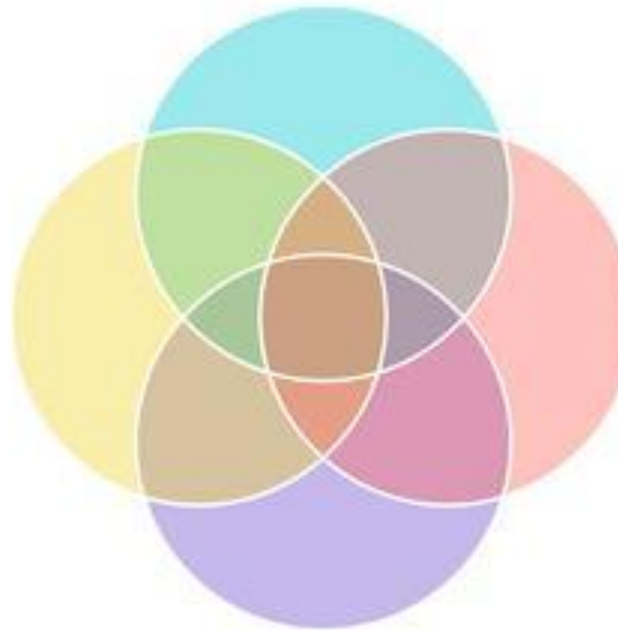
Inclusion–Exclusion …



Robin Whitty
LSBU Maths Study Group
September 11, 2025

… or should that be
Inclusion-
Exclusion

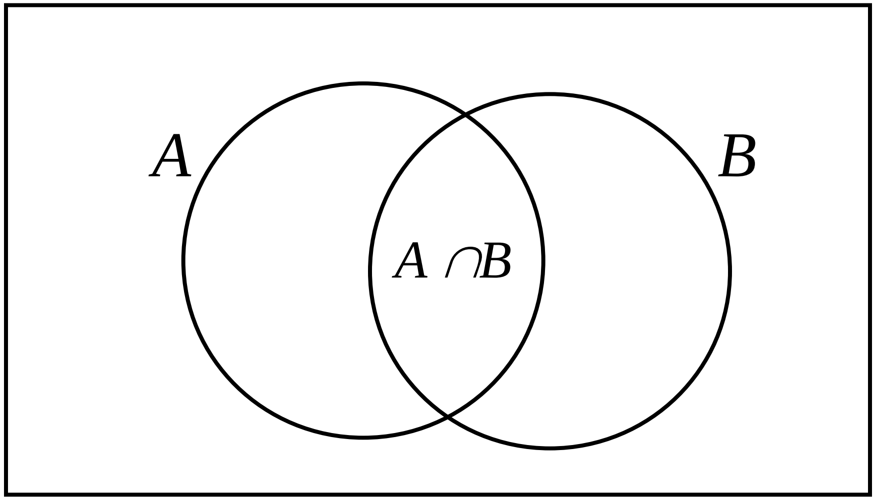
Contents

1. Inclusion–Exclusion – the idea
2. A serious example: counting prime numbers
3. A textbook example: counting surjective functions
4. What ought to be a textbook example (but I can’t see how): putting balls into bins so that some bin gets exactly one ball
5. Why I’m interested in putting balls into bins

Paul Erdős and THE BOOK

Rewley House, September 18 – 19, 2010

Inclusion Exclusion: the 2 set case



Example: roughly how many numbers in the range 1,…, 100 are divisible by either 2 or 3 (or both)?

Solution: there are

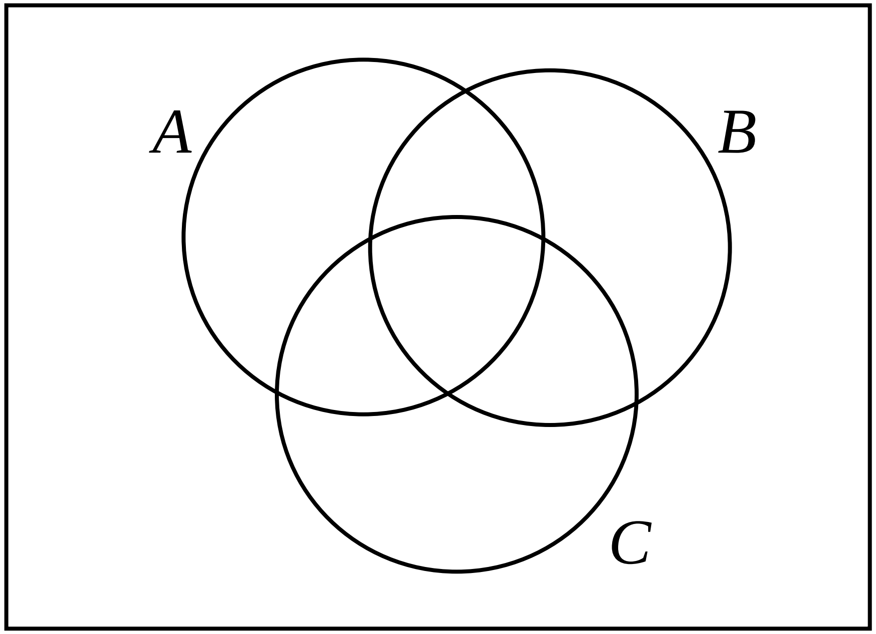
about 50 even numbers in the range;

about 33 multiples of 3, of which half are even;

So divisible by 2 or 3 ≈ 50 + 33 – 16 = 67.

Inclusion Exclusion: the 3 set case

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$



Example: An opinion poll reports that the percentage of voters who would be satisfied with each of three candidates A, B, C for President is 65%, 57%, 58% respectively. Further, 28% would accept A or B, 30% A or C, 27% B or C, and 12% would be content with any of the three. What do you conclude?

