HKUST – Department of Computer Science and Engineering

COMP 2711: Discrete Math Tools for Computer Science

Spring 2019 Final Examination

Date: Friday, 24 May 2019 Time: 08:30–11:30

Name: ______________________	Student ID: ______________________
Email: ______________________	Lecture: ______________________

Instructions

- This is a closed-book exam. It consists of 13 regular questions and one bonus question.
- Please write your name, student ID, email, lecture section in the space provided at the top of this page.
- Please sign the honor code statement on page 2.
- Answer all the questions within the space provided on the examination paper. Each question is on a separate page. This is for clarity and is not meant to imply that each question requires a full-page answer.
- Unless required otherwise, solutions can be written in terms of exponentiation, binomial coefficients, factorials, and the $C(n, k)$, $P(n, k)$ notation. For example, you can write $\binom{5}{3} + \binom{4}{2} + 3^5$ instead of 259. Avoid using nonstandard notation such as ${}_nP_k$ or ${}_nC_k$.
- Calculators may be used for the exam (but are not necessary).

Questions	1	2	3	4	5	6	7	8	9	10	11	12	13	Total	Bonus
Points	8	4	8	8	8	8	8	10	4	9	11	8	6	100	10
Score															

Student ID: ____________

As part of HKUST's introduction of an honor code, the HKUST Senate has recommended that all students be asked to sign a brief declaration printed on examination answer books that their answers are their own work, and that they are aware of the regulations relating to academic integrity. Following this, please read and sign the declaration below.

> I declare that the answers submitted for this examination are my own work.
>
> I understand that sanctions will be imposed, if I am found to have violated the University regulations governing academic integrity.
>
> Student's Name: ______________________________
>
> Student's Signature: __________________________

Student ID: _________

Problem 1: [8 pts] For each of the following statements, determine whether it is true or false, and justify your answer. The domain is all integers.

(a) $\exists x \forall y \; (2x - y > 1)$

(b) $\forall x \forall y \; ((x - y)^2 > 0)$

(c) $\forall x \exists y \; (2x - y \leq 1)$

(d) $\exists x \forall y \exists z \; (x + y + z = 0)$

Answer: (a) False. This is the negation of (c), which is true (shown below).

(b) False. Counter example: $x = y = 0$.

(c) True. For any x, setting $y = 2x$ makes the proposition true.

(d) True. Set $x = 0$. Then for any y, setting $z = -y$ makes the proposition true.

Grading scheme: 2 for each

Problem 2: [4 pts] Solve the following system of linear congruences by the Chinese remainder theorem. Show all steps except the extended Euclidean algorithm.

$$\begin{cases} x \equiv 4 \pmod{9} \\ x \equiv 8 \pmod{14} \\ x \equiv 3 \pmod{5} \end{cases}$$

Answer: Let $m = 9 \cdot 14 \cdot 5 = 630$. $M_1 = 14 \cdot 5 = 70$, $M_2 = 9 \cdot 5 = 45$, $M_3 = 9 \cdot 14 = 126$. By the extended Eucildean algorithm, we find that 4 is an inverse of M_1 (mod 9), 5 is an inverse of M_2 (mod 14) and 1 is an inverse of M_3 (mod 5). Hence, $x = 4 \cdot 70 \cdot 4 + 8 \cdot 45 \cdot 5 + 3 \cdot 126 \cdot 1 = 3298 \equiv 148$ (mod 630).

Student ID: ________

Problem 3: [8 pts] The algorithm below is the Euclidean algorithm for finding the greatest common divisor of two integers m and n (assuming $m \geq n$).

$\underline{\textbf{gcd}(m, n):}$
$x \leftarrow m$
$y \leftarrow n$
while $y \neq 0$
  $r \leftarrow x \bmod y$
  $x \leftarrow y$
  $y \leftarrow r$
return x

(a) Let y_i denote the value y at the beginning of the i-th iteration of the while loop. Let k denote the total number of iterations. Show that $y_i > 2y_{i+2}$ for $i = 1, \ldots, k-2$.

(b) Which one of the following functions is the tightest asymptotic upper bound of the running time of the Euclidean algorithm?

$$1, \log_2 n, \log_2^2 n, \log_2^3 n, \cdots, n, n^2, n^3, \cdots, 2^n, n!$$

Justify your answer.

Note: You may answer (b) even if you don't know how to prove (a).

