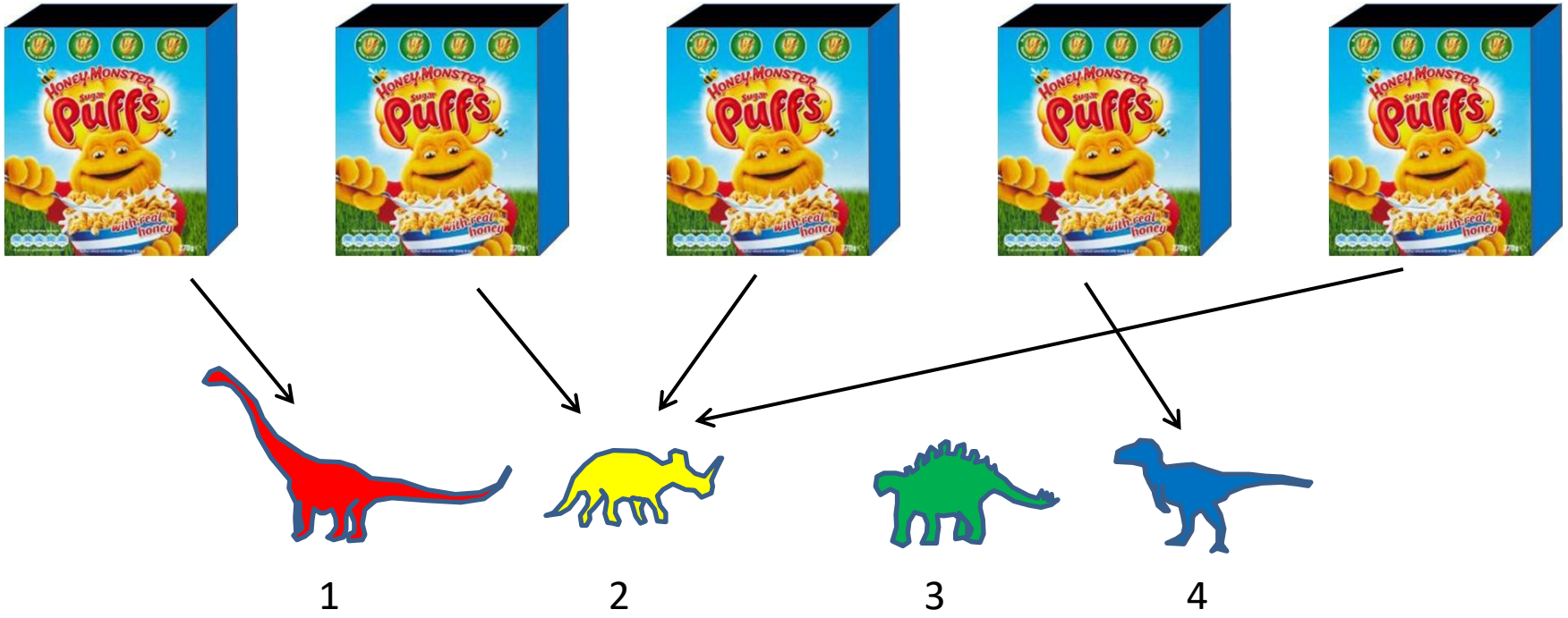
Bins, balls and pigeons

Robin Whitty
LSBU Maths Study Group
November 27, 2025

Photo: Stacalusa

The Coupon Collector's Problem



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

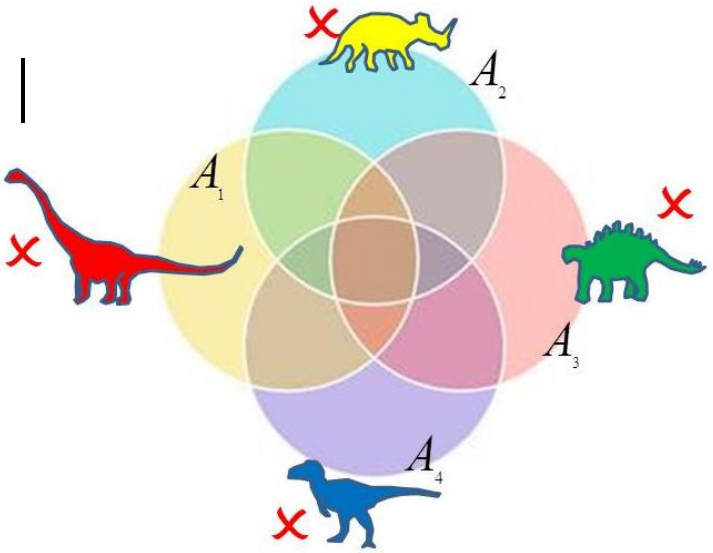
How many functions map a domain of size 5 to a range of size 4?

How many of these are surjective (onto)?

Coupon Collecting: Inclusion – Exclusion

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

A_i counts mappings NOT going to i



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

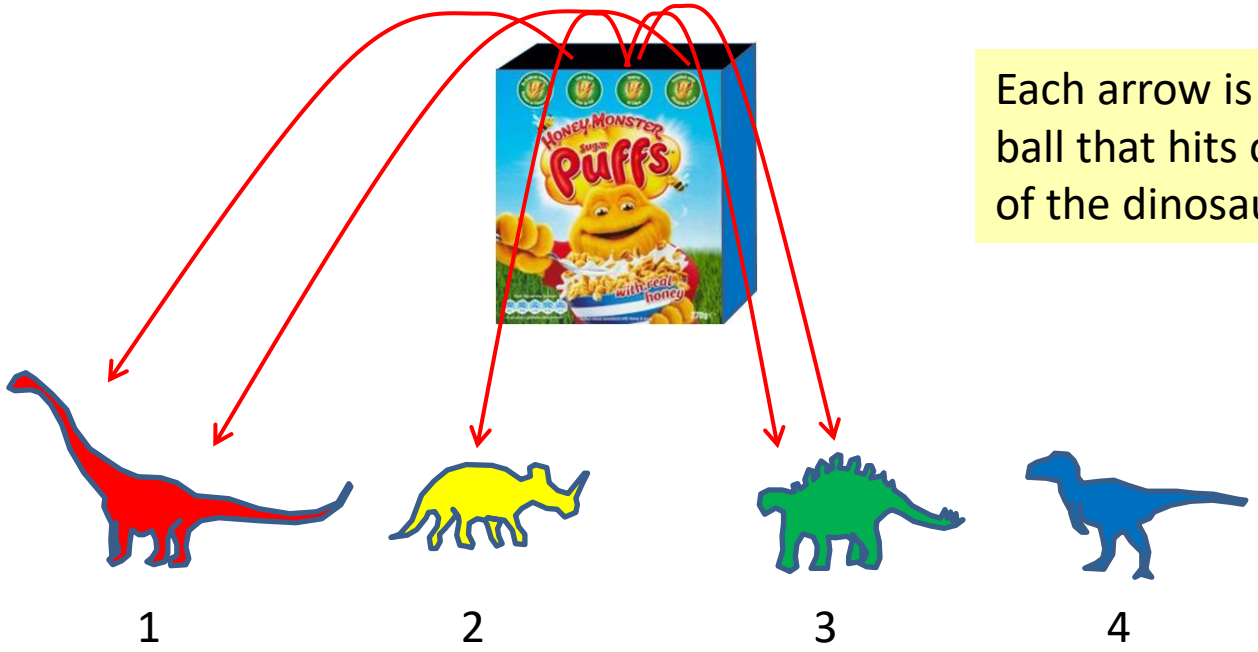
$$\binom{4}{1} \times 3^5 \qquad \binom{4}{2} \times 2^5 \qquad \binom{4}{3} \times 1^5 \qquad \binom{4}{4} \times 0^5$$

