CMPT 210: Probability and Computing

Lecture 4

Sharan Vaswani

January 18, 2024

Recap

Number of ways of choosing size k-subsets from a size n-set: $\binom{n}{k}$ (E.g. Number of n-bit sequences with exactly k ones).

Binomial Theorem: For all $n \in \mathbb{N}$ and $a, b \in \mathbb{R}$, $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$.

Generalization to Multinomials

We saw how to split a set into two subsets - one that contains some elements, while the other does not. Can generalize the arguments to split a set into more than two subsets.

A $(k_1, k_2, \ldots, k_m)$-split of set A is a sequence of sets $(A_1, A_2, \ldots A_m)$ s.t. sets A_i form a partition ($A_1 \cup A_2 \cup \ldots = A$ and for $i \neq j$, $A_i \cap A_j = \emptyset$) and $|A_i| = k_i$.

An example of a $(2, 1, 3)$-split of $A = \{1, 2, 3, 4, 5, 6\}$ is $(\{2, 4\}, \{1\}, \{3, 5, 6\})$. Here, $m = 3$, $A_1 = \{2, 4\}$, $A_2 = \{1\}$, $A_3 = \{3, 5, 6\}$ s.t. $|A_1| = 2$, $|A_2| = 1$, $|A_3| = 3$, $A_1 \cup A_2 \cup A_3 = A$ and for $i \neq j$, $A_i \cap A_j = \emptyset$.

Example: Consider strings of length 6 of a's, b's and c's such that number of a's $= 2$; number of b's $= 1$ and number of c's $= 3$. Possible strings: abaccc, ccbaac, bacacc, cbacac.

Each possible string, e.g. bacacc can be written as a $(2, 1, 3)$-split of $A = \{1, 2, 3, 4, 5, 6\}$ as $(\{2, 4\}, \{1\}, \{3, 5, 6\})$ where A_1 records the positions of a, A_2 records the positions of b and A_3 records the positions of c.

Generalization to Multinomials

Q: Show that the number of ways to obtain an $(k_1, k_2, \dots, k_m)$ split of A with $|A| = n$ is $\binom{n}{k_1, k_2, \dots k_m} = \frac{n!}{k_1!\, k_2! \dots k_m!}$ where $\sum_i k_i = n$.

Can map any permutation $(a_1, a_2, \dots a_n)$ into a split by selecting the first k_1 elements to form set A_1, next k_2 to form set A_2 and so on. For the same split, the order of the elements in each subset does not matter. Hence f : number of permutations $\to$ number of splits is a $k_1!\, k_2!\, \dots k_m!$-to-1 function.

Hence, $|\text{number of splits}| = \frac{|\text{number of permutations}|}{k_1!\, k_2! \dots k_m!} = \frac{n!}{k_1!\, k_2! \dots k_m!}$.

Generalization to Multinomials - Example

Q: Count the number of permutations of the letters in the word BOOKKEEPER.

We want to count sequences of the form $(1E, 1P, 2E, 1B, 1K, 1R, 2O, 1K) = EPEEBKROOK$. There is a bijection between such sequences and $(1, 2, 2, 3, 1, 1)$ split of $A = \{1, 2, \ldots, 10\}$ where A_1 is the set of positions of B's, A_2 is the set of positions of O's, A_3 is set of positions of K and so on.

For example, the above sequence maps to the following split:
({5}, {8,9}, { 6, 10}, { 1,3,4 }, { 2 }, { 7 })

Hence, the total number of sequences that can be formed from the letters in BOOKKEEPER = number of $(1, 2, 2, 3, 1, 1)$ splits of $A = [10] = \{1, 2, \ldots, 10\} = \frac{10!}{1!\,2!\,2!\,3!\,1!\,1!}$.

Q: Count the number of permutations of the letters in the word (i) ABBA (ii) A_1BBA_2 and (iii) $A_1B_1B_2A_2$? Ans: 6, 12, 24

Generalization to Multinomials - Example

Q: Suppose we are planning a 20 km walk, which should include 5 northward km, 5 eastward km, 5 southward km, and 5 westward km. We can move in steps of 1 km in any direction. For example, a valid walk is $(NENWSNSSENSWWESWEENW)$ that corresponds to 1 km north followed by 1 km east and so on. How many different walks are possible?

