Solutions to problem sheet Counting & Enumeration

Exercise 1. There are exactly 2^k binary words of length k. It follows that we need to choose

$$k \geq \lceil \log_2(38) \rceil = \lceil \log(38)/\log(2) \rceil = 6.$$

Here $\log_2$ is the binary logarithm whereas log without a base specified usually refers to either the natural logarithm or the logarithm base 10. The Change-of-Base formula $\log_b(\cdot) = \log(\cdot)/\log(b)$ is true for either. The brackets $\lceil \cdot \rceil$ mean that we are rounding up to the nearest integer.

Exercise 2. The difference from Exercise 1 is that the Morse code does not have a prescribed length for each letter. For instance, A is encoded as $\cdot -$, a word of length 2, whereas B is encoded as $- \cdot \cdot \cdot$, a word of length 4. There are exactly

$$2^1 + 2^2 + \cdots + 2^k = \sum_{i=1}^{k} 2^i$$

words of length *at most* k. We see that $\sum_{i=1}^{3} 2^i = 14$ and $\sum_{i=1}^{4} 2^i = 30$, so that we need to choose $k \geq 4$.

Bonus: for arbitrary sizes of the alphabet, use the closed formula

$$\sum_{i=1}^{k} 2^i = 2^{k+1} - 2.$$

One can easily prove this through an induction on k or by showing that this formula satisfies an appropriate recursion. Let us give a combinatorial proof instead:
The binary expansion of $\sum_{i=0}^{k} 2^i$ is

$$\sum_{i=0}^{k} 2^i = (\underbrace{111\ldots111}_{k+1})_2.$$

Adding $1 = (1)_2$ to this, we get

$$\sum_{i=0}^{k} 2^i + 1 = (\underbrace{111\ldots111}_{k+1})_2 + (1)_2 = (1\,\underbrace{000\ldots000}_{k+1})_2 = 2^{k+1}.$$

This implies

$$\sum_{i=1}^{k} 2^i = \sum_{i=0}^{k} 2^i - 1 = (2^{k+1} - 1) - 1 = 2^{k+1} - 2. \quad \square$$

To conclude, we have $\lceil \log_2(28 + 2) - 1 \rceil = 4$, which verifies the previously established answer.

Exercise 3. The first letter can be any of the four options A, C, T, or G. For every letter after that we have exactly three options: any of the four letters except for the one just used in the previous position. It follows that there are exactly $4 \cdot 3^{n-1}$ such words of length n.

Exercise 4. Fix $n \geq 0$ and let $\mathcal{W}$ be the set of words of length n with letters in $\{A, B\}$ and avoiding two consecutive A's. so that $a_n = |\mathcal{W}|$.

It is useful to consider that there exists a unique word of length zero: the *empty word* denoted by ϵ. Trivially ϵ does not have two consecutive A's, and $a_0 = 1$. Observe also that there are two words of length one (A and B) and that they also do not contain two consecutive A's. This means that $a_1 = 2$.

Let now $n \geq 2$ and consider the sets $\mathcal{W}_1$ consisting of all words in $\mathcal{W}$ ending with B, and $\mathcal{W}_2$ consisting of all words in $\mathcal{W}$ ending with BA. Both are subsets of $\mathcal{W}$. Furthermore, all possible endings of a word in $\mathcal{W}$ are covered by those two possibilities, as any other possible ending would violate the rule of not having two consecutive A's. Hence:

$$\mathcal{W} = \mathcal{W}_1 \cup \mathcal{W}_2. \tag{1}$$

Additionally, no word in $\mathcal{W}$ can have both endings at the same time. This implies:

$$\mathcal{W}_1 \cap \mathcal{W}_2 = \emptyset. \tag{2}$$

Both conditions (1) and (2) mean that $\mathcal{W}_1$ and $\mathcal{W}_2$ form a *partition* of $\mathcal{W}$ into two disjoint subsets (called the *parts*).

