Exercise 7

Task 1
Let $T = 001100$ and $P = 01$. Use the probabilistic algorithm of the lecture to compute the array MATCH$[1, \dots, 6]$, which encodes the occurrences of the pattern P in the string T.

Task 2
In this task we will consider an alternative class of fingerprint functions. For a word $w = a_1 \dots a_n \in \{0,1\}^*$ we define

$$h(a_1 \dots a_n) = \sum_{i=1}^{n} a_i 2^{n-i}.$$

Let $h_p(w) = h(w) \bmod p$ be the *fingerprint* of w with respect to a prime p.

(a) Construct a randomised pattern matching algorithm by using these fingerprint functions.

(b) What is the probability of an invalid match of your algorithm?

Task 3
For a given number $r \geq 1$ and a prime p let $x = (x_0, x_1, \dots, x_r)$ with $x_i \in \mathbb{F}_p$. Let $h_x : \mathbb{F}_p^{r+1} \to \mathbb{F}_p$ be the function defined by

$$h_x(a) = \sum_{i=0}^{r} a_i x_i \bmod p, \quad a = (a_0, \dots, a_r).$$

Show that $\mathcal{H} = \{h_x | x_i \in \mathbb{F}_p, 0 \leq i \leq r\}$ is a universal familiy of hash functions.
Is $\mathcal{H}$ also a familiy of pairwise independent hash functions?

Task 4
We generalize the definition on slide 121 in the following way: Let $\mathcal{H} \subseteq \{h \mid h : A \to B\}$ be a family of hash functions. We call $\mathcal{H}$ a *family of k-wise independent hash functions*, if for all $a_1, \dots, a_k \in A$ (pairwise different) and $b_1, \dots, b_k \in B$ we have

$$\text{Prob}[\bigwedge_{i=1}^{k} h(a_i) = b_i] = 1/|B|^k$$

for a randomly chosen $h \in \mathcal{H}$ (uniform distribution). Show that

$$\mathcal{H} = \{h_x : \mathbb{F}_p \to \mathbb{F}_p \mid h_x(a) = \sum_{i=0}^{k-1} x_i a^i, x = (x_0, \dots, x_{k-1}) \in \mathbb{F}_p^k\}$$

is such a k-wise independent family if $k \leq p$.

Task 5 (AMS algorithm)
Consider the stream $S = (101, 011, 010, 111, 011, 101, 000, 001)$ and the corresponding set A. Approximate the cardinality of A by using the hash functions $h_{x,y}(u) = xu + y$ over $\mathbb{F}_{2^3}$ with

1. $x = 101$ and $y = 001$,
2. $x = 100$ and $y = 101$.

Hint: You can use that $+$ over the field $\mathbb{F}_{2^3}$ works like a bitwise XOR and $x \cdot u$ is given by the following table:

u	000	001	010	011	100	101	110	111
$100 \cdot u$	000	100	011	111	110	010	101	001
$101 \cdot u$	000	101	001	100	010	111	011	110