Ans: The set $A = \{1, 2, \dots, 20\}$ needs to be split into 4 subsets N, S, E, W s.t. $|N| = |S| = |E| = |W| = 5$. Counting the number of walks = counting the number of sequences of the form $(3N, 5W, 4S, 4E, 2N, 1E, 1S)$ = number of ways to obtain an $(5, 5, 5, 5)$-split of set $\{1, 2, 3, \dots 20\}$. The total number of walks = $\frac{20!}{(5!)^4}$.

Multinomial Theorem

For all $m, n \in \mathbb{N}$ and $z_1, z_2, \ldots z_m \in \mathbb{R}$,

$$(z_1 + z_2 + \ldots + z_m)^n = \sum_{\substack{k_1, k_2, \ldots, k_m \\ k_1 + k_2 + \ldots k_m = n}} \binom{n}{k_1, k_2, \ldots, k_m} z_1^{k_1} z_2^{k_2} \ldots z_m^{k_m}$$

where $\binom{n}{k_1, k_2, \ldots, k_m} = \frac{n!}{k_1! k_2! \ldots k_m!}$.

Example 1: If $m = 2$, $k_1 = k$, $k_2 = n - k$ and $z_1 = a$, $z_2 = b$, recover the Binomial theorem.

Example 2: If $n = 4$, $m = 3$, then the coefficient of abc^2 in $(a + b + c)^4$ is $\binom{4}{1,1,2} = \frac{4!}{1!1!2!}$.

Questions?

Inclusion-Exclusion Principle

Recall that if A, B, C are disjoint subsets, then, $|A \cup B \cup C| = |A| + |B| + |C|$ (this is the Sum rule from Lecture 2).

For two general sets A, B, $|A \cup B| = |A| + |B| - |A \cap B|$. The last term fixes the "double counting".

Similarly, $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |B \cap C| - |A \cap C| + |A \cap B \cap C|$.

In general,

$$|\cup_{i=1,2,\ldots n} A_i| = \sum_i |A_i| - \sum_{i,j \text{ s.t. } 1 \le i < j \le n} |A_i \cap A_j| + \sum_{i,j,k \text{ s.t. } 1 \le i < j < k \le n} |A_i \cap A_j \cap A_k| + \ldots + (-1)^{n-1} |\cap_{i=1,2,\ldots n} A_i|$$

Inclusion-Exclusion Principle - Example

Q: Suppose there are 60 math majors, 200 EECS majors, and 40 physics majors. A student is allowed to double or even triple major. There are 4 math-EECS double majors, 3 math-physics double majors, 11 EECS-physics double majors and 2-triple majors. What is the total number of students across these three departments?

If M, E, P are the sets of students majoring in math, EECS and physics respectively, then we wish to compute $|M \cup E \cup P| = |M| + |E| + |P| - |M \cap E| - |M \cap P| - |E \cap P| + |M \cap E \cap P|$ $=$ 300 - $|M \cap E| - |M \cap P| - |E \cap P| + |M \cap E \cap P|$.

$|M \cap E| = 4 + 2 = 6$, $|M \cap P| = 3 + 2 = 5$, $|P \cap E| = 11 + 2 = 13$. $|M \cap E \cap P| = 2$

$|M \cup E \cup P| = 300 - 6 - 5 - 13 + 2 = 278$.

Inclusion-Exclusion Principle - Example

Q: In how many permutations of the set $\{0, 1, 2, \ldots, 9\}$ do either 4 and 2, 0 and 4, or 6 and 0 appear consecutively? For example, in the following permutation $\underline{42}067891235$, 4 and 2 appear consecutively, but 6 and 0 do not (the order matters).

Let P_{42} be the set of sequences such that 4 and 2 appear consecutively. Similarly, we define P_{60} and P_{04}. So we want to compute
$|P_{42} \cup P_{60} \cup P_{04}| = |P_{42}| + |P_{60}| + |P_{04}| - |P_{42} \cap P_{60}| - |P_{42} \cap P_{04}| - |P_{60} \cap P_{04}| + |P_{42} \cap P_{60} \cap P_{04}|$.

