INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR	
	Stamp / Signature of the Invigilator

EXAMINATION (Mid Semester)	SEMESTER (Autumn)

Roll Number									**Section**		**Name**
Subject Number	C	S	6	0	0	3	5	**Subject Name**	*Selected Topics in Algorithms*		
Department / Center of the Student								**Additional sheets**			

Instructions and Guidelines to Students Appearing in the Examination

1. Ensure that you have occupied the seat as per the examination schedule.
2. Ensure that you do not have a mobile phone or a similar gadget with you even in switched off mode. Note that loose papers, notes, books should not be in your possession, even if those are irrelevant to the paper you are writing.
3. Data book, codes or any other materials are allowed only under the instruction of the paper-setter.
4. Use of instrument box, pencil box and non-programmable calculator is allowed during the examination. However, exchange of these items is not permitted.
5. Additional sheets, graph papers and relevant tables will be provided on request.
6. Write on both sides of the answer script and do not tear off any page. Report to the invigilator if the answer script has torn page(s).
7. Show the admit card / identity card whenever asked for by the invigilator. It is your responsibility to ensure that your attendance is recorded by the invigilator.
8. You may leave the examination hall for wash room or for drinking water, but not before one hour after the commencement of the examination. Record your absence from the examination hall in the register provided. Smoking and consumption of any kind of beverages is not allowed inside the examination hall.
9. After the completion of the examination, do not leave the seat until the invigilator collects the answer script.
10. During the examination, either inside the examination hall or outside the examination hall, gathering information from any kind of sources or any such attempts, exchange or helping in exchange of information with others or any such attempts will be treated as adopting ‘unfair means’. Do not adopt ‘unfair means’ and do not indulge in unseemly behavior as well.

Violation of any of the instructions may lead to disciplinary action.

Signature of the Student

To be filled in by the examiner

Question Number	1	2	3	4	5	6	7	8	9	10	Total
Marks Obtained											

Marks obtained (in words)	Signature of the Examiner	Signature of the Scrutineer

CS60035/CS60086 Selected Topics in Algorithms, Autumn 2015–2016

Mid-Semester Test

21–September–2015 | CSE-107, 2:00–4:00pm $(+\,\varepsilon)$ | Maximum marks: 30

[*Write your answers in the question paper itself. Be brief and precise. Answer all questions.*]

1. [*Recap*: Let $\mathscr{H}$ be a finite family of hash functions mapping a universe $\mathscr{U}$ of keys to $\{0,1,2,\ldots,m-1\}$. We call $\mathscr{H}$ *universal* if for every $x,y \in \mathscr{U}$, $x \neq y$, we have $\Pr_{h \in_U \mathscr{H}}[h(x) = h(y)] \leqslant \frac{1}{m}$.]

Let $\mathscr{U}$ be the set of all n-bit binary strings (which are naturally identified as n-dimensional column vectors). Also, let $m = 2^b$, so the binary representation of a hash value can be identified as a binary b-dimensional column vector. Consider the family $\mathscr{H}$ parameterized by all $b \times n$ binary matrices. For such a matrix M, define the hash function $h_M \in \mathscr{H}$ as $h_M(x) = Mx$, where matrix multiplication is carried out modulo two. Prove that this family of hash functions is universal. **(8)**

Solution Let $x = (x_1, x_2, \ldots, x_n)^{\mathrm{t}}$ and $y = (y_1, y_2, \ldots, y_n)^{\mathrm{t}}$ be two different keys from $\mathscr{U}$. Since $x \neq y$, we have $x_i \neq y_i$ for at least one i. Fix one such i. Let $M = (m_{ij})$, $u = Mx = (u_1, u_2, \ldots, u_b)^{\mathrm{t}}$, and $v = My = (v_1, v_2, \ldots, v_b)^{\mathrm{t}}$. For $j \in \{1,2,\ldots,b\}$, the condition $u_j = v_j$ is equivalent to $\sum_{k=1}^{n} m_{jk} x_k \equiv \sum_{k=1}^{n} m_{jk} y_k \pmod 2$, that is,

$$m_{ji} \equiv \sum_{\substack{k=1\\k\neq i}}^{n} m_{jk}(x_k - y_k) \pmod 2.$$

For each choice of m_{jk}, $k = 1,2,\ldots,n$, $k \neq i$, there is a unique solution for m_{ji} to have the collision $u_j = v_j$ in the j-th component of the hash values. Since $m_{ji} \in_U \{0,1\}$, the chosen m_{ji} equals this solution with probability $\frac{1}{2}$. The possibility $u_j = v_j$ is influenced by the choice of the j-th row of M, that is, for two different values of j, the rows are different, and their choices are independent of one another. Therefore, the probability of a collision $h_M(x) = h_M(y)$ is the probability of $u_j = v_j$ for all $j = 1,2,\ldots,b$, which is $\frac{1}{2^b} = \frac{1}{m}$.

