CMPT 210: Probability and Computing

Lecture 5

Sharan Vaswani

September 19, 2024

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.

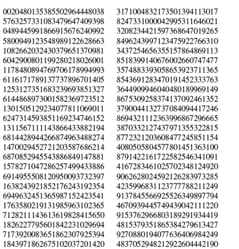
Pigeonhole principle - Example

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

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.

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 $\{113, 221\}$. $|A| = 2^n$

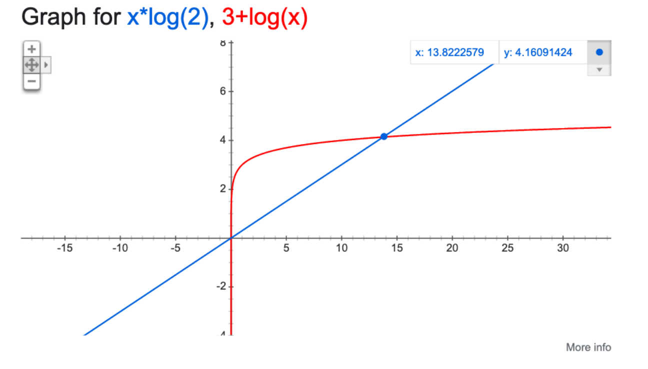
Let B be the set of possible sums of such subsets. f is a function that maps each subset to its corresponding sum. For example, if $b = 3$, $f(\{113, 221\}) = 334$.

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?

Introduction to Probability - Throwing dice

Q: Suppose we throw a standard dice. What is the probability that the number that comes up is 6?

What are the possible things that can happen? The dice comes up one of the numbers in $\{1, 2, 3, 4, 5, 6\}$.

What are the things that we care about? Getting a 6.

In how many ways can this happen? Just one.

Probability of getting a 6 $= \frac{\text{Number of ways in which the thing we care about happens}}{\text{Total number of ways in which something can happen}} = \frac{1}{6}$.

Introduction to Probability - Throwing dice

Q: Suppose we throw a standard dice. What is the probability that we get either a 3 or a 6?

What are the possible *outcomes* that can happen? The dice comes up one of the numbers in $\{1, 2, 3, 4, 5, 6\}$.

What is the *event* that we care about? Getting either a 3 or 6.

In how many ways can this *event* happen? Two (the dice comes 3 or 6).

Probability of getting either a 3 or a 6 $= \frac{\text{Number of ways in which the event we care about happens}}{\text{Total number of outcomes}} = \frac{2}{6}$.

Introduction to Probability - Throwing dice

Q: Suppose we throw two standard dice one after the other. What is the probability that we get two 6's in a row?

What are the possible outcomes that can happen? The first dice comes up one of the numbers in $1, 2, 3, 4, 5, 6$, the second dice comes up one of the numbers in $1, 2, 3, 4, 5, 6$.

If we consider both dice together, what are the possible outcomes – first dice is 1, second dice is 1; first is 1, second is 2, and so on. Let us write this compactly. The space of outcomes is $\{(1, 1), (1, 2), (1, 3), \ldots, (6, 6)\}$.

What is the size of this *outcome space*? 36 (By the product rule)

What is the event that we care about? Getting $(6, 6)$.

In how many ways can this happen? One (both die need to come up 6).

Probability of getting two 6's in a row $= \frac{\text{Number of ways in which the event we care about happens}}{|\text{outcome space}|} = \frac{1}{36}$.

Probability Basics

Sample (outcome) space $\mathcal{S}$: Nonempty (countable) set of possible outcomes. *Example*: When we threw one dice, the sample space is $\{1, 2, 3, 4, 5, 6\}$. When we threw two die, the sample space is $\{(1, 1), (1, 2), (1, 3), \ldots\} = \{1, 2, 3, 4, 5, 6\} \times \{1, 2, 3, 4, 5, 6\}$ (using the relation between sets and sequences).

The sample space is not necessarily numbers. *Example*: If we are randomly choosing colors from the rainbow, then $\mathcal{S} = \{\text{violet, indigo, blue, green, yellow, orange, red}\}$.

Outcome $\omega \in \mathcal{S}$: Possible "thing" that can happen. *Example*: When we threw one dice, a possible outcome is $\omega = 1$. For the rainbow example, the color "red" is a possible outcome.

Event E: Any subset of the sample space. *Example*: When we threw one dice, a possible event is $E = \{6\}$ (first example) or $E = \{3, 6\}$ (second example). When we threw two die, a possible event is $E = \{(6, 6)\}$.

An event E "happens" if the outcome ω (from some process) is in set E i.e. if $\omega \in E$.

Union of events

Since the event E is a set, all the set theory we learned is useful!

Suppose E, F are two events in $\mathcal{S}$. Define the union $E \cup F$ to consist of outcomes that are either in E or F (this is just the definition of the union of two sets). Formally,

$$G = E \cup F = \{\omega | \omega \in E \text{ OR } \omega \in F\}.$$

Another way to interpret this is to say event G occurs if either event E or event F occurs.

Example: We considered the case where we threw one dice and cared about getting *either* 3 or 6. In this case, event G happens if we get either 3 or 6. Formally, $E = \{3\}$, $F = \{6\}$, $G = E \cup F = \{3, 6\}$. And G occurs when the number that shows up is either 3 or 6.

