Review Questions

Mark the correct answer in each part of the following questions.

1. Let n be an arbitrary fixed even positive integer. We are interested in the collection $\mathcal{S}$ of subsets of $\{1, 2, \ldots, n\}$, including at most one of the two numbers 1 and n, at most one of the numbers 2 and $n-1$, ..., at most one of the numbers $n/2$ and $n/2+1$.

   (a) If we go over all subsets of $\{1, 2, \ldots, n\}$ according to the lexicographic order, then the number of subsets we encounter before we get for the first time a subset outside $\mathcal{S}$ is

   (i) $2^{n/2} + O(1)$.
   (ii) $3 \cdot 2^{n/2-1} + O(1)$.
   (iii) $2^{n/2+1} + O(1)$.
   (iv) $3 \cdot 2^{n/2} + O(1)$.
   (v) none of the above.

   (b) If we go over all subsets of $\{1, 2, \ldots, n\}$ according to the Gray code, then the number of subsets we encounter before we get for the first time a subset outside $\mathcal{S}$ is

   (i) $2^{n/2} + O(1)$.
   (ii) $3 \cdot 2^{n/2-1} + O(1)$.
   (iii) $2^{n/2+1} + O(1)$.
   (iv) $3 \cdot 2^{n/2} + O(1)$.
   (v) none of the above.

   (c) We consider now possible ways of enumerating the elements of $\mathcal{S}$ as fast as possible.

(i) We can enumerate the subsets in $\mathcal{S}$ with minimal changes. Moreover, it is possible to enumerate them along a Hamiltonian circuit. (Namely, from the last subset in the cycle it is possible to move to the first one by a minimal change.)

(ii) We can enumerate the subsets in $\mathcal{S}$ with minimal changes. Both $\emptyset$ and $\{1, 2, \ldots, n/2\}$ may be taken as the initial points. However, it is impossible to enumerate them along a Hamiltonian circuit.

(iii) We can enumerate the subsets in $\mathcal{S}$ with minimal changes. It is possible to have $\emptyset$ as the initial point, but not $\{1, 2, \ldots, n/2\}$.

(iv) We can enumerate the subsets in $\mathcal{S}$ with minimal changes. It is possible to have $\{1, 2, \ldots, n/2\}$ as the initial point, but not $\emptyset$.

(v) None of the above.

(d) We want to select a uniformly random element of $\mathcal{S}$. Consider the following two algorithms:

- $\mathcal{A}_1$ – select a uniformly random subset of $\{1, 2, \ldots, n\}$ until you get a set belonging to $\mathcal{S}$.
- $\mathcal{A}_1$ – select a uniformly random subset S of $\{1, 2, \ldots, n\}$. For each i between 1 and $n/2$, for which both i and $n+1-i$ belong to S, do the following: With probability $1/2$ remove i from S, and with the remaining probability remove $n+1-i$ from it.

(i) $\mathcal{A}_1$ is correct, but requires selecting on the average $(4/3)^{n/2}$ subsets to get a subset in $\mathcal{S}$. $\mathcal{A}_2$ selects subsets belonging to $\mathcal{S}$ only, and each such subset has a positive probability of being selected, but the probabilities are not as required.

(ii) $\mathcal{A}_1$ is correct, but requires selecting on the average select $3^{n/2}$ subsets to get a subset in $\mathcal{S}$. $\mathcal{A}_2$ selects subsets belonging to $\mathcal{S}$ only, and each such subset has a positive probability of being selected, but the probabilities are not as required.

(iii) Both $\mathcal{A}_1$ and $\mathcal{A}_2$ are correct, but $\mathcal{A}_1$ is much slower than $\mathcal{A}_2$ on the average.

(iv) $\mathcal{A}_1$ is much slower than $\mathcal{A}_2$ on the average, but it does not matter as both methods are incorrect.

(v) None of the above.

2. Here, we develop a recurrence formula for the number T_n of trees on n labeled vertices. (This formula may be used in turn to provide yet another proof of Cayley's formula, but we do not deal with this derivation here.)

