COMP 2711 Discrete Mathematical Tools for Computer Science
2021 Spring Semester – Final Exam (Part 2)

Note 1: For all non-proof questions, showing your steps is not necessary unless required otherwise. However, in case your answer is wrong, showing your steps may earn you some partial credits.
Note 2: You can express your answers using binomial coefficients, factorials, and D_n (derangement number). However, you should not have summation $\sum$ in your final answers. For example, $\binom{10}{3}D_9+4!$ and $1!+2!+3!+4!$ are valid, but $\sum_{i=0}^{n}\binom{n}{i}$ or $\binom{n}{0}+\binom{n}{1}+\cdots+\binom{n}{n}$ is not. The latter has to be simplified to 2^n.

Question 1: [10 pts] Recall the definition of Fibonacci number:

$$F(1)=1, F(2)=1, F(n)=F(n-1)+F(n-2) \text{ for } n\geq 2.$$

Give a combinatorial proof to the following identity:

$$F(n+2)=\sum_{i=0}^{(n+1)/2}\binom{n+1-i}{i}$$

You can assume $n \geq 1$ and n is odd.
[Hint: Consider bit strings with no consecutive 0*s.]*

Solution: Let S_n be the set of bit strings of length n with no consecutive 0s. We will show that both sides are equal to $|S_n|$. Let s_i denote any bit string in S_i.

Left-hand side: For $n=0$, the empty string is the only string and it has no consecutive 0s. So, $F(0+2)=1=|S_0|$ bit string that have no consecutive 0s. For $n=1$, there are $F(1+2)=F(3)=F(2)+F(1)=2=|S_1|$ bit strings of length 1 that have no consecutive 0s. For $n>1$, a bit string of length n that has no consecutive 0s can be obtained by appending a "1" to s_{n-1} and appending a "10" to s_{n-2}. The two strings must be different and these two are the only possible cases. By induction, we have $|S_n|=F(n+2)$.

Right-hand side: $\binom{n+1-i}{i}$ is counting the number of bit strings of length n that has exactly i 0s and no consecutive 0s. When a bit string of length n that has more than $(n+1)/2$ 0s, there must be at least two consecutive 0s. So, the expression sums over all mutually disjoint cases, and thus it equals to $|S_n|$.

Question 2: [10 pts] How many positive integers less than 100,000,000 have the sum of their digits equal to 19?

Solution: Let $d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8$ be the eight digits of the positive integer less than 100,000,000. Assuming each d_i can take any non-negative integers, there are $\binom{8+19-1}{19}$ ways to assign non-negative integers to d_i for every i so that their sum equals to 19. At most one d_i can be greater than 9, because $10 + 10 = 20 > 19$. The integer exceeding 9 can be any of the eight d_i. After selecting one, say d_1, to be exceeding 9, there are $\binom{8+9-1}{9}$ positive integers that the digits sum to 19.

So, the total number of positive integers less than 100,000,000 and have the sum of their digits equal to 19 is $\binom{8+19-1}{19} - 8 \cdot \binom{8+9-1}{9} = 566280$.

Question 3: [8 pts] Assume that all 7 days of the week are equally likely to be the birthday.

(a) What is the probability that two people chosen at random were born on the same day of the week?

(b) What is the probability that in a group of 5 people chosen at random, there are at least two born on the same day of the week?

(c) What is the minimum number of randomly chosen people needed so that the probability of having at least two people born on the same day of the week is greater than 1/2 ?

Solution: The probability of a birth in each day is 1/7.

(a) $7 \cdot (1/7) \cdot (1/7) = 1/7$.

(b) The probability that all the birth days-of-the-week are different is $\frac{7}{7} \cdot \frac{6}{7} \cdot \frac{5}{7} \cdot \frac{4}{7} \cdot \frac{3}{7}$. The probabilty that at least two born on the same day of the week is $1 - \frac{7}{7} \cdot \frac{6}{7} \cdot \frac{5}{7} \cdot \frac{4}{7} \cdot \frac{3}{7} \approx 0.85$.

(c) Follow the calculation, we know the probability for 3 people is about 0.3878 and the probability for 4 people is about 0.65. So the answer is 4.

Grading scheme: 2,3,3.

Question 4: [10 pts] *Randomised response* is a method used in survey interview that allows respondents to respond to sensitive issues (e.g., are you a smoker?) while maintaining confidentiality. It works as follows. The interviewer prepares two questions: one is the question to be surveyed, while the other is a dummy question whose answer is trivially true (e.g., is $1 + 1 = 2$?). Each respondent picks a question from the two randomly, and answers the question truthfully, but the interviewer doesn't know which question is picked. Suppose we have surveyed n people, and know a priori that the number of smokers, denoted m, is no more than $n/2$. Note that n is known but m is unknown (we only know $m \leq n/2$).

(a) Suppose a randomly chosen respondent has answered "yes" to the question. Show that the interviewer's confidence on the respondent being a smoker (i.e., the probability the respondent being a smoker given that his or her answer is "yes") is at most 2/3.

(b) Suppose X respondents have answered "yes". Design an unbiased estimator of the number of smokers, i.e., find a function $f(\cdot)$ such that $\mathsf{E}[f(X)] = m$. Justify your answer.

