Homework 6 Solutions
Fundamental Algorithms, Spring 2025, Professor Yap, Section Leader Dr. Bingwei Zhang

Due: Fri Apr 18, in GradeScope by 11:30pm.
HOMEWORK with SOLUTION

INSTRUCTIONS:

- We have a "NO LATE HOMEWORK" policy.
  Special permission must be obtained *in advance* if you have a valid reason.
- Any submitted solution must be fully your own (you must not look at a fellow student's solution, *even if you have discussed with him or her*). Likewise, you must not show your writeup to anyone. We take the academic integrity policies of NYU and our department seriously. When in doubt, ask.
- The official deadline is 11:30pm. Since you can resubmit as many times as you like before that time, we basically do not accept excuses about missing the deadline.

(Q1) (5+8 Points)
Exercise VIII.1.10, p. 12.
Picking black and red balls from two urns.

THE QUESTION We have two urns of balls:

- Urn (I) has 2 black balls and 3 red balls.
- Urn (II) has 3 black balls and 2 red balls.

We toss a coin to pick an urn U (so U=(I) or U=(II) with equal probability).
HINT: use a decision tree model.

(a) What is the probability that a random ball from U is black?

(b) After drawing the first ball, we draw from the same urn U again. If the first ball is black, what is the probability the second ball is also black?

SOLUTION: (a): $\Pr(B) = \Pr(I)\Pr(B|I) + \Pr(II)\Pr(B|II) = \frac{1}{2}\cdot\frac{2}{5} + \frac{1}{2}\cdot\frac{3}{5} = 1/2$

(b) See SOLUTION Figure 1

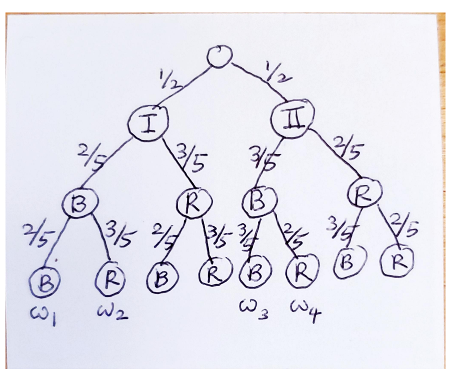


Figure 1: Decision Tree Model

SOLUTION: The sample space Ω corresponds to the 8 leaves of the decision tree in Figure 1. Let B_i be the event that the ith draw is black ($i = 1, 2$). From part(a), $\Pr(B_1) = \frac{1}{2}$. Thus $B_1 = \{\omega_1, \ldots, \omega_4\}$ in the figure.

$$\Pr(B_2|B_1) = \Pr(B_2B_1)/\Pr(B_1) = \Pr\{\omega_1, \omega_3\}/\tfrac{1}{2} = (\tfrac{1}{2}\tfrac{4}{25} + \tfrac{1}{2}\tfrac{9}{25})/\tfrac{1}{2} = \tfrac{4}{25} + \tfrac{9}{25} = \tfrac{13}{25}$$

NOTE: I had assumed that you put back the black ball you picked into its urn. This is called "choosing with replacement". It seems that some students assumed that the other model of "choosing without replacement". You will still get full credit with this assumption. In this case, the answer is $\Pr(B_2|B_1) = 2/5$.

Proof: using the same decision tree in Figure 1, we must now modify the probability of the edges into the leaves. Again, $\Pr(B_2B_1) = \Pr\{\omega_1, \omega_2\}$, but $\Pr(\omega_1) = \frac{1}{2}\frac{2}{5}\frac{1}{4} = \frac{1}{20}$ and $\Pr(\omega_2) = \frac{1}{2}\frac{3}{5}\frac{1}{2} = \frac{3}{20}$ (where the new probabilities are colored red). Thus

$$\Pr(B_2|B_1) = \Pr(B_2B_1)/\Pr(B_1) = (\tfrac{1}{20} + \tfrac{3}{20})/\tfrac{1}{2} = (\tfrac{1}{5})/\tfrac{1}{2} = \tfrac{2}{5}.$$

Comments: For part(b), $\Pr(B_2|B_1)$ is more than $\Pr(B_1) = \frac{1}{2}$, i.e., knowing that the first ball is black increases the probability of getting black again. We can iterate this process (keeping the same urn). What is the limit?