   Given a tree, and any edge in the tree, by removing this edge we obtain two non-empty trees – one on some set of k vertices containing the vertex 1 (where $1 \leq k \leq n-1$), and another on the complementary set of vertices. When going over all pairs (t, e), consisting of a tree t on the given n vertices and an edge e out of the $n-1$ edges of t, we get (by removing e from t) each pair of trees on complementary sets of vertices several times. By counting the occurrences of each pair, we obtain the recurrence

   (i)
   $$(n-1)T_n = \sum_{k=1}^{n-1} \binom{n-1}{k-1} (k-1)(n-k) T_k T_{n-k}, \qquad n \geq 2,$$
   where $T_1 = 1$.

   (ii)
   $$(n-1)T_n = \sum_{k=1}^{n-1} \binom{n-1}{k-1} (k-1)(n-k+1) T_k T_{n-k}, \qquad n \geq 2,$$
   where $T_1 = 1$.

   (iii)
   $$(n-1)T_n = \sum_{k=1}^{n-1} \binom{n-1}{k-1} k(n-k) T_k T_{n-k}, \qquad n \geq 2,$$
   where $T_1 = 1$.

   (iv)
   $$(n-1)T_n = \sum_{k=1}^{n-1} \binom{n-1}{k-1} k(n-k+1) T_k T_{n-k}, \qquad n \geq 2,$$
   where $T_1 = 1$.

(v) none of the above.

3. Recall that we denoted partitions of integers in two ways. The "lazy" notation is of the form

$$n = r_1 + r_1 + \ldots + r_1 + r_2 + r_2 + \ldots + r_2 + \ldots + r_l + r_l + \ldots + r_l,$$

and the "economic" form

$$n = m_1 r_1 + m_2 r_2 + \ldots + m_l r_l,$$

where each distinct number r_i in the partition is listed but once, along with its multiplicity m_i. The integers r_i are listed in decreasing order. We may measure the distance between two partitions of the same number in two ways, depending on the representation. When representing the partitions the lazy way, we drop equal r_i-s from the partitions, and count the remaining parts in the two partitions together. (For example, the distance between the partitions $35 = 10 + 10 + 10 + 5$ and $35 = 10+10+3+3+3+3+3$ is 7.) When representing the partitions the economic way, we drop equal pairs (r_i, m_i) and count the remaining parts in the two partitions together. (For example, the distance between the partitions $35 = 3 \cdot 10 + 1 \cdot 5$ and $35 = 2 \cdot 10 + 5 \cdot 3$ is 4.)

We would like to know how large the distance may be in the worst case between successive partitions in the dictionary order.

(i) The distance when representing the partitions either way is $O(1)$.

(ii) The distance when representing the partitions the lazy way is $O(1)$, and when representing them the economic way it is $\Omega(n)$. (Namely, there exists a fixed $\delta > 0$ such that, for each sufficiently large n, there are two successive partitions of n, the distance between which is at least δn.)

(iii) The distance when representing the partitions the lazy way is $\Omega(n)$, and when representing them the lazy way is $O(1)$.

(iv) The distance when representing the partitions either way is $\Omega(n)$.

(v) none of the above.

4. Consider the family of Young Tableaux, whose shape consists of one row of length $n+1$ and n rows of length 1. Unlike what we have done in class, we fill the shape with numbers from 0 to $2n$ (instead of 1 to $2n+1$). In the following, n will be an arbitrary fixed sufficiently large integer.

   Notice that, when going over all Young tableaux of the given shape, the set of numbers in the first column, excluding the 0 at the top left square, ranges over all subsets of size n of $\{1, 2, \ldots, 2n\}$.

   (a)

   (i) The number of tableaux in which the number 1 is in the first column is $\frac{1}{2} \cdot \binom{2n}{n}$, as is the number of tableaux in which the number $2n$ is in the first column. The number of tableaux in which both 1 and $2n$ are in the first column is $\frac{1}{4} \cdot \binom{2n}{n}$.