Solution: (a) Let S denote the event that the respondent is a smoker, and Y the event that her or his answer is "yes". We have $\Pr[S] \leq 1/2, \Pr[Y \mid S] = 1, \Pr[Y \mid \bar{S}] = 1/2$. By Bayes' theorem, we have

$$\begin{aligned}
\Pr[S \mid Y] &= \frac{\Pr[Y \mid S]\Pr[S]}{\Pr[Y \mid S]\Pr[S] + \Pr[Y \mid \bar{S}]\Pr[\bar{S}]} \\
&= \frac{\Pr[S]}{\Pr[S] + 1/2 \cdot (1 - \Pr[S])} \\
&= \frac{\Pr[S]}{1/2 + 1/2 \cdot \Pr[S]} \\
&= \frac{2(1/2 + 1/2 \cdot \Pr[S]) - 1}{1/2 + 1/2 \cdot \Pr[S]} \\
&= 2 - \frac{1}{1/2 + 1/2 \cdot \Pr[S]} \\
&\leq 2 - \frac{1}{1/2 + 1/2 \cdot 1/2} = 2/3.
\end{aligned}$$

(b) The m smokers always answer "yes", while each of the $n-m$ nonsmokers answers "yes" with probability 1/2. So $\mathsf{E}[X] = m + (n-m)/2 = (n+m)/2$. Rewrite it as $m = 2\mathsf{E}[X] - n = \mathsf{E}[2X - n]$ by linearity of expectation. So the unbiased estimator is $f(X) = 2X - n$.

Grading scheme: 5,5.

Question 5: [8 pts] For each of the following parts, either give an example graph or prove that such a graph does not exist.

(a) A simple undirected graph with 10 vertices, whose degrees are $2, 1, 0, 3, 8, 9, 7, 2, 1, 1$.

(b) A complete bipartite undirected graph with 10 vertices, whose degrees are $3, 3, 3, 7, 7, 7, 7, 7, 7, 7$.

(c) A strongly connected directed graph with 5 vertices, v_1, v_2, v_3, v_4 and v_5.

- In-degree of $v_1 = 1$, out-degree of $v_1 = 3$.
- In-degree of $v_2 = 2$, out-degree of $v_2 = 1$.

- In-degree of $v_3 = 3$, out-degree of $v_3 = 1$.
- In-degree of $v_4 = 0$, out-degree of $v_4 = 2$.
- In-degree of $v_5 = 2$, out-degree of $v_5 = 1$.

(d) A directed graph with 5 vertices, v_1, v_2, v_3, v_4 and v_5.

- In-degree of $v_1 = 1$, out-degree of $v_1 = 3$.
- In-degree of $v_2 = 2$, out-degree of $v_2 = 1$.
- In-degree of $v_3 = 3$, out-degree of $v_3 = 2$.
- In-degree of $v_4 = 1$, out-degree of $v_4 = 1$.
- In-degree of $v_5 = 2$, out-degree of $v_5 = 1$.

Solution: (a) None. It is not possible to have both a vertex of degree 9 and a vertex of degree 0.

(b) None. In a complete bipartite graph, we can partition the vertices into two sets V_1, V_2 so that there is an edge from every vertex in V_1 to every vertex in V_2. So, the degree of every vertex in V_1 should be $|V_2|$ and the degree of every vertex in V_2 should be $|V_1|$. When the two possible degrees are 3 and 7, we need 7 vertices with degree 3 and 3 vertices with degree 7 to make a complete bipartite graph. The given degrees are not in this case.

(c) None. There is no path from v_i to v_4 for $i \neq 4$, because in-degree of v_4 is 0.

(d) None. The sum of in-degrees is different with the sum of out-degrees.

Grading scheme: 2,2,2,2. No marks for answers without proofs. Points are all for the proofs.

Bonus: (10 pts) Let G be an undirected complete graph having n vertices. For a vertex v, $N(v)$ denotes the set of neighbors of v in G. A random walk on G is the following process, which occurs in a sequence of discrete steps: starting at a vertex v_0, we proceed at the first step to a neighbor of v_0 chosen uniformly at random. This may be thought of as choosing a random edge incident on v_0 and walking along it to a vertex $v_1 \in N(v_0)$. At the second step, we proceed to a randomly chosen neighbor of v_1, and so on.

(a) For two given vertices u and v of G, what's the expected number of steps of random walk that begins at u and ends upon first reaching v?

(b) Redo (a) for the case where G is a line graph of n vertices, and u is at one end and v is at the other end.

Solution: (a) The probability of visiting v at the next step is $p = \frac{1}{n-1}$ if the current vertex is not v. So the answer is $1/p = n - 1$.

(b) Number of the vertices as $1, \ldots, n$. Let T_i denote the number of steps to go from i to n. We have $T_n = 0$, so $\mathsf{E}[T_n] = 0$. The goal is to find $\mathsf{E}[T_1]$. From vertex 1, we can only go right, so $T_1 = 1 + T_2$ and $\mathsf{E}[T_1] = 1 + \mathsf{E}[T_2]$. Rewrite it as $\mathsf{E}[T_1] - \mathsf{E}[T_2] = 1$. For any $1 < i < n$, we go left or right with equal probability, so $\mathsf{E}[T_i] = 1 + \frac{1}{2}\mathsf{E}[T_{i-1}] + \frac{1}{2}\mathsf{E}[T_{i+1}]$. Rearrange it as $\mathsf{E}[T_i] - \mathsf{E}[T_{i+1}] = 2 + \mathsf{E}[T_{i-1}] - \mathsf{E}[T_i]$. Now we have

$$\begin{aligned}
\mathsf{E}[T_1] - \mathsf{E}[T_2] &= 1 \\
\mathsf{E}[T_2] - \mathsf{E}[T_3] &= 3 \\
&\cdots \\
\mathsf{E}[T_{n-1}] - \mathsf{E}[T_n] &= 2n - 3
\end{aligned}$$

Summing up, we have $\mathsf{E}[T_1] = \mathsf{E}[T_n] + 1 + 3 + \cdots + 2n - 3 = (n-1)^2$.

Grading scheme: 2,8.