(Q2) (5+8 Points)
Exercise VIII.2.4, p. 15.
Placing the dice game when you only have coins.

THE QUESTION Professor X likes to play a dice game in his probability class, but has no dice. To simulate a dice roll, he asks three students to each toss a fair coin, yielding a binary number between 0 and 7. If 0 or 7 are tossed, the three coins are tossed again. The process is repeated until a number between 1 and 6 is tossed.

(a) Prove that this process simulates a fair dice. HINT: Use the Craps principle (in Exercise VIII.2.3, p. 15).

(b) What is the expected number of individual coin tosses needed to get a dice roll?

SOLUTION: (a) This is an application of the Craps principle: for $i = 1, \ldots, 6$, we have $\Pr\{i|1-6\} = \Pr\{i\}/\Pr\{1-6\}$. Since $\Pr\{i\} = 1/8$ and $\Pr\{1-6\} = 6/8$. Therefore, $\Pr\{i|1-6\} = 1/6$.

SOLUTION: (b) If the expected number of coin tosses is C, then $C = 3 + \frac{1}{4}C$ since, after the first 3 coin tosses, there is $\frac{1}{4}$ chance that we have to repeat the process. Solving, $C = 4$.
Comments: Alternative derivation: let N be the number of attempts until we get a number between 1 and 6. Then $\mathtt{E}[N] = \sum_{i \geqslant 1} i \cdot \Pr(N = i) = \sum_{i \geqslant 1} i \cdot pq^{i-1}$ where $p = 3/4$ (prob. of success) and $q = 1/4$ (prob. of failure). A standard derivation gives $\sum_{i \geqslant 1} i \cdot pq^{i-1} = 1/p$. Thus $\mathtt{E}[N] = 1/p = 4/3$.

(Q3) (8+8+8 Points)
Exercise VIII.3.1, p. 22.
Simulation of random permutation generation.
HOWEVER, WE WANT YOU TO BEGIN YOUR RANDOM NUMBERS STARTING FROM THE SECOND ROW. So the first five numbers are

0.333389 0.286361 0.172394 0.156509 0.495376

THE QUESTION In this question, we want to compute a random 9-permutation to be stored in the array $A[1..9]$. For random hand simulations, please follow the guidelines described in the beginning of the current set of Exercises (page 21). You must use the random numbers in the guidelines. The first 10 random numbers in our list are:

0.721383 0.381787 0.659242 0.254401 0.210082
0.333389 0.286361 0.172394 0.156509 0.495376

However, we want you to begin your random numbers starting from the second row, i.e., starting with 0.333389.

(a) Please hand simulate the RandPerm algorithm to compute $A[1..9]$.

(b) Please hand simulate the IncRandPerm algorithm to compute $A[1..9]$.

(c) Generally, we can use just the first two digits of the random numbers in calculation. E.g., the first number is 0.721383, but you could just use 0.72. How can you justify this simplification, and when must you use more the first two digits?

Both parts(a) and (b) should use the same list of random numbers, i.e., both start from the second row of the list.

(a) Solution from Aditeya in Figure 2

Q3 (a) Let A = [1, 2, 3, 4, 5, 6, 7, 8, 9]

i	r	x	A → swap (A[i] ↔ A[x])
9	0.333389	$\lfloor 9 \times 0.333389 \rfloor + 1 = 4$	[1, 2, 3, 9, 5, 6, 7, 8, 4]
8	0.286361	$\lfloor 8 \times 0.286361 \rfloor + 1 = 3$	[1, 2, 8, 9, 5, 6, 7, 3, 4]
7	0.172394	$\lfloor 7 \times 0.172394 \rfloor + 1 = 2$	[1, 7, 8, 9, 5, 6, 2, 3, 4]
6	0.156509	$\lfloor 6 \times 0.156509 \rfloor + 1 = 1$	[6, 7, 8, 9, 5, 1, 2, 3, 4]
5	0.495376	$\lfloor 5 \times 0.495376 \rfloor + 1 = 3$	[6, 7, 5, 9, 8, 1, 2, 3, 4]
4	0.352533	$\lfloor 4 \times 0.352533 \rfloor + 1 = 2$	[6, 9, 5, 7, 8, 1, 2, 3, 4]
3	0.521332	$\lfloor 3 \times 0.521332 \rfloor + 1 = 2$	[6, 5, 9, 7, 8, 1, 2, 3, 4]
2	0.924328	$\lfloor 2 \times 0.924328 \rfloor + 1 = 2$	(no swap since i = x)

⇒ $\boxed{A = [6, 5, 9, 7, 8, 1, 2, 3, 4]}$

Figure 2: Simulating RandPerm on $A[1..9]$

(b) Solution from Aditeya in Figure 3

SOLUTION: (a) See Figure 2 for the solution of your classmate Aditeya .

