# Week 5

**Exercise 52** Suppose that we have $m$ different (disjoint) random variables $X_i$, where each has probability $p_i$ of occurring. Let $n$ be the number of attempts made and let $k_i$ be the number of times that the $i$th outcome happened out of the $n$ attempts. For $i < j$:

(1) What is the distribution of $X_i$?

(2) What is the distribution of $X_i + X_j$?

(3) What is the joint distribution of $X_i$, $X_j$, and $n - X_i - X_j$?

(Just give a formula for all parts.)
*Answer.* (1) This is basically asking compare $i$ with everything else. So there's a $p_i$ chance of success and $1 - p_i$ chance of failure. This is basically the binomial distribution. Since we know $i$ occurs $k_i$ times with probability $p_i$ we have:

$$P(X_i = k_i) = \binom{n}{k_i} (p_i)^{k_i} (1 - p_i)^{n-k_i}$$

(2) Notice that $X_i + X_j$ is just telling us when $X_i$ or $X_j$ occurs. This is given by the probability $p_i + p_j$ since they are disjoint. The chance of this not happening is given by $1 - p_i - p_j$, giving us the binomial distribution:

$$P(X_i + X_j = k_i + k_j) = \binom{n}{k_i + k_j} (p_i + p_j)^{k_i+k_j} (1 - p_i - p_j)^{n-k_i-k_j}$$

(3) This final one is just a multinomial distribution:

$$\binom{n}{k_i, k_j, (n - k_i - k_j)} (p_i)^{k_i} (p_j)^{k_j} (1 - p_i - p_j)^{1-k_i-k_j}$$

□

**Exercise 53** Let $X$ and $Y$ be two independent random variables which are each uniformly distributed on $\{1, 2, \dots, n\}$. Find:

(1) $P(X = Y)$

(2) $P(X < Y)$ and $P(X > Y)$

(3) $P(\max(X, Y) = k)$ for $1 \le k \le n$

(4) $P(\min(X, Y) = k)$ for $1 \le k \le n$

(5) $P(X + Y = k)$ for $2 \le k \le 2n$

*Answer.* We note that since $X$ and $Y$ are uniformly distributed, we know their probability for success is given by $\frac{1}{n}$. Additionally, since they're independent, we know $P(X = x, Y = y) = P(X = x) \cdot P(Y = y)$.

(1) We have:

$$P(X = Y) = \sum_{i=1}^{n} P(X = i, Y = i) = \sum_{i=1}^{n} P(X = i) \cdot P(Y = i) = n \cdot \frac{1}{n^2} = \frac{1}{n}$$

(2) By symmetry we have $P(X < Y) = P(X > Y)$.

$$1 = P(X > Y) + P(X < Y) + P(X = Y) = 2P(X < Y) + \frac{1}{n}$$

Therefore, $P(X < Y) = \frac{1-\frac{1}{n}}{2} = \frac{n-1}{2n}$

(3) Using partitions, we have:

$$\begin{aligned}
P(\max(X,Y)=k) &= P(X=k, Y<k) + P(X<k, Y=k) + P(X=k, Y=k) \\
&= P(X=k)P(Y<k) + P(X<k)P(Y=k) + P(X=k)P(Y=k) \\
&= \frac{1}{n}\cdot\frac{k-1}{n} + \frac{k-1}{n}\cdot\frac{1}{n} + \frac{1}{n}\cdot\frac{1}{n} \\
&= \frac{2(k-1)+1}{n^2} \\
&= \frac{2k-1}{n^2}
\end{aligned}$$

(4) Using partitions, we have:

$$\begin{aligned}
P(\min(X,Y)=k) &= P(X=k, Y>k) + P(X>k, Y=k) + P(X=k, Y=k) \\
&= P(X=k)P(Y>k) + P(X>k)P(Y=k) + P(X=k)P(Y=k) \\
&= \frac{1}{n}\cdot\frac{n-k}{n} + \frac{n-k}{n}\cdot\frac{1}{n} + \frac{1}{n}\cdot\frac{1}{n} \\
&= \frac{2(n-k)+1}{n^2} \\
&= \frac{2n-2k+1}{n^2}
\end{aligned}$$

Another way to do this is to notice that

$$P(\min(X,Y)=k) = P(\max(n+1-X, n+1-Y) = n-k+1) = \frac{2(n-k+1)-1}{n^2} = \frac{2n-2k+1}{n^2}$$

(5) Finally, we look at $P(X+Y=k)$. First thing we need to note is that both $X$ and $Y$ are symmetric about $\frac{n+1}{2}$. So $X+Y$ is symmetric about $2\cdot\frac{n+1}{2} = n+1$. So we first look at the first half.