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

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

Recap: Balls in bins I

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



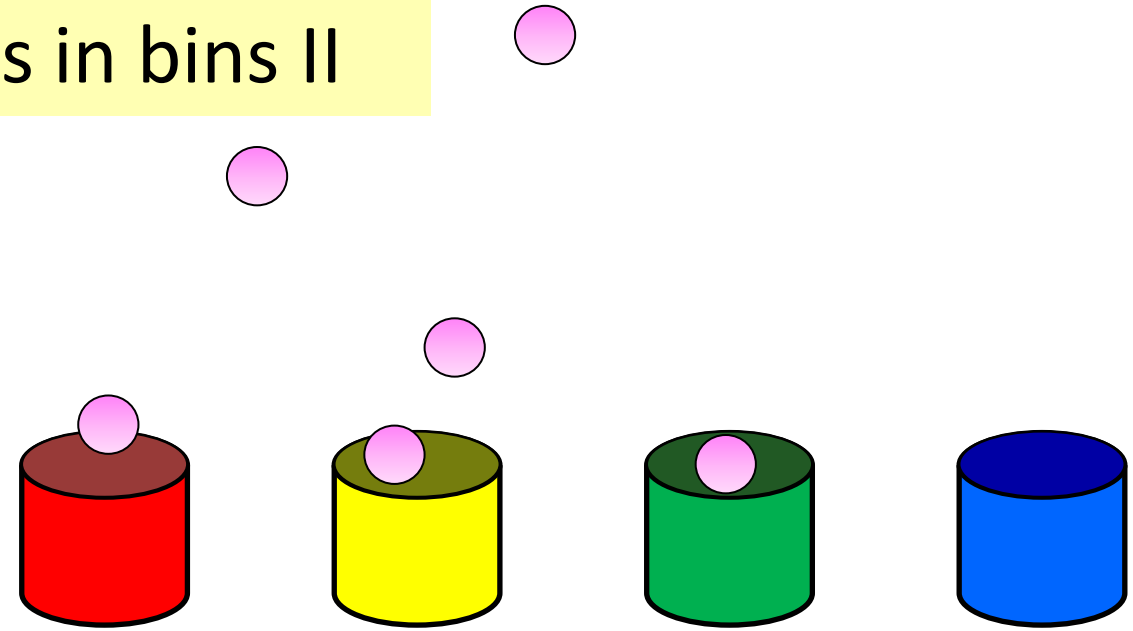
Each arrow is a ball that hits one of the dinosaurs

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?

Recap: 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 ball in each bin.

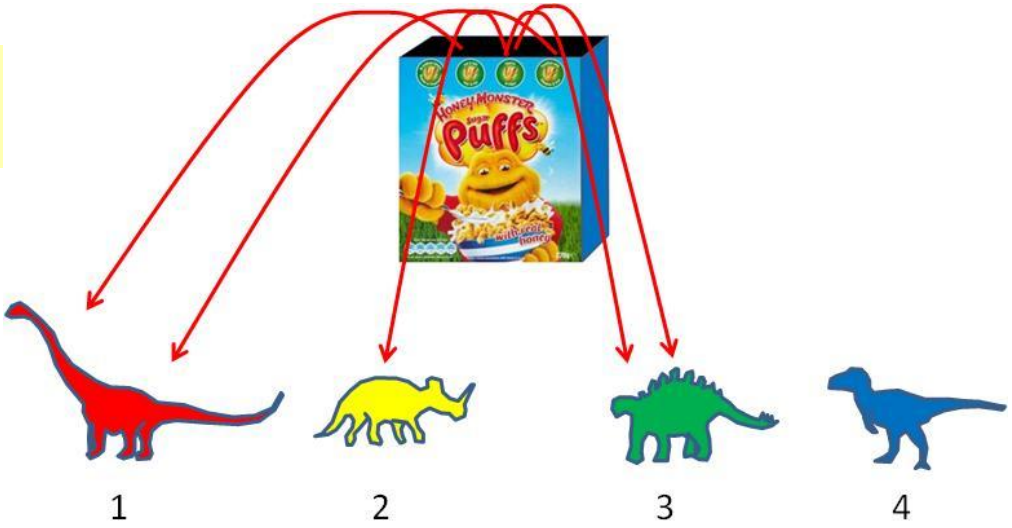
Basic count: number of ways to place n balls into m bins: $\binom{m-1+n}{m-1}$

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

Recap: 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 n identical balls into m 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.

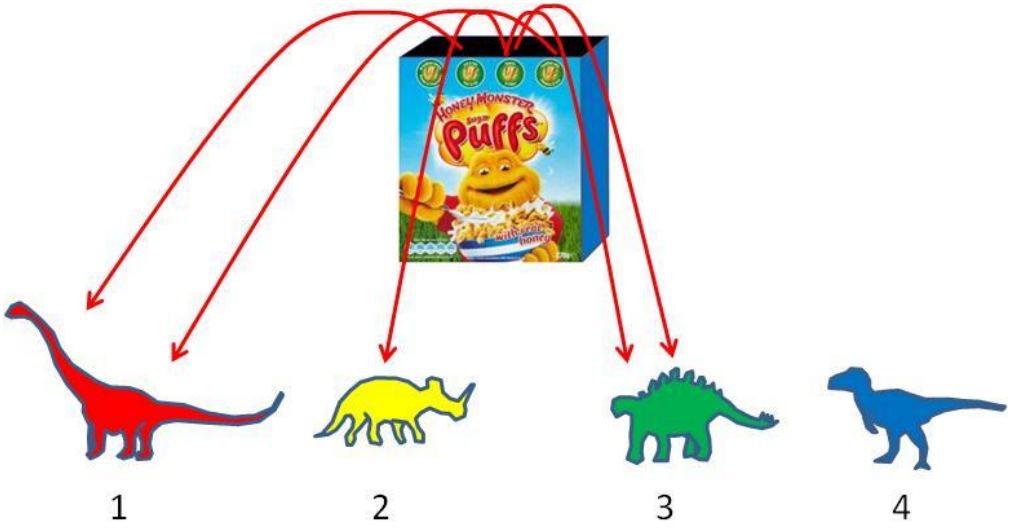
Now place the remaining $n-m$ balls in all possible ways: $\binom{m-1+n-m}{m-1}=\binom{n-1}{m-1}$

