Pigeonhole Principle

Robert Y. Lewis

CS 0220 2023

April 5, 2023

Overview

Sets of permutations

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?

Which of these permutations has this property?

- $(4, 6, 5, 0, 1, 8, 3, 2, 9, 7)$ nope.
- $(0, 4, 6, 1, 8, 5, 9, 3, 7, 2)$ 04!
- $(3, 4, 2, 0, 5, 6, 1, 9, 8, 7)$ 42!
- $(3, 9, 4, 1, 2, 7, 0, 5, 6, 8)$ nope.
- $(0, 2, 6, 3, 7, 8, 4, 9, 5, 1)$ nope.

P_{60}: permutations of 0 through 9 that contain 60.

P_{04}: permutations of 0 through 9 that contain 04.

P_{42}: permutations of 0 through 9 that contain 42.

Want: $|P_{60} \cup P_{04} \cup P_{42}|$.

Inclusion-exclusion, constrained permutation

$$\begin{aligned}&|P_{60} \cup P_{04} \cup P_{42}| \\ &= |P_{60}| + |P_{04}| + |P_{42}| \\ &\quad -|P_{60} \cap P_{04}| - |P_{42} \cap P_{04}| - |P_{60} \cap P_{42}| \\ &\quad +|P_{60} \cap P_{04} \cap P_{42}|\end{aligned}$$

$|P_{60}| = ?$

Clever trick: In P_{60}, can view “60” as a unit. So, each element of P_{60} is a permutation of $\{1, 2, 3, 4, 5, 7, 8, 9, 60\}$. Therefore, $|P_{60}| = 9!$. $|P_{04}| = 9!$. $|P_{42}| = 9!$.

Pairwise intersections

$$
\begin{aligned}
&|P_{60} \cup P_{04} \cup P_{42}| \\
&= |P_{60}| + |P_{04}| + |P_{42}| \\
&\quad -|P_{60} \cap P_{04}| - |P_{42} \cap P_{04}| - |P_{60} \cap P_{42}| \\
&\quad +|P_{60} \cap P_{04} \cap P_{42}| \\
&= 3 \times 9! \\
&\quad -|P_{60} \cap P_{04}| - |P_{42} \cap P_{04}| - |P_{60} \cap P_{42}| \\
&\quad +|P_{60} \cap P_{04} \cap P_{42}|
\end{aligned}
$$

$|P_{60} \cap P_{04}| =$? Trick works again! Can view “604” as a unit. So, each element is a permutation of $\{1, 2, 3, 5, 7, 8, 9, 604\}$. Therefore, 8!.

$|P_{42} \cap P_{04}| =$? Trick works again! Can view “042” as a unit. So, 8!.

$|P_{60} \cap P_{42}| =$? Trick fails! Wait, no, just changes. Now, each element is a permutation of $\{1, 3, 5, 7, 8, 9, 60, 42\}$. Still 8!.

Three-way intersection

$$
\begin{aligned}
&|P_{60} \cup P_{04} \cup P_{42}| \\
&= |P_{60}| + |P_{04}| + |P_{42}| \\
&\quad -|P_{60} \cap P_{04}| - |P_{42} \cap P_{04}| - |P_{60} \cap P_{42}| \\
&\quad +|P_{60} \cap P_{04} \cap P_{42}| \\
&= 3 \times 9! - 3 \times 8! \\
&\quad +|P_{60} \cap P_{04} \cap P_{42}|
\end{aligned}
$$

$|P_{60} \cap P_{04} \cap P_{42}| =$?. Yay, trick works again! Can view "6042" as a unit. So, each element is a permutation of $\{1, 3, 5, 7, 8, 9, 6042\}$. Therefore, 7!.

$|P_{60} \cup P_{04} \cup P_{42}| = 3 \times 9! - 3 \times 8! + 7! = 972720.$

n-way Inclusion-Exclusion

$|S_1 \cup S_2 \cup \cdots \cup S_n| =$

	the sum of the sizes of the individual sets
minus	the sizes of all two-way intersections
plus	the sizes of all three-way intersections
minus	the sizes of all four-way intersections
plus	the sizes of all five-way intersections, etc.

Hyper-mathy version:

$$\left|\bigcup_{i=1}^{n} S_i\right| = \sum_{X \in \mathcal{P}([1,n]) - \emptyset} (-1)^{|X|+1} \left|\bigcap_{i \in X} S_i\right|$$

Puzzle

A drawer in a dark room contains red socks, green socks, and blue socks. You are going to grab k socks, then go someplace light. How many socks must you take with you to be sure that you have at least one matching pair?

Two might be enough. But it's not guaranteed. After all, one might be one red, one green. Three is not enough. Might grab one of each. At four, one of the colors *must* repeat—you have a matching pair!