Answer: (a) For every $1 \leq i \leq k-2$, we have $x_{i+1} = y_i$, $y_{i+1} = x_i \bmod y_i$ and $y_{i+2} = x_{i+1} \bmod y_{i+1}$. So, $y_{i+2} = y_i \bmod y_{i+1}$. This implies that $y_i = y_{i+1}q + y_{i+2}$ for some integer $q \geq 1$ and $y_{i+2} < y_{i+1}$. Thus, $y_i \geq y_{i+1} + y_{i+2} > 2y_{i+2}$.

(b) The running time complexity is $O(\log n)$. This is because the value y reduces at least by half for every two iterations. y starts with n. When y divides x, the algorithm will terminate in the next iteration. So the maximum number of iterations of the while loop is $2\log_2 n+1$. The running time in each iteration is $O(1)$. Therefore, the running time of the Euclidean algorithm is $O(\log n)$.

Grading scheme: 4, 4

Problem 4: [8 pts]

(a) Arrange the functions $n^n, (\log n)^2, n^{1.0001}, (1.0001)^n, 2^{\sqrt{\log n}}$, and $n(\log n)^{1001}$ in asymptotic increasing order, i.e., each function is big-O of the next function.

(b) Give an example of two increasing functions $f(n)$ and $g(n)$ from the set of positive integers to the set of positive integers such that neither $f(n)$ is $O(g(n))$ nor $g(n)$ is $O(f(n))$. Justify your answer.

Answer: (a) $(\log n)^2, 2^{\sqrt{\log n}}, n(\log n)^{1001}, n^{1.0001}, (1.0001)^n$ and n^n

(b) We want the functions to play leap-frog, with first one much bigger than the other. Something like this will do the trick: Let $f(n) = n^{2\lfloor n/2 \rfloor + 1}$ and $g(n) = n^{2\lceil n/2 \rceil}$. Thus, $\{f(1) = 1^1, g(1) = 1^2\}$, $\{f(2) = 2^3, g(2) = 2^2\}$, $\{f(3) = 3^3, g(3) = 3^4\}$, $\{f(4) = 4^5, g(4) = 4^4\}$, and so on. Then for even n we have $f(n)/g(n) = n$, and for odd n we have $g(n)/f(n) = n$. Because both ratios are unbounded, neither function is big-O of the other.

Grading scheme: 4, 4

Student ID: ________

Problem 5: [8 pts]

Give a combinatorial proof to the following identity:

$$\sum_{i=0}^{n} \binom{n}{i} \sum_{j=0}^{n-i} \binom{n-i}{j} = 3^n.$$

Answer: The right side is easier to interpret, so we start with that. We take a set of size n, and partition them into 3 sets A, B and C. The RHS is the number of ways to split. For the left side, suppose we already know that we want i elements in A and j elements in B. Then, we have $\binom{n}{i}$ ways to choose the elements of A, and $\binom{n-i}{j}$ ways to choose the elements of B once we have already chosen the elements of A. After we have chosen the elements of A and B, the rest of the elements go to set C. Finally, we have to sum over all of the possible values of i and j.

Student ID: ________

Problem 6: [8 pts]

(a) Give a recursive definition of the function $ones(s)$, which counts the number of ones in a bit string s.

(b) Use structural induction to prove that $ones(st) = ones(s) + ones(t)$.

Answer: Let λ denote an empty bit string.

(a) $ones(\lambda) = 0$ and $ones(wx) = x + ones(w)$, where w is a bit string and x is a bit (viewed as an integer when being added).

(b) We prove by induction on the length of t. The basis step is when $t = \lambda$, in which case we have $ones(s\lambda) = ones(s) = ones(s) + 0 = ones(s) + ones(\lambda)$.

For the inductive step, write $t = wx$, where w is a bit string and x is a bit. Then we have

$$\begin{aligned} ones(s(wx)) &= ones((sw)x) \\ &= x + ones(sw) \qquad \text{(by definition)} \\ &= x + ones(s) + ones(w) \qquad \text{(by the inductive hypothesis)} \\ &= ones(s) + (x + ones(w)) \\ &= ones(s) + ones(wx) \qquad \text{(by definition)} \\ &= ones(s) + ones(t). \end{aligned}$$

Grading scheme: 2, 6

Problem 7: [8 pts] Let P be a set of 5 distinct points in the plane with integer coordinates. Draw a line segment between every two points of P (note that a total of 10 line segments will be drawn). Show that at least one of these line segments has its midpoint with integer coordinates.