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

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

Single occupancy I

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

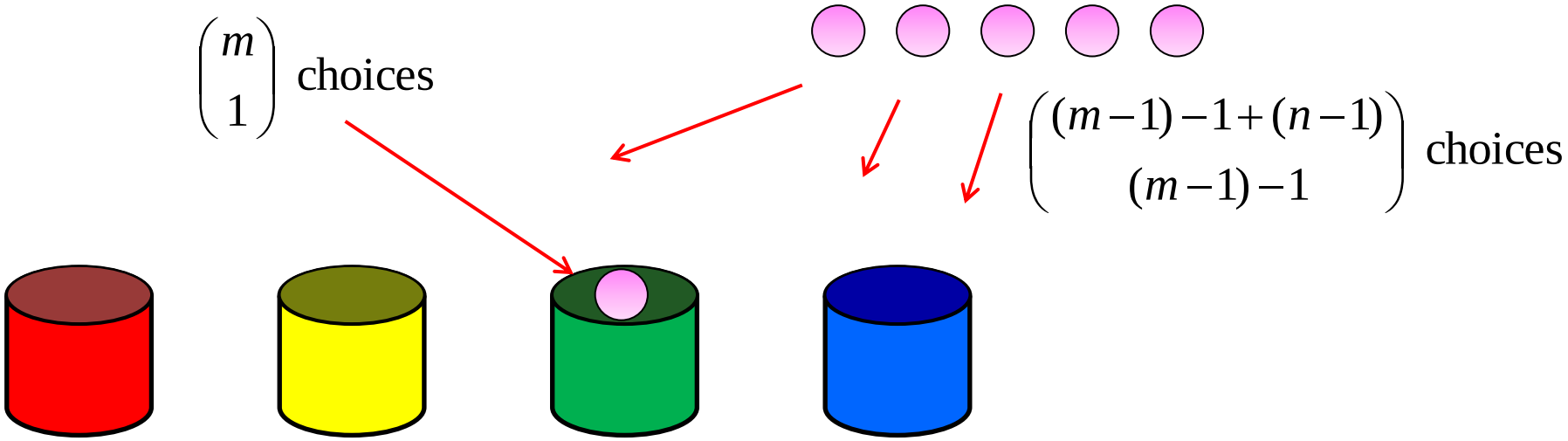
$$S_i = \text{number of placements of exactly 1 ball out of } n \text{ into exactly } i \text{ bins out of } m$$

Then I want

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

Single occupancy II

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



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

Single occupancy III

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. $m = 6, n = 4$

$$S_1: \; 310000 \rightarrow \frac{6!}{4!} = 30$$

$$S_2: \; 211000 \rightarrow \frac{6!}{3!2!} = 60$$

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

$$S_4: \; 111100 \rightarrow \frac{6!}{4!2!} = 15$$

Total

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

Single occupancy IV

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

S_1 : $61000 \to 20$

$42100 \to 60$

$33100 \to 30$

$22210 \to 20$

S_2 : $51100 \to 30$

$32110 \to 60$

S_3 : $41110 \to 20$

$22111 \to 10$

S_4 : $31111 \to 5$

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

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

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

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

Total

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

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

Single occupancy V

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

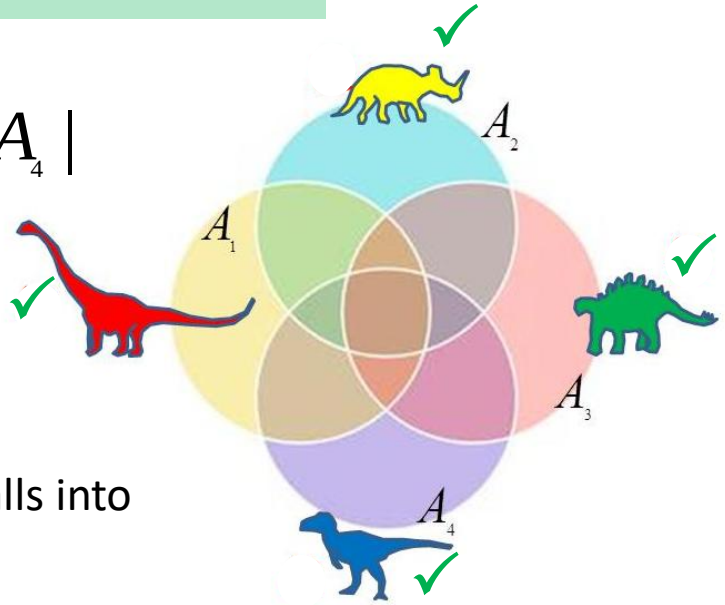
$i=1$	$\binom{1}{1}S_1$	$\binom{2}{1}S_2$	$\binom{3}{1}S_3$	$\binom{4}{1}S_4$ $+T_1$
$i=2$		$\binom{2}{2}S_2$	$\binom{3}{2}S_3$	$\binom{4}{2}S_4$ $-T_2$
$i=3$			$\binom{3}{3}S_3$	$\binom{4}{3}S_4$ $+T_3$
$i=4$				$\binom{4}{4}S_4$ $-T_4$
Totals :	$\binom{1}{0}S_1$	$\binom{2}{0}S_2$	$\binom{3}{0}S_3$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

