MSHSML Meet 5, Event C
Study Guide

5C Counting (Combinatorics) and Probability

Permutations, with and without replacement
Combinations, with and without replacement
Using the principle of inclusion - exclusion
Using the binomial and multinomial expansions
Nonnegative integer solutions to $x_1 + x_2 + \cdots + x_n = b$
Definition, simple applications of probability

1 Sets

1.1 Definitions

We start with a list of definitions for terms used in this section on sets.

A set is a collection of unordered, distinguishable objects called elements.

- Elements are unordered: $\{a, b\}$ and $\{b, a\}$ are the same set.
- Elements are distinguishable: $\{a, a, b\}$ is not a set because not all elements are distinguishable.

A multiset is a collection of unordered, but not necessarily distinguishable objects.

- Elements are unordered: $\{a, a, b\}, \{a, b, a\}$ and $\{b, a, a\}$ are the same multiset.

The number of elements in a set is called its cardinality or count. We use the notation $N(A)$ or $|A|$ to represent the cardinality of the set A.

A subset of the set A is a set B such that every element of B is also an element of A. We use the notation $B \subseteq A$ to denote that set B is a subset of set A. A subset B of set A does not necessarily contain any elements and a subset B of set A may contain all the elements of set A.

A set with no elements (*i.e.* with cardinality 0) is called the empty set and is denoted as $\emptyset$ or $\{\,\}$. The empty set is a subset of every set A.

A proper subset of a set A is a set B when B is a subset of A and the set A contains at least one element that is not in B (*i.e*. B is a subset but $B \not\equiv A$). We use the notation $B \subset A$ to denote that set B is a proper subset of set A.

Special Sets, Notation

Some sets are so common that they have their own dedicated symbol that you should be aware of.

$\mathbb{Z}$, the set of integers

$\mathbb{Q}$, the set of rational numbers

$\mathbb{R}$, the set of real numbers

$\mathbb{C}$, the set of complex numbers.

Note that $\mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.

1.2 Operations on Sets

Union of Two Sets A and B

Notation: $A \cup B$

Definition: The union of two sets A and B is the set C consisting of all objects that are in set A ***OR*** set B.

Intersection of Two Sets A and B

Notation: $A \cap B$

Definition: The intersection of two sets A and B is the set C consisting of all objects that are in set A ***AND*** set B

Union of Three Sets A, B, C

Notation: $A \cup B \cup C$

Definition: $A \cup B \cup C = (A \cup B) \cup C$

Intersection of Three Sets A, B, C

Notation: $A \cap B \cap C$

Definition: $A \cap B \cap C = (A \cap B) \cap C$

Note: These definitions extend in a natural way to unions or intersections of any number of sets. For example,

$$A \cup B \cup C \cup D = \big((A \cup B) \cup C\big) \cup D$$

and

$$A \cap B \cap C \cap D \cap E = \Big(\big((A \cap B) \cap C\big) \cap D\Big) \cap E.$$

In words, to find the union (intersection) of a number of sets, begin by finding the union (intersection) of the first two sets in the list. Then find the union (intersection) of this answer with the next set in the list. Then find the union (intersection) of this answer with the next set in the list, etc. (Operators defined in this way are called "**binary operators**".)

Translating "AND" and "OR"

In the language of set theory,

the word "AND" translates to INTERSECTION (symbol: ∩)

the word "OR" translates to UNION (symbol: ∪)

Inclusive and Exclusive "OR"

"We will get rain **OR** snow this week."

The use of "OR" in this sentence does not exclude the possibility that you will get BOTH rain and snow this week.

"I think that Mary **OR** John will take first place in the spelling contest."
The use of "OR" in this sentence does exclude the possibility that BOTH Mary and John will take first place in the spelling contest.

The use of the "OR" as in "event A will occur **OR** event B will occur" that includes the possibility that both event A and event B occur is called the **Inclusive OR.**

The use of the "OR" as in "event A will occur **OR** event B will occur" that excludes the possibility that both event A and event B occur is called the **Exclusive OR.**

> In the language of set theory, the word "OR" is by **default** the **Inclusive OR**.
>
> So when you want to find $A \cup B$ don't forget to include those situations where BOTH A and B occur.

Note: In the language of sets, if you want to use the word OR in its exclusive form you would have to say, "event A will occur OR event B will occur, but not both".

1.2.1 "Not" Operator for Sets

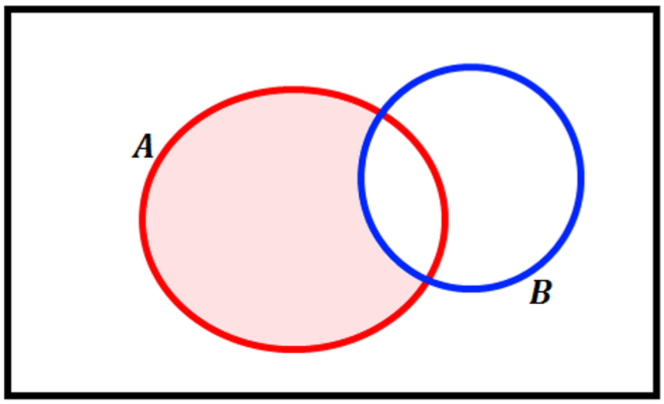
The idea of not being in a particular set comes up often enough that it has earned its own notation. The table below summarizes the various notations for "not A" in common used.

Various Notations for "not B"	
Written	Spoken
B'	B prime
$\overline{B}$	B bar
B^c	B complement

A But Not B

Define A as the set of all objects inside the red ellipse in the Venn Diagram below and define B as the set of all objects inside the blue ellipse.

Then the shaded region includes just that part of A which is not also in B. *i.e.* A but not B



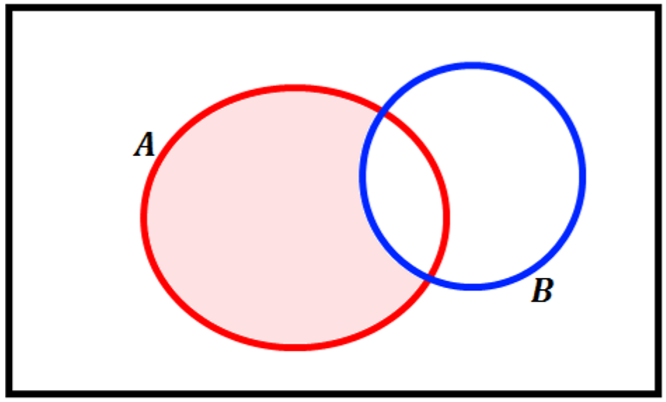
A but not B

Replacing "BUT" with "AND"

Notice that the phrase "A but not B" and the phrase "A and not B" are two ways of saying the same thing!

However, we've dealt with the "AND" before. In the language of sets, the word "AND" translates to the intersection (∩) operation.

So, in typical set notation, we would write $A \cap B'$ to represent the region where objects are "in the set A and (but) not in the set B"



$A \cap B'$

1.2.1.1 "NOT B" REQUIRES CONTEXT

If you are given that $B = \{1,7\}$ and are asked to find B', what would you say?

If the context is all integers from 0 to 9, then $B' = \{0,2,3,4,5,6,8,9\}$.

If the context is all numbers of the form 7^k with $k \in \{0,1,2,3\}$, then $B' = \{7^2, 7^3\}$.

If the context is all numbers x such that $|x - 4| = 3$, then $B' = \emptyset$.

The point is that you <u>cannot</u> specify B' without knowing the context.

1.2.1.2 Implicit Context for "Not B"

In many counting problems the context is *implicit* from the story line. If the problem comes with a story line, then the context is taken to be the set of all possible outcomes.

For example, if the story line asks for the count of all possible numbers which are <u>NOT</u> even when you roll a die, then from context we know that the set of all possible outcomes must be $\{1,2,3,4,5,6\}$. Therefore, *not*$\{\text{even}\} = \{1,3,5\}$ just from the context of the story line.

1.2.1.3 Sample Space

NOTATION: S for Sample Space

In set theory, the letter S, is (usually) reserved for the context (*i.e.* the set of all possible outcomes) in a given situation. The set of all possible outcomes is called the ***outcome space*** or ***sample space*** and the letter S refers back to the "s" in $\underline{s}$ample space.

Alternate Notation: In many books, the Greek letter **Ω** (capital omega) is used to represent the sample space.

NOTATION: ∅ for "not S"

Because S is the set of <u>all</u> possible outcomes, "not S" is the set with **<u>no</u>** possible elements. That is "not S" is a set without any elements.

The symbol ∅ is used to represent "not S". The set ∅ is called the **empty set**. In many books, the symbol { } is used to represent the empty set.

1.3 Partition of a Set

If the set A can be split up into disjoint pieces $B_1, B_2, \dots, B_n$, then we say that $B_1, B_2, \dots, B_n$ partitions A.

In the language of sets, we say that sets $B_1, B_2, \dots, B_n$ partition set A, if

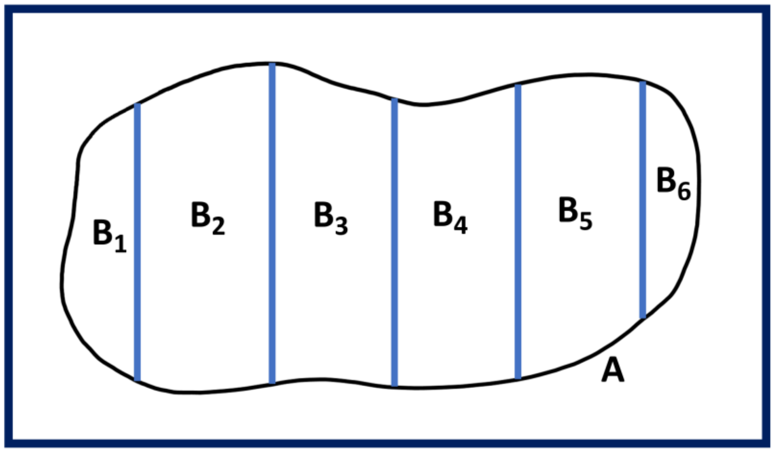
$$B_1 \cup B_2 \cup \cdots \cup B_n = A$$

<u>and</u>

$$B_i \cap B_j = \emptyset \text{ for all } 1 \le i < j \le n.$$

Note: This notation does not preclude the possibility that $B_i = \emptyset$ for some i .

Example: Sets $B_1, B_2, \dots, B_6$ partition set A.



That is, $B_1 \cup B_2 \cup \cdots \cup B_6 = A$ and $B_i \cap B_j = \emptyset$ for all $1 \leq i < j \leq 6$.

Disjoint (or Mutually Exclusive) Sets

Sets $A_1, A_2, \dots, A_n$ are disjoint (or mutually exclusive) if $A_i \cap A_j \equiv \emptyset$ for $i \neq j$ where $\emptyset$ is the empty set.