Answer: The midpoint of the segment whose endpoints are (a, b) and (c, d) is $((a+c)/2, (b+d)/2)$. We are concerned only with integer values of the original coordinates. Clearly the coordinates of these fractions will be integers as well if and only if a and c have the same parity (both odd or both even) and b and d have the same parity. Thus what matters in this problem is the parities of the coordinates. There are four possible pairs of parities: (odd, odd), (odd, even), (even, odd), and (even, even). Since we are given five points, the pigeonhole principle guarantees that at least two of them will have the same pair of parities. The midpoint of the segment joining these two points will therefore have integer coordinates.

Problem 8: [10 pts] How many strings of six lowercase letters from the English alphabet contain

(a) the letter a?

(b) the letters a and b?

(c) the letters a and b in consecutive positions with a preceding b, with all the letters distinct?

(d) the letters a and b, where a is somewhere to the left of b in the string, with all the letters distinct?

Answer: (a) The only reasonable way to do this is by subtracting from the number of strings with no restrictions the number of strings that do not contain the letter a. The answer is $26^6 - 25^6$.

(b) We use W_a to denote the set of words containing a, and similarly $W_b, W_{a\wedge b}, W_{a\vee b}$. By the inclusion-exclusion rule, we have

$$|W_{a\vee b}| = |W_a| + |W_b| - |W_{a\wedge b}|.$$

From (a), we know $|W_a| = |W_b| = 26^6 - 25^6$. Similarly, $|W_{a\vee b}| = 26^6 - 24^6$. So

$$|W_{a\wedge b}| = |W_a|+|W_b|-|W_{a\vee b}| = 2(26^6-25^6)-(26^6-24^6) = 26^6-2\cdot 25^6+24^6.$$

(c) First choose the position for the a; this can be done in 5 ways, since the b must follow it. There are four remaining positions, and these can be filled in $P(24,4)$ ways, since there are 24 letters left (no repetitions being allowed this time). Therefore the answer is $5P(24,4) = 1275120$.

(d) First choose the positions for the a and b; this can be done in $C(6,2)$ ways, since once we pick two positions, we put the a in the leftmost and the b in the other. There are four remaining positions, and these can be filled in $P(24,4)$ ways, since there are 24 letters left (no repetitions being allowed this time). Therefore the answer is $C(6,2)P(24,4) = 3825360$.

Grading scheme: 2, 3, 2, 3

Student ID: ________

Problem 9: [4 pts] A professor packs her collection of 40 (distinguishable) issues of a mathematics journal in four boxes with 10 issues per box. How many ways can she distribute the journals if

(a) each box is numbered, so that they are distinguishable?

(b) the boxes are identical, so that they cannot be distinguished?

Answer: (a) $40!/10!^4$.

(b) Each distribution into identical boxes gives rise to $4! = 24$ distributions into labeled boxes, since once we have made the distribution into unlabeled boxes we can arbitrarily label the boxes. Therefore the answer is the same as the answer in part (a) divided by 24, namely $(40!/10!^4)/4!$.

Grading scheme: 2, 2

Student ID: ________

Problem 10: [9 pts] Consider the equation $x_1 + x_2 + x_3 = 20$ where x_1, x_2, and x_3 are nonnegative integers.

(a) How many solutions are there?

(b) How many solutions are there so that $x_1 \geq 9$?

(c) How many solutions are there so that $x_1 \geq 9$ and $x_2 \geq 9$?

(d) How many solutions are there so that x_1, x_2 and x_3 are less than 9?

For this question, give exact numbers as the solution.

Answer: (a) $C(3+20-1, 3-1) = C(22, 2) = 231$.

(b) Let $x_1' = x_1 - 9$. The problem becomes the number of solutions to the equation $x_1' + x_2 + x_3 = 11$, which is $C(11+3-1, 3-1) = C(13, 2) = 78$.

(c) Same concept as part (b). The answer is $C(2+3-1, 3-1) = C(4, 2) = 6$.

(d) We use S_ϕ to denote the set of solutions subject to condition ϕ. By the inclusion-exclusion principle, we have

$$\begin{aligned} |S_{x_1 \geq 9 \vee x_2 \geq 9 \vee x_3 \geq 9}| &= 3 \cdot |S_{x_1 \geq 9}| - 3 \cdot |S_{x_1 \geq 9 \wedge x_2 \geq 9}| + |S_{x_1 \geq 9 \wedge x_2 \geq 9 \wedge x_3 \geq 9}| \\ &= 3 \cdot 78 - 3 \cdot 6 + 0 = 216. \end{aligned}$$

So, $|S_{x_1 < 9 \wedge x_2 < 9 \wedge x_3 < 9}| = |S| - |S_{x_1 \geq 9 \vee x_2 \geq 9 \vee x_3 \geq 9}| = 231 - 216 = 15$.
Alternative solution: Write $y_1 = 8 - x_1, y_2 = 8 - x_2, y_3 = 8 - x_3$. Then the equation becomes $24 - y_1 - y_2 - y_3 = 20$, or $y_1 + y_2 + y_3 = 4$, which has $C(4+3-1, 3-1) = C(6, 2) = 15$ solutions.