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

Single occupancy: Inclusion – Exclusion I

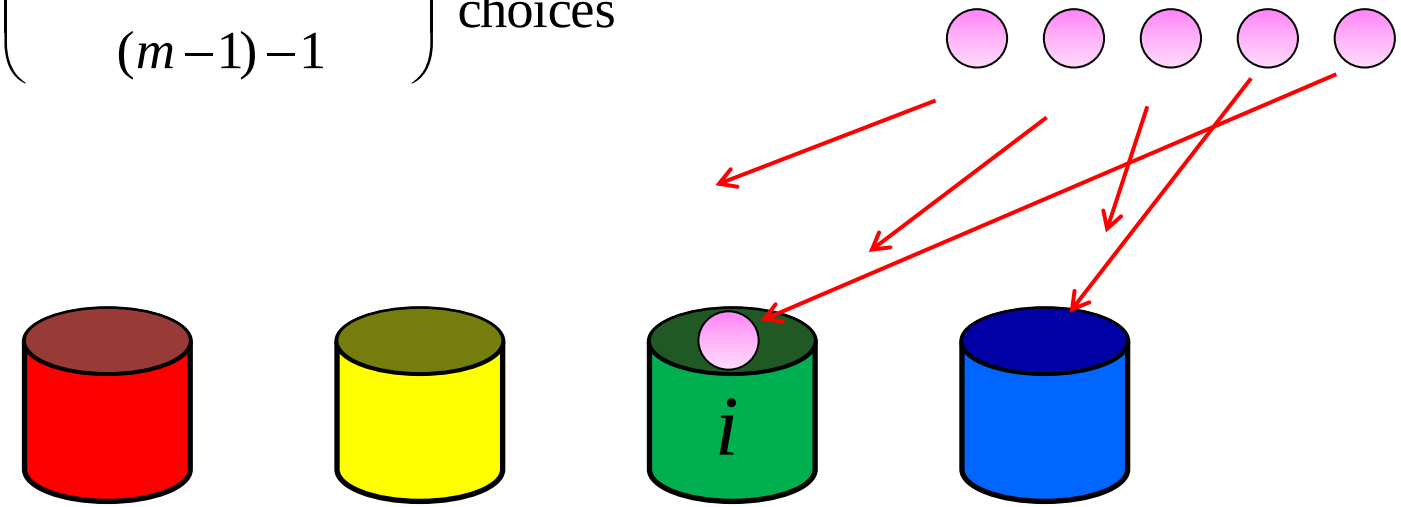
Number of single occupancies $= |A_1 \cup A_2 \cup A_3 \cup A_4|$

A_i counts placements with singleton at i



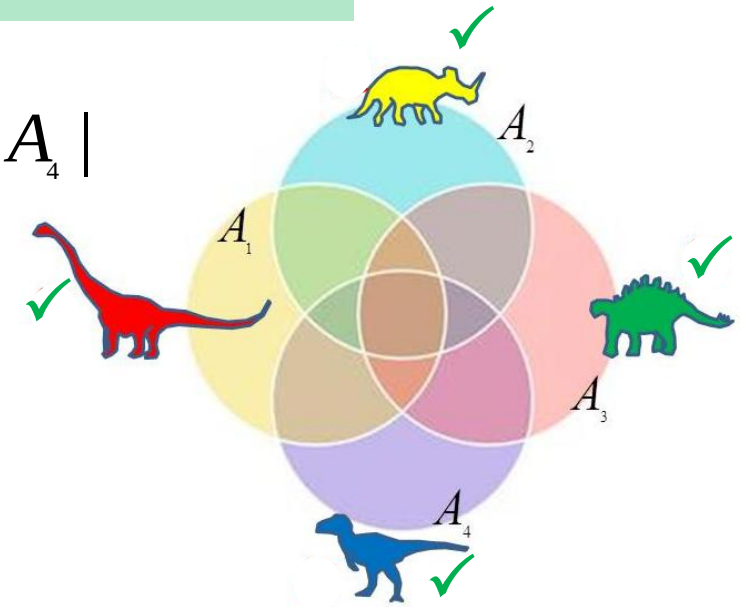
$A_i =$ Put one ball in bin i. Then put the remaining $n-1$ balls into the remaining $m-1$ bins

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



Single occupancy: Inclusion – Exclusion II

Number of single occupancies $= \left| A_1 \cup A_2 \cup A_3 \cup A_4 \right|$



A_i counts placements with singleton at i

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

$$\text{Then } \left| \bigcap_{i \in I, |I|=k} A_i \right| = \sum_{|I|=k} \binom{m-k-1+(n-k)}{m-k-1} = \binom{m}{k}\binom{m-2k-1+n}{m-k-1}$$

Recall T_k reserved k bins for single occupancy, and

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

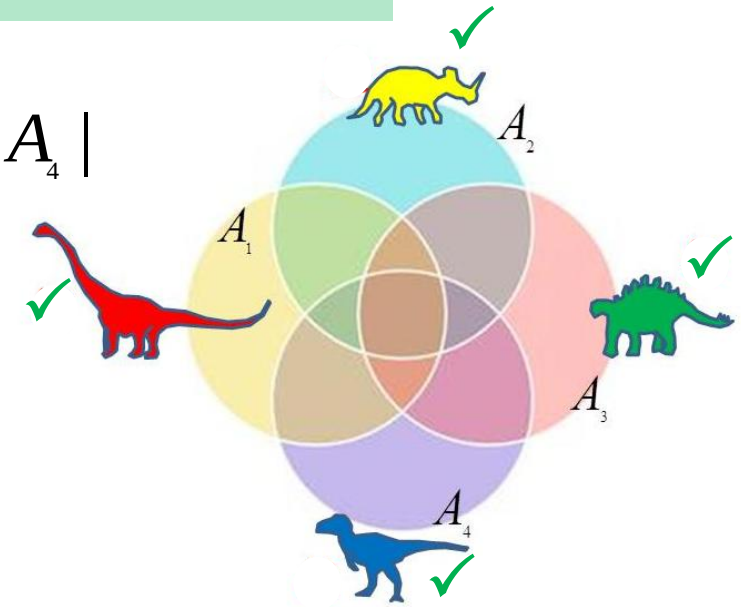
And recall our alternating sum :

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

Single occupancy: Inclusion – Exclusion III

Number of single occupancies $= |A_1 \cup A_2 \cup A_3 \cup A_4|$

A_i counts placements with singleton at i



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

$$= \binom{4}{1} \times \binom{(4-1)-1+5-1}{4-1-1} - \binom{4}{2} \times \binom{(4-2)-1+5-2}{4-2-1} + \binom{4}{3} \times \binom{(4-3)-1+5-3}{4-3-1}$$

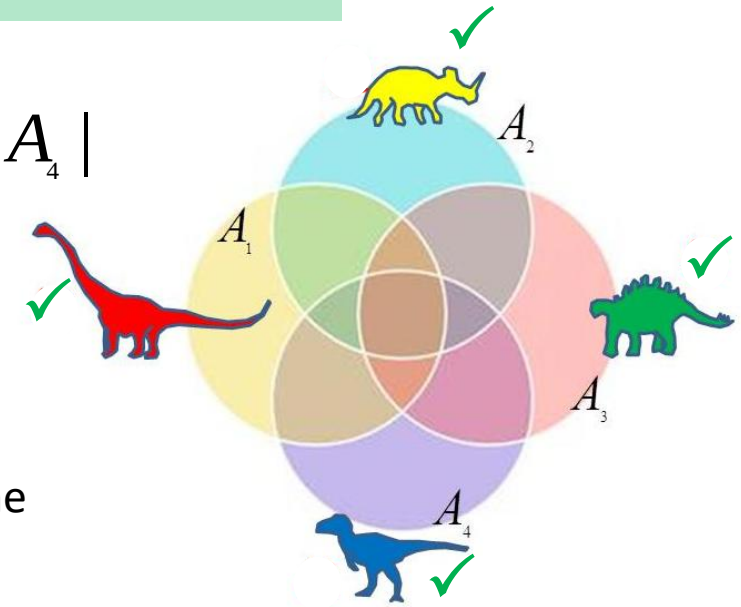
$$- \binom{4}{4} \times \binom{(4-4)-1+5-4}{4-4-1}$$

$$= 4 \times \binom{6}{2} - 6 \times \binom{4}{1} + 4 \times \binom{2}{0} - 0 = 40$$

So my clever algorithm was inclusion – exclusion all along!