Let us first compute $|P_{42}| = 9!$. Similarly, $|P_{60}| = |P_{04}| = 9!$.

What about intersections? $|P_{42} \cap P_{60}| =$ Number of sequences of the form $(42, 60, 1, 3, 5, 7, 8, 9) = 8!$. Similarly, $|P_{60} \cap P_{04}| = |P_{42} \cap P_{04}| = 8!$.

$|P_{42} \cap P_{60} \cap P_{04}| =$ Number of sequences of the form $(6042, 1, 3, 5, 7, 8, 9) = 7!$.

By the inclusion-exclusion principle, $|P_{42} \cup P_{60} \cup P_{04}| = 3 \times 9! - 3 \times 8! + 7!$.

Combinatorial Proofs

Recall that if we have to choose k elements out of a size n set. Number of ways to do this is $\binom{n}{k}$. But this is equivalent to saying, we want to find the number of ways to throw away $n-k$ elements $= \binom{n}{n-k}$. Hence, $\binom{n}{k} = \binom{n}{n-k}$. Can prove algebraic statements using combinatorial arguments.

Q: Prove Pascal's identity using a combinatorial proof: $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$

Consider n students in this class. What is the number of ways of selecting k students? $\binom{n}{k}$.

What is the number of ways of selecting k students if we have to ensure to include a particular student? $\binom{n-1}{k-1}$.

What is the number of ways of selecting k students if we have to ensure to NOT include a particular student? $\binom{n-1}{k}$.

Number of ways to select k students $=$ number of ways of selecting k students to include a particular student $+$ number of ways of selecting k students to NOT include a particular student. Hence, $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$.

Counting Practice

Q: In how many ways can we place (i) two identical black rooks (ii) a black rook and a white rook such that they do not share the same row or column?

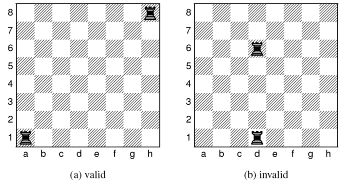


Figure 15.2 Two ways to place 2 rooks (♖) on a chessboard. The configuration in (b) is invalid because the rooks are in the same column.

Ans: The first rook can occupy 8×8 positions. After selecting the first rook, the number of valid remaining positions $= 7 \times 7$. Since two positions are equivalent (because these are two identical rooks), by the division rule, total number of ways to place the rooks $= \frac{8^2\,7^2}{2} = 32 \times 49$.

Ans: Same as before but since the two rooks are different, we are not double-counting. Hence, the number of ways $= 64 \times 49$.

Questions?

Pigeonhole principle

Q: A drawer in a dark room contains red socks, green socks, and blue socks. How many socks must you withdraw to be sure that you have a matching pair?

Such problems can be tackled using the Pigeonhole principle.

Pigeonhole Principle: If there are more pigeons than holes they occupy, then there must be at least two pigeons in the same hole.

Formally, if $|A| > |B|$, then for every total function (one that has an assignment for every element in A), $f : A \to B$, there exist two different elements of A that are mapped by f to the same element of B.

For the above problem, A = set of socks we picked = pigeons, B = set of colors {red, blue, green} = pigeonholes. $|A|$ = number of socks we picked. $|B| = 3$. $f : A \to B$ s.t. f(sock we picked) = it's color.

If there are more pigeons than holes (picked socks than colors), then at least two pigeons will be in the same hole (two of the picked socks will have the same color, and we get a matching pair). Hence, to ensure a matching pair, we need to pick 4 socks.

Pigeonhole principle - Example

Q: A class has 54 students. Prove that there exist at least 2 students with their birthday in the same week.

Ans: 54 students = pigeons. 52 weeks = pigeonholes.

Q: In the set of integers $\{1, 2, \ldots, 100\}$, use the pigeonhole principle to prove that there exist two numbers whose difference is a multiple of 41.