Notice now that if we remove the last letter B from a word in $\mathcal{W}_1$, we obtain a word of length $n-1$ that does not have any two consecutive A's. Conversely, we can concatenate a B at the end of any word of length $n-1$ that does not have any two consecutive A's to obtain a word in $\mathcal{W}_1$. This means that there are as many words in $\mathcal{W}_1$ as there are words of length $n-1$ that do not have any two consecutive A's, and by hypothesis there are a_{n-1} of them. That is $|\mathcal{W}_1| = a_{n-1}$. A similar argument, i.e. removing the last two letters BA from a word in $\mathcal{W}_2$, shows that $|\mathcal{W}_2| = a_{n-2}$. And (for $n \geq 2$) we have:

$$a_n = |\mathcal{W}| \overset{(1)}{=} |\mathcal{W}_1 \cup \mathcal{W}_2| = |\mathcal{W}_1| + |\mathcal{W}_2| - |\mathcal{W}_1 \cap \mathcal{W}_2| \overset{(2)}{=} |\mathcal{W}_1| + |\mathcal{W}_2| = a_{n-1} + a_{n-2}$$

It follows that a_n satisfies the recursion

$$a_n = a_{n-1} + a_{n-2} \quad \text{for all } n \geq 2, \qquad \text{with initial conditions } a_0 = 1 \text{ and } a_1 = 2.$$

We now consider the set $\mathcal{V}$ of words of length n, with letters in $\{A, B\}$ and avoiding *three* consecutive A's. If we let $b_n = |\mathcal{V}|$, then we still have trivially $b_0 = 1$ and $b_1 = 2$, but now $b_2 = 4$ as the word AA is allowed. We then partition $\mathcal{V}$ into three disjoint parts depending on the ending of each word. Namely, the words ending in B, the words ending in BA, and those ending in BAA. Arguing as before, we get that b_n satisfies the recursion

$$b_n = b_{n-1} + b_{n-2} + b_{n-3} \quad \text{for all } n \geq 3, \quad \text{with initial conditions } b_0 = 1,\ b_1 = 2 \text{ and } b_2 = 4.$$

Exercise 5. Similarly to Exercise 4, we partition the set of words of length n into subsets of words that end either in C, T, G, CA, TA or GA. Writing a_n for the number of such words, we see that a_n must satisfy the recurrence

$$a_n = 3a_{n-1} + 3a_{n-2} \quad \text{for all } n \geq 2, \qquad \text{with initial conditions } a_0 = 1 \text{ and } a_1 = 4.$$

Exercise 6. The word MISSISSIPPI has 11 letters. The letter M occurs once, I and S each occur four times and P occurs twice. We can consider the problem as choosing slots for the letters, that is we are drawing them unordered without repetition.

For M there are clearly 11 possibilities. Once M has been assigned a position, there are $11 - 1 = 10$ positions still available and therefore $\binom{10}{4}$ options for I. Following this, there are $\binom{6}{4}$ options for S. Lastly, exactly two positions remain for P. It follows that the answer is

$$\binom{11}{1}\binom{10}{4}\binom{6}{4}\binom{2}{2} = \frac{11! \cdot 10! \cdot 6! \cdot 2!}{1! \cdot 10! \cdot 4! \cdot 6! \cdot 4! \cdot 2! \cdot 2! \cdot 0!} = \frac{11!}{1! \cdot 4! \cdot 4! \cdot 2!} = 34\,650.$$

Applying the same arguments to the word GUADALQUIVIR, we get the answer

$$\binom{12}{1}\binom{11}{2}\binom{9}{2}\binom{7}{1}\binom{6}{1}\binom{5}{1}\binom{4}{2}\binom{2}{1}\binom{1}{1} = \frac{12!}{1! \cdot 2! \cdot 2! \cdot 1! \cdot 1! \cdot 1! \cdot 2! \cdot 1! \cdot 1!} = 59\,875\,200.$$

Exercise 7. The identity holds, as

$$\begin{aligned}\binom{n-1}{k}+\binom{n-1}{k-1} &= \frac{(n-1)!}{k!(n-1-k)!}+\frac{(n-1)!}{(k-1)!(n-k)!}\\ &= \frac{n!}{(k!(n-k)!}\cdot\left(\frac{n-k}{n}+\frac{k}{n}\right)=\binom{n}{k}\cdot\frac{n}{n}=\binom{n}{k}\end{aligned}$$

There is a combinatorial intuition behind this formula: $\binom{n}{k}$ counts the number of way of taking k balls out of a set of n (distinguishable) balls. We can then split the possible outcomes into two parts: fix some arbitrary ball B in the set, either the selection of k balls does not contain B or it does. The former is clearly counted by $\binom{n-1}{k}$ and the later by $\binom{n-1}{k-1}$, giving the formula.

Exercise 8. Let us prove the statement by induction on n, noting that we always have $n \geq k \geq 1$. We refer to the statement given by the formula as $P(n)$.

Base Case. Let us prove $P(1)$. In this case, we clearly must have $k = 1 = n$ and hence $\binom{k}{k} = 1 = \binom{k+1}{k+1}$.