Single occupancy: Distinguished balls I

Number of single occupancies $= |A_1 \cup A_2 \cup A_3 \cup A_4|$

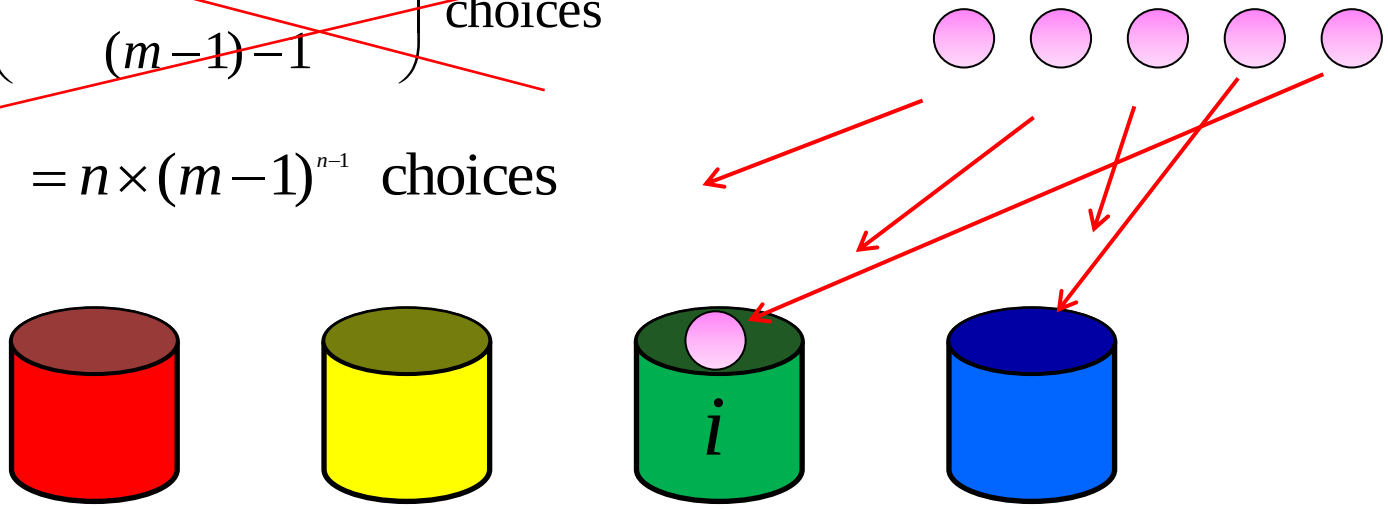


A_i counts placements with singleton at i

$A_i =$ Mappings that assign one ball to bin i. Then map the remaining $n-1$ balls to the remaining $m-1$ bins

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

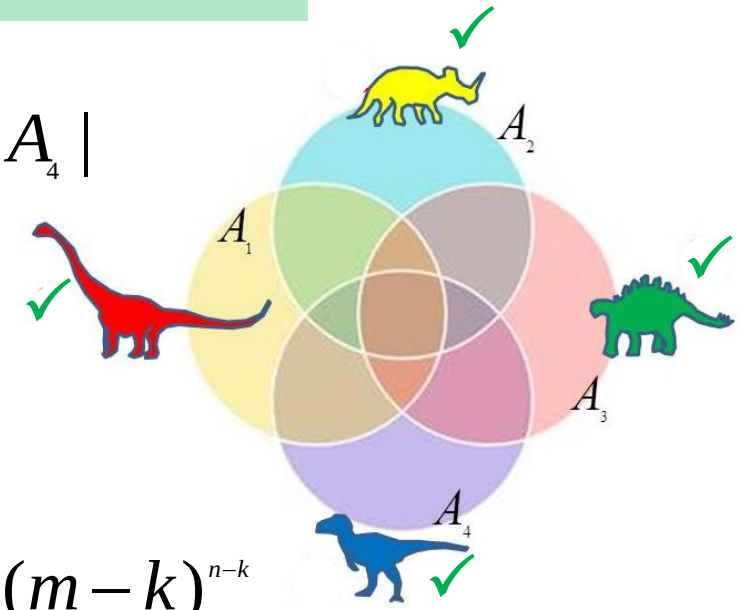
$$= n \times (m-1)^{n-1} \text{ choices}$$



Single occupancy: Distinguished balls II

Number of single occupancies $= \left| A_1 \cup A_2 \cup A_3 \cup A_4 \right|$

A_i counts placements with singleton at i



$$A_i = n \times (m-1)^{n-1}$$

$$\left| \bigcap_{i \in I, |I|=k} A_i \right| = \sum_{|I|=k} n \times (n-1) \times \cdots \times (n-k+1) \times (m-k)^{n-k}$$

$$\text{And } \left| \bigcup_{i=1}^{n} A_i \right| = \sum_{k=1}^{n} (-1)^{k+1} \frac{n!}{(n-k)!} (m-k)^{n-k}$$

Placing balls can confuse (people like me)

The words “place 5 **identical** balls into 4 bins” means it happens **simultaneously**. The balls are not identical if they arrive one by one.

E.g. toss a coin 5 times. This is putting 5 balls into a bin called H and a bin called T.

The outcomes are:

H	**T**
5	0
4	1
3	2
2	3
1	4
0	5

What is the probability of 5 heads? From the table it is 1/6.

Coin tossing: observed or not?

E.g. toss a coin 5 times. This is putting 5 balls into a bin called H and a bin called T.

But what if we **watch** as the coin is tossed. Then the outcomes we may observe are:

H H H H H

H H H H T

H H H T H

⋮

H T T T T

T T T T T

What is the probability of 5 heads? From the table it is 1/32. The table is a **function** from a sequence of 5 tosses to two outcomes.

So if we don't watch the coin tossing we are much more likely to get 5 heads!?