Ans: $\{1, 2, \ldots, 100\}$ = pigeons, $\{0, 1, 2, \ldots 40\}$ = holes, $f : \{1, 2, \ldots, 100\} \rightarrow \{0, 1, 2, \ldots 40\}$ s.t. $f(n) = n \bmod 41$ i.e. $f(n)$ returns the remainder after dividing by 41. Since $|\text{pigeons}| > |\text{holes}|$, there exist 2 numbers a, b that have the same remainder after dividing by 41. Let the remainder be r, then $a = 41m_1 + r$ and $b = 41m_2 + r$ where m_1, m_2 are integers.
$a - b = 41(m_1 - m_2)$. Hence, $a - b$ is a multiple of 41.

Pigeonhole principle - Example

A kind of problem that arises in cryptography is to find different subsets of numbers with the same sum. For example, in this list of 25-digit numbers, find a subset of numbers that have the same sum. For example, maybe the sum of the last ten numbers in the first column is equal to the sum of the first eleven numbers in the second column.

0020480135385502964448038	3171004832173501394113017
5763257331083479647409398	8247331000042995311646021
0489445991866915676240992	3208234421597368647019265
5800949123548989122628663	8496243997123475922766310
1082662032430379651370981	3437254656355157864869113
6042900801199280218026001	8518399140676002660747477
1178480894769706178994993	3574883393058653923711365
6116171789137737896701405	8543691283470191452333763
1253127351683239693851327	3644909946040480189969149
6144868973001582369723512	8675309258374137092461352
1301505129234077811069011	3790044132737084094417246
6247314593851169234746152	8694321112363996867296665
1311567111143866433882194	3870332127437971355322815
6814428944266874963488274	8772321203608477245851154
1470029452721203587686214	4080505804577801451363100
6870852945543886849147881	8791422161722582546341091
1578271047286257499433886	4167283461025702348124920
6914955508120950093732397	9062628024592126283973285
1638243921852176243192354	4235996831123777788211249
6949632451365987152423541	9137845566925526349897794
1763580219131985963102365	4670939445749439042111220
7128211143613619828415650	9153762966803189291934419
1826227795601842231029694	4815379351865384279613427
7173920083651862307925394	9270880194077636406984249
1843971862675102037201420	4837052948212922604442190

This is a hard problem which is why it is used in cryptography. The first step to figure out is whether there even exists such a subset of numbers. We can do this using the pigeonhole principle!

Pigeonhole principle - Example

Q: More generally, in a list of n b-digit numbers, are there two different subsets of numbers that have the same sum?

Let A = set of all subsets of the n numbers. For example, if $b = 3$, an element of A is $\{113, 221\}$. $|A| = 2^n$

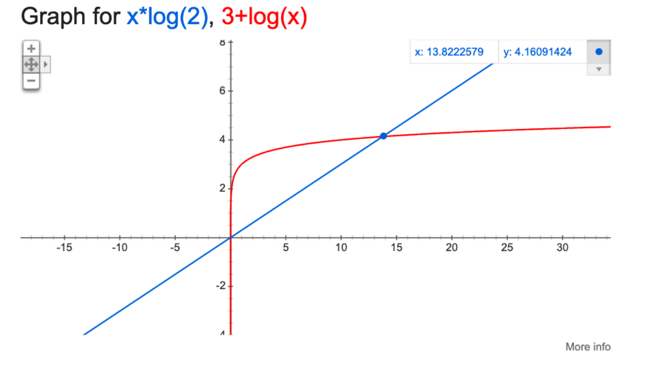
Let B be the set of possible sums of such subsets. f is a function that maps each subset to its corresponding sum. For example, if $b = 3$, $f(\{113, 221\}) = 334$.

Let us compute $|B|$. For any list of n numbers, the minimum possible sum $= 0$ and the max possible sum $< 10^b \times n$. For example, if $b = 3$ and $n = 5$, then the maximum possible sum $=$ $999 \times 5 < 1000 \times 5$. Hence, $|B| \leq 10^b \times n$.

By the pigeonhole principle, for any list of n b-digit numbers, there definitely exist different subsets with the same sum if $|A| > |B|$ i.e. if $2^n > 10^b \times n$.

For $b = 3$, this is possible if $2^n > 1000n$, meaning this is possible if $n \log(2) > 3 + \log(n)$ (since log is a monotonic function). Let's plot.

Pigeonhole - Example



Hence, it is possible when $n > 15$. Similarly, for a general b, there exist different subsets with the same sum if $n \log(2) > b + \log(n)$.

Questions?