment to model the situation. Is your answer the same as in the previous part?
   (b) Repeat the above two questions for arbitrary values of $p$ and $q$. Express your answers in terms of $p$ and $q$.

3. DNA Sequencing
The DNA of a bacteria is a length 1 million circular sequence containing the nucleotides A, G, C and T. We sequence the DNA by sampling 100,000 reads randomly from the DNA, where each read is a length 100 substring of the genome. We are interested in particular in a specific gene, which is a specific length 1000 substring of the DNA.

(a) What is the probability that the first two reads overlap?
(b) What is the probability that a specific nucleotide of the DNA is covered by at least one read?
(c) What is the probability that there is at least one read that is contained in the gene of interest?

Solve these problems by systematically defining a probability model as we did in class.
4. A new game

You have two quarters and a table with a row of squares marked like this:

<p>| | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

Before the game begins, you get to place each quarter on one square. You can put either both quarters on the same square, or you can put them on two different squares: your choice.

Then, you roll two fair dice, sum up the numbers showing on the dice to get a number from 2–12, and if there’s a quarter on the square labelled with that number, remove it from the table. (If there are two quarters on that square, remove only one of them.) Now roll the two fair dice a second time, again getting a number from 2–12, and again removing a single quarter from the square with that number, if there’s a quarter there. At this point, the game is over. If you removed both quarters, you win; if any quarter remains on the table, you lose.

(a) What’s the probability of winning, if you put two quarters on the square labelled 5?

(b) What’s your best strategy? In other words, what’s the best place to put your two quarters, if you want to maximize the probability of winning? State where you should put your two quarters. Then, calculate the probability that you win, if you put your two quarters there.

5. Monty Hall Again

In the three-door Monty Hall problem, there are two stages to the decision, the initial pick followed by the decision to stick with it or switch to the only other remaining alternative after the host has shown an incorrect door. An extension of the basic problem to multiple stages goes as follow.

Suppose there are four doors, one of which is a winner. The host says: “You point to one of the doors, and then I will open one of the other non-winners. Then you decide whether to stick with your original pick or switch to one of the remaining doors. Then I will open another (other than the current pick) non-winner. You will then make your final decision by sticking with the door picked on the previous decision or by switching to the only other remaining door.”

(a) How many possible strategies are there?

(b) For each of the possible strategies, calculate the probability of winning. Proceed by systematically defining a probability model as we did in class. What is the best strategy?