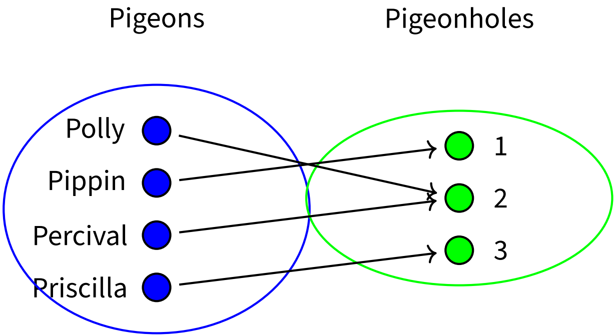
Pigeonhole principle

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

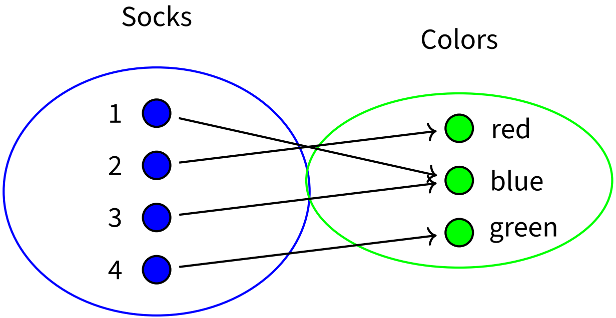
Why pigeons? A linguistic artifact. «Principe des tiroirs », “ladenprincipe”: think “drawers” or “cabinets.”

We will use it as a tool for reasoning about the relationships between the sizes of sets.

Pigeonhole principle as a relation



Sock example



Pigeonhole principle

Rule: If $|A| > |B|$, then, for every function $f : A \to B$, there exist two different elements of A that are mapped by f to the same element of B.

Recall: If a function from A to B is injective (no collisions), then $|A| \leq |B|$.

By the contrapositive of this statement, if $|A| > |B|$ (it's not the case that $|A| \leq |B|$), then a function from A to B has collisions (is not injective).

Australian sheep

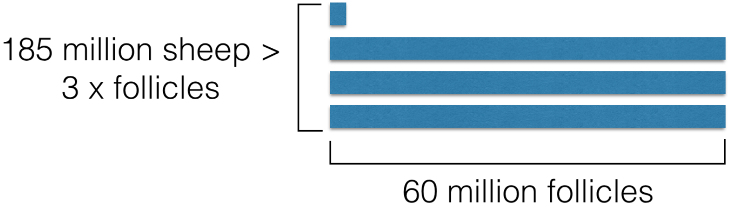
There’s a pair of sheep in Australia that have precisely the same number of hairs.

Sheep have up to 60 million wool follicles. Australia has 70 million sheep. Pigeonhole principle says there must be (at least!) two that match up.

All the sheep in China

China has 185 million sheep, each with up to 60 million follicles. Same argument as Australia applies. But, we can go further.

There must be a *foursome* of sheep with precisely the same number of hairs on their body. Argument: Even if we spread things out maximally, there’s some follicle count with 4 sheep in it.



Generalized Pigeonhole Principle

Rule: If $|A| > k|B|$, then every function $f : A \to B$ maps at least $k + 1$ different elements of A to the same element of B.

Lossless compression

An example from CS!

Let B be the set of bit sequences of length at most n. A *compression algorithm* is a function $f : B \to B$ that intends to map "interesting" sequences to shorter sequences. A *lossless* compression algorithm is one that is invertible.

By the pigeonhole principle: every lossless compression algorithm that compresses at least one input must expand some other input.

25 2-digit numbers

68	34	41	89	99
0	95	56	90	43
41	37	94	94	69
15	96	85	76	30
27	24	75	63	1

Two pairs with the same sum?

$$76 + 30 = 106 = 43 + 63$$
$$37 + 94 = 131 = 41 + 90$$

Lots of others.

Analysis

List of 25 2-digit numbers. Must two different pairs have the same sum?

A: How many possible pairs? $\binom{25}{2} = 25 \times 24/2 = 300$. In general, n numbers can be paired in $\binom{n}{2}$ ways.

B: How many possible pairwise sums? Smallest is $0 + 0 = 0$, largest is $99 + 99 = 198$. So, 199 possible values. In general, pairs of d-digit numbers can sum to $2 \times 10^d - 1$ possible values.

Since $300 > 199$, by the pigeonhole principle, some possible sum *must* be the result of two different pairs.

40 10-digit numbers

372995585	2234938293	7149708291	7060913492
2931952606	3111391181	951202341	735405394
9217967649	824907338	5014746657	1237286631
522671716	8407467172	8707242896	683510550
4879527699	5950653401	5161166931	816710053
1132327190	5154516759	2790970967	2560087807
8425382298	4088307493	3641040020	2603650330
3153589691	8144289385	4677056137	4058664392
9202688143	144564604	6259750302	2210904697
6037630341	498877221	186119657	71015872

Two *subsets* with the same sum?

Analysis

List of 40 10-digit numbers. Must two different subsets have the same sum?

A: How many possible subsets? $2^{40} = 1.1e + 12$. In general, n numbers can produce 2^n subets.

B: How many possible subset sums? Smallest is $0 + 0 = 0$, largest is $40 \times 9999999999 = 4.0e + 11$. In general, up to n d-digit numbers can sum to $n \times 10^d$ possible values.

Since $1.1e + 12 > 4.0e + 11$, by the pigeonhole principle, some possible sum *must* be the result of two different subsets.