1.4 DeMorgan's Laws

Consider an experiment where the set of all possible outcomes is denoted by S. Furthermore, let $A_1, A_2, \dots, A_n$ be some subsets of S. Then

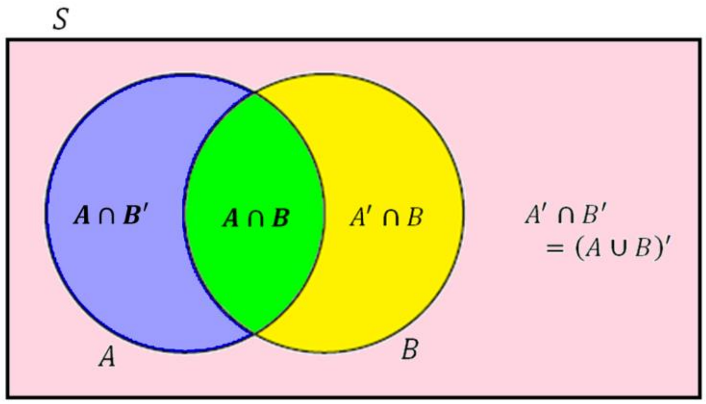
$$(A_1 \cup A_2 \cup \cdots \cup A_n)' = A_1' \cap A_2' \cap \cdots \cap A_n'$$

and

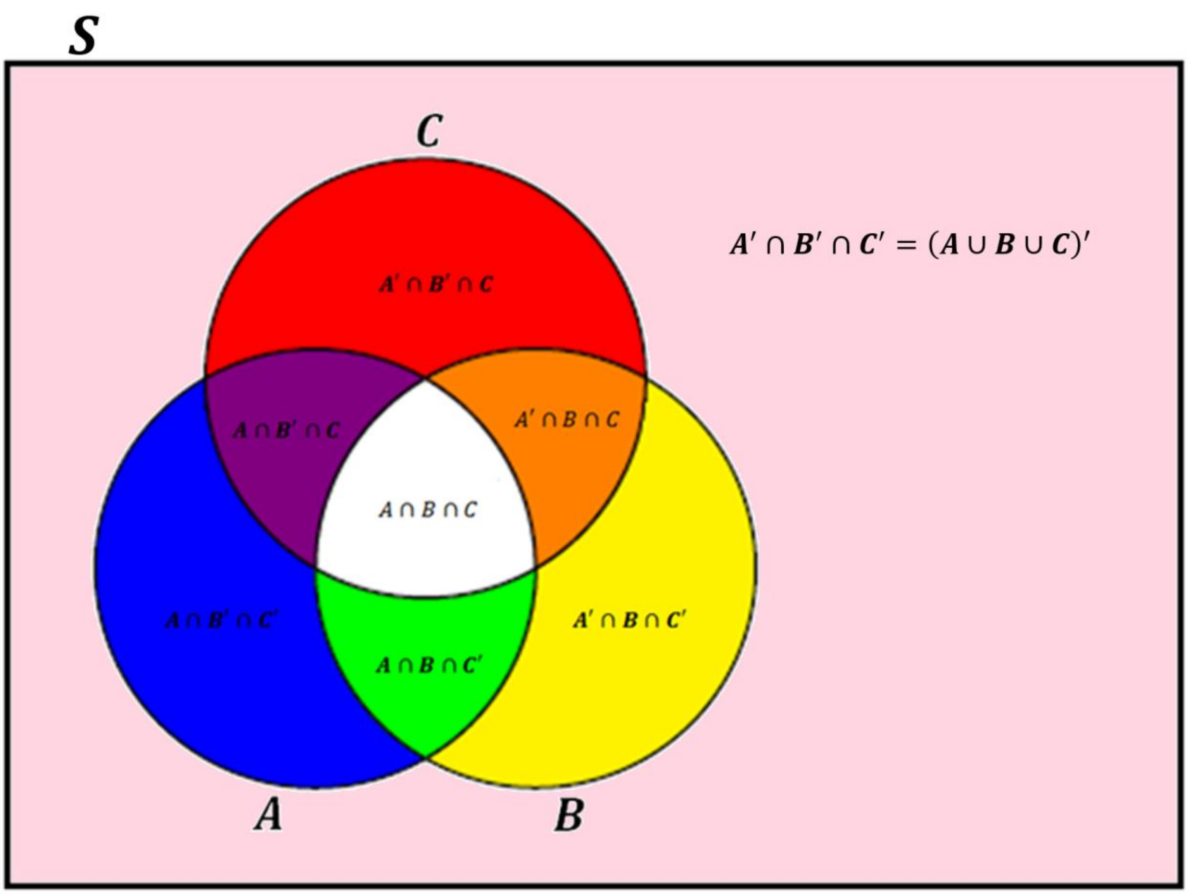
$$(A_1 \cap A_2 \cap \cdots \cap A_n)' = A_1' \cup A_2' \cup \cdots \cup A_n'.$$

1.5 Venn Diagrams

Labeling the Parts in a Two Set Venn Diagram



Labeling the Parts in a Three Set Venn Diagram



1.6 Counting Rules

Notation for Counting

We use the notation $\boldsymbol{N(A)}$ or $|\boldsymbol{A}|$ to represent the cardinality of (*i.e.* the number of objects in) set A.

1.6.1 Rule of Sum

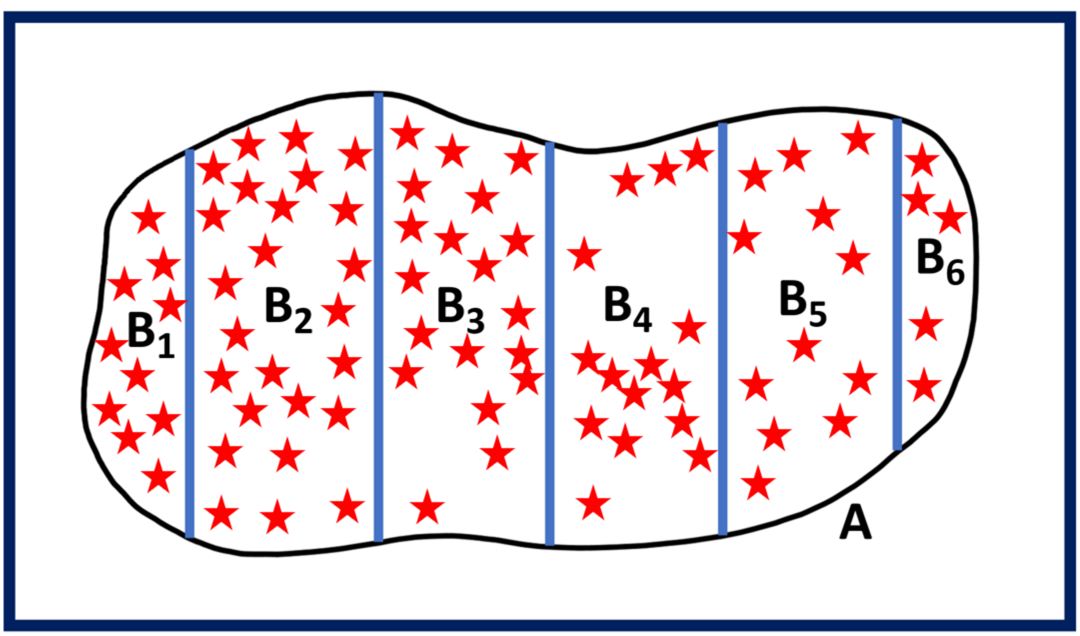
If sets $B_1, B_2, \ldots, B_n$ partition set A, then

$$N(A) = N(B_1) + N(B_2) + \cdots + N(B_n).$$

Example:

Let $N(A)$ equal the number of stars in set A and let $N(B_i)$ equal the number of stars in set B_i, $i = 1,2,3,4,5,6$. The Rule of Sum states that

$$N(A) = N(B_1) + N(B_2) + N(B_3) + N(B_4) + N(B_5) + N(B_6).$$



Note: By the definition of a partition no star can be in more than one partition.

1.6.2 Counting Rule for Not B

$$N(B') = N(S) - N(B).$$

This follows immediately from the Rule of Sum and the definition of B'. Because B and B' partition the sample space S it follows from the Rule of Sum that $N(S) = N(B) + N(B')$.

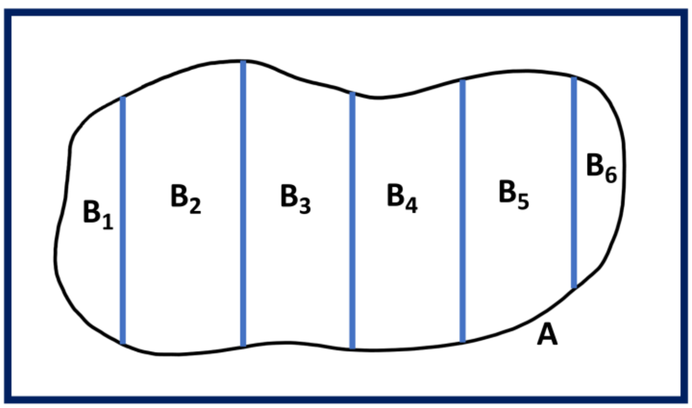
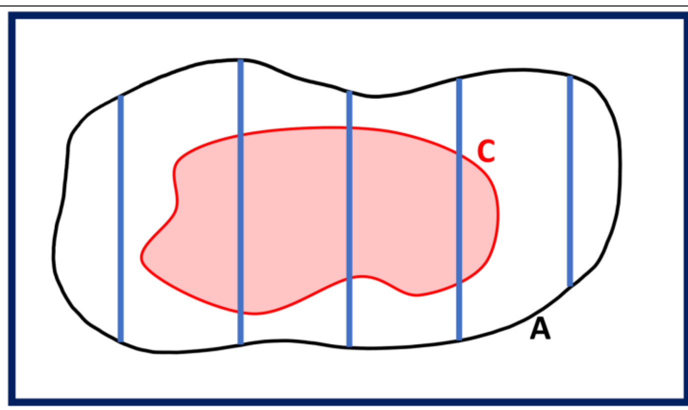
1.6.3 Rule of Total Count

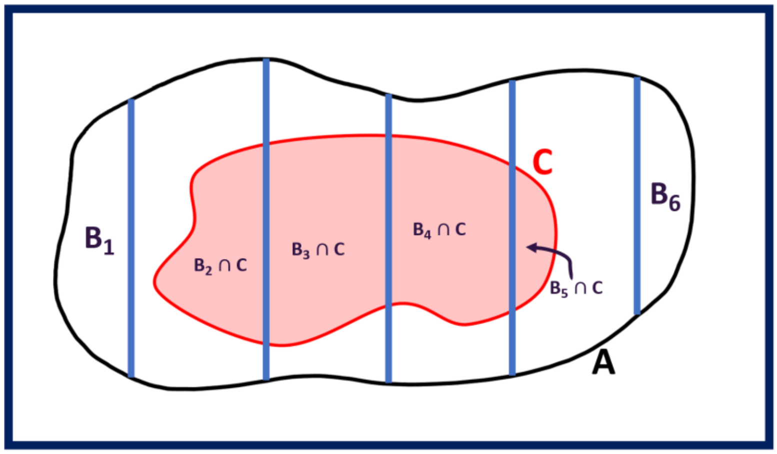
Suppose that set A is partitioned by the sets $B_1, B_2, \dots, B_n$. Furthermore, suppose that $C \subseteq A$. Then

$$N(C) = N(C \cap B_1) + N(C \cap B_2) + \cdots + N(C \cap B_n).$$

(The figures shown below will help to make this intuitive.)

<u>Proof</u>

	If $B_1, B_2, \dots, B_n$ partition A
	and if $C \subseteq A$



then

$B_1 \cap C, B_2 \cap C, \dots, B_n \cap C$

partition C.