Solution: the percentage of voters who reject all candidates is

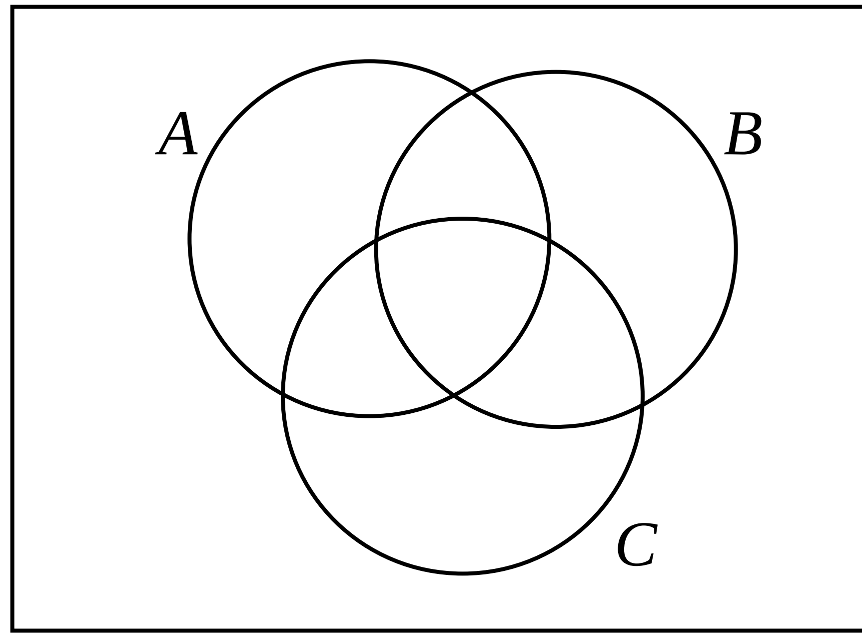
$$100 - 65 - 57 - 58 + 28 + 30 + 27 - 12 = -7;$$

so there must be a mistake.

(from webspace.maths.qmul.ac.uk/p.j.cameron/comb/ch5s.pdf)

Three sets and counting primes

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$



Question: roughly how many prime numbers are there less than 50?

$|A|$: multiples of 2 ≈ 25
$|B|$: multiples of 3 ≈ 16
$|C|$: multiples of 5 ≈ 10
$|A \cap B|$: multiples of 6 ≈ 8
$|A \cap C|$: multiples of 10 ≈ 5
$|B \cap C|$: multiples of 15 ≈ 3
$|A \cap B \cap C|$: multiples of 30 ≈ 1

Solution: the number of composite numbers up to 50 is about

$$25 + 16 + 10 - 8 - 5 - 3 + 1 = 36;$$ (includes 2,3,5 so **subtract 3**, omits 1 so **add 1**)

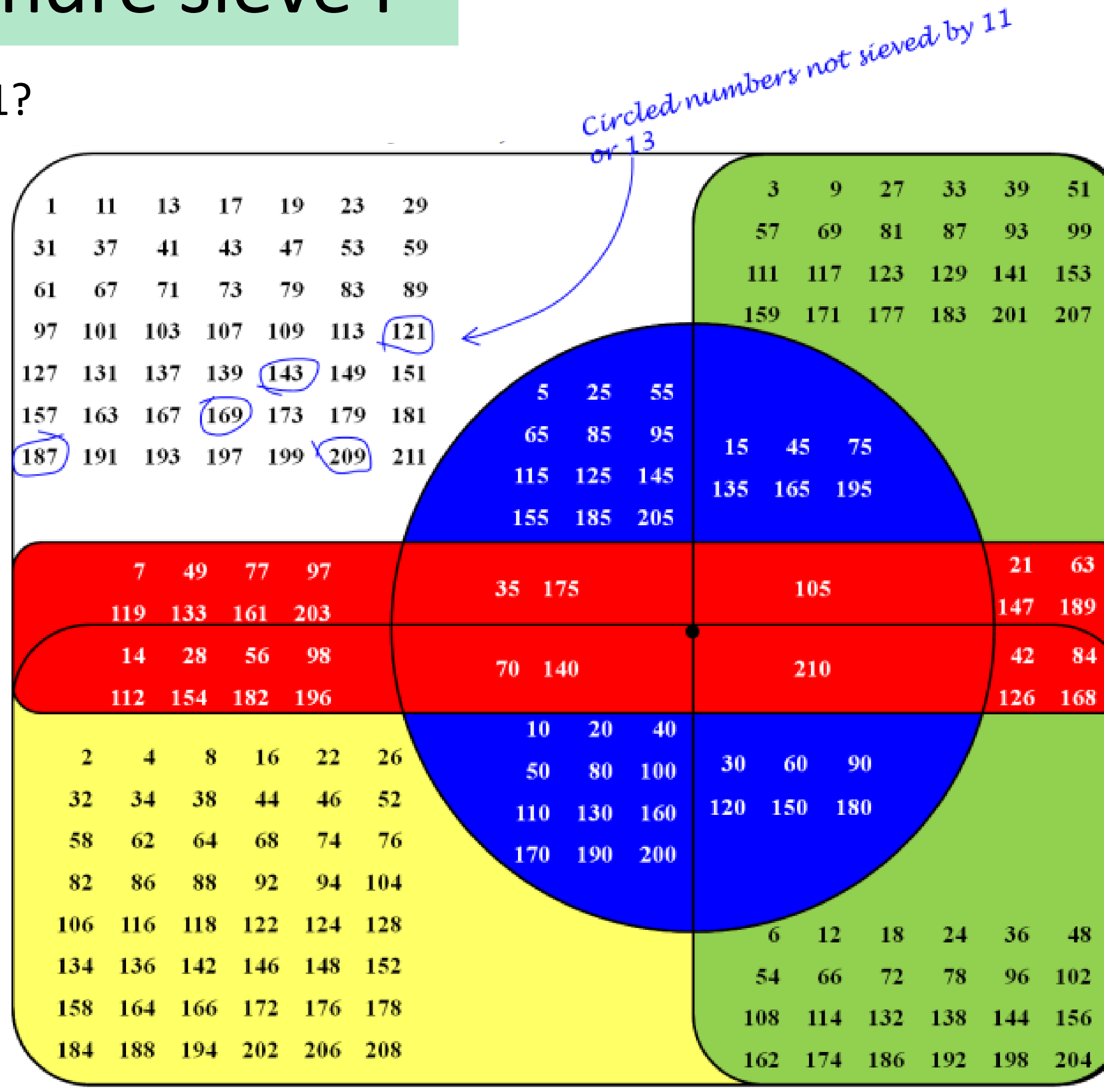
so there are about $50 - (36 - 3 + 1) = 16$ primes.