2. [*Coupon collector's problem*] The UVW Chocolate Company prepares large numbers of n different types of coupons, and inserts them in chocolate packets to be sold in a city. The company makes sure that it has distributed an equal number of coupons of each type. If a customer can produce all of the n types of coupons to the company, she will receive a sedan as a gift from the company. In order that the company does not end up gifting too many sedans, it adopts a trick. It distributes the coupons in such a way that people from any given locality experience scarcity of some coupon types. Ms. Randoma is determined to win a gift. She guesses that the company may have played some tricks with its customers. She makes a whole-day tour in the entire city, chooses shops at random locations, and buys random chocolate packets. Your task is to help her by estimating the expected number of chocolate packets she should buy in order to win the gift.

(a) Let an event E occur with probability $\alpha \in (0,1]$ in each trial of a random experiment. If the trials are repeated independently, what is the expected number of trials necessary to have the first occurrence of E? **(4)**

Solution For $i \in \mathbb{N} = \{1,2,3,\ldots\}$, the probability that the first occurrence of E is in the i-th trial is $p_i = (1-\alpha)^{i-1}\alpha$. Therefore, the expected number of trials is

$$\sum_{i\in\mathbb{N}} ip_i = \alpha\Big[1+2(1-\alpha)+3(1-\alpha)^2+4(1-\alpha)^3+\cdots\Big] = \frac{\alpha}{\Big(1-(1-\alpha)\Big)^2} = \frac{1}{\alpha}.$$

Here, the geometric series converges, because $1-\alpha \in [0,1)$.

(b) Solve the problem of Ms. Randoma, that is, find the expected number of chocolate packets she should buy in order that all of the n types of coupons are available to her. Because Ms. Randoma picks chocolate packets randomly, assume that in each packet, each of the n types of coupons is equally likely to occur. **(4)**

Solution Let Z denote the random variable standing for the number of packets needed for winning the gift. We need to compute $\mathrm{E}[Z]$. For $i = 1,2,3,\ldots,n$, define Z_i to be the random variable that stands for the number of packets to be bought to get the i-th type of coupon *after* exactly $i-1$ different types of coupons are available. We have $Z = Z_1+Z_2+\cdots+Z_n$, so by linearity of expectation, $\mathrm{E}[Z] = E[Z_1]+\mathrm{E}[Z_2]+\cdots+\mathrm{E}[Z_n]$. In order to determine $\mathrm{E}[Z_i]$, we note that Ms. Randoma has $i-1$ different types of coupons, and she needs any one of the remaining $n-i+1$ types of coupons. The probability that a randomly chosen packet contains a new type of coupon is $\alpha_i = \frac{n-i+1}{n}$. By Part (a), $\mathrm{E}[Z_i] = \frac{1}{\alpha_i} = \frac{n}{n-i+1}$, and it follows that

$$\mathrm{E}[Z] = n\sum_{i=1}^{n}\frac{1}{n-i+1} = nH_n,$$

where H_n is the n-th harmonic number. Since $H_n \approx \ln n$ (indeed $\ln(n+1) \leqslant H_n \leqslant \ln n + 1$ for all n), about $n\ln n$ packets should be bought to expect a win with significant probability.

3. Let $C = (v_1, v_2, \ldots, v_n)$ be a cycle of length $n \geqslant 3$ in an undirected graph. The goal is to produce a *proper 3-coloring* of C (that is, a coloring of $v_1, v_2, \ldots, v_n$ by three colors R, G, and B such that no edge of the cycle has two endpoints of the same color). The following algorithm produces a random 3-coloring of C. Let c_i denote the color that the vertex v_i gets.

> 1. Randomly color v_1 by $c_1 \in_U \{R, G, B\}$.
> 2. For $i = 2, 3, \ldots, n$, randomly color v_i by $c_i \in_U \{R, G, B\} \setminus \{c_{i-1}\}$.

(a) Prove that the above algorithm produces a proper 3-coloring of C with probability $\geqslant \frac{1}{2}$. **(4)**

Solution If the random coloring produces an improper coloring of C, this means that $c_1 = c_n$ (the algorithm already insures that $c_i \neq c_{i-1}$ for all $i = 2, 3, \ldots, n$). For each such improper coloring, we can recolor v_n by the unique color in $\{R, G, B\} \setminus \{c_1, c_{n-1}\}$. That is, for each improper coloring, there is one proper coloring. Moreover, for two different improper colorings, the corresponding proper colorings derived as above are distinct too. This implies the given probability bound.