Note, in the example shown to the left, $B_1 \cap C = \emptyset$ and $B_6 \cap C = \emptyset$.

By the Rule of Sum, because $B_1 \cap C, B_2 \cap C, \dots, B_n \cap C$ partition C, we know that

$$N(C) = N(C \cap B_1) + N(C \cap B_2) + \cdots + N(C \cap B_n).$$

1.6.4 Counting Rule for "A But Not B"

$$N(A \cap B') = N(A) - N(A \cap B)$$

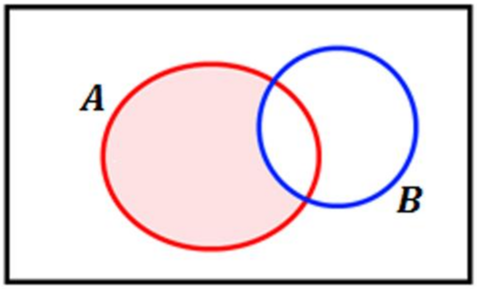
<u>Proof</u>

Provided we assume some context, that is, some set S of all possible outcomes, such that $A \subseteq S$ and $B \subseteq S$, then by definition B and B' partition S and again by definition $B \cap B' = \emptyset$.

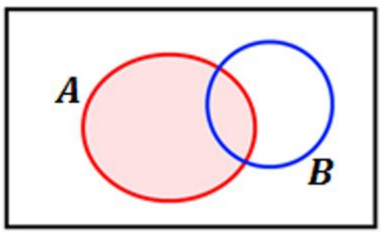
Therefore, by the Rule of Total Count,

$$N(A) = N(A \cap B') + N(A \cap B)$$

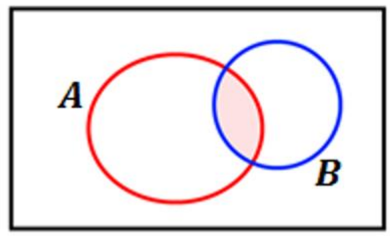
from which the result follows immediately. We can visualize this result with Venn Diagrams in the following way.



$\boldsymbol{N(A \cap B')}$

$=$ 

$\boldsymbol{N(A)}$

$-$ 

$\boldsymbol{N(A \cap B)}$

1.6.5 DeMorgan's Laws and Counting

Consider an experiment where the set of all possible outcomes is denoted by S. Furthermore, let $A_1, A_2, \ldots, A_n$ be some subsets of S. Then

$$N(A_1 \cup A_2 \cup \cdots \cup A_n) = N(S) - N(A_1' \cap A_2' \cap \cdots \cap A_n')$$

and

$$N(A_1 \cap A_2 \cap \cdots \cap A_n) = N(S) - N(A_1' \cup A_2' \cup \cdots \cup A_n').$$

<u>Proof</u>

$N(A_1 \cup A_2 \cup \cdots \cup A_n)$

$= N(S) - N((A_1 \cup A_2 \cup \cdots \cup A_n)')$ Counting Rule for Not B

$= N(S) - N(A_1' \cap A_2' \cap \cdots \cap A_n')$ DeMorgan's Laws

and

$$
\begin{aligned}
&N(A_1 \cap A_2 \cap \cdots \cap A_n) \\
&\quad = N(S) - N\big((A_1 \cap A_2 \cap \cdots \cap A_n)'\big) && \text{Counting Rule for Not } B \\
&\quad = N(S) - N(A_1' \cup A_2' \cup \cdots \cup A_n') && \text{DeMorgan's Laws.}
\end{aligned}
$$

1.6.6 Rule of Sum for a Two Stage Experiment

Two experiments are to be performed. Suppose that the first can result in any one of m possible outcomes. Furthermore, suppose that if the first experiment results in outcome i, then the second experiment can result in any of n_i possible outcomes, $i = 1,2, \dots, m$. Then the number of possible outcomes of the two experiments equals

$$n_1 + n_2 + \cdots + n_m.$$

1.6.7 Count Independent Experiments (Tasks, Steps, …)

Suppose that two experiments (or tasks) are to be performed and suppose that the first experiment (or task) can result in one of a number of different outcomes.

Furthermore, suppose that regardless of the particular outcome of the first experiment, the ***number*** of possible outcomes to the second experiment (task) is the same.

Then we will say that these two experiments (tasks) are **count independent**.

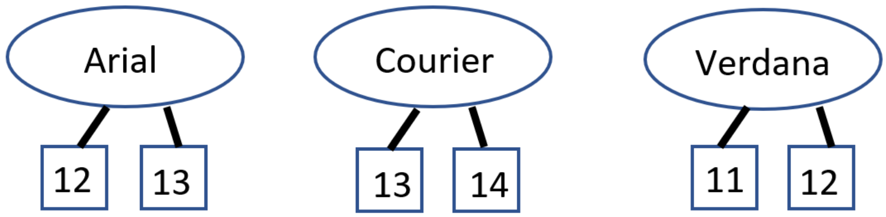
Referring back to the previous paragraph, suppose that the first experiment can result in any one of m possible outcomes and if the first experiment results in outcome i, then the second experiment can result in any of n possible outcomes, $i = 1,2, \dots, m$. In this case these two experiments would be count independent.

Furthermore, from the Rule of Sum for a Two Stage Experiment, the number of possible outcomes of these two experiments would equal

$$n_1 + n_2 + \cdots + n_m = n + n + \cdots + n = n \cdot m.$$

Example:

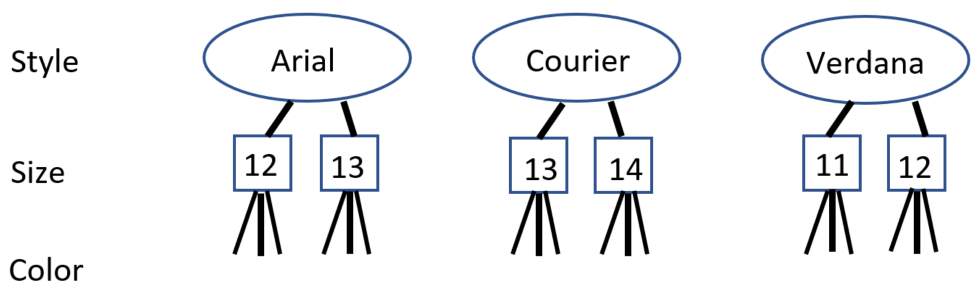
(1) Suppose that you are typing a paper and need to decide on an appropriate font style and then to decide on an appropriate font size. The choices you are satisfied with are delineated below.



The tasks (experiment) of picking a font style and a font size are count independent (in this situation) because regardless of which font style you pick the number of font sizes you can pick from is the same (namely two).

It is important to notice that the particular font sizes choices varies from font to font, but the **number** of font size choices remains the same (namely two) regardless of which font style you pick.

(2) Continuing with the first example, now suppose that for each of the six different (font style, font size) pairs as shown above, there are three different font colors you have decided would be appropriate.



In this situation, we can say that the tasks of picking a font style, size and color are count independent because the number of different ways to do Task 2 (pick a font size) is always the same no matter what the outcome of Task 1 (pick a font style) and the number of different ways to do Task 3 (pick a font color) is always the same no matter the outcome of (Task 1, Task 2) pair.

Note that the three colors you will pick from if (for example) you go with (Arial, 13) might be entirely different than the three colors you will pick from if you go with (Courier, 14). This would not change the fact that these three tasks (pick style, pick size, pick color) are count independent. It only matters that the **number** of colors you will pick from if you go with (Arial, 13) is the same as the **number** of colors you will pick from if you go with (Courier, 14).

1.6.8 Rule of Product

Suppose that each outcome in event C can be *constructed* by performing Step 1 and Step 2 in that order, where,

Step 1. Perform Task 1.
Step 2. Perform Task 2 on the outcome resulting from Step 1.

Suppose that there are m ways of performing Task 1 and n ways of performing Task 2. Furthermore, suppose that Tasks 1 and 2 are count independent.

Then event C consists of $m \times n$ outcomes.

Generalized Rule of Product

Suppose that each outcome in event C can be *constructed* by performing Step 1, Step 2 , … , Step r in that order, where,

Step 1. Perform Task 1.

Step 2. Perform Task 2 on the outcome resulting from Step 1.

Step 3. Perform Task 3 on the outcome resulting from Step 2.

$\vdots$

Step r. Perform Task k on the outcome resulting from Step $(r-1)$.

Furthermore, suppose that there are n_i ways of performing Task i and that Tasks 1,2, … , r are all count independent.

Then event C consists of $n_1 \times n_2 \times \cdots \times n_r$ outcomes.

Considerations When Using the Rule of Sum

The basic idea of the Rule of Sum is to partition a “difficult” problem into a number of “not as difficult” special **cases**.

Remember that to be a valid partition of a set C

(i) each outcome in C must belong to one and only one special case

(ii) there cannot be any outcomes in the union of the outcomes in the special cases that does not belong in C.

Considerations When Using the Rule of Product

In the initial description of the Rule of Product, we said

> "suppose that each outcome in event C can be *constructed* by performing Steps $1, 2, \ldots, r$"

Implicit in that statement was that all the outcomes in C were **correctly** constructed. In this section we will elaborate on what it means to correct construct the outcomes in an event.

To check to see if a construction is in fact "correct" you need to mentally answer the following questions.

1. Does the construction create any outcomes that do not belong in C?
2. Does the construction fail to create some of the outcomes in C?
3. Does the construction create some (or all) of the outcomes in C more than once?
4. Are the construction steps really count independent?

If the answer to each of these questions is "NO", then your construction is a correct one and the Rule of Product does apply.

Is there a way to tell whether to use the Rule of Sum or a Rule of Product on a problem?

KEY: Ask yourself, "Are the phrases of the problem connected with **OR**'s or **AND**'s?"

“OR’s” are an indicator of cases as in

Case 1 “**OR**” Case 2 “**OR**” ⋯

which suggests the <u>Rule of Sum</u>.

while

“AND’s” are an indicator of steps as in

Step 1 “**AND**” then Step 2 “**AND**” then ⋯

which suggests the <u>Rule of Product</u>.

But be cautious here!

The presence (explicitly or implicitly) of connecting “OR’s” is an *indicator* but not a guarantee that you should use the Rule of Sum. You still need to go through the mental exercise of checking if the Rule of Sum is appropriate.

Similarly, the presence (explicitly or implicitly) of connecting “AND’s” is an *indicator* but not a guarantee that you should use the Rule of Product. You still need to go through the mental exercise of checking if the Rule of Sum is appropriate.

Do you Ever Need to use <u>Both</u> Rules in the Same Problem?

The answer is a definite “YES”.

A typical situation of this type occurs when you begin by

using the Rule of Sum to break a problem into a bunch of (hopefully) easier special cases,

and then

use the Rule of Product to get the correct count for each of these special cases.