   (ii) The number of tableaux in which the number 1 is in the first column is $\frac{1}{2} \cdot \binom{2n}{n}$, as is the number of tableaux in which the number $2n$ is in the first column. The number of tableaux in which both 1 and $2n$ are in the first column is less than $\frac{1}{4} \cdot \binom{2n}{n}$.

   (iii) The number of tableaux in which the number 1 is in the first column is $\frac{1}{2} \cdot \binom{2n}{n}$, as is the number of tableaux in which the number $2n$ is in the first column. The number of tableaux in which both 1 and $2n$ are in the first column is greater than $\frac{1}{4} \cdot \binom{2n}{n}$.

   (iv) There exists a unique pair of integers k, l, with $1 \le k < l \le 2n$, such that both k and l are in the first column in exactly $\frac{1}{4} \cdot \binom{2n}{n}$ tableaux. However, $(k, l) \ne (1, 2n)$.

   (v) None of the above.

   (b) When we go over all Young tableaux of the given shape according to the order suggested in class, the set of numbers in the first row goes over all subsets of size n of $\{1, 2, \ldots, 2n\}$

   (i) according to the lexicographic order.

   (ii) with minimal changes, in the same order we went over all subsets of size k of $\{1, 2, \ldots, n\}$.

   (iii) with minimal changes, but not in the same order we went over all subsets of size k of $\{1, 2, \ldots, n\}$.

(iv) with changes very far from being minimal. In fact, it may happen that all $2n$ numbers from 1 to $2n$ change their location when passing from some table to the next.

(v) none of the above.

Solutions

1. (a) During the first $2^{n/2}$ steps, we go over all subsets of $\{1, 2, \ldots, n/2\}$, so that we have sets in $\mathcal{S}$. At that step, the number $n/2+1$ appears for the first time in the sets we encounter. It will take another $2^{n/2-1}$ steps until the number $n/2$ joins the set again. This will be the first set we get which is outside $\mathcal{S}$.

   Thus, (ii) is true.

   (b) Recall that, when using the Gray code, the sequence determining which bit changes at each step consists of 1-s at all odd locations, of 2-s at all locations divisible by 2 but not by 4, of 3-s at all locations divisible by 4 but not by 8, and so forth. The first time a number k joins the set is at the 2^{k-1}-th step. At this point, all bits from 1 to $k-2$ have been changed an even number of times, and hence the corresponding numbers are not in the set. However, bit $k-1$ has been changed exactly once, so that the number $k-1$ is in the set. For $k = n/2 + 1$, this means that the the set we have after $2^{n/2}$ steps is $\{n/2, n/2 + 1\}$. This is the first set we see outside $\mathcal{S}$.

   Thus, (i) is true.

   (c) For each pair of i and $n + 1 - i$, a set in $\mathcal{S}$ may contain either none of them, or only i or only $n+1-i$, altogether 3 possibilities. Thus, $\mathcal{S}$ consists of $3^{n/2}$ sets. We claim that, structurally, $\mathcal{S}$ is a discrete $n/2$-dimensional cube with sides of length 3, as follows (Remark: Compare the situation with that you had in homework problem 9.) Denote $B = \{-1, 0, 1\}^{n/2}$. Define a 1-1 mapping f

from $\mathcal{S}$ onto B as follows. For $1 \le i \le n/2$, if $i \in S$ then the i-th component of $f(S)$ is -1, if $n+1-i \in S$ then it is 1, and if $i, n+1-i \notin S$ then it is 0. A minimal change of a set $S \in \mathcal{S}$ corresponds to changing the corresponding component of $f(S)$ by 1. Hence the question may be rephrased in terms of going over B with minimal changes.

