Solutions to MASD (Probability Part) Assignment 5

Problem 1 (Modelling Probabilities): Provide a suitable model (set of outcomes S and probability distribution $\mathbb{P}$) for the following probabilistic experiments. Answer the questions asked in each case.

1. We toss a coin and afterwards we roll a die. What is the probability that we observe heads and then roll a number smaller or equal to 4?

2. We are given an unfair die. We roll it 1000 times and record how often we see each number:

1	2	3	4	5	6
107	195	52	492	112	42

   How would you model the experiment of rolling this die again twice? Let X_1, X_2 be the outcome of the first and second roll, respectively. Compute the probability that the roll sum $X_1 + X_2$ equals 4.

3. We are given a deck of 52 (distinct) cards. They are well organised, starting with hearts $2, 3, 4, 5$. We shuffle them as best we can. What is the probability that after shuffling the first four cards of the deck are again the same, but in possibly different order?

Solutions to Problem 1: For 1: We encode the outcome of the coin toss by $\{0, 1\}$ (heads is 1 and tails is 0). The die roll we encode in the number of eyes we observe, i.e. by $\{1, 2, 3, 4, 5, 6\}$. Overall the possible outcomes of the experiment is an element of

$$S = \{0, 1\} \times \{1, 2, 3, 4, 5, 6\} = \{(s_1, s_2) : s_1 \in \{0, 1\}\,,\ s_2 \in \{1, 2, 3, 4, 5, 6\}\}\,.$$

As probability distribution $\mathbb{P}$ we choose the uniform distribution since none of these outcomes is more or less likely than any other. Now we compute

$$\begin{aligned}
&\mathbb{P}(\text{coin toss shows heads and die roll shows a number } \leq 4) \\
=&\,\mathbb{P}(\{(s_1, s_2) \in S : s_1 = 1, s_2 \leq 4\}) \\
=&\,\mathbb{P}(\{(1, 1), (1, 2), (1, 3), (1, 4)\}) \\
=&\,\frac{4}{|S|} \\
=&\,\frac{4}{12} = \frac{1}{3}\,.
\end{aligned}$$

For 2: Let us first model the experiment of rolling the unfair dice again once. In this case the probability space is $(S_1, \mathbb{P}_1)$, where $S_1 = \{1, 2, 3, 4, 5, 6\}$ is the outcome of one dice roll and it is reasonable to assign the probabilities

$$\mathbb{P}_1(\{i\}) := p_i\,, \quad \text{where } p_1 := \frac{107}{1000}\,,\ p_2 := \frac{195}{1000}\,,\ p_3 := \frac{52}{1000}\,,\ p_4 := \frac{492}{1000}\,,\ p_5 := \frac{112}{1000}\,,\ p_6 := \frac{42}{1000}\,.$$

for $i = 1, 2, 3, 4, 5, 6$. Now we model two such dice rolls. Then the underlying probability space is $(S_2, \mathbb{P}_2)$ where $S_2 = \{1, 2, 3, 4, 5, 6\}^2$ is the outcome of both rolls and we assign the probabilities

$$\mathbb{P}(\{(i, j)\}) := p_i p_j\,, \qquad i, j = 1, 2, 3, 4, 5, 6\,.$$

The two projections X_1 and X_2 model the outcome of each roll

$$X_1 : S \to \{1,2,3,4,5,6\}, (s_1, s_2) \mapsto s_1\,, \qquad X_2 : S \to \{1,2,3,4,5,6\}, (s_1, s_2) \mapsto s_2\,.$$

Put differently, $X_i((s_1, s_2)) := s_i$ for $i = 1, 2$. Then we compute the probability

$$\begin{aligned}\mathbb{P}(X_1 + X_2 = 4) &= \mathbb{P}(\{(1,3),(3,1),(2,2)\})\\ &\overset{(1)}{=} \mathbb{P}(\{(1,3)\}) + \mathbb{P}(\{(3,1)\}) + \mathbb{P}(\{(2,2)\})\\ &= 2p_1p_3 + p_2^2 = 2 \times \frac{107}{1000} \times \frac{52}{1000} + \frac{195^2}{1000^2}\,.\end{aligned}$$

In (1) we used the additivity of probabilities on disjoint sets (in this case each with one element).