1.7 Inclusion-Exclusion for Two Sets

$$N(A \cup B) = N(A) + N(B) - N(A \cap B)$$

Proof

By the Rule of Total Count

$$N(A) = N(A \cap B') + N(A \cap B)$$

and it follows by symmetry that

$$N(B) = N(B \cap A') + N(B \cap A).$$

This allows us to see that

$$N(A \cap B') = N(A) - N(A \cap B)$$

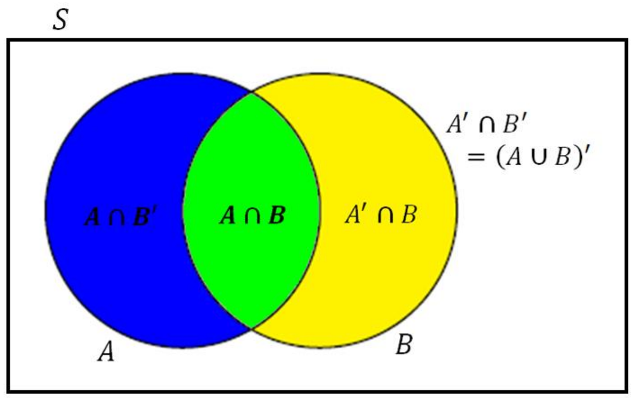
and

$$N(A' \cap B) = N(B) - N(A \cap B).$$

Furthermore,

$$N(A \cup B) = N(A \cap B') + N(A \cap B) + N(A' \cap B).$$

Why? Look back at the diagram:



The sets $A \cap B'$, $A \cap B$, and $A' \cap B$ partition $A \cup B$, so by the Rule of Sum

$$N(A \cup B) = N(A \cap B') + N(A \cap B) + N(A' \cap B).$$

So

$$\begin{aligned} &N(A \cup B) \\ &\quad = \boldsymbol{N(A \cap B')} + \boldsymbol{N(A \cap B)} + \boldsymbol{N(A' \cap B)} \\ &\quad = \big(\boldsymbol{N(A) - N(A \cap B)}\big) + \boldsymbol{N(A \cap B)} + \big(\boldsymbol{N(B) - N(A \cap B)}\big) \\ &\quad = N(A) + N(B) - N(A \cap B). \end{aligned}$$

Comments

In Chapter 4 we will extend two-set inclusion-exclusion in two directions. First, we will learn how it applies to counting $N(A_1 \cup A_2 \cup \cdots \cup A_n)$. Then we will learn that we can replace the counting set function $N(\ )$ with any additive set function such as probability, area and volume.

1.8 Boole's Inequality

$$N(A \cup B) \le N(A) + N(B)$$

Proof

$$N(A \cup B) = N(A) + N(B) - N(A \cap B) \qquad \text{(Rule of Two Set Inclusion-Exclusion)}$$
$$\le \big(N(A) + N(B) - N(A \cap B)\big) + N(A \cap B)$$
$$= N(A) + N(B)$$

Boole's Inequality, Generalized

$$N(A_1 \cup A_2 \cup \cdots \cup A_n) \le N(A_1) + N(A_2) + \cdots + N(A_n)$$

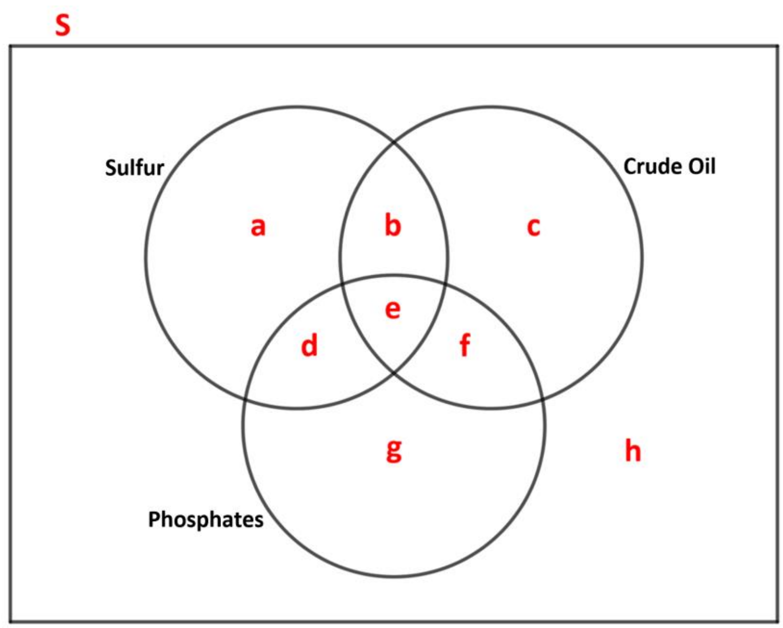
1.9 Practice Problems with Full Solutions

Problem 1.

A study was made of 1000 rivers to determine what pollutants were in them.

- 177 rivers were clean
- 101 rivers were polluted only with crude oil
- 439 rivers were polluted with phosphates
- 72 rivers were polluted with sulfur compound and crude oil, but not with phosphates
- 289 rivers were polluted with phosphates, but not with crude oil
- 463 rivers were polluted with sulfur compounds
- 137 rivers were polluted with only phosphates

Find the value of S, a, b, c, d, e, f, g and h based on the given data.



Source: https://www.math.tamu.edu/~kahlig/venn/pollutants/pollutants.html

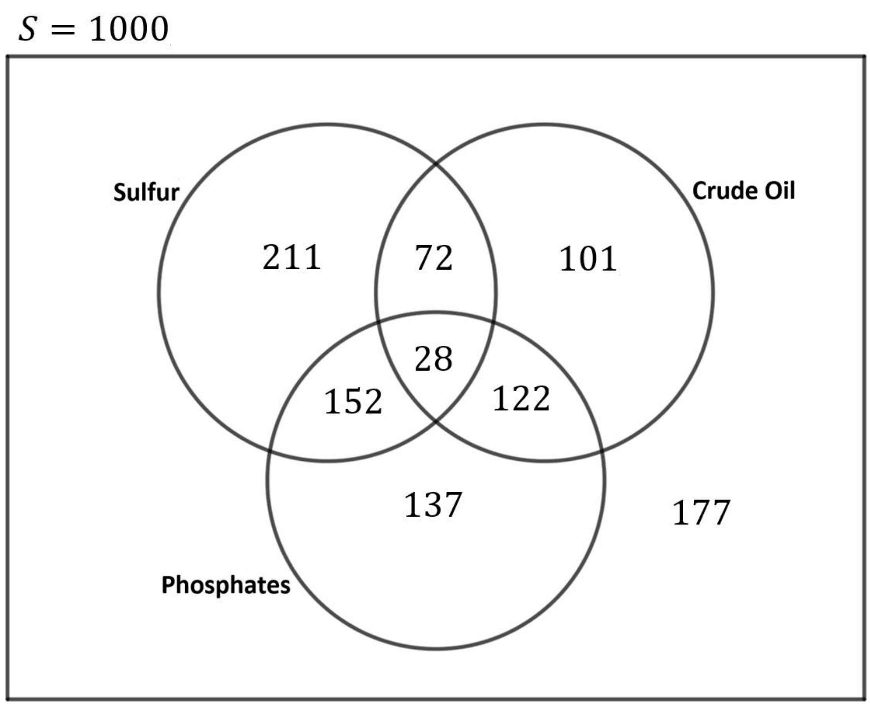
Solution

(1)	study of 1000 rivers	$S = 1000$
(2)	177 rivers clean	$h = 177$
(3)	101 with crude oil only	$c = 101$
(4)	439 with phosphates	$d + e + f + g = 439$
(5)	72 with sulfur and crude oil but not phosphates	$b = 72$
(6)	289 with phosphates but not crude oil	$d + g = 289$
(7)	463 with sulfur compounds	$a + b + d + e = 463$
(8)	137 with phosphates only	$g = 137$

$g = 137, d + g = 289$ $\Longrightarrow d = 289 - 137 = 152$
$439 = d + e + f + g = 152 + e + f + 137$ $\Longrightarrow e + f = 439 - 152 - 137 = 150$
$463 = a + b + d + e = a + 72 + 152 + e$ $\Longrightarrow a + e = 463 - 72 - 152 = 239$

$S = 1000 = a + b + c + d + e + f + g + h$ $= (a + e) + b + c + d + f + g + h$ $= 239 + 72 + 101 + 152 + f + 137 + 177$ $\Longrightarrow f = 1000 - (239 + 72 + 101 + 152 + 137 + 177) = 122$
$f = 122, e + f = 150$ $\Longrightarrow e = 150 - 122 = 28$
$e = 28, a + e = 239$ $\Longrightarrow a = 239 - 28 = 211$

Therefore, $S = 1000, a = 211, b = 72, c = 101, d = 152, e = 28, f = 122, g = 137, h = 177$.



■

Problem 2.

Sometimes it is instructive to examine why a mistake is a mistake. Consider the following *incorrect* construction for a 5 card subset with at least one ace taken from a standard deck of 52 cards.

Step 1. Select an ace from the four aces in a deck
and then

Step 2. Select a 4 card subset from the 51 remaining cards (3 remaining aces $+\ 48$ non-aces).

The actual number of ways to accomplish Step 1 and Step 2 are not relevant to understanding why this is a faulty construction. So, let's just say that there are m ways to accomplish Step 1 and n ways to accomplish Step 2.

Would $m \times n$ be the correct answer to this problem? Consider each of the thought questions you have been recommended to consider each time you prepare to use the Rule of Product.

1.	Do we really want to do Step 1 ***and*** then Step 2? Should it be do Step 1 ***or*** Step 2?
2.	Are our two steps ***really*** count independent? Or does the number of ways to do Step 2 depend on *what* ace out of the four we selected in Step 1?
3.	If we go through this two-step construction in all possible ways would we create any five card hands that don't have at least one ace?
4.	If we go through this two-step construction in all possible ways would we create all possible five card hands with at least one ace or would we miss some?
5.	If we go through this two-step construction in all possible ways would we end up creating some five card hands with at least one ace more than once?

Hint: The Rule of Product breaks down in *exactly one* of these five thought questions. Which one and how?

Solution

The mistake is in Thought Question 5. This construction will create some (but not all) five card hands with at least one ace more than one time.

Consider the following two *distinct* ways to go through the two steps.

Step 1. Select the Ace of Hearts from the four aces in a deck
Step 2. Select the four-card subset $\{$3 Clubs, 6 Spades, Ace of Diamonds, Jack of Diamonds$\}$ from the 51 cards remaining after Step 1.

and

Step 1. Select the Ace of Diamonds from the four aces in a deck
Step 2. Select the four-card subset $\{$3 Clubs, 6 Spades, Ace of Hearts, Jack of Diamonds$\}$ from the 51 cards remaining after Step 1.

These two *distinct* constructions create the same hand – we are creating duplicates.

You should notice that *if* every possible five card hand with at least one ace was created twice (such as above) then we could fix the faulty answer by dividing our final answer by 2.

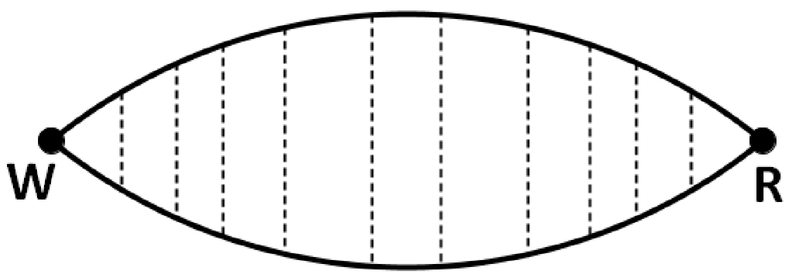
But unfortunately, it is not that simple here. For example, our two-step construction plan will create the hand $\{\text{Ace of Hearts}, 3 \text{ Clubs}, 6 \text{ Spades}, 3 \text{ Hearts}, \text{Jack of Diamonds}\}$ once and only once.