Grading scheme: 2,2,2,3

Student ID: ________

Problem 11: [11 pts]

Recall that we gave two explanations for the Monty Hall game. The first is an intuitive one: Our initial guess is correct with probability $\frac{1}{3}$, which is not affected by Monty opening another door, so switching will give us a winning probability of $\frac{2}{3}$. Then we gave a formal proof that $p(W = 1 \mid M = 2) = p(W = 1) = \frac{1}{3}$, where W denotes the winning door and M the door Monty opens. Below, we will see that the intuitive explanation only "happens" to be correct in this case, and can easily go wrong if we slightly perturb the probabilities.

Now, suppose that the probabilities of the 3 doors having the prize are $0.5, 0.3$, and 0.2, respectively. Then obviously we should pick door 1 as our initial guess. Then Monty will open one of door 2 or 3 without the prize. If neither has the prize, he will pick one of them randomly with equal probability.

(a) Suppose $M = 2$, i.e., Monty opens door 2. What is $p(W = 1 \mid M = 2)$ and $p(W = 3 \mid M = 2)$? Based on your calculation, should you switch to door 3?

(b) Now suppose $M = 3$, i.e., Monty opens door 3. What is $p(W = 1 \mid M = 3)$ and $p(W = 2 \mid M = 3)$? In this case, should you switch to door 2?

(c) What's the winning probability of your strategy above?

(d) Can you design a better strategy? What's its winning probability? [Hint: What is "obvious" is not obvious.]

Answer: (a) By the Bayes' theorem,

$$p(W = w \mid M = 2) = \frac{p(M = 2 \mid W = w)p(W = w)}{p(M = 2)},$$

where $p(M = 2) = \sum_{w=1}^{3} p(M = 2 \mid W = w)p(W = w)$. We have $p(W = 1) = 0.5, p(W = 2) = 0.3, p(W = 3) = 0.2$, $p(M = 2 \mid W = 1) = \frac{1}{2}, p(M = 2 \mid W = 2) = 0, p(M = 2 \mid W = 3) = 1$. Plugging in these numbers, we obtain $p(M = 2) = 0.45$ and $p(W = 1 \mid M = 2) = \frac{5}{9}$, $p(W = 3 \mid M = 2) = \frac{4}{9}$. So, we should not switch. Contrary to our intuition, the probability that our initial guess is correct has increased after door 2 is opened!

(b) Same as above, but with $p(M = 3 \mid W = 1) = \frac{1}{2}, p(M = 3 \mid W = 2) = 1, p(M = 3 \mid W = 3) = 0$. Plugging in these numbers, we obtain $p(M = 3) = 0.55$ and $p(W = 1 \mid M = 3) = \frac{5}{11}$, $p(W = 2 \mid M = 3) = \frac{6}{11}$. So, we should switch to door 2. This time, the probability that our initial guess is correct has decreased.

(c) Let $\mathcal{E}$ denote the event that you win. We have $p(\mathcal{E}) = \sum_{m=2}^{3} p(M = m)p(\mathcal{E} \mid M = m) = 0.45 \cdot \frac{5}{9} + 0.55 \cdot \frac{6}{11} = 0.55$.

Alternatively, $p(\mathcal{E}) = p(W = 1, M = 2) + p(W = 2, M = 3) = 0.5 \cdot 0.5 + 0.3 \cdot 1 = 0.55$.

(d) We pick door 3 first, then always switch. The winning probability is $p(\mathcal{E}) = p(W = 1, M = 2) + p(W = 2, M = 1) = 0.5 \cdot 1 + 0.3 \cdot 1 = 0.8.$

Grading Scheme: 3,3,2,3

Student ID: ________

Problem 12: [8 pts] Suppose the time line is discretized into intervals, with the length of each interval being 1 second. A web server has probability p $(0 < p < 1)$ of getting 1 visit in each time interval and probability $1 - p$ of getting 0 visit.

(a) How long is the server expected to wait until it gets 100 visits?

(b) How many visits is the server expected to get during 1 hour?

(c) What is the variance of the number of visits that the server gets during 1 hour?