(b) See Figure 3 for the solution of your classmate Aditeya .

(c) Justification of using 2 digits: we approximate a random number x by using $\widetilde{x}$ with only 2 digits. The error in $\widetilde{x}$ is $|x - \widetilde{x}| \leqslant 0.00999\cdots = 0.01$. So the error in $n \times \widetilde{x}$ is at most $0.01n$. In our above examples, $n \leqslant 9$ and so the error is at most $0.01n \leqslant 0.09$. This can affect our value of $\lfloor n \times x \rfloor$ only if the fractional part (call it θ) of our approximation to $\lfloor n \times x \rfloor$ is more than 0.91 (since $0.91 + 0.09 = 1.00$). In summary: you can do parts (a) and (b) using just 2 digits approximations, but use more digits if you found that the $\theta \geqslant 0.91$. You can do a similar analysis if you use k digit approximations of x.

Example: in part(a), the first approximation $\lfloor 9 \times 0.33 \rfloor + 1 = \lfloor 2.97 \rfloor + 1 = 2 + 1 = 3$ is wrong since $\theta = 0.97 > 0.91$. But the next approximation, $\lfloor 8 \times 0.28 \rfloor + 1 = \lfloor 2.24 \rfloor + 1 = 2 + 1 = 3$ is correct since $\theta = 0.24 < 0.91$.

(Q4) (20 Points
Exercise VIII.4.1, p. 26.

(b)

$A[i]=i$	r	X	$A \to$ swap $(A[i] \leftrightarrow A[X])$
1			[1]
2	0.333389	$\lfloor 2 \times 0.333389 \rfloor + 1 = 1$	[2, 1]
3	0.286361	$\lfloor 3 \times 0.286361 \rfloor + 1 = 1$	[3, 1, 2]
4	0.172394	$\lfloor 4 \times 0.172394 \rfloor + 1 = 1$	[4, 1, 2, 3]
5	0.156509	$\lfloor 5 \times 0.156509 \rfloor + 1 = 1$	[5, 1, 2, 3, 4]
6	0.495376	$\lfloor 6 \times 0.495376 \rfloor + 1 = 3$	[5, 1, 6, 3, 4, 2]
7	0.352533	$\lfloor 7 \times 0.352533 \rfloor + 1 = 3$	[5, 1, 7, 3, 4, 2, 6]
8	0.521332	$\lfloor 8 \times 0.521332 \rfloor + 1 = 5$	[5, 1, 7, 3, 8, 2, 6, 4]
9	0.924328	$\lfloor 9 \times 0.924328 \rfloor + 1 = 9$	[5, 1, 7, 3, 8, 2, 6, 4, 9]

$\Rightarrow \boxed{A = [5, 1, 7, 3, 8, 2, 6, 4, 9]}$

Figure 3: Simulating IncRandPerm on $A[1..9]$

Play until bust in Las Vegas.
NOTE: I do not know how compute the value (but let me know if you think you know).

THE QUESTION (Play until Bust in Las Vegas)
On a trip to Las Vegas, you stopped over at a casino to play a game of chance. You have \$10, and each play costs \$2. You have 1/3 chance to win \$5, and 2/3 chance to win nothing. You just want to play as long as you have enough money. How many games do you expect to play in total? *Be careful! If you cannot give the exact answer, give upper and lower bounds.*