The Eratothenes–Legendre sieve I

What is $\pi(211)$, number of primes ≤ 211?

How might we count the primes up to $211 = 1 + 2 \times 3 \times 5 \times 7$? A first approximation is to count all the integers from 1 to 211 which are excluded from the shaded regions of the 4-set Venn diagram on the right: bottom = multiples of 2; right = multiples of 3; circle=multiples of 5; central = multiples of 7. Inclusion-exclusion 'sieves out' products of just the first four primes (# denotes 'number of'):

			211
−	# multiples of 2, 3, 5, 7	$\lfloor \frac{211}{2} \rfloor + \ldots + \lfloor \frac{211}{7} \rfloor$	247
+	# multiples of $2 \times 3, 2 \times 5, \ldots, 5 \times 7$	$\lfloor \frac{211}{6} \rfloor + \ldots + \lfloor \frac{211}{35} \rfloor$	101
−	# multiples of $2 \times 3 \times 5, \ldots, 3 \times 5 \times 7$	$\lfloor \frac{211}{30} \rfloor + \ldots + \lfloor \frac{211}{105} \rfloor$	17
+	# multiples of $2 \times 3 \times 5 \times 7$	$\lfloor \frac{211}{210} \rfloor$	1
	Total (as shown in top-left of Venn diagram):		49



Inclusion–exclusion counts 162 numbers in the coloured regions.

Of these, 4 are the primes we are taking multiples of ; and non-prime-non-composite 1 has been missed out.

So our best estimate for $\pi(211)$ is:

$$211 - 162 + 4 - 1 = 52.$$

But we are out by 5 because we did not include – exclude two of the primes $\leq\sqrt{211} \approx 14.5$.

The Eratothenes–Legendre sieve II

The Eratosthenes–Legendre Sieve *Let $\pi(x)$ denote the number of primes not exceeding x, and $P(x)$ denote the product of all primes not exceeding x. Then*

$$\pi(x) = \sum_{d \mid P(\sqrt{x})} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor + \pi(\sqrt{x}) - 1$$

where $\mu(n)$ is the Möbius function defined for positive integers n by

$$\mu(n) = \begin{cases} (-1)^r & \text{if } n \text{ is a product of } r \text{ distinct primes (with } r = 0 \text{ if } n = 1) \\ 0 & \text{if } n \text{ has a square factor.} \end{cases}$$

$+1$ for multiples of even numbers of primes, and for $d = 1$

-1 for multiples of odd numbers of primes

$$\pi(x) = \sum_{d \mid P(\sqrt{x})} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor + \pi(\sqrt{x}) - 1$$