Let $k$ be between $1+1=2$ and $n+1$, *i.e.*, $2 \le k \le n+1$. Then

$$P(X+Y=k) = \sum_{i=1}^{k-1} P(X=i)P(Y=k-i) = (k-1)\cdot\frac{1}{n}\frac{1}{n} = \frac{k-1}{n^2}$$

We can now use symmetry to solve when $n+1 \le k \le 2n$:

$$\begin{aligned}
P(X+Y=k) &= P\left(X+Y=(n+1)+(k-(n+1))\right) \\
&= P\left(X+Y=(n+1)-(k-(n+1))\right) \\
&= P\left(X+Y=2(n+1)-k\right) \\
&= \frac{2(n+1)-k-1}{n^2} \\
&= \frac{2n-k+1}{n^2}
\end{aligned}$$

□

**Exercise 54** Suppose that you throw three (fair six-sided) dice and look at the sum. What are the chances that the sum of the numbers is 11 or more? (You should be able to do this without a calculator.)

*Answer.* We know that the numbers on each dice is symmetric about 3.5. We have three dice, so the sum of the numbers is symmetric about $3\cdot 3.5 = 10.5$. In other words, the probability of getting 11 or more is the same as getting 10 or less. Therefore, since the two sets are complements, the probability is $\frac{1}{2}$. □

**Exercise 55** Suppose I'm looking at the grades for the Monday update and the mastery quiz in class (of 250 people). For the Monday update 10% of the students got a 0 and 90% got a 1. For the mastery quiz 20% of the students got a 0 and 80% got a 1. I want to see how well my students are doing in two different ways. First I create one list where I add each student's grade from the two quizzes. On another list I multiply each student's grade from the two quizzes.

(1) Can I calculate the average grade for students on the first list? (the one I added them together) If so, what's the average?

(2) Can I calculate the average grade for students on the second list? (the one I multiplied them together) If so, what's the average?

*Answer.* The average of the Monday update is $0 \cdot 0.1 + 1 \cdot 0.9 = 0.9$. The average of the mastery quiz is $0 \cdot 0.2 + 1 \cdot 0.8 = 0.8$.

To help explain, we'll look at a small example with two students: $s_1$ and $s_2$. $s_1$ gets a score of $u_1$ (which is equal to 0 or 1) for the Monday update and a $q_1$ for the mastery quiz. Similarly $s_2$ gets a score of $u_2$ for the Monday update and a $q_2$ for the mastery quiz.

(1) Looking at our small example, this first list is taken by doing the following:

$$\frac{(u_1 + q_1) + (u_2 + q_2)}{2}$$

But, addition is distributive! So we have

$$\begin{aligned}\frac{(u_1 + q_1) + (u_2 + q_2)}{2} &= \frac{u_1 + u_2 + q_1 + q_2}{2} \\ &= \frac{u_1 + u_2}{2} + \frac{q_1 + q_2}{2}\end{aligned}$$

In other words, the average of the overall grade is just the sum of the averages of the individual grades!

So all we have to do is: $0.9 + 0.8 = 1.7$, giving us the average grade for students on the first list.

(2) Looking at our small example, this second list is taken by doing the following:

$$\frac{(u_1 \cdot q_1) + (u_2 \cdot q_2)}{2}$$

Notice how at this point, there is no way for me to split this apart like we did in the first exercise. In other words, I *can not* calculate the average grade for students on the second list.

□

**Exercise 56** What is the expected value of the number of spades drawn if you pull 7 cards randomly from a (well-shuffled, normal) deck of 52 cards?

*Answer.* We know that 13 out of 52 cards are spades (or exactly $\frac{1}{4}$). Since we're looking at success/failure, indicator functions are a great way to approach this. Let $X_i$ denote the $i$th card where $X_i = 1$ if the $i$th card is a spade, or 0 otherwise. Then we're asking what is $E(X)$ where $X = X_1 + X_2 + \ldots + X_7$.

$$E(X) = \sum_{i=1}^{7} P(X_i = 1) = 7P(X_1 = 1) = 7 \cdot \frac{1}{4} = \frac{7}{4}$$

Note here we are using the fact that $P(X_i = 1) = P(X_1 = 1)$ for all $i$. Here is an example for when $i = 2$

and you can expand from there.

$$\begin{aligned}
P(X_2 = 1) &= P(X_2 = 1 \cap X_1 = 1) + P(X_2 = 1 \cap X_1 = 0) \\
&= P(X_2 = 1 \mid X_1 = 1)P(X_1 = 1) + P(X_2 = 1 \mid X_1 = 0)P(X_1 = 0) \\
&= \frac{12}{51} \cdot \frac{13}{52} + \frac{13}{51} \cdot \frac{39}{52} \\
&= \frac{(12 + 39)(13)}{51 \cdot 52} \\
&= \frac{51 \cdot 13}{51 \cdot 52} \\
&= \frac{13}{52}
\end{aligned}$$

□