SOLUTION: Let B be your balance when bust. Note that B is either 0 or 1. It follows that

$$0 \leqslant \mathtt{E}\big[B\big] \leqslant 1.$$

We show that both these inequalities are actually strict:

$$0 < \mathtt{E}\big[B\big] < 1. \tag{1}$$

Note that $\Pr\{B=0\} + \Pr\{B=1\} = 1$ and $\mathtt{E}\big[B\big] = \Pr(B=1)$. From $\Pr\{B=0\} \geqslant \Pr\{TTTTT\} = (2/3)^5 > 0$, we conclude that $\mathtt{E}\big[B\big] = \Pr(B=1) < 1$. From $\Pr\{B=1\} \geqslant \Pr\{TTTTHTTT\} = (2/3)^6(1/3) > 0$, we conclude that $\mathtt{E}\big[B\big] = \Pr(B=1) > 0$. Since the expected win per game is $(1/3)5 - 2 = -1/3$, it follows that $\mathtt{E}\big[B\big] = 10 - \mathtt{E}\big[N\big]/3$. Combined with (1), we conclude that $0 < 10 - \mathtt{E}\big[N\big]/3 < 1$ or

$$27 < \mathtt{E}\big[N\big] < 30 \tag{2}$$

Students in our class reported that $\mathtt{E}\big[N\big] \sim 28.41$ by doing a million random simulations, which is of course consistent with our bound.

What if we want an exact solution?

We can set up a linear recurrence,

$$E_n = -3 + 3E_{n-3} - 2E_{n-5}$$

with $E_0 = E_1 = 0$. Using the transformation (from Hw3),

$$t(n) = -3n + E_n$$

we can obtain the homogeneous linear recurrence

$$t(n) = 3t(n-3) - 2t(n-5)$$

which has roots

$$(x_1, \ldots, x_5) = (1, 1.12, -0.76, -0.68 + \boldsymbol{i}1.37, -0.68 - \boldsymbol{i}1.37).$$

Note that 3 are real roots, but we have a pair of complex conjugate roots. So we know the general solution is

$$t(n) = \sum_{i=1}^{5} A_i x_i^n$$

where A_i's are constants to be determined from initial conditions. Only 4 of these constants are independent, because A_4 and A_5 are conjugates. UNFORTUNATELY, we have only two initial conditions. How to proceed?
Some students propose to use random simulations to obtain two or three more constants, $E_2 = 5.28552191, E_3 = 6.5444775$ and possibly $E_4 = 10.87600969$. Then we can solve the recurrences. Of course, this is still an approximate numerical solution. My Matlab code gives $E_10 \sim 28.3771$.
Note that this is really an infinite Markov chain and there seems to be no general solutions.

(Q5) (18 Points)
Exercise VIII.15.1, p. 87.
Basic treapifying.

THE QUESTION Let $n = 5$ and $\sigma = (3, 1, 5, 2, 4)$ *i.e.*, $\sigma(3) = 5$ and $\sigma(4) = 1$. Draw the treap. Next, change the priority of key4 to 10 (from 1) and treapify. Next, change the priority of key 3 to 0 (from 5) and treapify.

SOLUTION: See Figure 4 for the solution of your classmate Aditeya.

Treapify solution in Figure 4

(Q6) (18 Points)

Treap simulation using digits of π.

To simulate Treaps, you must assign a random priority to each key k. Please use the list of random numbers in p. 22 (before Ex. VIII.3.1).

Please insert or delete the first 15 digits of π starting from an initially empty treap:

$$3141'5926'5358'979$$

For each digit d, if that digit is not in the current treap, do $Insert(d)$; otherwise, do $Delete(d)$.

Show successive treaps at the end of each 15 operations; but be sure to show intermediate almost-treaps. To keep track of priorities of the current set of keys in the treap, you may represent them as a priority list. Note that after you delete a digit, its priority is deleted. So if it is reinserted, you must give it a new priority.

(Solution in Figure 5)