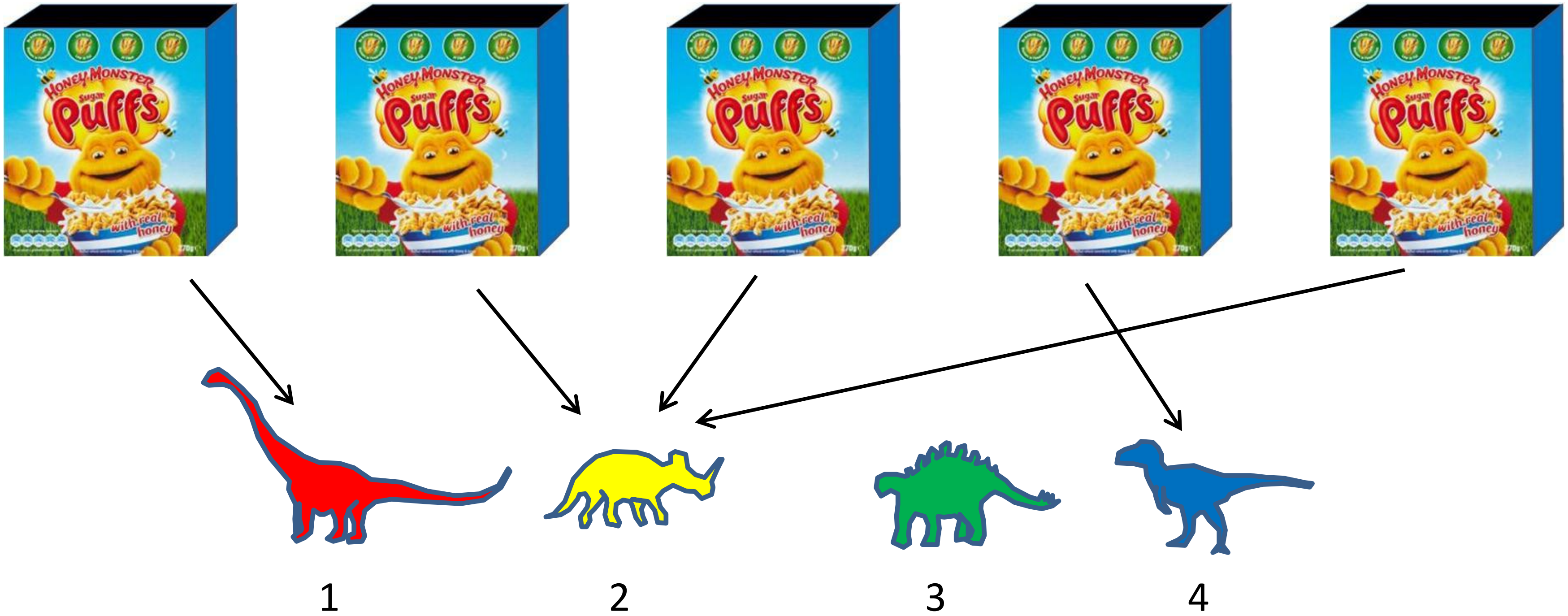
What we want

How many multiples

Put back primes we needed to sieve by

Non-prime-non-composite

The Coupon Collector’s Problem I

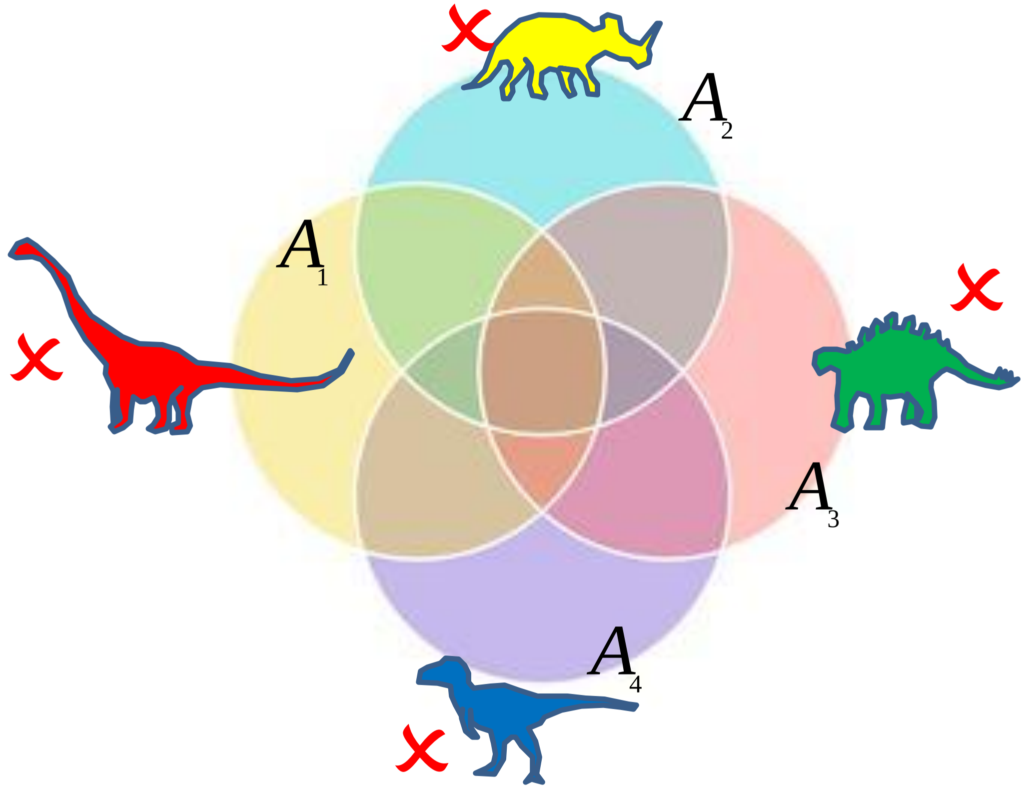


What is the probability of getting all four dinosaurs by opening five packets of cereal?

How many functions map a domain of size 5 to a range of size 4?

How many of these are surjective (onto)?

The Coupon Collector's Problem II



Number of non-surjections $= |A_1 \cup A_2 \cup A_3 \cup A_4|$

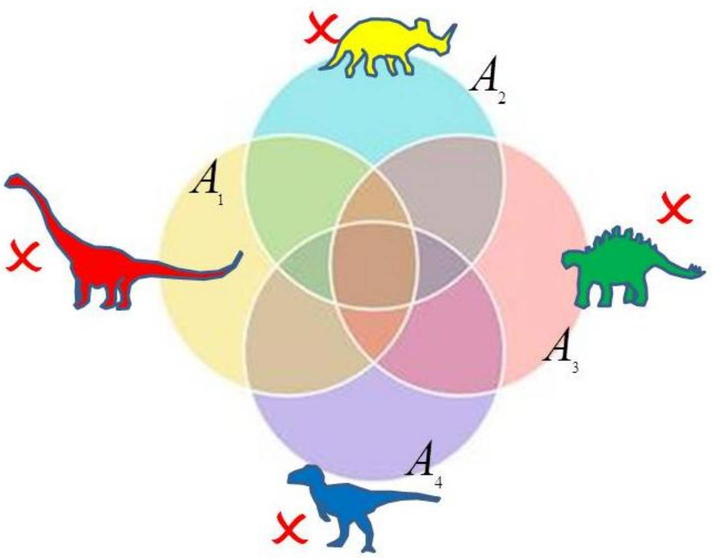
$$= \sum |A_i| - \sum |A_i \cap A_j| + \sum |A_i \cap A_j \cap A_k| - \sum |A_i \cap A_j \cap A_k \cap A_l|$$

The Coupon Collector's Problem III

A_1 is same as functions to {2,3,4}

A_2 is same as functions to {1,3,4}

etc



Number of functions from domain of size *m* to a range of size *n*:

$$n^m$$

So

$$|A_i| = 3^5$$

Similarly

$A_1 \cap A_2$ is same as functions to {3,4}

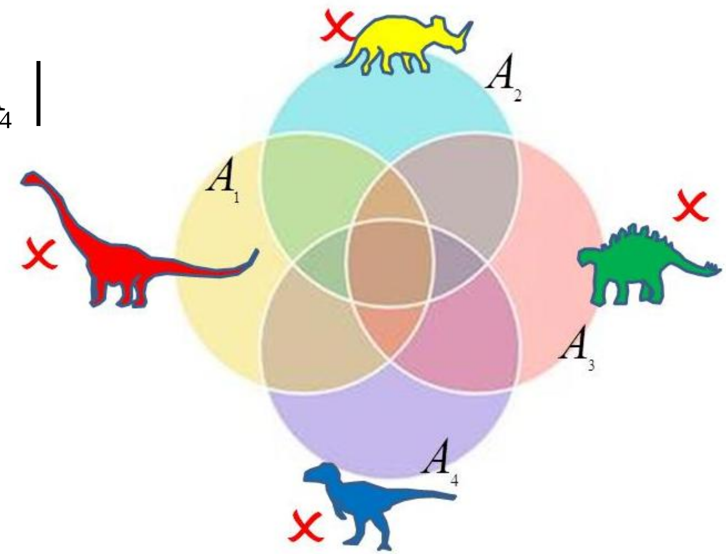
etc

So

$$|A_i \cap A_j| = 2^5$$

The Coupon Collector's Problem IV

Number of non-surjections $=|A_1 \cup A_2 \cup A_3 \cup A_4|$



$$= \sum |A_i| - \sum |A_i \cap A_j| + \sum |A_i \cap A_j \cap A_k| - \sum |A_i \cap A_j \cap A_k \cap A_l|$$

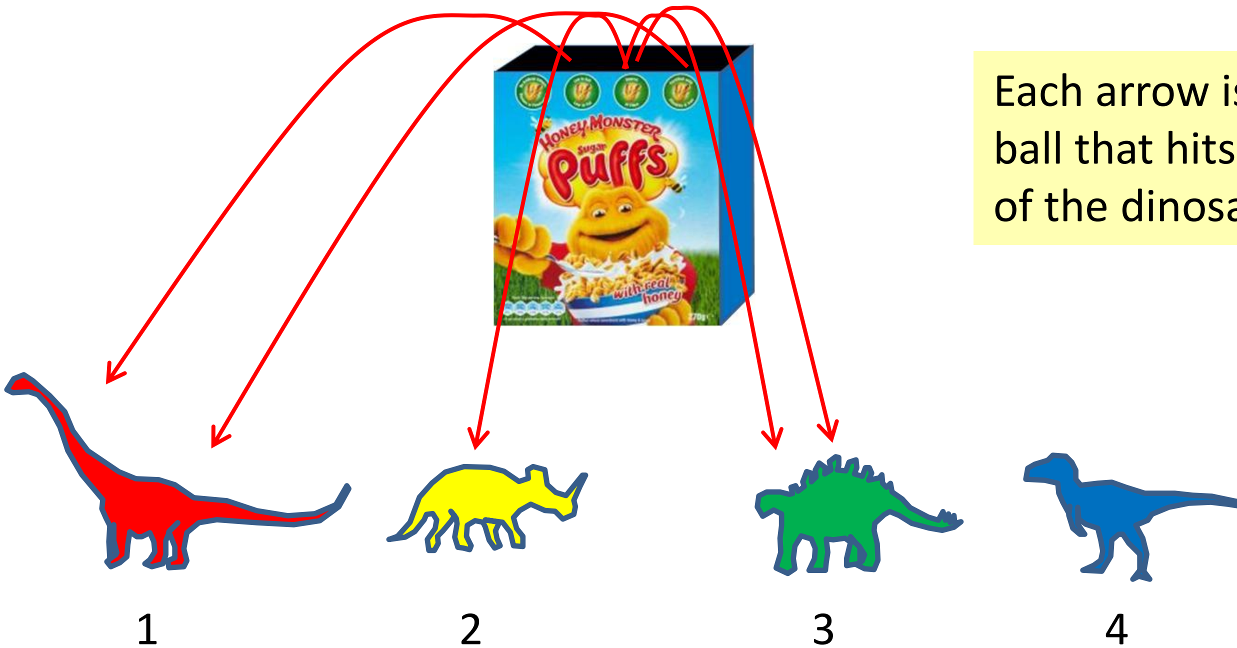
$$\binom{4}{1} \times 3^5 \qquad \binom{4}{2} \times 2^5 \qquad \binom{4}{3} \times 1^5 \qquad \binom{4}{4} \times 0^5$$

This gives no. of non-surjections = 784. So probability of a surjection with 5 packets of cereal is

$$\frac{4^5 - 784}{4^5} \approx 0.23$$

Balls in bins I

Putting n identical balls into m bins: coupon collector's but with a magic sugar puffs packet that generates a stream of dinosaurs



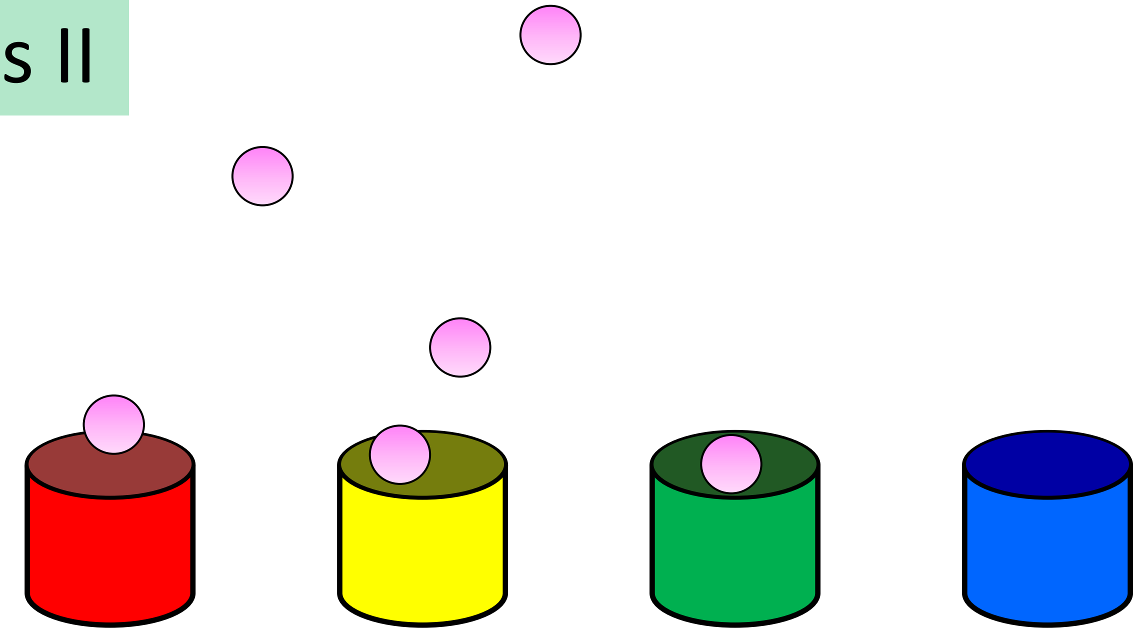
Each arrow is a ball that hits one of the dinosaurs

What is the probability of getting every one of m dinosaurs in a stream of n from the packet?