Note: We will offer two different *correct* constructions for this problem in the next section where we take up "Urn Models".

■

Problem 3. (Vilenkin, *Combinatorics*, 1971)

There are two highways connecting the cities of Winona and Rochester and there are ten county roads that connect these two highways along the way (see the map below). How many routes are there from Winona to Rochester that don't loop back on themselves?



<u>Solution</u>

At first glance, this problem seems difficult because not all admissible paths from W (Winona) to R (Rochester) are of the same length. However, consider any admissible path from W to R. It will consist of exactly 11 binary decision points, regardless of the length of this path. The location of these decision points varies from path to path, but there are always exactly 11. Hence, there are 2^{11} different routes.

■

Problem 4. (Mathematics Magazine (1985), variation on AMC 12B (2012))

Imagine a billiard table game where any of the balls numbered from 1 to 15 can be pocketed *first* but thereafter the next ball to go must be numbered consecutively to a ball already pocketed. We do *not* consider the "1" ball and the "15" ball as consecutively numbered in this game.

So, for example, if the "3" ball is pocketed *first* then the next ball pocketed must either be the "2" ball or the "4" ball. If the first two balls pocketed are the "3" and the "2", then the third ball pocketed must be either the "1" or "4". If the first four balls pocketed were $\{3,2,4,1\}$, then the

"5" ball is the *only* choice for the next ball pocketed (remember, we do not consider the "1" ball and the "15" ball as consecutively numbered in this game) and thereafter the balls must come off in the order "6", "7", "8", … , "15".

How many different allowable sequences are there for removing all fifteen balls off this billiard table?

Solution

As is true for many puzzles which consist of a finite number of moves, the analysis is easier if we look at the *last* move in the puzzle and work our way backwards.

In this problem a simple contradiction argument shows that the last ball that is pocketed *must* be either the "1" ball or the "15" ball.

Why? Suppose not. Suppose for example the *last* ball pocketed is the "9" ball. In this case the first ball pocketed must be either less than 9 or greater than 9.

If the last ball is the "9" and the first ball pocketed is less than 9, then the rules of the game would never allow us to pocket a ball greater than "9" – which contradicts the assumption that the "9" ball was the last to go. If the last ball is the "9" and the first ball pocketed is greater than 9, then we could never get to any of the balls $\{1,2,\ldots,8\}$ – which is again a contradiction.

Let's think of picking a ball to go *last* as our "Step 1". So, there are 2 possible ways to accomplish Step 1. Consider picking a ball to go *second to last* as our "Step 2".

If we picked the "1" ball in "Step 1" then (continuing to think backwards) for "Step 2" we just have a reduced version (14 balls numbered "2" to "15") of what we dealt with in "Step 1". It follows from the above discussion for "Step 1" that the only possibilities for "Step 2" would either be the "2" or "15" ball.

That is, if we pick the "1" ball in "Step 1" then there are 2 possible ways to accomplish Step 2.

What if we picked the "15" ball in "Step 1". Once again we would dealing with a reduced version (14 balls numbered "1" to "14") of what we dealt with in "Step 1". So once again it follows from the above discussion for "Step 1" that the only possibilities for "Step 2" would either be the "1" or "14" ball.

That is, if we pick the "15" ball in "Step 1" then there are 2 possible ways to accomplish "Step 2".

So, no matter what happens in "Step 1" there will be 2 possible ways to accomplish "Step 2". This tells us that "Step 1" and "Step 2" are *count independent steps*.

We can continue looking backward to "Step 3", pocketed the third to last ball in the sequence. The same type of analysis of "Step 3" will show that there 2 possible ways to accomplish "Step

3", *regardless of what happened in Steps 1 and 2*. That is, "Step 3" is count independent of all previous steps.

Continuing to work our way backwards we will find that each step is count independent of all previous steps and that there are 2 ways to accomplish each Step 1 to Step 14 but only 1 way to accomplish Step 15 (because at that point there is only one ball left to pick from).

Therefore, by the Rule of Product, there are

$$2 \times 2 \times \cdots \times 2 \times 1 = 2^{14}$$

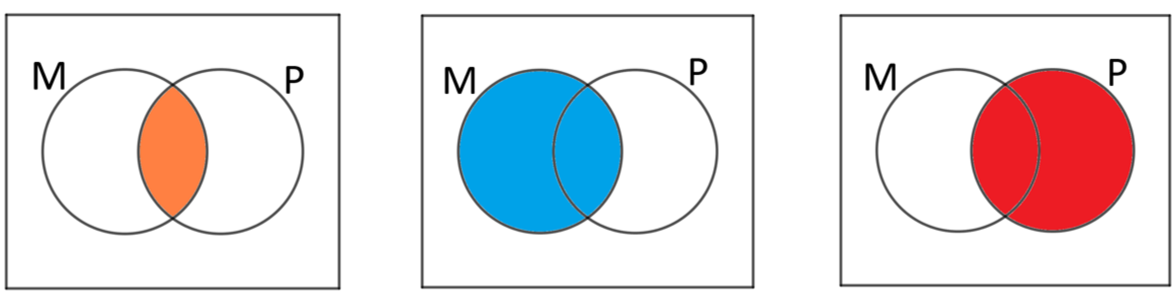
permissible ways to knock the 15 billiard balls off the table.

■

Problem 5. (Manhattan Mathematical Olympiad 2002, Problem 2)

10% of all mathematicians are also philosophers and 1% of all philosophers are also mathematicians. Are there more philosophers or mathematicians?

Solution



Brown region makes up 10% of the blue region.

Brown region makes up 1% of the red region.

$$N(M \cap P) = 0.10 \cdot N(M) \Longrightarrow N(M) = 10 \cdot N(M \cap P)$$

$$N(M \cap P) = 0.01 \cdot N(P) \Longrightarrow N(P) = 100 \cdot N(M \cap P).$$

So

$$N(P) = 100 \cdot N(M \cap P) = 10 \cdot \big(10 \cdot N(M \cap P)\big) = 10 \cdot N(M).$$

There are 10 times as many philosophers as mathematicians.

■

Problem 6.

Pick Me! Pet Store surveyed 500 families about their pets. 400 stated they own a dog, 200 stated they own a cat and 50 stated they own both a dog and a cat. Looking over the survey numbers the store owner noticed something must be wrong. What did the owner notice?

<u>Solution</u>

$N(S) = 500, N(D) = 400, N(C) = 200, N(D \cap C) = 50.$

$N(D \cup C) = N(D) + N(C) - N(D \cap C) = 400 + 200 - 50 = 550 > 500 = N(S).$

But this is impossible because $(D \cup C) \subset S$ so $N(D \cup C) \leq N(S)$.

So, the store owner noticed something must be wrong with this data!

■

Problem 7. (MSHSML, 2015-16 State Tournament, Invitational)

Determine the number of elements of S that are in simplest form if

$$S = \left\{\frac{1}{144}, \frac{2}{144}, \frac{3}{144}, \cdots, \frac{142}{144}, \frac{143}{144}\right\}.$$

<u>Solution</u>

The fraction will be in simplest form if and only if the numerator and denominator are coprime (have no common prime factors). The denominator 144 has prime factors of 2 and 3. So, the problem comes down to counting the numbers in $\{1,2,3, \ldots ,143\}$ that are not divisible by 2 or 3.

Define the events

A_1: numbers in $\{1,2,3, \ldots ,143\}$ that are divisible by 2

A_2: numbers in $\{1,2,3, \ldots ,143\}$ that are divisible by 3.

We want

$$\begin{aligned} N(A_1' \cap A_2') &= N(S) - N(A_1 \cup A_2) \\ &= N(S) - \big(N(A_1) + N(A_2) - N(A_1 \cap A_2)\big) \\ &= 143 - \left(\text{floor}\left(\frac{143}{2}\right) + \text{floor}\left(\frac{143}{3}\right) - \text{floor}\left(\frac{143}{6}\right)\right). \end{aligned}$$

■

Floor Function, floor(x)

Recall that the floor(x) function equals the largest integer less than or equal to x. For example,

$$\text{floor}\left(\frac{5}{2}\right) = 2, \text{floor}(3) = 3, \text{floor}\left(\frac{26}{4}\right) = 6.$$

In the above solution we use two results from number theory.

It is easily seen that the number of integers in the range 1 to n inclusive that are divisible by the positive integer $a \leq n$ is just

$$\text{floor}\left(\frac{n}{a}\right)$$

and the number of integers in the range 1 to n inclusive that are divisible by the positive integers a and b, both less than or equal to n, is just

$$\text{floor}\left(\frac{n}{\text{lcm}(a, b)}\right)$$

where lcm(a, b) is the least common multiple of the two integers a and b.

Problem 7. (*A Tangled Tale*, *Knot X*, 1885, Lewis Carroll)

In a riddle posed by Lewis Carroll, he tells of 100 soldiers after a battle where 85 lost a leg, 80 lost an arm, 75 lost an ear and 70 lost an eye. What is the minimum number of soldiers that lost all four?

<u>Solution</u>

A: set of all soldiers who lost a leg
B: set of all soldiers who lost an arm
C: set of all soldiers who lost an ear

D: set of all soldiers who lost an eye
S: set of all soldiers

The problem is asking for a lower bound on $N(A \cap B \cap C \cap D)$. From the corollary to Boole's inequality we have

$$N(A \cap B \cap C \cap D) \geq N(A) + N(B) + N(C) + N(D) - (4 - 1) \cdot N(S).$$

Therefore, we immediately have

$$N(A \cap B \cap C \cap D) \geq 85 + 80 + 75 + 70 - (4 - 1) \cdot 100 = 10$$

for all possible sets A, B, C, D with the given cardinalities.

But is the best we can do? While it is true that $N(A \cap B \cap C \cap D) \geq 10$, by itself this statement does not logically preclude the possibility that $N(A \cap B \cap C \cap D) \geq 11$ or something higher for all possible sets A, B, C, D with the given cardinalities. In the language of mathematics, we need to show that the inequality $N(A \cap B \cap C \cap D) \geq 10$ is ***sharp***. One way to do this is to describe show sets A, B, C and D where $N(A \cap B \cap C \cap D) = 10$.

We know that $N(A') = 100 - 85 = 15$, $N(B') = 100 - 80 = 20$, $N(C') = 100 - 75 = 25$ and $N(D') = 100 - 70 = 30$. You might notice that

$$N(A') + N(B') + N(C') + N(D') = 90 < 100.$$

This means that we could take any 15 soldiers and assign them to the set A'. Take a different set (disjoint set) of 20 soldiers and assign them to the set B', assign a disjoint set of 25 soldiers to set C' and a disjoint of soldiers to the set D' *without running out of soldiers*.

Now we know from the Rule of Sum that when sets A', B', C' and D' are disjoint then

$$N(A' \cup B' \cup C' \cup D') = N(A') + N(B') + N(C') + N(D') = 90.$$

Furthermore, we know from DeMorgan's Law that (in general)

$$N(A \cap B \cap C \cap D) = N(S) - N(A' \cup B' \cup C' \cup D').$$

Thus, if we construct sets A', B', C' and D' so they are disjoint (which we have just reasoned is necessarily doable), then

$$N(A \cap B \cap C \cap D) = N(S) - N(A' \cup B' \cup C' \cup D') = 100 - 90 = 10.$$

So, the inequality $N(A \cap B \cap C \cap D) \geq 10$ is *sharp*. It's the best we can do.