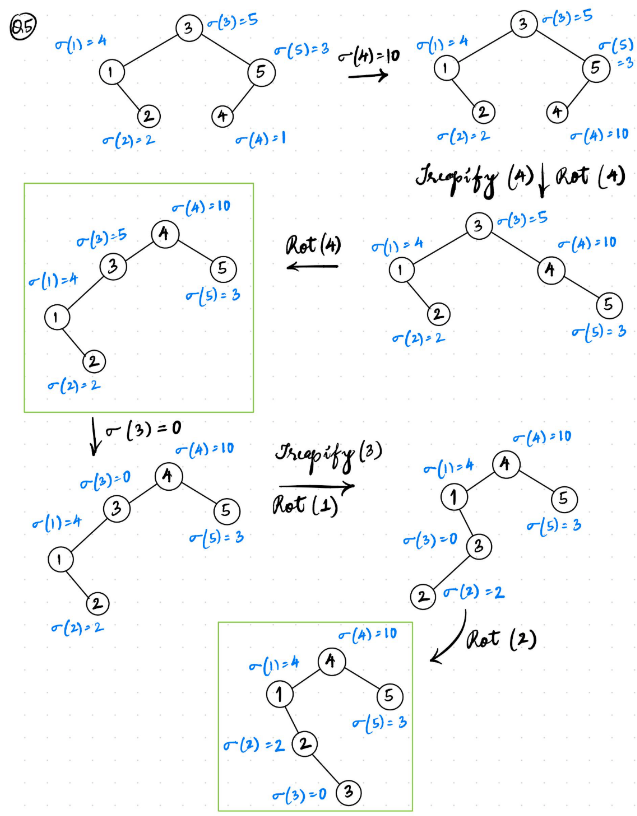


Figure 4: Treapifying an overtreap and undertreap on 5 keys

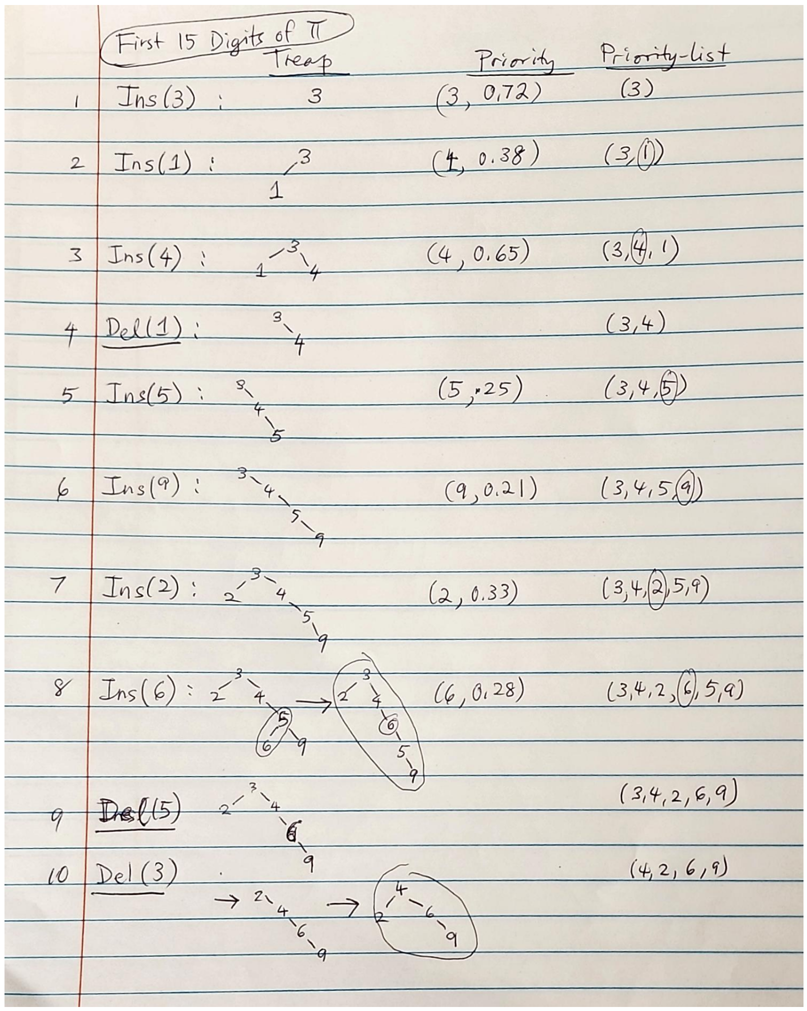


Figure 5: Inserting/Deleting the first 10 digits of π

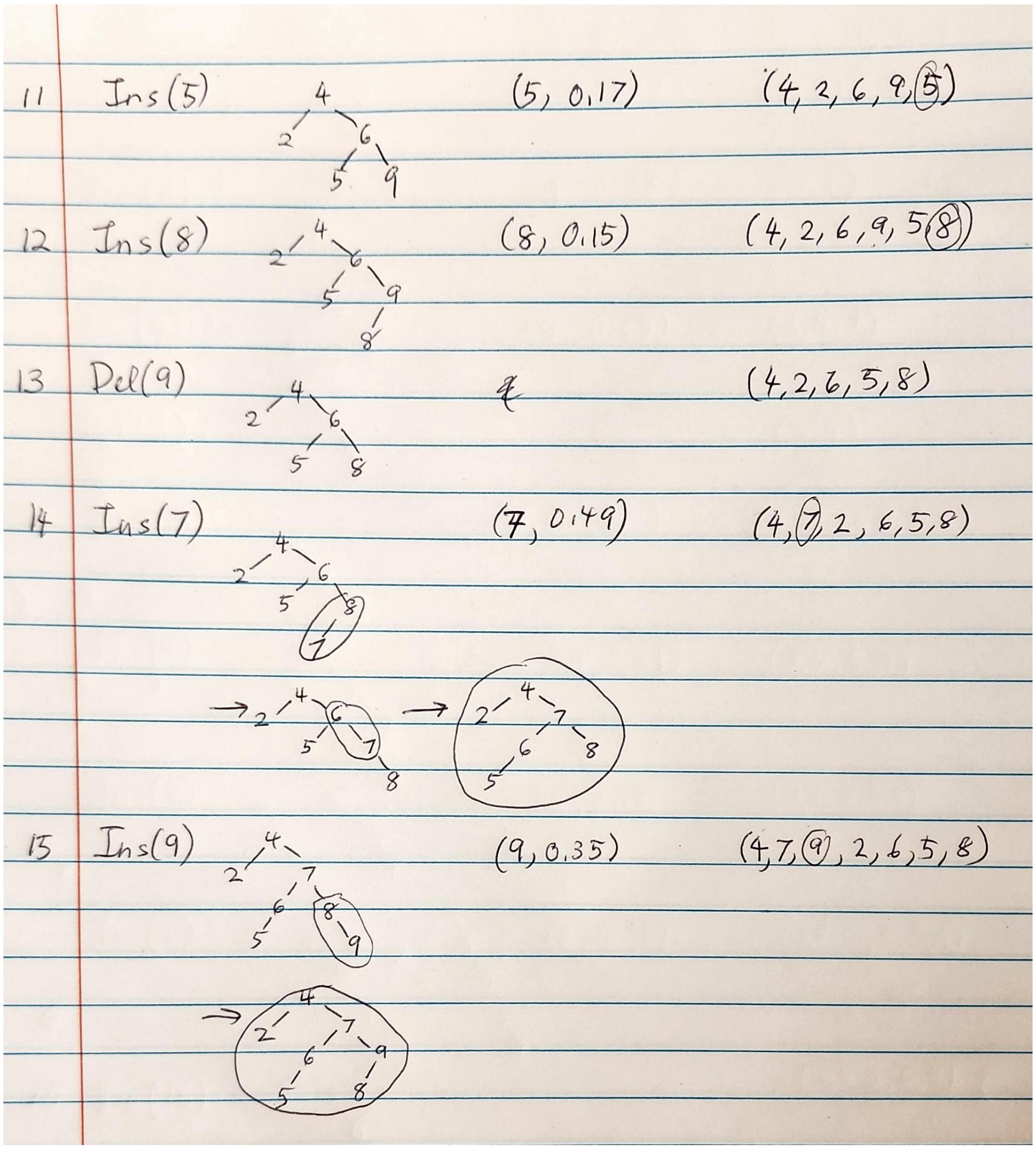


Figure 6: Inserting/Deleting the next 5 digits of π

(Q7) (16 Points)

Simulation of universal hashing.
Please read ¶XI.23, p. 25.

Let K be the set of the following 9 key words:

$$\begin{array}{c}\texttt{fig}, \texttt{plum}, \texttt{kiwi}, \texttt{pear}, \texttt{lime}, \texttt{date}, \\ \texttt{apple}, \texttt{melon}, \texttt{mango},\end{array} \tag{3}$$

We consider K to be a subset of $U := \mathbb{Z}_{29}^5$ where each letter $[a..z]$ is viewed as a number $[1..26]$, and a word with less than 5 letters is padded with 0's. E.g., `lime` is $(12, 9, 13, 5, 0) \in U$.

