1 Kolmogorov Complexity

Compression of a bit string x of length n involves creating a program shorter than n bits that returns x. The Kolmogorov complexity of a string $K(x)$ is the length of shortest program that returns x (i.e. the length of a maximally compressed version of x).

(a) Explain why "the smallest positive integer not definable in under 100 characters" is paradoxical.

(b) Prove that for any length n, there must be at least one bit string that cannot be compressed to fewer than n bits.

(c) Imagine you had the program K, which outputs the Kolmogorov complexity of string. Design a program P that when given integer n outputs the bit string of length n with the highest Kolmogorov complexity. If there are multiple strings with the highest complexity, output the lexicographically first (i.e. the one that would come first in a dictionary).

(d) Suppose the program P you just wrote can be written in m bits. Show that P and by extension, K, cannot exist, for a sufficiently large input n.

Solution:

(a) Since there are only a finite number of characters then there are only a finite number of positive integers that can be defined in under 100 characters. Therefore there must be positive integers that are not definable in 100 characters and by the well-ordering principle there is a smallest member of that set. However the statement "the smallest positive integer not definable in under 100 characters" defines the smallest such an integer using only 67 characters (including spaces). Hence, we have a paradox (called the Berry Paradox).

(b) The number of strings of length n is 2^n. The number of strings shorter than length n is $\sum_{i=0}^{n-1} 2^i$. We know that sum is equal to $2^n - 1$ (remember how binary works). Therefore the cardinality of the set of strings shorter than n is smaller than the cardinality of strings of length n. Therefore there must be strings of length n that cannot be compressed to shorter strings.

(c) We write such a program as follows:

```
def P(n):
  complex_string = "0" * n
  for j in range(1, 2 ** n):
```

```
    # some fancy Python to convert j into binary
    bit_string = "0:b".format(j)
    # length should now be n characters
    bit_string = (n - len(bit_string)) * "0" + bit_string
    if K(bit_string) > K(complex_string):
      complex_string = bit_string
  return complex_string
```

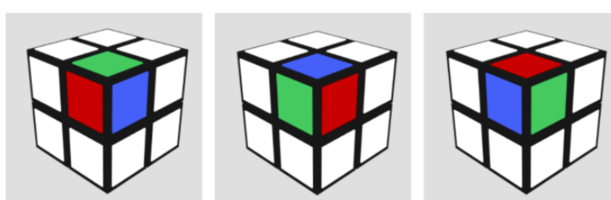
(d) We know that for every value of n there must be an incompressible string. Such an incompressible string would have a Kolmogorov complexity greater than or equal to its actual length. Therefore our program P must return an incompressible string. However, suppose we choose size n_k such that $n_k \gg m$. Our program $P(n_k)$ will output a string x of length n_k that is not compressible meaning $K(x) \geq n_k$. However we have designed a program that outputs x using fewer bits than n_k. This is a contradiction. Therefore K cannot exist.

2 Rubik's Cube Scrambles

We wish to count the number of ways to scramble a $2 \times 2 \times 2$ Rubik's Cube, and take a quick look at the $3 \times 3 \times 3$ cube. Leave your answer as an expression (rather than trying to evaluate it to get a specific number).

(a) The $2 \times 2 \times 2$ Rubik's Cube is assembled from 8 "corner pieces" arranged in a $2 \times 2 \times 2$ cube. How many ways can we assign all the corner pieces a position?

(b) Each corner piece has three distinct colors on it, and so can also be oriented three different ways once it is assigned a position (see figure below). How many ways can we *assemble* (assign each piece a position and orientation) a $2 \times 2 \times 2$ Rubik's Cube?



Three orientations of a corner piece

(c) The previous part assumed we can take apart pieces and assemble them as we wish. But certain configurations are unreachable if we restrict ourselves to just turning the sides of the cube. What this means for us is that if the orientations of 7 out of 8 of the corner pieces are determined, there is only 1 valid orientation for the eighth piece. Given this, how many ways are there to *scramble* (as opposed to *assemble*) a $2 \times 2 \times 2$ Rubik's Cube?

(d) We decide to treat scrambles that differ only in overall positioning (in other words, the entire cube is flipped upside-down or rotated but otherwise unaltered) as the same scramble. Then we overcounted in the previous part! How does this new condition change your answer to the previous part?

(e) Now consider the $3 \times 3 \times 3$ Rubik's Cube. In addition to 8 corner pieces, we now have 12 "edge" pieces, each of which can take 2 orientations. How many ways can we *assemble* a $3 \times 3 \times 3$ Rubik's Cube?

Solution:

(a) 8! ways. This is just the number of permutations of 8 objects.