By induction, one sees that it is possible to go over B with minimal changes, starting at $(-1, -1, \dots, -1)$ and ending at $(1, 1, \dots, 1)$. The proof is by induction on the dimension m (which, in our case, is $n/2$). For $m = 1$ we have a path of length 3, and the claim is trivial. Assuming the claim is correct for dimension $m-1$, we go over $B = \{-1, 0, 1\}^m$ as follows. Starting at $(-1, -1, \dots, -1)$, we go over $\{-1, 0, 1\}^{m-1} \times \{-1\}$, ending at $(1, 1, \dots, 1, -1)$. Then we move to $(1, 1, \dots, 1, 0)$. Using again the induction hypothesis, we go to $(-1, -1, \dots, -1, 0)$. We move to $(-1, -1, \dots, -1, 1)$, and then using the induction hypothesis a third time, complete the traversal at $(1, 1, \dots, 1)$. Note that we started from $f(\{1, 2, \dots, n/2\})$.

However, we cannot start from $\emptyset$. In fact, by induction we show that, out of the $3^{n/2}$ points of B, there are $(3^{n/2}+1)/2$ whose coordinate sum is odd, and only $(3^{n/2}-1)/2$ whose coordinate sum is even. As we move, we always move from a point with even coordinate sum to one with an odd coordinate sum and vice versa. Hence we must start at the larger set.

Thus, (iv) is true.

(d) $\mathcal{A}_1$ is correct in general. Given an algorithm for drawing uniformly from some set S, and being asked to draw uniformly from some subset T of S, one may draw from S until obtaining an element of T. The random variable, counting the drawings until an element of T is drawn, is $G(|T|/|S|)$-distributed. Hence the average number of drawings is $|S|/|T|$, which in our case is $2^n/3^{n/2} = (4/3)^{n/2}$.

$\mathcal{A}_2$ indeed may select any set in $\mathcal{S}$ already at its first stage. However, it selects sets with incorrect probabilities. For example, $\emptyset$ is drawn by $\mathcal{A}_2$ with probability $1/2^n$ instead of $1/3^{n/2}$.

Thus, (i) is true.

2. The other $k-1$ vertices in the connected component of vertex 1 may be chosen in $\binom{n-1}{k-1}$ ways. For each of these, the connected component of vertex 1 may be chosen in T_k ways, and the other connected component in T_{n-k} ways. The edge deleted to get these two components may be any edge connecting a vertex in one component with a vertex in the other, and hence may be chosen in $k(n-k)$ ways.

   Thus, (iii) is true.

3. The distance when representing the partitions the lazy way is $\Omega(n)$. For example, for even n, the partition following

   $$n = 2 + 2 + \ldots + 2$$

   is

   $$n = 3 + 1 + 1 + \ldots + 1.$$

   The distance between the two partitions is $3n/2 - 2$.

   The distance when representing the partitions the economic way is $O(1)$. In fact, either the last pair is removed, or the last two are. Instead, we get at most two other pairs. Hence, the distance is at most 4.

   Thus, (iii) is true.

4. (a) By symmetry, each number between 1 and $2n$ appears the same number of times in the first row and in the first column. Hence it appears $\frac{1}{2} \cdot \binom{2n}{n}$ times in the first column.

   When we take any k, l between 1 and $2n$, the number of times they appear simultaneously in the first column is the number of subsets of size n of $\{1, 2, \ldots, 2n\}$, containing both of k and l, namely $\binom{2n-2}{n-2}$. Now

$$\begin{aligned}\binom{2n-2}{n-2} &= \frac{(2n-2)!}{n!(n-2)!} \\ &= \frac{n(n-1)}{2n(2n-1)} \cdot \frac{(2n)!}{n!^2} \\ &= \frac{n-1}{2(2n-1)} \cdot \binom{2n}{n} \\ &< \frac{1}{4} \cdot \binom{2n}{n}.\end{aligned}$$

Thus, (ii) is true.

(b) We do not go over the sets according to the lexicographic order. If we did, then during all the first $\frac{1}{2} \cdot \binom{2n}{n}$ steps, the number 1 would be in the first column, and thereafter would never be there. However, in all $\binom{2n-2}{n-1}$ steps when 1 is in the first row and 2 in the first column, 1 moves right away to the first column.

The changes are far from being minimal. When the first row contains the numbers $1, 2, \ldots, n-1, 2n$ and the first column contains $n, n+1, \ldots, 2n-1$, all numbers (except for 0) change their locations.

Thus, (iv) is true.