Let $\boldsymbol{a} = (a_0, \ldots, a_5) \in \Lambda := \mathbb{Z}_{29}^6$. If $\boldsymbol{x} = (x_1, \ldots, x_5) \in U$, then define $h_{\boldsymbol{a}}(\boldsymbol{x}) = (a_0 + \sum_{i=1}^5 a_i x_i) \,\mathbf{mod}\, m$ for some fixed m to be determined. Thus,

$$H_m = \{h_{\boldsymbol{a}} : \boldsymbol{a} \in \Lambda\} \subseteq [U \to \mathbb{Z}_m]$$

is a universal hash set (see ¶XI.23, p. 23). Let us pick a random $\boldsymbol{a} \in \Lambda$. Then, for each $m = 9, m = 10, m = 11, \ldots$, we test if $h_{\boldsymbol{a}}$ is perfect for the set K in (3). We stop at the first m for which it is perfect. Note that the same $\boldsymbol{a}$ is fixed once and for all. To find $\boldsymbol{a}$, use the list of random numbers from Exercise VIII.3.1, p. 23. Please tell us how you do the computation (you may use calculators or programs).

> Here is a precise instructions of what to submit for your computations:
> (1) What is $\boldsymbol{a} = (a_0, \ldots, a_5)$?
> (2) For each $k \in K$ in (3), if $k = (x_1, \ldots, x_5)$, please show the value $\sum_{i=1}^5 a_i x_i$. Call this value $g_{\boldsymbol{a}}(k)$. Note that $h_{\boldsymbol{a}}(k) = g_{\boldsymbol{a}}(k) \,\mathbf{mod}\, m$.
> (3) For each $m = 9, 10, \ldots$, if $h_{\boldsymbol{a}}$ is not perferct for m, show us the **total conflict** value. This is the value $C_m = \sum_{i=0}^{m-1} \binom{b_i}{2}$ where b_i is the size of the ith bucket.
> (4) Finally, when found the m for which $h_{\boldsymbol{a}}$ is perfect on K, show the values of $h_{\boldsymbol{a}}(k)$, for each $k \in K$.

Q7 Padded Mapping $(\mathbb{Z}_{29}^5) \to$

fig = (6, 9, 7, 0, 0)

plum = (16, 12, 21, 13, 0)

kiwi = (11, 9, 23, 9, 0)

pear = (16, 5, 1, 18, 0)

lime = (12, 9, 13, 5, 0)

date = (4, 1, 20, 5, 0)

apple = (1, 16, 16, 12, 5)

melon = (13, 5, 12, 15, 14)

mango = (13, 1, 14, 7, 1, 5)

Computing $a = (a_0 \ldots a_5) \to$

$a_0 = \lfloor 29 \times 0.721383 \rfloor = 20$

$a_1 = \lfloor 29 \times 0.381787 \rfloor = 11$

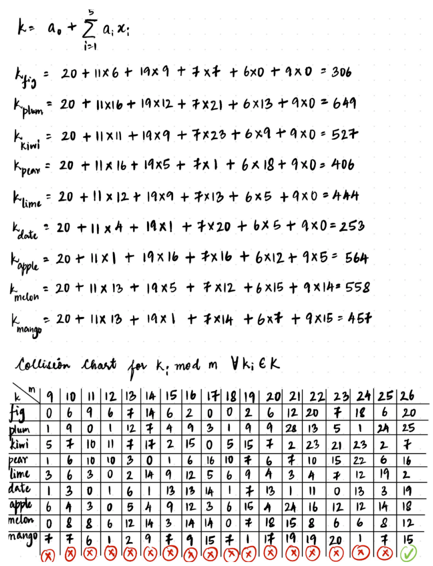
$a_2 = \lfloor 29 \times 0.659242 \rfloor = 19$

$a_3 = \lfloor 29 \times 0.254401 \rfloor = 7$

$a_4 = \lfloor 29 \times 0.210082 \rfloor = 6$

$a_5 = \lfloor 29 \times 0.333389 \rfloor = 9$

Figure 7: The random hash function



$k = a_0 + \sum_{i=1}^{5} a_i x_i$

$k_{fig} = 20 + 11\times 6 + 19\times 9 + 7\times 7 + 6\times 0 + 9\times 0 = 306$

$k_{plum} = 20 + 11\times 16 + 19\times 12 + 7\times 21 + 6\times 13 + 9\times 0 = 649$

$k_{kiwi} = 20 + 11\times 11 + 19\times 9 + 7\times 23 + 6\times 9 + 9\times 0 = 527$

$k_{pear} = 20 + 11\times 16 + 19\times 5 + 7\times 1 + 6\times 18 + 9\times 0 = 406$

$k_{lime} = 20 + 11\times 12 + 19\times 9 + 7\times 13 + 6\times 5 + 9\times 0 = 444$