(b) We need to repeat the above algorithm a constant expected number of times to get a proper coloring of C. You instead modify the algorithm slightly such that only one run of the modified algorithm produces a random proper 3-coloring of C. **(4)**

Solution The following algorithm uses a modified strategy for coloring v_n. If $c_1 = c_{n-1}$, there are two choices for c_n. If $c_1 \neq c_{n-1}$, there is a unique choice for c_n.

> 1. Randomly color v_1 by $c_1 \in_U \{R, G, B\}$.
> 2. For $i = 2, 3, \ldots, n-1$, randomly color v_i by $c_i \in_U \{R, G, B\} \setminus \{c_{i-1}\}$.
> 3. Randomly color v_n by $c_n \in_U \{R, G, B\} \setminus \{c_{n-1}, c_1\}$.

4. Let S be the unit square in the plane (that is, the square with corners at $(0,0)$, $(0,1)$, $(1,1)$ and $(1,0)$), and let C be the circle of radius $\frac{1}{2}$ and center at $(\frac{1}{2},\frac{1}{2})$. Notice that S has area one, whereas C has area $\frac{\pi}{4}$.

(a) Propose a randomized algorithm for approximating the value of π. The algorithm should be based upon randomly choosing t points inside S. **(4)**

Solution

1. Set $Z = 0$.
2. Repeat t times:
   (a) Choose real numbers $x, y \in_U [0,1]$.
   (b) If $\left(x-\frac{1}{2}\right)^2 + \left(y-\frac{1}{2}\right)^2 \leqslant \frac{1}{4}$, set $Z = Z+1$.
3. Return $\frac{4Z}{t}$.

(b) Let π' be an approximate value of π returned by the algorithm of Part (a). Let $\varepsilon \in (0,1)$ be an upper bound on the relative approximation error. Estimate the number t of points to be chosen in Part (a) to ensure that $\Pr\left[\left|\pi' - \pi\right| \leqslant \varepsilon\pi\right] \geqslant \frac{1}{2}$.

(**Hint:** Let $Z_1, Z_2, \ldots, Z_t$ be t independent and identically distributed indicator variables, each with mean μ and standard deviation σ. Then, $\Pr\left[\left|\frac{1}{t}\sum_{i=1}^{t} Z_i - \mu\right| \leqslant \frac{3\sigma}{\sqrt{t}}\right]$ is a positive constant (close to one, for t large enough). This follows readily from the *central limit theorem*. You may use this result without proving it.) **(4)**

Solution Let Z_i denote the indicator variable denoting whether $(x,y) \in C$ in the i-th iteration of the loop. For each i, we have $\Pr[Z_i = 1] = \frac{\pi}{4}$, and $\Pr[Z_i = 0] = 1 - \frac{\pi}{4}$. Consequently, $\mathrm{E}[Z_i] = 1 \times (\frac{\pi}{4}) + 0 \times (1 - \frac{\pi}{4}) = \frac{\pi}{4}$ for each i. Let $Z = \frac{1}{t}(Z_1 + Z_2 + \cdots + Z_t)$. By linearity of expectation, we have $\mathrm{E}[Z] = \frac{\pi}{4}$. Moreover, $E[Z_i^2] = 1^2 \times \frac{\pi}{4} + 0^2 \times (1 - \frac{\pi}{4}) = \frac{\pi}{4}$. Therefore $\sigma = \sqrt{E[Z_i^2] - E[Z_i]^2} = \sqrt{\frac{\pi}{4}(1 - \frac{\pi}{4})}$. By the central limit theorem, it follows that

$$\Pr\left[\left|Z - \frac{\pi}{4}\right| \leqslant \frac{3\sqrt{\frac{\pi}{4}(1-\frac{\pi}{4})}}{\sqrt{t}}\right] \geqslant \delta$$

for some constant $\delta \in (0,1]$. Therefore, if we insure that

$$\frac{12\sqrt{\frac{\pi}{4}(1-\frac{\pi}{4})}}{\sqrt{t}} \leqslant \varepsilon\pi,$$

we get a maximum relative error of ε. This requires

$$t \geqslant \frac{9(4-\pi)}{\pi\varepsilon^2}.$$

If $\delta \geqslant 0.5$, these many trials suffice. Otherwise, $\Theta(\frac{1}{2\delta})$ times these many trials suffice.

Sequel: Ms. Randoma is driving her new sedan to meet her endodontist. She needs to treat her abscessed teeth caused by eating too much chocolate.