For 3: We assign a number to each of the 52 cards starting with hearts 2,3,4,5. Every reshuffled deck of cards is modelled by an assignment of the position in the deck to the card at that position, i.e. by a one to one map of $\{1, 2, \ldots, 52\}$ to itself. Mathematically this means that we choose $S = S_{52}$ the set of permutations of the numbers $\{1, 2, \ldots, 52\}$. For example the map (upstairs is the argument and below that the value)

$$s = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & \ldots & 52 \\ 3 & 2 & 1 & 4 & 5 & \ldots & 52 \end{pmatrix}$$

represents switching the first and third card. Since no shuffling is special we equip S with the uniform distribution $\mathbb{P}$. We compute the probability

$$\begin{aligned}&\mathbb{P}(\text{hearts } 2,3,4,5 \text{ in any order are the first four cards})\\ =&\,\mathbb{P}(\{s \in S_{52} : \text{there is a permutation in } \sigma \in S_4 \text{ such that } s(i) = \sigma(i) \text{ for all } i = 1,2,3,4\})\\ =&\sum_{\sigma \in S_4} \mathbb{P}(\{s \in S_{52} : s(i) = \sigma(i) \text{ for all } i = 1,2,3,4\})\,,\end{aligned}$$

where we used the additivity of probabilities in the last step. Now we compute for some fixed $\sigma \in S_4$ the probability

$$\mathbb{P}(\{s \in S_{52} : s(i) = \sigma(i) \text{ for all } i = 1,2,3,4\}) = \frac{|S_{52-4}|}{|S_{52}|} = \frac{48!}{52!}\,.$$

Altogether we find

$$\mathbb{P}(\text{hearts } 2,3,4,5 \text{ in any order are the first four cards}) = |S_4| \times \frac{48!}{52!} = \frac{4! \times 48!}{52!}\,.$$

Problem 2 (Exclusion-Inclusion Principle): Let $A = \{1, 2, \ldots, 50\}$. For any $n \in \mathbb{N}$ let A_n be the set of integers in A that are divisible by n. For example, $A_3 = \{3, 6, 9, \ldots, 48\}$ and $A_4 = \{4, 8, 12, \ldots, 48\}$.

1. Compute the number of elements in the sets A_2, A_3, A_4, i.e. the values of $|A_2|, |A_3|, |A_4|$.

2. Compute $|A_2 \cup A_3 \cup A_7|$.

Solutions to Problem 2: For 1: In general $|A_i|$ is the integer part of $50/i$. In particular, $|A_2| = 25$, $|A_3| = 16$ and $|A_4| = 12$.

For 2: To compute $|A_2 \cup A_3 \cup A_7|$ we use the Exclusion-Inclusion Principle, i.e. the formula

$$|A_2 \cup A_3 \cup A_7| = |A_2| + |A_3| + |A_7| - |A_2 \cap A_3| - |A_2 \cap A_7| - |A_3 \cap A_7| + |A_2 \cap A_3 \cap A_7|\,.$$

We insert $|A_2| = 25$, $|A_3| = 16$, $|A_7| = 7$, as well as

$$\begin{aligned}
|A_2 \cap A_3| &= |A_6| = 8\,,\\
|A_2 \cap A_7| &= |A_{14}| = 3\,,\\
|A_3 \cap A_7| &= |A_{21}| = 2\,,\\
|A_2 \cap A_3 \cap A_7| &= |A_{42}| = 1\,,
\end{aligned}$$

to get $|A_2 \cup A_3 \cup A_7| = 25 + 16 + 7 - 8 - 3 - 2 + 1 = 36$.