$\blacksquare$

2 Urn Models

The tradition of modeling combinatorics (counting) and probability problems as an exercise in removing or distributing balls in one or more urns dates to Jacob Bernoulli's seminal book in combinatorics, *Ars Conjectandi* (Latin for "The Art of Conjecturing"), which was written around 1684. Bernoulli's terminology is based on the practice in ancient Greek and Roman times of election by lots (*i.e.* through a lottery) which involved drawing different colored balls from a Roman *urna*, a three handled vessel for storing water. The characteristic shape of this vessel was ideal for preventing someone from seeing what color ball they were pulling out. The common practice of using this particular type of vessel for casting lots explains how the word *urna* came to mean "ballot box" in present day Italian.

The purpose of "urn models" is to solve a problem in as generic of form as possible. This problem and solution then constitutes a blank form that we can then "plug into" to answer a particular problem *without having to reinvent the wheel*.

Consider, for example, the problem of selecting three students from a probability class. In this situation the "urn" is the classroom and the "balls" are the students.

This is a situation where we are taking balls out of an urn.

In a problem where we are distributing 5 identical pieces of candy amongst 3 kids, the "urns" are the kids and the "balls" are the pieces of candy.

This is a situation where we are putting balls into a group of urns.

In these notes we will refer to urn models where we are ***taking objects out of a single urn*** according to some given set of rules as ***Type I* urn models**. We will refer to urn models where we are ***putting objects into a row of urns*** according to some given set of rules as ***Type II* urn models**.. [But this is for our own convenience. It is not a standard term used throughout combinatorics.]

We will introduce eight different urn models, four Type I and four Type II urn models. However, there is a bijection between Type I "put objects into an urn" and Type II "take objects out of urns" models so there are only four formulas to remember instead of eight.

We follow this introduction with a lengthy list of solved problems with discussion. For the most part these are "one step" problems once you realize which of the eight urn models applies. (This contrasts with AMC and AIME problems which are rarely "one step".)

But do not overlook these "simple" problems. There are many ways to disguise which, *if any*, of these eight urn models is appropriate. These problems will introduce many of the keywords that can guide you towards the correct model. Pay close attention to the wording as you read and study through these solved problems. Think about what phrase or words in a problem are associated with each urn model.

Urn models are versatile. You may find it surprising just how many situations can be successfully translated into an equivalent urn model problem.

Learning how to translate a problem from "words into math" is a learned skill and one you need to focus on as part of becoming a successful problem solver.

2.1 Type I Urn Model (Taking Objects Out of an Urn)

Sequences, Type I

The number of ways to select r objects taken with replacement from an urn containing n distinguishable objects when the order in which objects are selected is important is equal to

$$n^r.$$

Permutations, Type I

The number of ways to select r objects taken without replacement from an urn containing n distinguishable objects when the order in which objects are selected is important is equal to

$$\frac{n!}{(n-r)!}.$$

Combinations, Type I

The number of ways to select r objects taken without replacement from an urn containing n distinguishable objects when the order in which objects are selected is not important is equal to

$$\frac{n!}{r!\,(n-r)!}.$$

Multisets, Type I

The number of ways to select r objects taken with replacement from an urn containing n distinguishable objects when the order in which objects are selected is not important is equal to

$$\frac{(n+r-1)!}{r!\,(n-1)!}.$$

Look carefully at what is the same and what varies in the description of these four models.

Same in All Four Type I Models

n, the number of distinguishable objects initially in the urn

r, the number of objects we take out of the urn

Varies from Model to Model

whether we sample with or without replacement after each draw

whether the order the objects are selected is important or not.

2.2 Type II Urn Model (Putting Objects Into Urns)

Sequences, Type II

The number of ways to distribute r distinguishable objects into a row of n distinguishable urns with any number of objects per urn is equal to

$$n^r.$$

Permutations, Type II

The number of ways to distribute r distinguishable objects into a row of n distinguishable urns with at most one object per urn is equal to

$$\frac{n!}{(n-r)!}.$$

Combinations, Type II

The number of ways to distribute r indistinguishable objects into a row of n distinguishable urns with at most one object per urn is equal to

$$\frac{n!}{r!\,(n-r)!}.$$

Multisets, Type II

The number of ways to distribute r indistinguishable objects into a row of n distinguishable urns with any number of objects per urn is equal to

$$\frac{(n+r-1)!}{r!\,(n-1)!}.$$

Look carefully at what is the same and what varies in the description of these four models.

Same in All Four Type II Models

n, the number of distinguishable urns

r, the number of objects we are putting into these urns

Varies from Model to Model

rules for the maximum number of objects we can put into a given urn

whether the objects we are putting into the urns are distinguishable or indistinguishable.

2.3 Proving the 8 Urn Model Formulas

The proofs of these urn model formulas are highly instructive and working through them will help you understand when and how to apply them. Of particular importance is the reasoning connecting what I call “Type I” and “Type II” models.

You can find these proofs and discussions in the 2018 SMI Class Notes posted at this website.

2.4 Binomial Coefficients

From here on in these notes we will adopt the ***binomial coefficient notation***

$$\binom{n}{r} = \begin{cases} \dfrac{n!}{r!\,(n-r)!} & \text{for all integers } n \geq r \geq 0 \\ 0 & \text{else} \end{cases}$$

for combinations of n objects taken r at a time. But you will find many other notations of this number in use in other sources including C_r^n, ${}_nC_r$, ${}^nC_r$, $C(n,k)$ and $C_{n,r}$.

When speaking this symbol $\binom{n}{r}$ is read as “n choose r”. Be sure not to include a fraction bar between the numbers n and r when using this notation. It does not belong there!

We will mention, but not elaborate on, the fact that this definition for the binomial coefficient can be extend to cases other than integers $n \geq r \geq 0$. But we will not need to consider this in these notes.

Notice that we can also express the formula for multisets in terms of the binomial coefficient notation. That is,

$$\binom{n+r-1}{r} = \frac{(n+r-1)!}{r!\,(n-1)!}.$$

We do not use a special symbol for permutations of n objects taken r at a time in these notes. Rather we just denote it by its formula

$$\frac{n!}{(n-r)!}$$

for all integers $n \geq r \geq 0$. But as for the case of combinations you will find many notations for this formula including P_r^n, ${}_nP_r$, ${}^nP_r$, $P(n,k)$ and $P_{n,r}$.

Terminology

Notice the terms "sequences" and "multisets" that we have used in both Type I and Type II urn models. These are the names we will use throughout these notes, but we caution (*caveat lector*) that other names are in common in use in "contest math" as well. In particular, note that

Sequences $\equiv$ Permutations with Repetition

Multisets $\equiv$ Combinations with Repetition.

2.4.1 Special Cases of Binomial Coefficients

It will be helpful to remember the following special cases for binomial coefficients as they come up frequently.

$$\binom{a}{0} = 1, \qquad \binom{a}{1} = a, \qquad \binom{a}{2} = \frac{a(a-1)}{2}.$$

<u>Proof</u>

$$\binom{a}{0} = \frac{a!}{0!\,(a-0)!} = \frac{a!}{0!\,a!} = \frac{a!}{1 \cdot a!} = 1$$

$$\binom{a}{1} = \frac{a!}{1!\,(a-1)!} = \frac{a \times (a-1)!}{1!\,(a-1)!} = a$$

$$\binom{a}{2} = \frac{a!}{2!\,(a-2)!} = \frac{a(a-1) \times (a-2)!}{2!\,(a-2)!} = \frac{a(a-1)}{2}.$$

2.5 Symmetry Rule for Binomial Coefficients

For integers $a \geq b \geq 0$,

$$\binom{a}{b} = \binom{a}{a-b}.$$

So, for example

$$\binom{13}{5} = \binom{13}{8}.$$

Proof

$$\binom{\boldsymbol{a}}{\boldsymbol{a}-\boldsymbol{b}} = \frac{\boldsymbol{a}!}{(\boldsymbol{a}-\boldsymbol{b})!\,(\boldsymbol{a}-(\boldsymbol{a}-\boldsymbol{b}))!} = \frac{a!}{(a-b)!\,b!} = \frac{a!}{b!\,(a-b)!} = \binom{a}{b}.$$

2.6 Permutations with Repeated Letters

There are

$$\boxed{\frac{(a_1 + a_2 + \cdots + a_r)!}{a_1!\,a_2!\cdots a_{r-1}!\,a_r!}}$$

ways to permute $a_1 + a_2 + \cdots + a_r$ letters consisting of

a_1 Type 1 letters (for example "A's")

a_2 Type 2 letters (for example "B's")

$\vdots$

a_r Type r letters.

Of course, if all the letters are distinct ($a_1 = a_2 = \cdots = a_r = 1$) then there are

$$\frac{(a_1 + a_2 + \cdots + a_r)!}{a_1!\, a_2! \cdots a_{r-1}!\, a_r!} = \frac{r!}{1!\, 1! \cdots 1!\, 1!} = r!$$

permutations of r distinct letters.

Explanation

We can think of this problem as asking for the number of distinct ways to fill n distinct slots arranged in a row with a_1 type 1 letters, a_2 type 2 letters, ... , a_r type r letters.

To simplify the notation in what follows we will let $n = a_1 + a_2 + \cdots + a_r$. We can construct a permutation of these $n = a_1 + a_2 + \cdots + a_r$ letters in r steps.

Step 1. Select a_1 of the r slots as the locations for the type 1 letters
Step 2. Select a_2 of the remaining $r - a_1$ unused slots as the locations for the type 2 letters
Step 3. Select a_3 of the remaining $r - a_1 - a_2$ unused slots as the locations for the type 3 letters

$\vdots$

Step r. Select a_r of the remaining $n - a_1 - a_2 - \cdots - a_{r-1}$ unused slots as the locations for the type r letters

Each step is a **combinations - type I** problem (selecting distinct slots, without replacement, order of selection not important) from the all slots remaining after the previous steps. Furthermore, these r steps are count independent. (For example, the number of ways to locate the a_2 type 2 letters does not depend on where the a_1 type 1 are located.)