(b) $8! \cdot 3^8$. Each of eight pieces has 3 possible orientations, to add onto the 8! permutations from part (a).

(c) $8! \cdot 3^7$. We divide the previous part by three. We can still choose the orientation of 7 of 8 corner pieces, but the eighth is then fixed.

(d) For each scramble, there are $6 \cdot 4 = 24$ ways to position/rotate it. We can see this by first choosing one of 6 faces to be on "top", then choosing one of four rotations with that particular face on top. So now we have $\frac{1}{24} \cdot 8! \cdot 3^7$.

(e) $8! \cdot 3^8 \cdot 12! \cdot 2^{12}$. We use the same procedure on the edge pieces as we do the corner pieces.

3 The Count

(a) How many permutations of COSTUME contain "COME" as a substring? How about as a subsequence (meaning the letters of "COME" have to appear in that order, but not necessarily next to each other)?

(b) How many of the first 100 positive integers are divisible by 2, 3, or 5?

(c) How many ways are there to choose five nonnegative integers x_0, x_1, x_2, x_3, x_4 such that $x_0 + x_1 + x_2 + x_3 + x_4 = 100$, and $x_i \equiv i \pmod 5$?

(d) The Count is trying to choose his new 7-digit phone number. Since he is picky about his numbers, he wants it to have the property that the digits are non-increasing when read from left to right. For example, 9973220 is a valid phone number, but 9876545 is not. How many choices for a new phone number does he have?

Solution:

(a) If we need "COME" to be a substring, then this is equivalent to finding the number of ways to arrange 4 items in a row ("COME", "S", "T", and "U"), or $4! = 24$ ways.

If we need "COME" to be a subsequence, let's first choose which 4 positions contain the letters of COME; then, that uniquely determines the positions of C, O, M, E, since they have to be in that order. There are $\binom{7}{4}$ ways to do this, and 3! ways to choose the positions of the remaining three letters, so there are a total of $\binom{7}{4}3! = 210$ valid subsequences.

(b) We use inclusion-exclusion to calculate the number of numbers that satisfy this property. Let A be the set of numbers divisible by 2, B be the set of numbers divisible by 3, and C be the set of numbers divisible by 5. Then, we calculate

$$\begin{aligned}&|A|+|B|+|C|-|A\cap B|-|A\cap C|-|B\cap C|+|A\cap B\cap C|\\ &\quad=\left\lfloor\frac{100}{2}\right\rfloor+\left\lfloor\frac{100}{3}\right\rfloor+\left\lfloor\frac{100}{5}\right\rfloor-\left\lfloor\frac{100}{6}\right\rfloor-\left\lfloor\frac{100}{10}\right\rfloor-\left\lfloor\frac{100}{15}\right\rfloor+\left\lfloor\frac{100}{30}\right\rfloor\\ &=50+33+20-16-10-6+3=74\end{aligned}$$

numbers.

(c) Let $x_i = 5y_i + i$ for nonnegative integers y_i (we do this because of the modulo conditions). Then, the equation we must satisfy becomes $(5y_0+0)+(5y_1+1)+(5y_2+2)+(5y_3+3)+(5y_4+4)=100$, which simplifies down to $y_0+y_1+y_2+y_3+y_4=18$ for nonnegative integers y_i. This is a standard stars and bars problem, with 18 objects and 4 dividers. We want to choose how to put the 4 dividers out of the 22 objects, so our answer is $\binom{22}{4}$.

Why is $y_0+y_1+y_2+y_3+y_4=18$ a stars and bars problem? y_i is the number of stars in bin i, and 18 is the number of stars in total. Here are a few possible arrangements:

$$\underbrace{\star\star\star\star\star\star}_{y_0=6}\,|\,\underbrace{\quad}_{y_1=0}\,|\,\underbrace{\star\star\star\star\star\star\star}_{y_2=7}\,|\,\underbrace{\star\star\star\star\star}_{y_3=5}\,|\,\underbrace{\quad}_{y_4=0}$$

$$\underbrace{\star\star}_{y_0=2}\,|\,\underbrace{\star\star\star\star}_{y_1=4}\,|\,\underbrace{\star\star\star\star}_{y_2=4}\,|\,\underbrace{\star\star\star\star\star\star}_{y_3=6}\,|\,\underbrace{\star\star}_{y_4=2}$$

Notice that we arrange the stars into <u>5 bins</u> using <u>4 bars</u>. The total number of ways we can arrange 18 stars and 4 bars side-by-side in a line is $\binom{22}{4}$. Why?

- 22 is the total number of objects in the line, 18 stars + 4 bars.
- 4 is the number of bars to be placed in the line.

"22 choose 4", or $\binom{22}{4}$, is the number of ways we can position 4 bars on a line <u>with</u> 18 indistinguishable stars.