How many ways of putting n identical balls into m bins?

How many of these place at least one ball into each bin?

Balls in bins II



So now we have *m* bins into which we place *n* balls with repetition allowed and we want at least one bin to get exactly 1 ball.

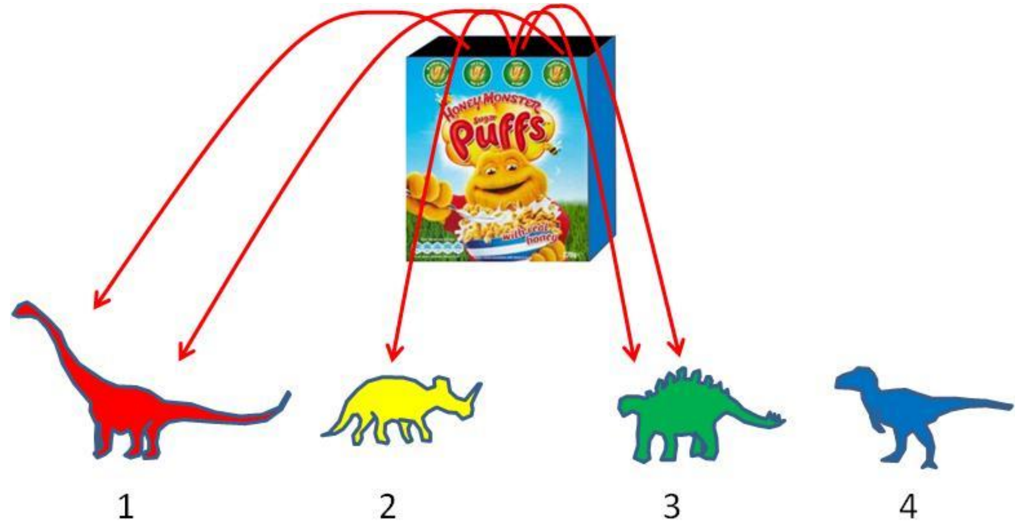
Basic count: number of ways to place *n* balls into *m* bins: $\binom{m-1+n}{m-1}$

E.g. 6 balls into 4 bins: take $4-1+6$ places. Choose $4-1$ bin markers and the remaining 6 places are occupied by the balls.

__ | | __ __ __ | __ __ = 1 ball, 0 balls, 3 balls, 2 balls

Balls in bins III

What is the probability of getting every one of n dinosaurs in a stream of m from the packet?



How many ways of putting m identical balls into n bins? $\binom{m-1+n}{m-1}$

How many of these place at least one ball into each bin?

Take m of the n balls and place one in each bin.

Now place the remaining $n-m$ balls in all possible ways: $\binom{m-1+n-m}{m-1} = \binom{n-1}{m-1}$

So probability of a surjection with 5 balls from the magic packet is

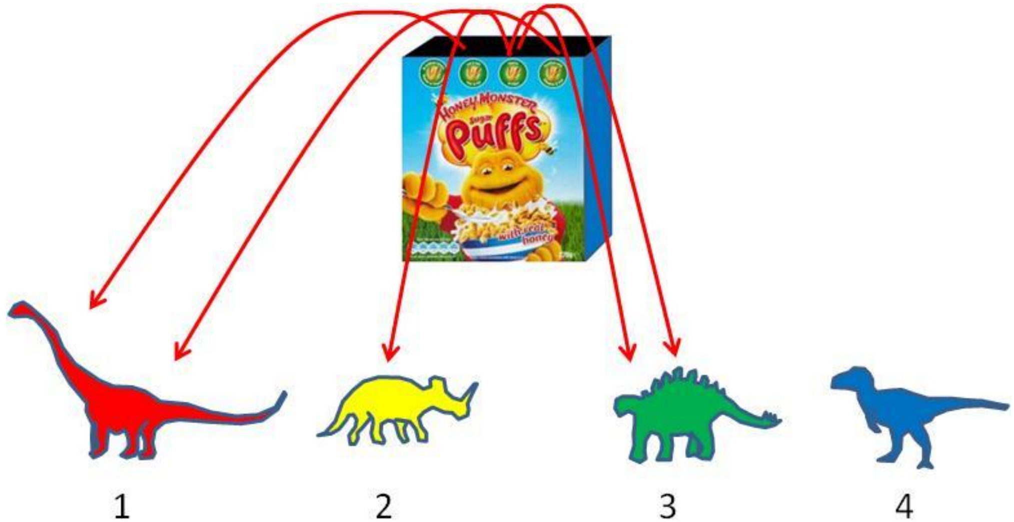
$$\frac{\binom{4}{3}}{\binom{8}{3}} = \frac{4}{56} \approx 0.07$$

No inclusion–exclusion needed…

But you're worse off with the magic cereal packet!

Balls in bins IV

What is the probability of getting at least one of the *n* dinosaurs **exactly once** in a stream of *m* from the packet?



How many ways of putting *m* identical balls into *n* bins? $\binom{m-1+n}{m-1}$

How many of these place exactly one ball into at least one of the bins?