Coin tossing – the penny drops I

thanks to thecatherinehuang.github.io/2020/08/18/balls-and-bins.html

I've been having some trouble with identifying distinguishable and indinguishable objects in probability problems and wanted to write a blog to clear up some concepts.

post on how order is related to permutations, combinations, and common pitfalls I've encountered.

We have four possible problems involving m balls and n bins:

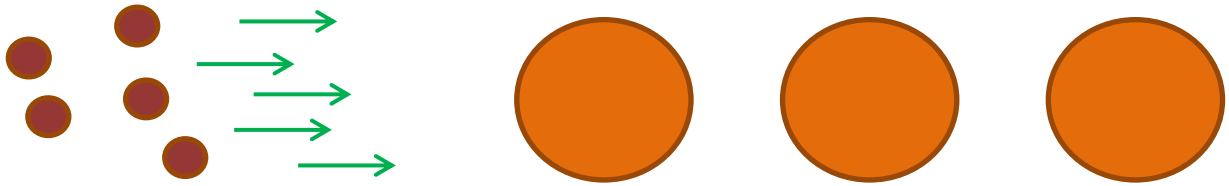
-Distinguishable balls, distinguishable bins: n^m
-Indistinguishable balls, distinguishable bins: $(m+n-1)C(n-1)$
-Distinguishable balls, indistinguishable bins: casework on each of the possible partitions and duplicate sizes need to be accounted for
-Indistinguishable balls, indistinguishable bins: number of possible partitions

What happens when we move into probability? When we look at probability as `\frac{# of successes}{all possibilities},` it doesn't matter if we make order matter or if we don't care about order as long as we're doing the same thing in the numerator and denominator. Does this translate when we have the idea of balls and bins? The key lies in having a very clear idea about what your sample space is.

Coin tossing – the penny drops II

thecatherinehuang.github.io/2020/08/18/balls-and-bins.html

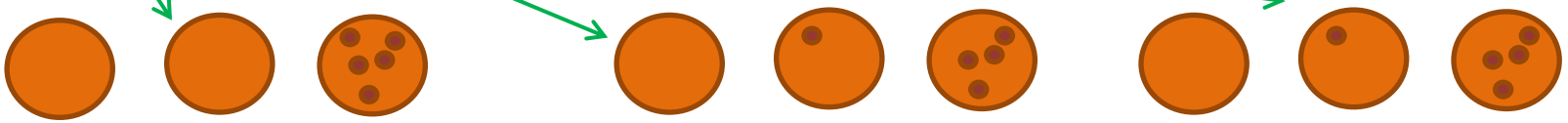
Consider the problem where we have 5 chocolate chips and 3 cookies. What is the probability that each cookie has at least one chocolate chip per cookie? What is our sample space? We want to define our sample space as things that are equally likely to happen, so counting the number of times something happens is meaningful.



What is the sample space if we look at this problem from the perspective of distinguishable cookies (bins) and distinguishable chips (balls)? enumerate chocolate chips 1, …, 5 and the cookies a, b, c. We expect there to be 3^5 possibilities The sample space

$\Omega = \{||12345, |1|2345, |2|1345, |3|1245, |4|1235, |5|1234, |12|345, |13|245, |14|235, |15|234, |23|145, |24|135, |25|134, |34|125, |35$

. Each of these events happens with probability 1/243. How many of them have at least one chocolate chip per cookie? Do complementary counting: the number of elements in the sample space that have at least one cookie with no chips in it is (243-(3+30+60))/243 = (81-31) /81 = 50/81.



How about distinguishable cookies and indistinguishable chips? The sample space contains 7C2 = 21 elements. If we want to ensure that all the cookies have chips, we can prefill them with the indistinguishable chips. Now 2 chips remain to be inserted into any cookie. This can be done 4C2=6 ways. So the probability that each cookie has at least one chip is 6/21 = 2/7.

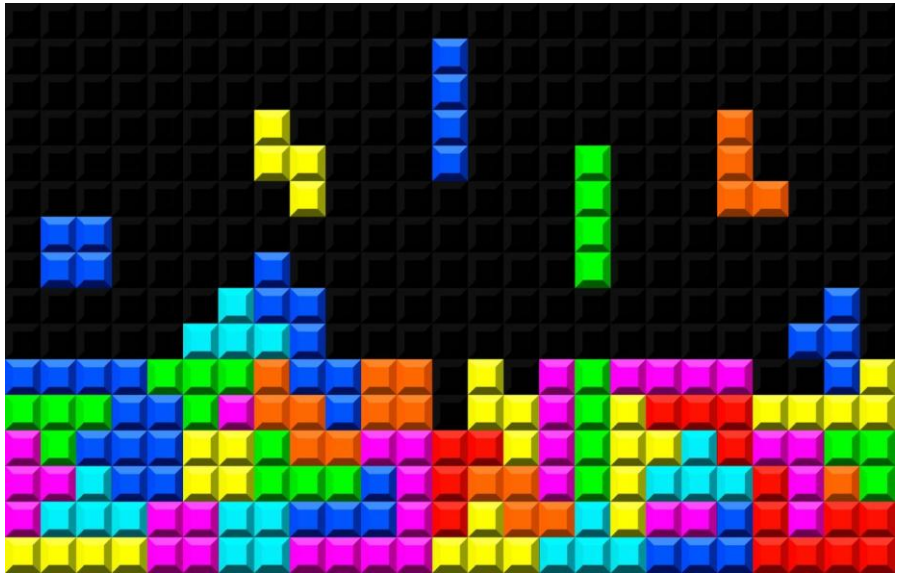
Hold on, 50/81 != 2/7. What went wrong? It turns out that with distinguishable cookies and indistinguishable chips, each event in the sample space doesn't have the same probability of occuring, so we can't count the number of successes and divide by the number of total possibilities. For example, if the chips are ppppp and the cookies are abc, the probability that $|ppppp|$ occurs is less than the probability that $|pp|ppp$ happens.

Distinguishable vs indistinguishable I

Conclusions

1. The instruction ‘place *n* indistinguishable balls into *m* boxes’ is meaningless.

I tried to convince myself it meant **simultaneous placements**. But this is just like saying preconfigured placements: they arrive like Tetris patterns, already formed.



www.futura-sciences.com

Maybe there is a some ball sifting mechanism that distributes *n* balls among *m* boxes so as to achieve indistinguishability of balls?

en.wikipedia.org/wiki/Galton_board

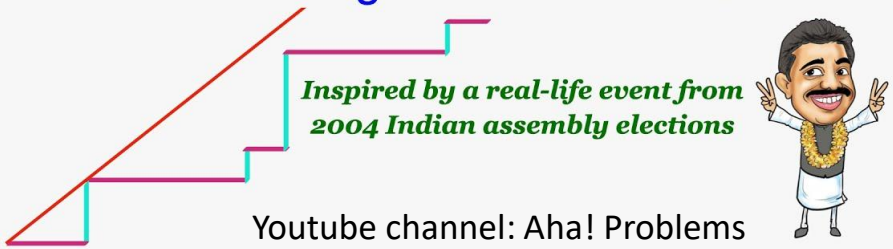
Distinguishable vs indistinguishable II

Conclusions

2. The phrase 'the probability that *i* indistinguishable balls are placed in box *j*' is without practical application.

The Ballot Problem

In an election, candidate **A** receives **40,572** votes and candidate **B** receives just **ONE** less vote. What is the probability that **A will be strictly ahead of B** throughout the ballot count?