(d) This is actually a stars and bars problem in disguise! We have seven positions for digits, and nine dividers to partition these positions into places for nines, places for eights, etc. This is because we know that the digits are non-increasing, so all the nines (if any) must come first, then all the eights (if any), and so on. That means there are a total of 16 objects and dividers, and we are looking for where to put the nine dividers, so our answer is $\binom{16}{9}$.

4 Binomial Beads

(a) Satish is making school spirit keychains, which consist of a sequence of n beads on a string. He has blue beads and gold beads. How many unique keychains can he make with exactly $k \le n$ blue beads?

(b) Satish decides to sell his keychains! He decides on the following pricing scheme:

- Blue beads have a value of x
- Gold beads have a value of y
- The price of a keychain is the product of the values of all of its beads.

What is the price of a keychain with exactly $k \leq n$ blue beads?

(c) Satish decides to make exactly one of every possible unique keychain. If he sells every keychain he creates, how much revenue will he make? Use parts (a) and (b), and leave your answer in summation form.

(d) Draw a connection between part (c) and the binomial theorem.

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Hint: How do you calculate the product (x + y)(x + y)?

Solution:

(a) $\binom{n}{k}$. We choose k locations for the blue beads to go.

(b) $x^k y^{n-k}$.

(c) Satish can place $0-n$ blue beads on a string. Using the above parts, we have

$$\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

(d) The expansion of the LHS results in the sum of many terms, with each term representing the product of a string of x's and y's. We can think of the LHS as summing up the prices of all the keychains individually. The RHS, as calculated in part (c), is also the sum of the prices of each unique bracelet, but grouped by the number of "blue beads".

5 Probability Practice

(a) If we put 5 math, 6 biology, 8 engineering, and 3 physics books on a bookshelf at random, what is the probability that all the math books are together?

(b) A message source M of a digital communication system outputs a word of length 8 characters, with the characters drawn from the ternary alphabet $\{0,1,2\}$, and all such words are equally probable. What is the probability that M produces a word that looks like a byte (*i.e.*, no appearance of '2')?

(c) If five numbers are selected at random from the set $\{1,2,3,\ldots,20\}$, what is the probability that their minimum is larger than 5? (A number can be chosen more than once, and the order in which you select the numbers matters)

Solution:

(a) $18!5!/22! = 1/1463$. The 18! comes from 18 "units": 3 physics books, 8 engineering books, 6 biology books and 1 block of math books. The 5! comes from number of ways to arrange the 5 math books within the same block. 22! is just the total number of ways to arrange the books.

(b) $\left(\frac{2^8}{3^8}\right) = 256/6561$. There are 2^8 possible binary-like strings out of 3^8 total possible ternary strings.

(c) $\left(\frac{15^5}{20^5}\right) = 243/1024$. There are 20^5 total possible sequences of numbers we might select. In order for the minimum to be larger than 5, we need to have our entire sequence made up only of the numbers $\{6,7,...,20\}$. Since there are 15 of these numbers, there are 15^5 possible sequences of them.

6 Shooting Range

You and your friend are at a shooting range. You ran out of bullets. Your friend still has two bullets left but magically lost his gun. Somehow you both agree to put the two bullets into your six-chambered revolver in successive order, spin the revolver, and then take turns shooting. Your first shot was a blank. You want your friend to shoot a blank too; should you spin the revolver again before you hand it to your friend?

Solution:

No, you shouldn't.

The first chamber fired was one of the four empty chambers. Since the bullets were placed in consecutive order, one of the empty chambers is followed by a bullet, and the other three empty chambers are followed by another empty chamber. So the probability that a bullet will be fired is 1/4.

If you spins the chamber again, the probability that a real bullet is shot would be 2/6, or 1/3, since there are two possible bullets that would be in firing position out of the six possible chambers that would be in position.

7 Cliques in Random Graphs

In last week's homework you worked on a graph $G = (V,E)$ on n vertices which is generated by the following random process: for each pair of vertices u and v, we flip a fair coin and place an (undirected) edge between u and v if and only if the coin comes up heads. Now consider:

(a) What is the size of the sample space?

(b) A k-clique in graph is a set of k vertices which are pairwise adjacent (every pair of vertices is connected by an edge). For example a 3-clique is a triangle. What is the probability that a particular set of k vertices forms a k-clique?

Solution:

(a) There are two choices for each of the $\binom{n}{2}$ pairs of vertices, so the size of the sample space is $2^{\binom{n}{2}}$.

(b) For a fixed set of k vertices to be a k-clique, all of the $\binom{k}{2}$ pairs of those vertices have to be connected by an edge. The probability of this event is $1/2^{\binom{k}{2}}$.