Let

$$S_i = \text{number of placements of exactly 1 ball out of } n \text{ into exactly } i \text{ bins out of } m$$

Then I want

$$S(m,n) = \sum_{i=1}^{m} S_i$$

Balls in bins V

Want

$$S(m,n)=\sum_{i=1}^{m} S_i$$

where

S_i = number of placements of exactly 1 ball out of n into exactly i bins out of m

E.g. $m=6, n=4$

$$S_1: \; 3\,1\,0\,0\,0\,0 \rightarrow \frac{6!}{4!}=30$$

$$S_2: \; 2\,1\,1\,0\,0\,0 \rightarrow \frac{6!}{3!2!}=60$$

$$S_3: \; \text{none possible} \rightarrow =0$$

$$S_4: \; 1\,1\,1\,1\,0\,0 \rightarrow \frac{6!}{4!2!}=15$$

Total

$$S(6,4)=105$$

Balls in bins VI

E.g. $m = 5, n = 7$

S_1 : $6\,1\,0\,0\,0 \to 20$

$4\,2\,1\,0\,0 \to 60$

$3\,3\,1\,0\,0 \to 30$

$2\,2\,2\,1\,0 \to 20$

S_2 : $5\,1\,1\,0\,0 \to 30$

$3\,2\,1\,1\,0 \to 60$

S_3 : $4\,1\,1\,1\,0 \to 20$

$2\,2\,1\,1\,1 \to 10$

S_4 : $3\,1\,1\,1\,1 \to 5$

Total

$$S(5,7) = 255$$

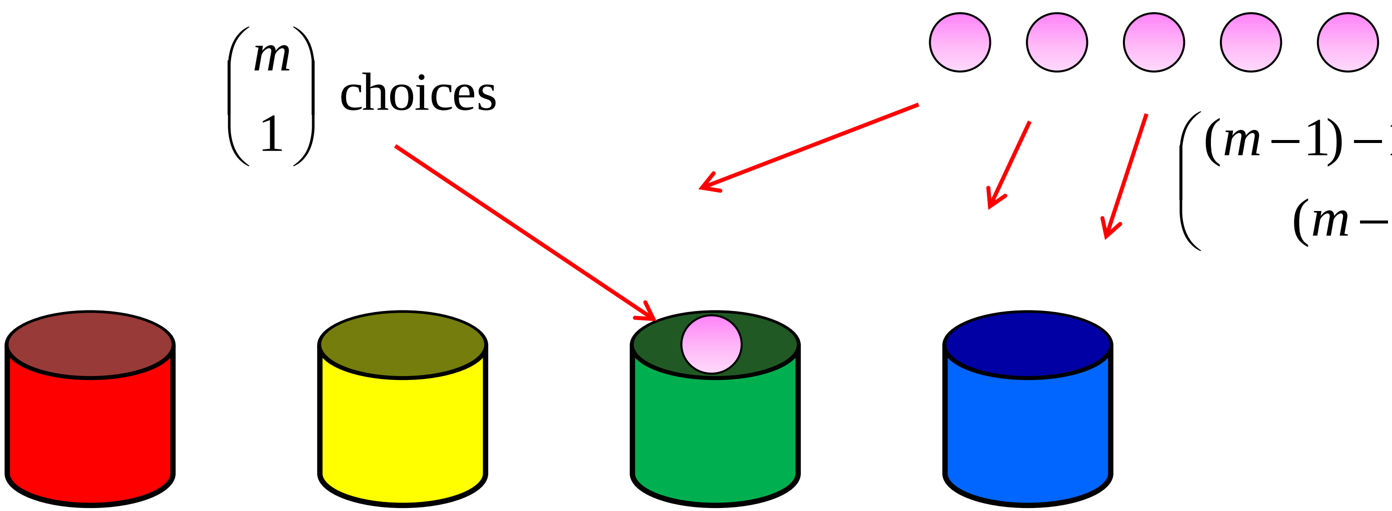
Calculating $S(m,n) = \sum_{i=1}^{m} S_i$ is presumably an exercise in counting partitions, e.g. using Bell numbers. But we want something easier.

Balls in bins VII

We can try and do something similar to the approach to counting 'surjective ' placements of balls in bins.
We'll choose a bin and put one ball in it. Then we'll put the remaining $n-1$ balls into the remaining $m-1$ bins

$\binom{m}{1}$ choices

$\binom{(m-1)-1+(n-1)}{(m-1)-1}$ choices



Denote by T_i the number of ways we can do this with i bins. Then

$$T_i = \binom{m}{i} \times \binom{(m-i)-1+(n-i)}{(m-i)-1}$$

BUT this overcounts because many remaining choices will also create single occupancy bins

Balls in bins VIII

E.g. $m=5, n=7$ continued

$S_1: \; 6\,1\,0\,0\,0 \rightarrow 20$

$4\,2\,1\,0\,0 \rightarrow 60$

$3\,3\,1\,0\,0 \rightarrow 30$

$2\,2\,2\,1\,0 \rightarrow 20$

$S_2: \; 5\,1\,1\,0\,0 \rightarrow 30$

$3\,2\,1\,1\,0 \rightarrow 60$

$S_3: \; 4\,1\,1\,1\,0 \rightarrow 20$

$2\,2\,1\,1\,1 \rightarrow 10$

$S_4: \; 3\,1\,1\,1\,1 \rightarrow 5$

Total

$S(5,7) = 255$

$$T_1 = \binom{5}{1} \times \binom{(5-1)-1+(7-1)}{(5-1)-1} = 420$$

$$T_2 = \binom{5}{2} \times \binom{(5-2)-1+(7-2)}{(5-2)-1} = 210$$

$$T_3 = \binom{5}{3} \times \binom{(5-3)-1+(7-3)}{(5-3)-1} = 50$$

$$T_4 = \binom{5}{4} \times \binom{(5-4)-1+(7-4)}{(5-4)-1} = 5$$

$$420 - 210 + 50 - 5 = 255$$

Balls in bins IX

Indeed, we have $$\sum_{i=1}^{m-1}(-1)^{i+1}T_i = \sum_{i=1}^{m-1}S_i = S(m,n) \qquad (*)$$