Problem 3 (Combinatorics):

1. How many subsets of the numbers from 1 to 100 are there that have 10 elements and also contain the numbers 5 and 10?
2. An anagram is a word created from the letters of another word (with exactly the same number of appearances of each letter). How many (possibly nonsensical) anagrams are there of the word "STATISTICS"?
3. You have an even number $n \geq 20$ of friends, k of them are female and $n - k$ of them male, where $10 \leq k \leq n$. You invite half of them to a dinner party. How many different dinner party configurations can you organise with exactly 10 female friends present? How many can you organise with at least 10 female friends present? Express the outcome in terms of binomial coefficients.
4. How many distinct ways are there to write 50 as a sum of 6 non-negative integers, i.e. how many possible solutions are there to $x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 50$ with $x_i \in \mathbb{N}$?

Solutions to Problem 3: For 1: Since we are required to have the numbers 5 and 10 in our 10-element subset, we are effectively computing the number of subsets of $\{1, \ldots, 100\} \setminus \{5, 10\}$ with 8 elements. There are '98 choose 8' of those, or in mathematical terms

$$\#\{I : I \subset \{1, \ldots, 100\} \setminus \{5, 10\},\ \#I = 8\} = \binom{98}{8}\,.$$

For 2: To count the number of anagrams we count the number of positions (10 of them at the beginning) where we can place each letter. Then we have to multiply the number of all these options. We start with "A". We can place "A" at any of the 10 positions. Let us keep track of the factors by defining $N_A = 10$. For the letter "C" we are left with 9 positions since "A" already took up one of them. Thus, we set $N_C = 9$. The letter "I" comes up twice. Since we cannot distinguish the two "I"'s from each other the number of choices here is the same as the number of subsets of the remaining 8 positions with 2 elements, i.e. $N_I = \binom{8}{2}$. We are left with 6 positions for the three letters "S", so the number of choices is the same as all 3-element subsets of $\{1, 2, 3, 4, 5, 6\}$, i.e. $N_S = \binom{6}{3}$. The positions of the remaining "T"'s are fixed, because there are only three positions left, i.e. $N_T = 1$. The answer is thus that there are

$$N_A N_C N_I N_S N_T = 10 \times 9 \times \binom{8}{2} \times \binom{6}{3} = 50400$$

anagrams of the word "STATISTICS".

For 3: We determine the number $N_{n,k,l}$ of dinner parties you can organise with exactly l female friends present, where $10 \leq l \leq k$ and $l \leq n/2$. Since you can choose l female invitees arbitrarily out of k female friends, there are $\binom{k}{l}$ options of choosing the configuration of female participants. For the male participants you can choose $n/2 - l$ out of your $n - k$ male friends., provided $n - k \geq n/2 - l$. Thus, you have $\binom{n-k}{n/2-l}$ possibilities for the male participation in this case. Altogether we have

$$N_{n,k,10} = \binom{k}{10}\binom{n-k}{n/2-10}$$

the number of dinner arrangements with exactly 10 female friends invited, as long as $n - k \geq n/2 - 10$. In case $n - k < n/2 - 10$ it is not possible to organise a dinner party with $n/2$ friends of which 10 are female, i.e.

$$N_{n,k,l} = 0 \qquad n - k < n/2 - l\,,$$

holds for all l and in particular for $l = 10$.

Finally, to compute the number of dinner configuration with at least 10 female friends invited, we sum up all $N_{n,k,l}$ with $l \geq 10$, i.e. we get

$$\sum_{l=10}^{n/2} N_{n,k,l}\,,$$

possibilities, where

$$N_{n,k,l} = \begin{cases} \binom{k}{l}\binom{n-k}{n/2-l} & \text{if } n - k \geq n/2 - l \\ 0 & \text{if } n - k < n/2 - l \end{cases}.$$

For 4: This problem is the same as the example of distributing coins from the lecture. It is an example of unordered sampling with replacement. Since $x_i \geq 1$ here we can write $x_i = 1 + y_i$ with non-negative integers y_i. Then the question becomes: How many solutions to $y_1 + y_2 + \cdots + y_6 = 44$ are there with $y_i \geq 0$. By the formula for unordered sampling with replacement we get for the number of solutions:

$$\binom{n+k-1}{n-1} = \binom{49}{5} = 1906884\,, \qquad n = 6\,, \qquad k = 44\,.$$