Problem 1

(6 points total)

Answer with explanations and proofs:

(a) If $f(n) = 4^{(\log n)^3}$ and $g(n) = n^4$, is $f(n) = o(g(n))$ or $g(n) = o(f(n))$? **(2 points)**

(b) All the DNA sequences of length k over the alphabet $\{A, C, G, T\}$ i.e., all strings of length k comprised of characters $'A', 'C', 'G', 'T'$ are distributed among $g(n) = n^4$ groups. How many DNA strings of length k exist? When $k = (\log n)^3$, can all these length k strings be distributed into a distinct group each i.e., can all the strings be distributed among $g(n) = n^4$ groups such that no group has more than one string ? **(2 points)**

Hint: *For the second part, focus on what happens as $n \to \infty$. Try to use the answer to the previous part, and use Pigeonhole principle to reason about this.*

(c) How many DNA sequences/strings of length k exist where at least two of the four characters from $\{A, C, G, T\}$ appear? **(2 points)**

a) First note that $(\log n)^3 > 3\log n$ for sufficiently large n, hence $4^{3\log n} = O(4^{(\log n)^3})$. More over

$4^{3\log n} = 2^{6\log n} = n^6$ and $\lim_{n\to\infty} \frac{n^4}{n^6} = 0$

hence. $g(n) = n^4 = o(n^6)$. Combining the above gives $g(n) = o(f(n))$

b) The number sequences of length k is 4^k since we make k selections and there are 4 possibilities for each selection.

It follows that if $k = (\log n)^3$, then the total number of DNA strings of length n is $4^{(\log n)^3} = f(n)$. From part a) we have $g(n) = o(f(n))$ hence $f(n) > g(n)$ for n sufficictly large. Using the pigeon hole principal then allows us to conclude that at least one group has more than one string when n is large.

c) Observe that the number of strings with at least two characters from $\{A, G, C, T\}$ is equal to the total number of strings minus the number of strings which have only one letter in them. The only strings which have exacthy one letter are AAA....AA, CCC...CC, GGG...GG, and TTT...TT. Thus the number of strings with at least two characters is $4^k - 4$.

Problem 2

(5 points) Use Induction to prove that for any $r > 0$ (and $r \neq 1$), $a \in \mathbb{R}$, and any natural number $n \geq 1$, we have

$$a + 2ar + 3ar^2 + \ldots nar^{n-1} = \frac{a\Big(nr^{n+1} - (n+1)r^n + 1\Big)}{(r-1)^2}.$$

Let $P(k)$ be the predicate that

$$a + 2ar + \cdots + kar^{k-1} = \frac{a(kr^{k+1} - (k+1)r^k + 1)}{(r-1)^2}$$

We first prove $P(1)$ is true. Here the right hand side becomes

$$\frac{a(r^2 - 2r + 1)}{r^2 - 2r + 1} = a,$$ while the LHS is a.

Thus $P(1)$ is true. Now we assume $P(k)$ is true and prove $P(k+1)$ is true
Using our induction hypothesis we have

$$a + \cdots + kar^{k-1} + (k+1)ar^k = \frac{a(kr^{k+1} - (k+1)r^k + 1)}{(r-1)^2} + (k+1)ar^k$$

$$= \frac{a(kr^{k+1} - (k+1)r^k + 1)}{(r-1)^2} + \frac{(k+1)ar^k(r-1)^2}{(r-1)^2}$$

$$= \frac{akr^{k+1} - a(k+1)r^k + a + (k+1)ar^{k+2} - 2(k+1)ar^{k+1} + a(k+1)r^k}{(r-1)^2}$$

$$= \frac{(k+1)ar^{k+2} + a(k-2k-2)ar^{k+1} + a}{(r-1)^2}$$

$$= \frac{a\left((k+1)r^{k+2} - (k+2)r^{k+1} + 1\right)}{(r-1)^2}$$

That is $P(k+1)$ is true. We conclude using induction that $P(k)$ is true for all $k \geq 1$.

Problem 3

(4 points) Given two numbers $n \geq r$, we want to calculate the number of ways of distributing n identical objects into r distinct boxes such no box is empty.

1. Suppose we represent the n identical objects with n zeroes (n of them) and use $r-1$ ones to represent the partitioning of the boxes. Explain in English (succinctly), what properties are satisfied by these $n+r-1$ long bit strings of zeros and ones. (**2 points**)

The first entry of the string must be a zero. Then, every time a one occurs in the sequence, the immediatly following digit must be a zero. That is, if the lth digit is 1, the the $l+1$ th digit is 0. Furthermore the final digit is zero.

Problem 3

2. Using the previous part, prove that the number of ways of distributing n identical objects into r distinct boxes such no box is empty is $\binom{n-1}{r-1}$. Remember that $\binom{n}{k} = \frac{n!}{k!(n-k)!}$. **(2 points)**

By associating the string 10 to a symbol X, the number of strings satisfying the above properties is equal to the number of strings consiting of 0's and X's with length $n+(r-1)-(r-1)=n$ which have exactly $n-(r-1)$ zeros and $(r-1)$ X's and have their first entry equal to zero. This is inturn equal to the number of strings of 0's and X's of length $n-1$ which have exactly $n-(r-1)-1=n-r$ zeros and $(r-1)$ X's. This quantity is equal to $\binom{n-1}{r-1}$, as claimed

Problem 4

(5 points total) For this problem, let n be an integer at least 3.

(a) A permutation of a set S is a sequence consisting of all the elements of S with no repetitions. Let S be the set of permutations of $[n]$, and consider the uniform distribution on S. For $i \in [n]$ let $E_i \subseteq S = \{s \mid s \in S, s_i = i\}$, that is, E_i is the set of permutations for which the ith element of each permutation is i. Give with proof the value of $\Pr[E_1]$ and $\Pr[E_2]$.

For example, if $n = 3$, then $E_2 = \{(1,2,3),(3,2,1)\}$.

(b) Give with proof the value of $\Pr[E_1 \cap E_2]$. Determine if E_1 and E_2 are independent, with a brief explanation why.

We have $|S| = n!$. On the other hand $|E_1| = |E_2| = (n-1)!$. Therefore

$$\Pr(E_1) = \Pr(E_2) = \frac{(n-1)!}{n!} = \frac{1}{n}$$

(b) We have $\Pr[E_1 \cap E_2] = \frac{|E_1 \cap E_2|}{|S|}$

$|E_1 \cap E_2|$ is equal to the total number of permutations which have 1 in there first position, 2 in the second position and then have the remaining $n-2$ integers in any order, Therefore

$$|E_1 \cap E_2| = (n-2)!.$$

We conclude that $P(E_1 \cap E_2) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$

On the otherhand $P(E_1)P(E_2) = \frac{1}{n^2} \neq \frac{1}{n(n-1)} = P(E_1 \cap E_2)$

Thus these events are not independent.