Exercise 6

Task 1
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 $a_1, \ldots, a_k \in A$ (pairwise different) and $b_1, \ldots, b_k \in B$ 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

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

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

Solution
The proof generalizes exactly like on slides 125 to 127. We choose $a_1, \ldots, a_k, b_1, \ldots, b_k \in \mathbb{F}_p$, where the a_i are pairwise different. Then the system

$$\begin{aligned} x_0 + a_1 x_1 + \cdots + a_1^{k-1} x_{k-1} &\equiv b_1 \bmod p \\ x_0 + a_2 x_1 + \cdots + a_2^{k-1} x_{k-1} &\equiv b_2 \bmod p \\ &\vdots \\ x_0 + a_k x_1 + \cdots + a_k^{k-1} x_{k-1} &\equiv b_k \bmod p \end{aligned}$$

has a unique solution $(x_0, \ldots, x_{k-1}) \in \mathbb{F}_p^k$: The system is equivalent to

$$\begin{pmatrix} 1 & a_1 & \cdots & a_1^{k-1} \\ 1 & a_2 & \cdots & a_2^{k-1} \\ \vdots & \vdots & & \vdots \\ 1 & a_k & \cdots & a_k^{k-1} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ \vdots \\ x_{k-1} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_k \end{pmatrix} \quad \text{in } \mathbb{F}_p.$$

The matrix is the Vandermonde matrix and hence invertible, since the a_i are pairwise different (shown on Sheet 1). We argue as on slide 126 and 127 to obtain a probability of

$$\text{Prob}[\bigwedge_{i=1}^{k} h_x(a_i) = b_i] = 1/p^k.$$

Task 2 (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

Solution
We use the table in order to get $h_{x,000}(u)$ and then perform some bitflips.

1.

u	101	011	010	111	011	101	000	001
$h_{101,000}(u)$	111	100	001	110	100	111	000	101
$h_{101,001}(u)$	110	101	000	111	101	110	001	100
$\rho(h_{101,001}(u))$	0	0	3	0	0	0	2	0

Hence, after step 3 of the AMS algorithm we have $z = 3$ and thus we return $2^{3.5} = 11.3\ldots > 8$. In practice we could stop after reading $u = 010$, since $z = 3$ is the maximal value.

2.

u	101	011	010	111	011	101	000	001
$h_{100,000}(u)$	010	111	011	001	111	010	000	100
$h_{100,101}(u)$	111	010	110	100	010	111	101	001
$\rho(h_{100,101}(u))$	0	1	0	0	1	0	0	2

Hence, after step 3 of the AMS algorithm we have $z = 2$ and thus we return $2^{2.5} = 5.65685\ldots \approx 6$.

The stream has 6 different values, so the second hash function yields a better approximation.

Task 3 (Average height of binary search trees)

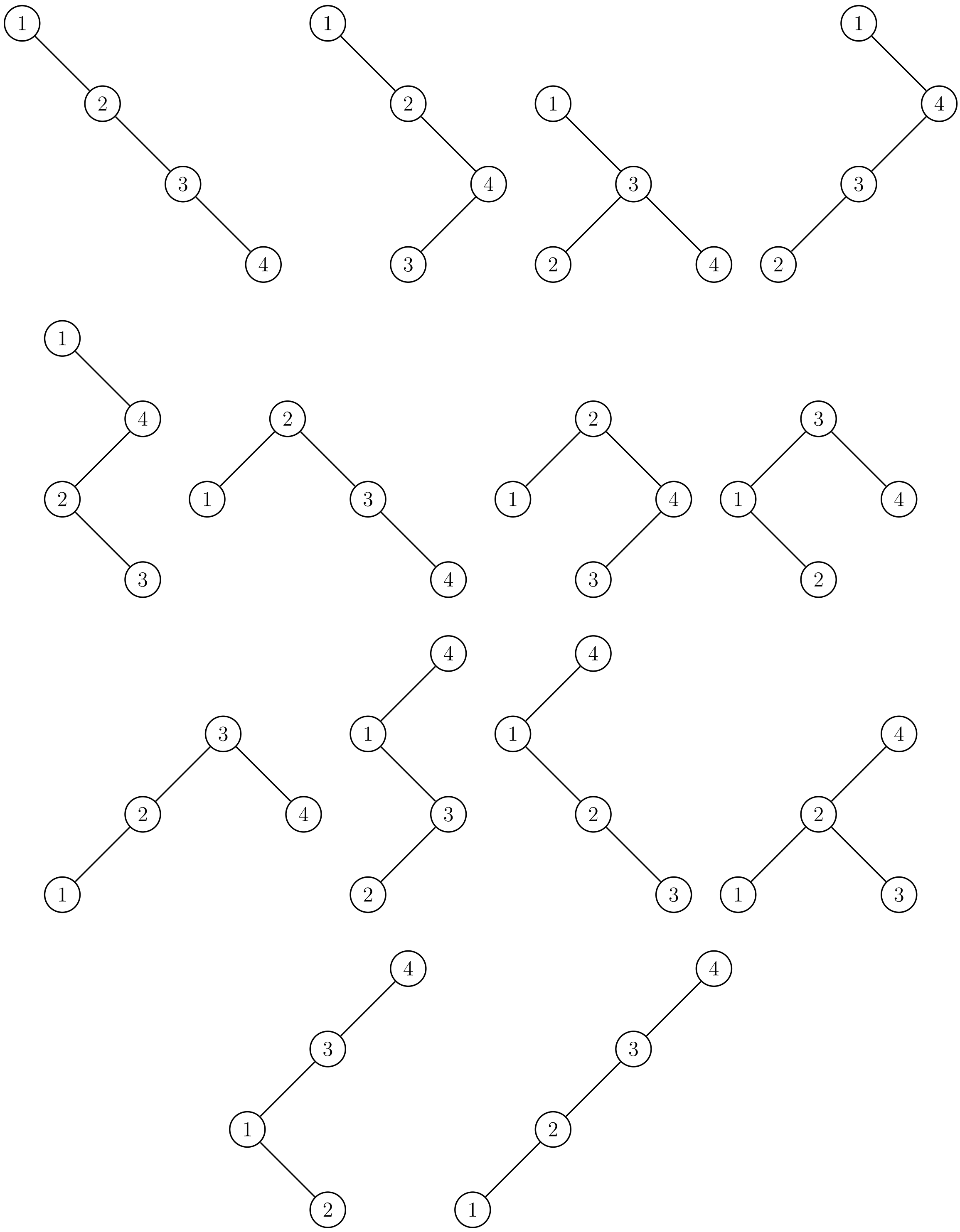
(a) Write down all binary search trees (BSTs) with 4 nodes.

(b) Compute the average height of a BST with 4 nodes (uniform distribution).

(c) Compute the expected value $E[H_4]$ (slide 153).

Solution

(a) In total we obtain 14 trees:



(b) For this part of the task, we indeed consider the uniform distribution on these 14 trees. We have 6 trees of height 2 and 8 trees of height 3. Hence the average height is $\frac{1}{14}(6 \cdot 2 + 8 \cdot 3) = \frac{18}{7} = 2\frac{4}{7}$.

(c) The expected value of the height of the BSTs chosen by the uniform distribution on the permutations S_4 yields a different number. We obtain

$$E[H_4] = \frac{2}{24}(3 + 3 + 2 \cdot 2 + 3 + 3 + 2 \cdot 3 \cdot 2) = \frac{7}{3} = 2\frac{1}{3}.$$

The flatter trees have more weight than with the uniform distribution and hence the expected value is smaller.