Bertrand's Ballot Theorem :

if A gets m votes and B gets $n < m$ votes then the probability that A always has the lead is $\dfrac{m-n}{m+n}$

$m = 4, \mathrm{n} = 2, \ (4-2)/(4+2) = 1/3$

Number of AB voting sequences :

$$\binom{6}{2} = 15$$

Sequences with A leading

AAAABB
AAABAB
AAABBA
AABAAB
AABABA

But these are indistinguishable voters. There is no such thing!

Pigeonholes: where distinguishability is no threat

The Pigeonhole Principle *Let $q_1, q_2, \ldots, q_n$ be positive integers and let Q be a set of size $q_1 + \ldots + q_n + 1$. Take a mapping from Q to a set $P = \{p_1, \ldots, p_n\}$ of size n. Then under this mapping*

1. *for some i, $1 \leq i \leq n$, p_i is the image of more than q_i points in Q;*
2. *if any of the q_i are allowed to be infinity then some p_i in P is the image of infinitely many points in Q.*

Part (1) is sometimes referred to as the 'strong' Pigeonhole Principle

If $q_i = 1$ for all i then we have the classic Pigeonhole Principle.

Sometimes this strong form puts $q_1 + q_2 + \ldots + q_n - n + 1$ balls into the n boxes (not sure why)

Example 2.3. Show that every sequence $a_1, a_2, \ldots, a_{n^2+1}$ of $n^2 + 1$ real numbers contains either an increasing subsequence of length $n + 1$ or a decreasing subsequence of length $n + 1$.

Proof. Suppose there is no increasing subsequence of length $n + 1$. We suffices to show that there must be a decreasing subsequence of length $n + 1$.

Let ℓ_k be the length of the longest increasing subsequence which begins with a_k, $1 \leq k \leq n^2 + 1$. Since it is assumed that there is no increasing subsequence of length $n + 1$, we have $1 \leq \ell_k \leq n$ for all k. By the strong form of the pigeonhole principle, $n + 1$ of the $n^2 + 1$ integers $\ell_1, \ell_2, \ldots, \ell_{n^2+1}$ must be equal, say,

$$\ell_{k_1} = \ell_{k_2} = \cdots = \ell_{k_{n+1}},$$

where $1 \leq k_1 < k_2 < \cdots < k_{n+1} \leq n^2 + 1$. If there is one k_i $(1 \leq i \leq n)$ such that $a_{k_i} < a_{k_{i+1}}$, then any increasing subsequence of length $\ell_{k_{i+1}}$ beginning with $a_{k_{i+1}}$ will result a subsequence of length $\ell_{k_{i+1}} + 1$ beginning with a_{k_i} by adding a_{k_i} in the front; so $\ell_{k_i} > \ell_{k_{i+1}}$, which is contradictory to $\ell_{k_i} = \ell_{k_{i+1}}$. Thus we must have

$$a_{k_1} \geq a_{k_{2+1}} \geq \cdots \geq a_{k_{n+1}},$$

which is a decreasing subsequence of length $n + 1$. □

Infinitely many (distinguishable) balls

§1. RAMSEY'S THEOREM

DEFINITION 7.1: Let $[N]^2$ be the set of all *unordered* pairs of natural numbers, and partition $[N]^2$ into two sets A and B. A subset $X \subseteq N$ is said to be **homogeneous for the partition** $\{A, B\}$ if $[X]^2 \subseteq A$ or $[X]^2 \subseteq B$.

Ramsey's Theorem states that for any partition $\{A, B\}$ of $[N]^2$ there is an infinite homogeneous subset X of N. (This is just a special case of Ramsey's Theorem; a more general statement will be presented shortly.)

Another way of looking at this situation is to imagine a countably infinite set P of points on the plane, no three of which are collinear. Each pair of points is joined by either a red line or a blue line. Then Ramsey's Theorem says that at least one of the following statements is correct:

(1) There is an infinite subset $P_0 \subseteq P$ such that the line joining any two points of P_0 is red.

(2) There is an infinite subset $P_0 \subseteq P$ such that the line joining any two points of P_0 is blue.

Joseph G. Rosenstein, Linear Orderings, Academic Press, 1982.

Ramsey's infinite theorem in action

EXERCISE 7.2: (1) In each of the following cases find a subset $X \subseteq N$ which is homogeneous for the given partition of $[N]^2$:

(a) $\{a,b\} \in A$ if $a \equiv b \pmod 2$, and $\{a,b\} \in B$ otherwise.
(b) $\{a,b\} \in A$ if a and b are relatively prime, and $\{a,b\} \in B$ otherwise.
(c) $\{a,b\} \in A$ if $a = 2n+1$, $b = 2m$, and $n \equiv m \pmod 2$ *or* $a = 2n$, $b = 2m$, and $n \equiv m \pmod 2$ *or* $a = 2n+1$, $b = 2m+1$, and $n \not\equiv m \pmod 2$, and $\{a,b\} \in B$ otherwise.

(2) Show that in case (c), there is no infinite homogeneous subset X for the partition which contains more than one even number and more than one odd number.

From the exercises above, the reader might get the impression that Ramsey's Theorem is a triviality. If so, he should consider the following partition of $[N]^2$ and attempt to describe a set which is homogeneous for it. Linearly order $[N]^2$ as follows: $\{a,b\} \ll \{a',b'\}$ if $a+b < a'+b'$ or if $a+b = a'+b'$ and $\min\{a,b\} < \min\{a',b'\}$. Then $\langle [N]^2, \ll \rangle$ has order type ω; the first few elements are $\{0,1\} \ll \{0,2\} \ll \{0,3\} \ll \{1,2\} \ll \{0,4\} \ll \{1,3\} \ll \{0,5\} \ll \{1,4\} \ll \{2,3\} \ll \{0,6\} \ll \{1,5\} \ll \{2,4\} \ll \cdots$. Now define A to be those elements of $[N]^2$ which are in even places in this ordering and define B to be those elements of $[N]^2$ which are in odd places in this ordering, and try to find a subset $X \subseteq N$ which is homogeneous for $\{A, B\}$.