Hence, there are

$$\binom{n}{a_1}\binom{n-a_1}{a_2}\binom{n-a_1-a_2}{a_3}\cdots\binom{n-a_1-a_2-\cdots-a_{r-1}}{a_r}$$

distinct ways to permute the $n = a_1 + a_2 + \cdots + a_r$ letters. Fortunately this messy product simplifies. We note that

$$\binom{n}{a_1}\binom{n-a_1}{a_2}\binom{n-a_1-a_2}{a_3}\cdots\binom{n-a_1-a_2-\cdots-a_{r-1}}{a_r}$$

$$= \left(\frac{n!}{a_1!\,(n-a_1)!}\right)\left(\frac{(n-a_1)!}{a_2!\,(n-a_1-a_2)!}\right)\cdots\left(\frac{(n-a_1-a_2-\cdots-a_{r-1})!}{a_r!\,(n-a_1-a_2-\cdots-a_{r-1}-a_r)!}\right)$$

$$= \left(\frac{n!}{a_1!\,\cancel{(n-a_1)}!}\right)\left(\frac{\cancel{(n-a_1)}!}{a_2!\,\cancel{(n-a_1-a_2)}!}\right)\left(\frac{\cancel{(n-a_1-a_2)}!}{a_3!\,\cancel{(n-a_1-a_2-a_3)}!}\right)$$

$$\cdots\left(\frac{(n-a_1-a_2-\cdots-a_{r-1}-a_{r-2})!}{a_{r-1}!\,\cancel{(n-a_1-a_2-\cdots-a_{r-1})}!}\right)\left(\frac{\cancel{(n-a_1-a_2-\cdots-a_{r-1})}!}{a_r!\,(n-a_1-a_2-\cdots-a_{r-1}-a_r)!}\right)$$

$$= \frac{n!}{a_1!\,a_2!\,a_3!\cdots a_r!\,(n-a_1-a_2-\cdots-a_{r-1}-a_r)!}$$

$$= \frac{n!}{a_1!\,a_2!\,a_3!\cdots a_r!\,0!} = \frac{(a_1+a_2+\cdots+a_r)!}{a_1!\,a_2!\cdots a_{r-1}!\,a_r!}$$

because $(n-a_1-a_2-\cdots-a_{r-1}-a_r)! = 0! = 1.$

■

Example:

The classic example used to illustrate this result is to count the number of permutations of the letters in the word MISSISSIPPI. In this case we have

$a_1 = 1$ M's (type 1 letters)

$a_2 = 4$ I's (type 2 letters)

$a_3 = 4$ S's (type 3 letters)

$a_4 = 2$ P's (type 4 letters)

So, by our general formula there are

$$\frac{(1+4+4+2)!}{1!\,4!\,4!\,2!}$$

possible permutations of these letters.

2.7 Practice Problems with Full Solutions

Problem 1a

How many ways can you distribute 7 different toys among 4 children if you place no restrictions on the number of toys any child can receive?

Answer

4^7

Explanation

Sequences – type II. Distributing distinguishable objects (toys) into distinguishable urns (children) with any number of objects (toys) per urn (child).

■

Problem 1b

How many ways can you distribute 7 identical toys among 4 children if every child must receive at least one toy?

Answer

$\binom{6}{3}$

Explanation

Step 1. First give each child one toy, leaving 3 toys. There is only 1 way to do this.

Step 2. Then you are left with the **multiset – type II** problem of distributing the remaining $r = 3$ indistinguishable objects (toys) into the $n = 4$ distinguishable urns (children) with any number objects (toys) per urn (child).

$$\binom{n+r-1}{r} = \binom{4+3-1}{3} = \binom{6}{3}.$$

The number of ways to do Step 2 does not depend on what happened in Step 1. So by the Rule of Product the final answer must be

$$1 \times \binom{6}{3} = \binom{6}{3}.$$

■

Problem 1c

How many ways can you distribute 4 identical toys among 7 children if no child is to receive more than one toy?

Answer

$\binom{7}{4}$

Explanation

Combinations – type II. Distributing $r = 4$ identical objects (toys) into $n = 7$ distinguishable urns (children) with at most one object (toy) per urn (child).

■

Problem 2a (MSHM, 2013-14 State Tournament, Individual Event C)

A convenience store serves four kinds of ice cream: Butter Pecan, Chocolate, Strawberry, and Vanilla. If the order in which scoops are placed on the cone doesn't matter, and if scoops can be the same flavor, how many different 2-scoop cones are possible?

<u>Solution</u>

Multisets – type I (sampling with replacement, order not important)

$$\binom{n+r-1}{r} = \binom{4+2-1}{2} = \binom{5}{2}$$

■

Problem 2b (MSHM, 2014-15, Meet 5, Individual Event C)

In how many ways can 9 identical candy bars be distributed to 4 children? *(The candy bars are only distributed in whole units, without being cut.)*

<u>Solution</u>

Multisets – type II (distributing identical objects into distinguishable urns)

$$\binom{4+9-1}{9} = \binom{12}{9}$$

■

Problem 3a

How many strings of $r = 5$ letters can you form using the $n = 26$ letters in the English language if the letters must be in alphabetic order and if the letters cannot be repeated?

<u>Answer</u>

$$\binom{26}{5} = \frac{26!}{5!\,(26-5)!} = \frac{26!}{5!\,21!}.$$

Explanation

The restriction that the letters be in alphabetical order is equivalent to specifying that only one ordering of any given letters should be counted. This is equivalent to specifying that order does not matter. Hence, the problem is one of unordered sampling, $r = 5$ out of $n = 26$ letters without replacement, *i.e.* **combinations – type I**.

$$\binom{n}{r} = \binom{26}{5}.$$

Detailed Explanation

The phrase "order not important" means that we do not want to count different orderings of the same set of objects as distinct selections.

For example, if we select three letters from the alphabet and we direct that "order is not important", then

$$(A, B, C), (A, C, B), (B, A, C), (B, C, A), (C, A, B), (C, B, A)$$

should be viewed as the same selection and hence should only add 1 (and not 6) to the final count.

When a problem requires the elements selected be listed in alphabetic order, only the first of the six possible selections

$$(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}), (A, C, B), (B, A, C), (B, C, A), (C, A, B), (C, B, A)$$

is included in the final count (because it is the only one in **alphabetic order**.)

So, when you say "order is not important" the above six selections only add 1 to the final count because the six selections are considered to be the same (indistinguishable) and we only want to count distinguishable selections.

And when you say “in alphabetic order” the above selections only add 1 to the final count because only 1 of the six meets the required alphabetic order requirement.

So, even though it sounds a bit odd, “**order is not important**” and “**in alphabetic order**” lead to the same final count (but for different reasons).

Therefore, for this problem, we want to count all possible samples of size $r = 5$ from an urn containing $n = 26$ distinct object when we sample *without replacement* (we are not allowing letters to be repeated) and where *order is not important* (because this leads to the same count as requiring the letters are in alphabetic order.

Hence, the problem is one of unordered sampling, $r = 5$ out of $n = 26$ letters without replacement, *i.e.* combinations.

$$\binom{n}{r} = \binom{26}{5}$$

■

Problem 3b

How many strings of $r = 5$ letters can you form using the $n = 26$ letters in the English language if the letters must be in alphabetic order and if the letters can be repeated?

Answer

$\binom{30}{5}$

Explanation

The restriction that the letters be in alphabetical order is equivalent to specifying that only one ordering of any given letters should be counted. This is equivalent to specifying that order does not matter. Hence, the problem is one of unordered sampling, $r = 5$ out of $n = 26$ letters with replacement, *i.e.* **multisets – type I**.

$$\binom{n+r-1}{r} = \binom{26+5-1}{5} = \binom{30}{5}.$$

■

Problem 3c

How many strings of $r = 5$ letters can you form using the $n = 26$ letters in the English language if the letters do *not* have to be in alphabetic order and if letters cannot be repeated?

Answer

$$\frac{26!}{(26-5)!}$$

Explanation

Removing the restriction that the letters must be in alphabetical order is equivalent to specifying that all ordering of any given five letters should be counted. This is, specifying that order does matter. Hence, the problem is one of ordered sampling, $r = 5$ out of $n = 26$ letters without replacement, *i.e.* **permutations – type I**.

■

Problem 3d

How many strings of $r = 5$ letters can you form using the $n = 26$ letters in the English language if the letters do not have to be in alphabetic order and if letters can be repeated?

Answer

26^5

Explanation

Removing the restriction that the letters must be in alphabetical order is equivalent to specifying that all ordering of any given five letters should be counted. This is, specifying that order does matter. Hence, the problem is one of ordered sampling, $r = 5$ out of $n = 26$ letters with replacement, *i.e.* **sequences – type I**.

■

Problem 4a

How many strings of 5 digits can you form using the digits from $\{0,1,\ldots,9\}$ if no digit appears more than once in the string?

Answer

$\frac{10!}{5!}$

Explanation

This problem is one of sampling $r = 5$ times from the set of $n = 10$ distinct elements $\{0,1,\ldots,9\}$ without replacement (we required that no digit appears more than once) and when order important (23 is not the same as 32, for example).

So, this is just **permutations, type I**.

$$\frac{n!}{(n-r)!} = \frac{10!}{(10-5)!}.$$

■

Problem 4b

How many 5 digit numbers can you form using the digits from $\{0,1,\ldots,9\}$ if no digit appears more than once in the number?

Answer

$9 \times 9 \times 8 \times 7 \times 6$

Explanation

The plan is to account for the problem that the first digit cannot be a "0". If we pull apart the answer to part (a) we can see the answer equals $10 \times 9 \times 8 \times 7 \times 6$, which follows from viewing part (a) as a five-step construction held together by the **rule of product**.

Taking this latter approach, we only need to modify that first two steps. Because the first digit cannot be a 0 there are only 9 possibilities for that first digit. The second step is filling in the second digit. Because no digit may appear more than once, we cannot reuse whatever (nonzero) digit was used in Step 1. This leaves us 8 nonzero digits that can be used in Step 2 but don't forget that in Step 2 we are allowed to use the "0". So, there are 9 possible digit choices for Step 2. Continuing, there would be 8, 7 and 6 possibilities respectively for the third, fourth and fifth digits in our number.

■

> In Problem 4a the question asks for "strings of 5 digits" and in 4b the question asks for "5 digit numbers". The difference is that (by convention) 0 is allowed as a first digit in a "string" but not in a "number". Be on the lookout for this subtle difference in wording on contest problems.

Problem 4c

How many 4 digit numbers can you form using only the digits from $\{1,2,3,4,5,6\}$ such that each digit (starting from the right) is at least as large as the previous digit? e.g. 5332 or 5321 or 6664.

Answer

$\binom{9}{4}$

Explanation

This uses the same reasoning we developed in (3b) where counted all possible 5 letter strings when letters could be repeated and when the letters had to be in alphabetic order.

The same exact idea holds in this problem but instead of letters in alphabetic order we have numbers in nonincreasing order. It is still a **multiset – type I** problem with $r = 4, n = 6$.

■

Notice that 0 is not one of the six allowed digits in this problem so a 4 digit number and a string of 4 digits amounts to the same thing.

Problem 5a

Suppose 5 indistinguishable dice are tossed. How many distinguishable outcomes are there? Note: Because the dice are indistinguishable, the results

$\boxed{1}$ $\boxed{3}$ $\boxed{1}$ $\boxed{5}$ $\boxed{6}$ and $\boxed{3}$ $\boxed{5}$ $\boxed{1}$ $\boxed{6}$ $\boxed{1}$

are not considered as distinguishable outcomes.

<u>Answer</u>

$\binom{10}{5}$

<u>Explanation</u>

Think of an urn with balls numbered from 1 to 6. Because the balls are numbered differently, the balls are distinguishable.

Do you need to replace the ball you pull out of the urn each time? Yes. Because it is possible for two (or more) dice to show the same number within the story line of this problem.

Do different orderings of the same set of selected balls make for distinct outcomes? No. This was the point of saying the dice are indistinguishable (*i.e.* identical).