This is because we have $$T_i = \sum_{k=i}^{m-1}\binom{k}{i}S_k$$

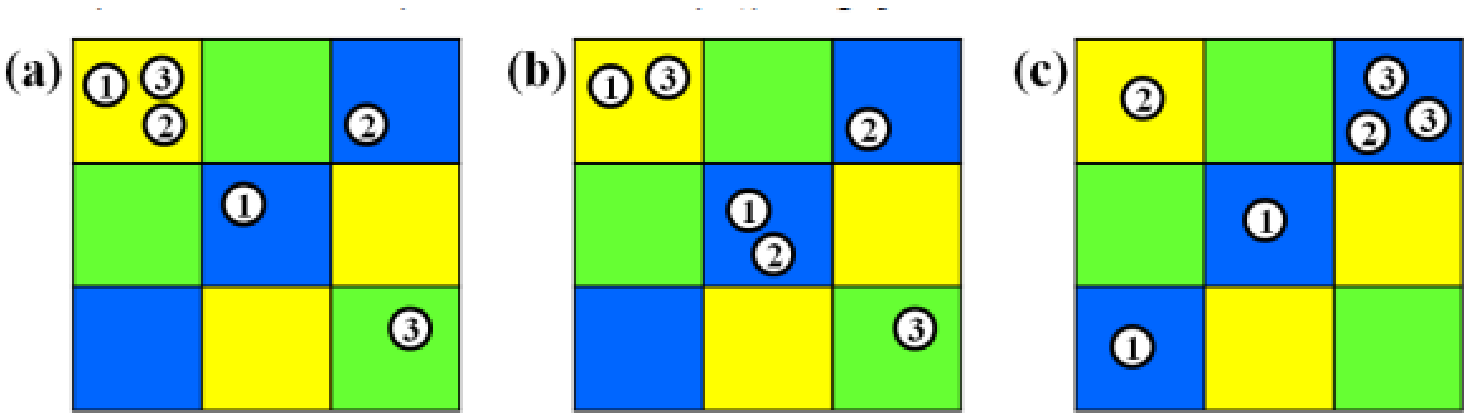
and the alternating sum in (*) takes an alternating sum along rows of Pascal's triangle

$i=1$	$\binom{1}{1}S_1$	$\binom{2}{1}S_2$	$\binom{3}{1}S_3$	$\binom{4}{1}S_4$	$+T_1$
$i=2$		$\binom{2}{2}S_2$	$\binom{3}{2}S_3$	$\binom{4}{2}S_4$	$-T_2$
$i=3$			$\binom{3}{3}S_3$	$\binom{4}{3}S_4$	$+T_3$
$i=4$				$\binom{4}{4}S_4$	$-T_4$
Totals :	$\binom{1}{0}S_1$	$\binom{2}{0}S_2$	$\binom{3}{0}S_3$	$\binom{4}{0}S_4$	$\sum S_i$

This looks like inclusion – exclusion! If so then that would be the preferred way to count single ball occupancy placements.

An application I

The Lovász Local Lemma *Let $A_1, \ldots A_t$ be events in a probability space, with every event being independent of all except at most d others. Suppose, for some non-negative real number p satisfying $p \leq 1/e(d+1)$ $(e = 2.718\ldots)$, we have $\mathbb{P}(A_i) < p$ for all i, $1 \leq i \leq t$. Then $\mathbb{P}(\cap_{i=1}^{t} \overline{A}_i) > 0$.*



n players commit privately to a placement of 2 tokens on the cells of an $\sqrt{m} \times \sqrt{m}$ grid.

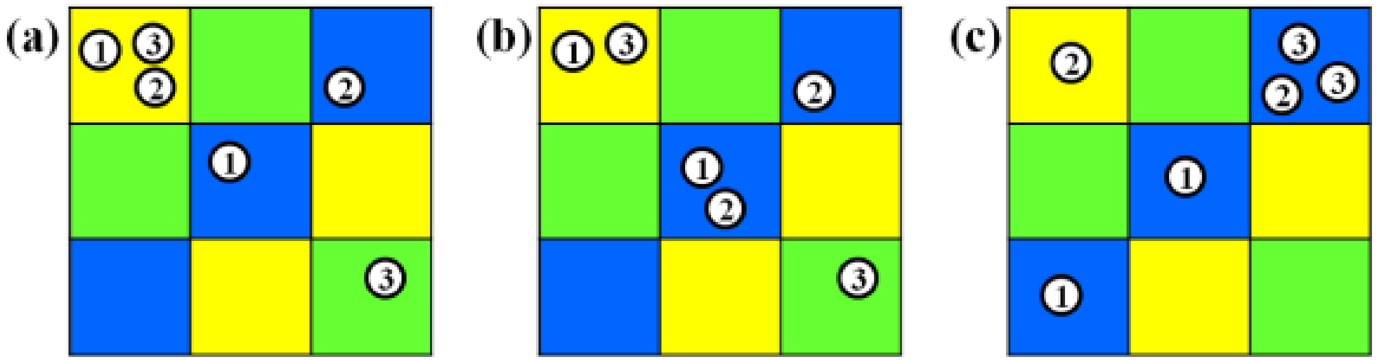
Here, $n = 3$, $m = 9$.

They reveal their choices and their stake money is shared equally according to the tokens which single-occupy cells.

E.g. in game (a), players 1,2 and 3 share the stake. In game (c), players 1 and 2 share the stake in the ration 2:1.

Question: if there are no singly-occupied cells then 'house' takes the stake. What values of n and m make this profitable for the house?

An application II



$$\Pr(\text{no single occupancies}) = \binom{m-1+2n}{m-1}^{-1} \sum_{i=0}^{m-1} (-1)^i \binom{m}{i} \binom{m-2i-1+2n}{m-i-1}$$

If $2n$ approaches m (it's $2n$ because each player has 2 tokens) then there is less chance of single occupany so House wins.

But the larger the value of n, the bigger the stake, so more incentive to play.