$k_{date} = 20 + 11\times 4 + 19\times 1 + 7\times 20 + 6\times 5 + 9\times 0 = 253$

$k_{apple} = 20 + 11\times 1 + 19\times 16 + 7\times 16 + 6\times 12 + 9\times 5 = 564$

$k_{melon} = 20 + 11\times 13 + 19\times 5 + 7\times 12 + 6\times 15 + 9\times 14 = 558$

$k_{mango} = 20 + 11\times 13 + 19\times 1 + 7\times 14 + 6\times 7 + 9\times 15 = 457$

Collision Chart for k_i mod m $\forall k_i \in K$

k \ m	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
fig	0	6	9	6	7	14	6	2	0	0	2	6	12	20	7	18	6	20
plum	1	9	0	1	12	7	4	9	3	1	9	9	28	13	5	1	24	25
kiwi	5	7	10	11	7	17	2	15	0	5	15	7	2	23	21	23	2	7
pear	1	6	10	10	3	0	1	6	16	10	7	6	7	10	15	22	6	16
lime	3	6	3	0	2	14	9	12	5	6	9	4	3	4	7	12	19	2
date	1	3	0	1	6	1	13	13	14	1	7	13	1	11	0	13	3	19
apple	6	4	3	0	5	4	9	12	3	6	15	4	24	16	12	12	14	18
melon	0	8	8	6	12	14	3	14	14	0	7	18	15	8	6	6	8	12
mango	7	7	6	1	2	9	7	9	15	7	1	17	19	19	20	1	7	15
	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓

Figure 8: The hashing of fruits using table sizes $m = 9, 10, \ldots, 26$

SOLUTION:

(1) See the solution of your classmate Aditeya in Figure 7.

(2) See the solution of your classmate Aditeya in Figure 8.

(3) See Figure 8. We see that the first value of m such that $h_{\boldsymbol{a}}$ is perfect on K is $m = 26$.

Comments: How did we know that the perfect m exists in part(3) above? Well, the set $G = \{g_{\boldsymbol{a}}(k) : k \in K\}$ has a maximum value M, and $m \leqslant M$. That is true provided $|G| = 9$. It is highly unlikely that $|G| < 9$, otherwise we would have to look for another $\boldsymbol{a}$, but that is not part of my question.

(Q8) (20 Points)

Simulation of FKS optimal static hashing.
Please read ¶XI.29, p. 34.

Assume that we are using the universal hash family $H_m = \{h_a : a \in \mathbb{Z}_p\}$ where $h_a(x) = (ax \,\mathbf{mod}\, p) \,\mathbf{mod}\, m)$ where $p = 31$ but m will be variously chosen according to the FKS scheme.

Construct a FKS scheme for the following input: $K = \{2, 4, 9, 18, 20, 28, 30\}$.

To choose your hash function h_a, please choose a from successive values in the list $2, 3, 4, 5, \ldots$ until you find a suitable a (pretend that this list is "random").

Solution Figure 9

SOLUTION: Since $n = |K| = 7$, the first hash function we choose is $h_2(x) = ((2x \,\mathbf{mod}\, 31) \,\mathbf{mod}\, 7)$ as given in this table:

$x \in K:$	2	4	9	18	20	28	30
$h(x)$	4	1	4	1	5	4	4

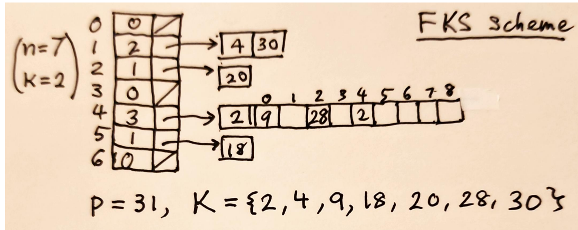


Figure 9: A FKS scheme for $p = 31, K = \{2, 4, 9, 18, 20, 28, 30\}$.

The only secondary key we need to provide is for the bin $B_4 = \{2, 9, 28\}$. So $b_4 = |B_4| = 3$.

To h "randomly choose the secondary key k_4 to be 2 and check that it is perfect. So $h_4(x) = ((2x \textbf{ mod } p) \textbf{ mod } 9)$ and we see from the next table that it is perfect:

$x \in B_4$:	2	9	28
$h_4(x)$	4	0	2

Finally, the FKS scheme is shown in Figure 9.