Induction Step. Let $n > 1$ and suppose $P(n-1)$ is true. Note first that by the same argument as in the base case, $P(k)$ (that is, the case $n = k$) holds for any $k \geq 1$, so from now on we can assume that $k \leq n-1$. Hence we can use the induction hypothesis and the previous exercise to see that

$$\binom{n+1}{k+1} \overset{(7)}{=} \binom{n}{k+1}+\binom{n+1}{k} \overset{P(n-1)}{=} \binom{k}{k}+\cdots+\binom{n}{k}+\binom{n+1}{k},$$

so $P(n)$ holds.

Exercise 9. There is a total of 2^{2n} possible outcomes. Out of these, exactly $\binom{2n}{n}$ represent a tie. It follows that the probability is given by $\binom{2n}{n}/2^{2n}$, which for $n = 3\,761\,500$ is about 0.03%.

Exercise 10. The recurrence was given by

$$a_n = a_{n-1} + a_{n-2} \quad \text{for all } n \geq 2, \qquad \text{with initial conditions } a_0 = 1 \text{ and } a_1 = 2.$$

Note that this looks just like the Fibonacci sequence from the lecture, at the exception of the initial conditions. The *characteristic polynomial equation* of this recurrence is

$$x^2 - x - 1 = 0.$$

Its roots are given by

$$\varphi = \frac{1+\sqrt{5}}{2} \quad \text{and} \quad \bar{\varphi} = \frac{1-\sqrt{5}}{2}.$$

Using the theorem for linear recurrence relations from the lecture, we have that

$$a_n = A\,\varphi^n + B\,\bar{\varphi}^n \qquad \text{for all } n \geq 0, \tag{3}$$

and for some appropriate constants A and B that we determine next. In particular, as (3) is true for all $n \geq 0$, it is true when setting $n = 0$ and $n = 1$. From this we get

$$\begin{aligned} a_0 &= A\,\varphi^0 + B\,\bar{\varphi}^0 \iff 1 = A + B\\ a_1 &= A\,\varphi^1 + B\,\bar{\varphi}^1 \iff 2 = A\,\varphi + B\,\bar{\varphi}\end{aligned}$$

The above linear system admits the following unique solution

$$A = \frac{2-\bar{\varphi}}{\varphi-\bar{\varphi}} = \frac{3+\sqrt{5}}{2\sqrt{5}} \quad \text{and} \quad B = 1 - \frac{3+\sqrt{5}}{2\sqrt{5}} = \frac{\sqrt{5}-3}{2\sqrt{5}}$$

And plugging the above equalities in (3) we obtain

$$a_n = \frac{3+\sqrt{5}}{2\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^n + \frac{\sqrt{5}-3}{2\sqrt{5}}\left(\frac{1-\sqrt{5}}{2}\right)^n \qquad \text{for all } n \geq 0.$$

We can verify the above statement by checking that $a_2 = 3$.

Exercise 11. The recurrence was given by

$$a_1 = 4,\ a_2 = 15 \quad \text{and} \quad a_n = 3a_{n-1} + 3a_{n-2} \text{ for all } n \geq 3.$$

The characteristic polynomial of this recurrence is $x^2 - 3x - 3$ which has roots

$$\lambda_1 = (3+\sqrt{21})/2 \quad \text{and} \quad \lambda_2 = (3-\sqrt{21})/2.$$

We therefore know that the general form of the recurrence is

$$a_n = A\,\lambda_1^n + B\,\lambda_2^n$$

for some appropriate constants A and B. Note that by setting $a_0 = 1$ we have $3a_0+3a_1 = 15 = a_2$, so we can use this to make our lives easier. It follows that A and B must satisfy

$$\begin{aligned} A + B &= 1 \\ A\lambda_1 + B\lambda_2 &= 4 \end{aligned}$$

Solving this we get

$$A = \frac{5+\sqrt{21}}{2\sqrt{21}} \quad \text{and} \quad B = \frac{\sqrt{21}-5}{2\sqrt{21}}$$

And therefore

$$a_n = \frac{5+\sqrt{21}}{2\sqrt{21}}\left(\frac{3+\sqrt{21}}{2}\right)^n + \frac{\sqrt{21}-5}{2\sqrt{21}}\left(\frac{3-\sqrt{21}}{2}\right)^n.$$

Exercise 12. We note that for $n = 1$ there is clearly just one of these rectangles (consisting of one 2×1 block). And for $n = 2$ there are three (one 2×2 block, two 1×2 blocks or two 2×1 blocks). For $n \geq 3$, a $2 \times n$ rectangle may end in a 2×1 block, a 2×2 block or two 1×2 blocks.