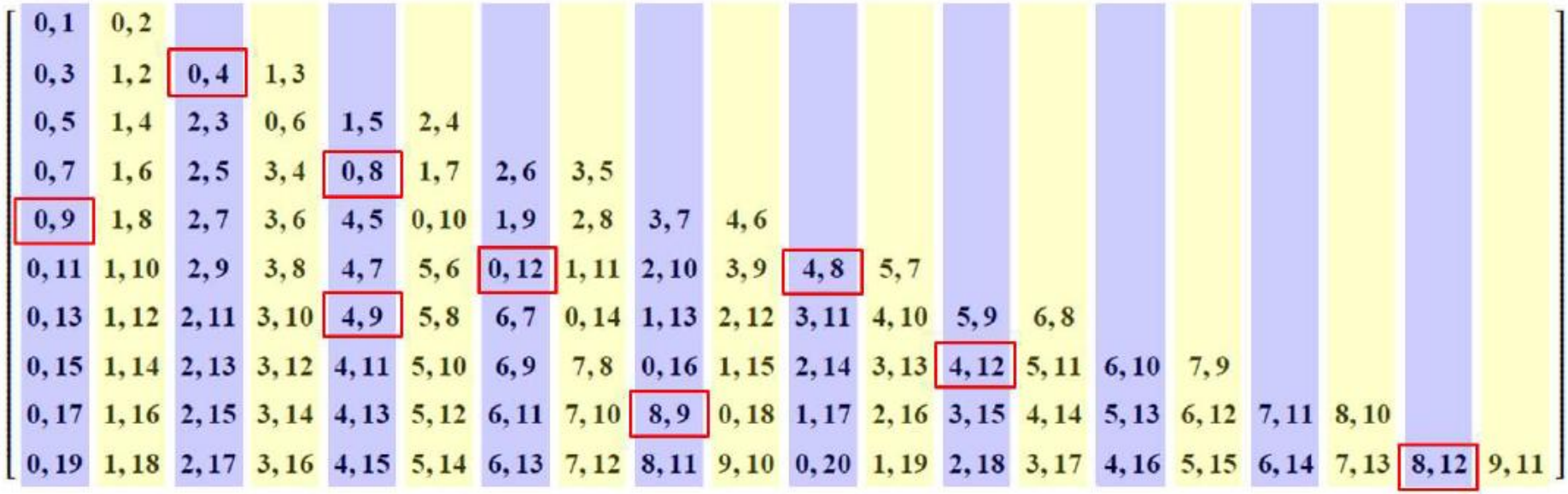
Joseph G. Rosenstein, Linear Orderings, Academic Press, 1982

Rosenstein's Ramsey theorem problem

0,1	0,2																		
0,3	1,2	0,4	1,3																
0,5	1,4	2,3	0,6	1,5	2,4														
0,7	1,6	2,5	3,4	0,8	1,7	2,6	3,5												
0,9	1,8	2,7	3,6	4,5	0,10	1,9	2,8	3,7	4,6										
0,11	1,10	2,9	3,8	4,7	5,6	0,12	1,11	2,10	3,9	4,8	5,7								
0,13	1,12	2,11	3,10	4,9	5,8	6,7	0,14	1,13	2,12	3,11	4,10	5,9	6,8						
0,15	1,14	2,13	3,12	4,11	5,10	6,9	7,8	0,16	1,15	2,14	3,13	4,12	5,11	6,10	7,9				
0,17	1,16	2,15	3,14	4,13	5,12	6,11	7,10	8,9	0,18	1,17	2,16	3,15	4,14	5,13	6,12	7,11	8,10		
0,19	1,18	2,17	3,16	4,15	5,14	6,13	7,12	8,11	9,10	0,20	1,19	2,18	3,17	4,16	5,15	6,14	7,13	8,12	9,11

Table listing the sets $\{i, m-i\}$ for $i = 0 \ldots \lfloor (m-1/2) \rfloor$, with $m = 1 \ldots n$. We have chosen $n = 20$ and each table row covers two values of m, so there are 10 rows in the table. The sets are partitioned as belonging in odd or even columns of the table. The boxed entries seem to promise a set of integers $\{0, 4, 8, 9, 12, \ldots\}$ that will only ever contribute odd column entries, but this fails when $n = 21$ for $\{9, 12\}$ which is even.

Rosenstein's Ramsey theorem problem



0,1	0,2																	
0,3	1,2	0,4	1,3															
0,5	1,4	2,3	0,6	1,5	2,4													
0,7	1,6	2,5	3,4	0,8	1,7	2,6	3,5											
0,9	1,8	2,7	3,6	4,5	0,10	1,9	2,8	3,7	4,6									
0,11	1,10	2,9	3,8	4,7	5,6	0,12	1,11	2,10	3,9	4,8	5,7							
0,13	1,12	2,11	3,10	4,9	5,8	6,7	0,14	1,13	2,12	3,11	4,10	5,9	6,8					
0,15	1,14	2,13	3,12	4,11	5,10	6,9	7,8	0,16	1,15	2,14	3,13	4,12	5,11	6,10	7,9			
0,17	1,16	2,15	3,14	4,13	5,12	6,11	7,10	8,9	0,18	1,17	2,16	3,15	4,14	5,13	6,12	7,11	8,10	
0,19	1,18	2,17	3,16	4,15	5,14	6,13	7,12	8,11	9,10	0,20	1,19	2,18	3,17	4,16	5,15	6,14	7,13	8,12	9,11

Infinite Ramsey Theorem. Suppose that each of the two-element subsets of N is coloured with one of two colours, B or Y, say. Then there is an infinite subset X of N all of whose two-element subsets have the same colour.

Proof. We will specify an infinite sequence, $v_0, v_1, \ldots$, of natural numbers from which we can construct an infinite sequence of two-elements subsets all coloured B or all Y. Let $V_0 = \mathbb{N}$ and let v_0 be its least element, i.e. 0. For all $i > 0$, $\{v_0, i\}$ is either B or Y. By the Pigeonhole Principle there is an infinite subset V_1 of $\mathbb{N}$ with $\{v_0, i\}$ having the same colour for all $i \in V_1$. Let v_1 be the least element of V_1 and repeat this process with V_1 in place of $\mathbb{N}$ to specify $v_2 \in V_2$ and then $v_3 \in V_3$, and so on. Now observe that, for any i, if $j > i$ then v_j lies in V_{i+1} so $\{v_i, v_j\}$ has the same colour. For short, let us say that v_i is either B or Y. This colours the whole sequence of v_is. Apply the Pigeonhole Principle again: there must be an infinite subsequence of the v_i having the same colour, say B. Write this sequence $w_0, w_1, \ldots$. Then every two-element subset $\{w_i, w_j\}$ of $\{w_0, w_1, \ldots\}$ is B.