Can define union between more than two events in the same way we defined union between more than two sets. $G = E_1 \cup E_2 \cup \ldots E_n$. G happens when *at least one* of the events E_i happen.

Intersection of events

Suppose E, F are two events in $\mathcal{S}$. Define the intersection $E \cap F$ to consist of outcomes that are in both E and F (this is just the definition of the intersection of two sets). Formally,

$$G = E \cap F = \{\omega | \omega \in E \text{ AND } \omega \in F\}$$

.

Another way to interpret this is to say event G occurs if both events E and F occur.

Example: We threw two dice and cared about getting 6 in the first throw *and* 6 in the second throw. In this case, E is the event we get a 6 for the first dice. $E = \{(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}$, F is the event we get a 6 for the second dice. $F = \{(1,6),(2,6),(3,6),(4,6),(5,6),(6,6)\}$, $G = E \cap F = \{(6,6)\}$. G happens when both E and F happen i.e. the first dice has a 6 and the second dice has 6.

Can define intersection between more than two events in the same way we defined intersection between more than two sets. $G = E_1 \cap E_2 \cap \ldots E_n$. G happens when *all* of the events E_i happen.

Mutually exclusive and complement events

Mutually exclusive events: If E and F are two events such that $E \cap F = \{\}$, then events E and F are mutually exclusive.

Example: We threw one dice and want to get both 3 *and* 6. This is not possible. Formally, $E = \{6\}$, $F = \{3\}$ and $E \cap F = \{\}$, hence, events E and F are mutually exclusive.

Complement of an event: If E is an event, then its complement E^c is defined such that $E \cap E^c = \{\}$ and $E \cup E^c = \mathcal{S}$. Event E^c will occur if and only if event E does not occur.

Example: We threw one dice and want to get a 6 i.e. we define $E = \{6\}$. $E^c = \{1, 2, 3, 4, 5\}$.

Two complement events are mutually exclusive, but two mutually exclusive events need not be the complements of each other. *Example*: $E = \{6\}$ and $F = \{3\}$ are mutually exclusive, but not complements.

Subset: If $E \subset F$, then if E happens F will happen. *Example*: When we throw one dice, if $E = \{3\}$ and $F = \{1, 2, 3\}$ i.e. E is the event that we get 3 and F is the event that we can either $1, 2, 3$. Clearly, if E happens, F will happen.

Axioms of Probability

Probability function on a sample space $\mathcal{S}$ is a total function $\text{Pr} : \mathcal{S} \to [0,1]$.

For any $\omega \in \mathcal{S}$, $0 \leq Pr[\omega] \leq 1 \quad ; \quad \sum_{\omega \in \mathcal{S}} \text{Pr}[\omega] = 1$

Probability space: The outcome space $\mathcal{S}$ together with the probability function.

Recall that we can define functions on sets. In this case, for an event E, $\text{Pr}[E] = \sum_{\omega \in E} \text{Pr}[\omega]$.

Union: For mutually exclusive events $E_1, E_2, \ldots, E_n$ (sets $E_1, E_2, \ldots, E_n$ are disjoint), $\text{Pr}[E_1 \cup E_2 \cup \ldots E_n] = \text{Pr}[E_1] + \text{Pr}[E_2] + \ldots + \text{Pr}[E_n]$.

Proof:

$$\text{Pr}[E_1 \cup E_2 \cup \ldots E_n] = \sum_{\omega \in \{E_\mathbf{1} \cup E_\mathbf{2} \cup \ldots E_n\}} \text{Pr}[\omega]$$

Since E_i's are disjoint, any ω can only be in one of $E_1, E_2, \ldots E_n$

$$= \sum_{\omega \in E_\mathbf{1}} \text{Pr}[\omega] + \sum_{\omega \in E_\mathbf{2}} \text{Pr}[\omega] + \ldots + \sum_{\omega \in E_n} \text{Pr}[\omega] = \text{Pr}[E_1] + \text{Pr}[E_2] + \ldots + \text{Pr}[E_n].$$

Back to throwing dice

Q: Suppose we throw a standard dice. What is the probability that the number that comes up is 6?

$\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. Since the dice is "standard", each outcome is equally likely, i.e. $\Pr[1] = \Pr[2] = \ldots = \Pr[6]$.

Since $\Pr[\mathcal{S}] = 1 \implies \sum_{\omega \in \mathcal{S}} \Pr[\omega] = 1 \implies \Pr[1] + \Pr[2] + \ldots \Pr[6] = 1$ $\implies \Pr[6] = \frac{1}{6}$.

Back to throwing dice

Q: Suppose we throw a standard dice. What is the probability that we get either a 3 or a 6?

$E = \{3\}$, $F = \{6\}$, $G = \{3, 6\}$. Since $E \cap F = \{\}$, E and F are mutually exclusive events, implying that $\Pr[G] = \Pr[E] + \Pr[F] = \Pr[\{3\}] + \Pr[\{6\}] = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Hence, probability of getting either a 3 or a 6 is equal to $\frac{1}{3}$.

Q: Compute the probability of getting either 1, 2 or 3.

Q: Compute the probability of getting an even number.

Q: Compute the probability of getting either 1, 2, 3, 4, 5, 6