It follows that, if a_n counts the number of such $2 \times n$ rectangles we get the recursion

$$a_1 = 1,\ a_2 = 3 \quad \text{and} \quad a_n = a_{n-1} + 2a_{n-2} \text{ for all } n \geq 3.$$

Since $a_1 + 2 \cdot 1 = 3 = a_2$, we may set $a_0 = 1$ to make calculations easier. The characteristic polynomial of this recurrence relation is $x^2 - x - 2$, which has roots $\lambda_1 = 2$ and $\lambda_2 = -1$. Solving

$$\begin{aligned} A + B &= 1 \\ 2A - B &= 1 \end{aligned}$$

gives $A = 2/3$ and $B = 1/3$. Therefore,

$$a_n = \frac{2}{3} \cdot 2^n + \frac{1}{3} \cdot (-1)^n \text{ for all } n \geq 0.$$

And we verify our result by computing $a_2 = 3$.

Exercise 13. The characteristic polynomial is $x^3 + x^2 - 4x - 4 = (x+1)(x+2)(x-2)$ which has roots $\lambda_1 = -1$, $\lambda_2 = -2$, and $\lambda_3 = 2$. The general form of the recurrence is hence

$$a_n = A \cdot (-1)^n + B \cdot (-2)^n + C \cdot 2^n$$

for some appropriate constants A, B, and C. Using the initial conditions, we solve

$$\begin{aligned} A + B + C &= 8 \\ -A - 2B + 2C &= 6 \\ A + 4B + 4C &= 26 \end{aligned}$$

to obtain $A = 2$, $B = 1$ and $C = 5$. And the general form of the homogeneous recurrence is

$$a_n = 2 \cdot (-1)^n + (-2)^n + 5 \cdot 2^n.$$

To solve the inhomogeneous case, we remark that the general homogeneous solution (ignoring initial conditions) is of the form $b_n = A \cdot (-1)^n + B \cdot (-2)^n + C \cdot 2^n$. Next, we guess a particular solution c_n of the inhomogeneous recurrence (again ignoring the initial conditions). As the inhomogeneous part is a constant, we guess that $c_n = D$ for some constant D. It follows that D must satisfy

$$D = c_n = -c_{n-1} + 4c_{n-2} + 4c_{n-3} + 12 = -D + D \cdot D + 4 \cdot D + 12$$

so that $D = -2$. Finally, a solution to the inhomogeneous recurrence must be of the form $a_n = b_n + c_n$. We now determine A, B, C in order to satisfy the initial conditions, that is

$$\begin{aligned} A + B + C - 2 &= 8, \\ -A - 2B + 2C - 2 &= 6, \\ A + 4B + 4C - 2 &= 26. \end{aligned}$$

Thus $A = 4$, $B = 0$ and $C = 6$. And the general form of the inhomogeneous recurrence is

$$a_n = 4 \cdot (-1)^n + 6 \cdot 2^n - 2.$$

Exercise 14. Let a_n denote the number of bacteria at the end of month n, then $a_0 = 30$ and $a_1 = 30$. Denote by M_n the number of mature bacteria, and by Y_n the number of young bacteria, both at the end of month n. Clearly, $a_n = M_n + Y_n$. Now notice that any mature bacterium was either already mature in the previous month, or it was young, and hence

$$M_n = M_{n-1} + Y_{n-1} = a_{n-1}.$$

On the other hand, a young bacterium was produced by a mature one at a rate of 2 to 1, and hence

$$Y_n = 2M_{n-1} = 2a_{n-2},$$

where the last equality follows from our previous argument. Putting these observations together, we get the recurrence relation

$$a_0 = 30,\ a_1 = 30,\ \text{and } a_n = a_{n-1} + 2a_{n-2} \text{ for all } n \geq 2.$$

The characteristic polynomial is $x^2 - x - 2$, which by Exercise 12 has roots $\lambda_1 = 2$ and $\lambda_2 = -1$. Hence the general form of the recurrence is

$$a_n = A \cdot 2^n + B \cdot (-1)^n,$$

where

$$\begin{aligned} A + B &= 30 \\ 2A - B &= 30 \end{aligned}$$

Solving this, we get $A = 20$ and $B = 10$. Thus

$$a_n = 20 \cdot 2^n + 10 \cdot (-1)^n.$$