Answer: (a) Let X_i be the number of seconds the server waits until it gets the i-th visit after $i - 1$ visits. This is a geometric random variable, so $E(X_i) = \frac{1}{p}$. The number of seconds for which the server is expected to wait until it gets the 100-th visit is then

$$\sum_{i=1}^{100} E(X_i) = 100/p.$$

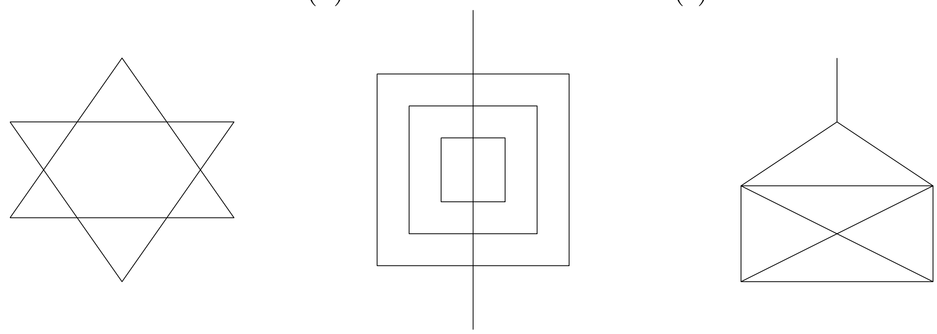
(b) 1 hour has 3600 seconds. There are 3600 independent Bernoulli trials. So the expected number of visits is $3600p$.

(c) Let X_i be the indicator random variable that indicates the event of getting a visit at the i-th second during the hour. The variance of a Bernoulli trial, $V(X_i)$, is $(1 - p)p$. There are 3600 independent Bernoulli trials. So, the variance is $\sum_{i=1}^{3600} V(X_i) = 3600(1 - p)p$.

Grading scheme: 3, 2, 3

Problem 13: [6 pts] For each of the following graphs, determine whether the graph has an Euler path and/or Euler circuit. Briefly justify your answers.

(a) (b) (c)



Answer:

(a) There is an Euler path and Euler circuit because the graph has all vertices of even degree.

(b) There is an Euler path, but no Euler circuit because the graph has two vertices of odd drgree.

(c) There is no Euler path or Euler circuit because the graph has 4 vertices of odd degree.

Grading scheme: 2, 2, 2

Bonus Problem: [10 pts]

Consider the following recurrence: $f_1 = 1, f_n = 2f_{n-1} + n$ for $n \geq 2$. The following is a purported induction proof that $f_n = O(n)$.

Proof. Base case: When $n = 1$, $f(n) = O(n)$ because $f(1) = O(1)$.

Induction step: Assuming $f_{n-1} = O(n-1)$, we will prove $f_n = O(n)$:

$$\begin{aligned} f_n &= 2f_{n-1} + n \\ &= 2 \cdot O(n-1) + n \qquad \text{(induction hypothesis)} \\ &= O(3n-2) \\ &= O(n) \qquad \text{(dropping lower order term and constant coefficient)} \end{aligned}$$

as desired. □

(a) Identify the mistake in the proof.

(b) Prove that $f(n) = O(2^n)$ by induction.

Answer: (a) The hidden constant in the induction conclusion $f_n = O(n)$ is different (larger) from the one in the induction hypothesis $f_{n-1} = O(n-1)$. Generally, one cannot use asymptotic notation in induction proofs, because the induction step is implicitly applied many times, so the constant coefficient will no longer be a constant.

(b) We will show that $f(n) \leq 2 \cdot 2^n - n - 2$.

Base case: When $n = 1$, $f(n) = 1 \leq 2 \cdot 2^1 - 1 - 2$.

Induction step: Assuming $f_{n-1} \leq 2 \cdot 2^{n-1} - (n-1) - 2$, we will prove $f_n \leq 2 \cdot 2^n - n - 2$:

$$\begin{aligned} f_n &= 2f_{n-1} + n \\ &\leq 2(2 \cdot 2^{n-1} - (n-1) - 2) + n \qquad \text{(induction hypothesis)} \\ &= 2 \cdot 2^n - n - 2. \end{aligned}$$

Comment: It is impossible to prove $f_n \leq c \cdot 2^n$ by induction, no matter how you choose the constant c. Interestingly, and somehow counter-intuitively, it is possible to prove the stronger claim $f(n) \leq c \cdot 2^n - n - 2$. Fundamentally, this is because a stronger claim will also make the induction hypothesis stronger, which sometimes makes the induction step easier to go through. This technique is known as "strengthening the induction hypothesis".

Grading scheme: 3, 7.