So, this is a **multisets – type I** situation with $n = 6$ (number of distinct objects in the urn) and $r = 5$ (number of with replacement draws from that urn) when order is not important.

$$\binom{n+r-1}{r} = \binom{6+5-1}{5} = \binom{10}{5}.$$

■

Problem 5b

Suppose 5 *distinguishable* (*i.e.* different colored) dice are tossed. How many distinguishable outcomes are there?

<u>Answer</u>

6^5

<u>Explanation</u>

Think of an urn with balls numbered from 1 to 6. Because the balls are numbered differently, the balls are distinguishable.

Do you need to replace the ball you pull out of the urn each time? Yes. Because it is possible for two (or more) dice to show the same number within the story line of this problem.

Do different orderings of the same set of selected balls make for distinct outcomes? Yes. This was the point of saying the dice are distinguishable (*i.e.* different colors).

So, this is a **sequences – type I** situation with $n = 6$ (number of distinct objects in the urn) and $r = 5$ (number of with replacement draws from that urn) when order is important.

$$n^r = 6^5.$$

■

Problem 5c

Suppose 5 *indistinguishable* dice are tossed. How many distinguishable outcomes are there where none of the numbers are repeated?

<u>Answer</u>

$\binom{6}{5}$

Explanation

Think of an urn with balls numbered from 1 to 6. Because the balls are numbered differently, the balls are distinguishable.

Do you need to replace the ball you pull out of the urn each time? No. Because it is not possible for two (or more) dice to show the same number within the story line of this problem.

Do different orderings of the same set of selected balls make for distinct outcomes? No. This was the point of saying the dice are indistinguishable (*i.e.* identical).

So, this is a **combinations – type I** situation with $n = 6$ (number of distinct objects in the urn) and $r = 5$ (number of without replacement draws from that urn) when order is not important.

$$\binom{n}{r} = \binom{6}{5} = \frac{6!}{(6-5)!\ 5!} = 6$$

■

Problem 5d

Suppose 5 *distinguishable* dice are tossed. How many distinguishable outcomes are there where none of the numbers are repeated?

Answer

$$\frac{6!}{(6-5)!}$$

Explanation

Think of an urn with balls numbered from 1 to 6. Because the balls are numbered differently, the balls are distinguishable.

Do you need to replace the ball you pull out of the urn each time? No. Because it is not possible for two (or more) dice to show the same number within the story line of this problem.

Do different orderings of the same set of selected balls make for distinct outcomes? Yes. This was the point of saying the dice are distinguishable (different colors).

So, this is a **permutations – type I** situation with $n = 6$ (number of distinct objects in the urn) and $r = 5$ (number of without replacement draws from that urn) when order is important.

$$\frac{n!}{(n-r)!} = \frac{6!}{(6-5)!} = 6! = 720.$$

■

Problem 6a

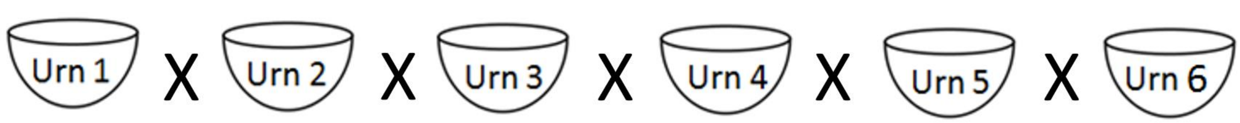
How many arrangements of five "X's" and ten "O's" in a line are there with no consecutive "X's"?

Answer

$$\binom{11}{6}$$

Explanation

In this problem we start by putting the five X's in a line with an urn before and after each X. Then we distribute the identical O's into the six urns with the only restriction being that except for the urn on either end (Urns 1 and 6), no urn can be left empty.



Step 1. Arrange the five X's. There is only 1 distinct way to do this.

Step 2. Distribute the identical O's.

We start by putting a single X in Urns $2, 3, 4,$ and $5.$

Note: Either (or both) of urns 1 and 6 could be left empty without violating the conditions set up in the storyline. But urns 2,3,4 and 5 cannot be left empty.

There is only 1 way to distribute a single X in each of urns 2,3,4 and 5.

The remaining $r = 10 - 4 = 6$ identical O's can be distributed without restriction into the $n = 6$ urns. Distributing $r = 6$ identical objects into $n = 6$ distinguishable urns with any number of objects per urn is a **multisets – type II** situation.

This leads to a total of

$$1 \cdot \binom{n + r - 1}{r} = \binom{6 + 6 - 1}{6} = \binom{11}{6}$$

ways to do Step 2.

Taking Steps 1 and 2 together there are

$$1 \cdot \binom{11}{6} = \binom{11}{6}$$

arrangements of five "X's" and ten "O's" in a line with no consecutive "X's".

■

This problem should look familiar to you. It is just a different way of wording the problem of forming 5 element subsets from $\{1,2, \dots ,15\}$ with no two consecutive integers which we worked on as part of establishing bijections in Chapter 1 on sets.

Problem 6b

How many ways can five children be seated in a row of five chairs if two of the children (Tom and Jerry) refuse to sit next to each other?

Answer

$5! - 4! - 4!$

Explanation

It is easier to count the number of ways where Tom and Jerry do sit next to each other and then subtract this count from the total number of ways to seat the five children without restricting whether Tom and Jerry do or do not sit next to each other. (**Counting Rule for "Not" $\boldsymbol{B}$**).

The total number of unrestricted ways to seat five distinct people in a row of five distinct chairs is 5!. [Distribute five distinct balls (children) into a row of five empty and distinct urns (chairs) with at most ball (child) per urn (chair). This is **permutations – type II** with $n = 5$ and $r = 5$.

So, we get

$$\frac{n!}{(n-r)!} = \frac{5!}{(5-5)!} = \frac{5!}{0!} = \frac{5!}{1} = 5!$$

for the total unrestricted count.

Now we need to count the number of ways where Tom and Jerry do sit next to each other. We can break this into two cases which are connected with "or". Tom sits next to Jerry with Tom on the left of Jerry OR Tom sits next to Jerry with Tom on the right of Jerry.

These two cases partition the set of ways where Tom and Jerry sit next to each other. So, we need to get the count of each case and ADD the two counts. (**Rule of Sum**).

So first count the number of ways where Tom sits to the immediate left of Jerry. To help see how to get the right count imagine that we "glue" Tom and Jerry together so we can think of them as a "single" child that moves together as a single unit.

(Tom, Jerry), (Child 2), (Child 3), (Child 4)
"Child" 1

When we do this we have reduced the problem to just four children that need to be seated, without restriction. But in the above argument we just showed there would be $4!$ ways to seat four children unrestricted.

By the same argument there would be $4!$ ways to seat

(Jerry, Tom), (Child 2), (Child 3), (Child 4).
"Child" 1

So, in total there are $4! + 4!$ ways to seat the five children so that Tom and Jerry <u>do</u> sit next to each other. Hence, there must be

$$5! - 4! - 4!$$

ways to seat the five children if we require that Tom and Jerry <u>do not</u> sit next to each other.

■

> The idea of "gluing" objects together and treating them as a single unit is a common "trick" you need to have in your toolbox.

Problem 6c (MSHSML, 2005-06 Event 5C)

Five pennies made in 2000, 2002, 2004, 2005, and 2006 are to be put in a pile, face up. In how many ways can this be done so that the 2005 and 2006 pennies are not touching each other?

<u>Solution</u>

$$N(\text{piles without restrictions}) - N(\text{piles with "glued" 2005 and 2006 together})$$

$$= 5! - (4! + 4!)$$

This is just the Tom and Jerry problem with new names.

Problem 7a

How many ways are there to distribute r identical balls into n distinguishable boxes with exactly k balls in the first box?

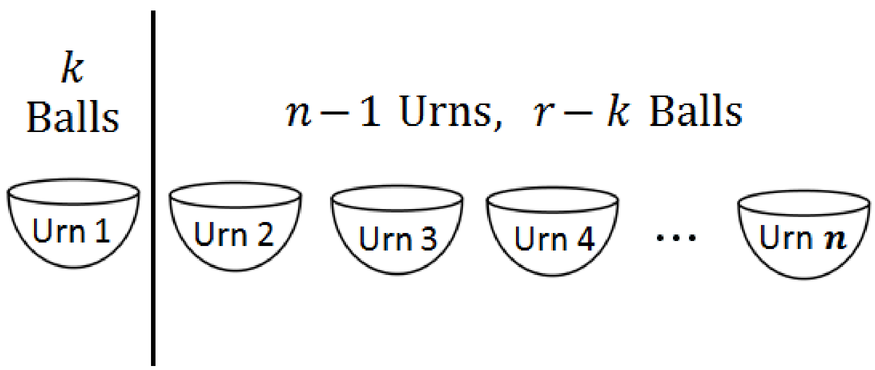
Answer

$$\binom{n+r-k-2}{r-k}$$

Explanation

Step 1. Distribute k identical balls in the first urn, unrestricted

Step 2. Distribute the remaining $r-k$ balls into the remaining $n-1$ urns, without restriction.



Both steps are **multisets – type II** situations. The multisets – type II count formula is

$$\binom{\text{\# urns} + \text{\# balls} - 1}{\text{\# balls}}.$$

In Step1, intuitively there is only 1 distinct way to put k identical balls into Urn 1. Plugging into our **multisets – type II** formula we do indeed get 1, as suspected.

$$\binom{\text{\# urns} + \text{\# balls} - 1}{\text{\# balls}} = \binom{1+k-1}{k} = \binom{k}{k} = 1.$$

In Step 2, we see the count is

$$\binom{\text{\# urns} + \text{\# balls} - 1}{\text{\# balls}} = \binom{(n-1)+(r-k)-1}{r-k} = \binom{n+r-k-2}{r-k}.$$

Thus, the final count is

$$1 \cdot \binom{n+r-k-2}{r-k} = \binom{n+r-k-2}{r-k}.$$

■

Problem 7b

How many ways are there to distribute r identical balls into n distinguishable boxes with at least one ball per box?

Answer

$\binom{r-1}{n-1}$

Explanation

Step 1. Distribute exactly one ball in each of the n urns.

and then do

Step 2. Distribute the remaining $r-n$ balls into any of the n urns unrestricted.

There is only 1 way to do Step 1. Step 2 is a **multisets – type II** situation with $r-n$ identical balls being distributed into n urns without restriction.

This gives a total of

$$\text{Step One} \times \text{Step Two} = 1 \cdot \binom{\text{\# urns} + \text{\# balls} - 1}{\text{\# balls}} = \binom{n+(r-n)-1}{r-n} = \binom{r-1}{r-n}.$$

However, we know from the symmetry property of binomial coefficients that for general positive integers $a \geq b$,

$$\binom{a}{b} = \binom{a}{a-b}.$$

Thus,

$$\binom{r-1}{r-n} = \binom{r-1}{(r-1)-(r-n)} = \binom{r-1}{n-1}.$$

■

There is nothing wrong or unsimplified with the answer in the form $\binom{r-1}{r-n}$ but in contests where you are given answers to choose from you should be prepared to switch from $\binom{a}{b}$ to $\binom{a}{a-b}$, or vice-versa, to see if your answer matches one of the given choices.

Also, this formula pops up often enough in MSHSML, AMC 10/12 and AIME problems that you might want to commit it to memory.

Problem 7c

Suppose 10 indistinguishable dice are tossed. How many distinguishable outcomes are there for which each of the integers $1,2,\ldots,6$ appears on at least one die?

Answer

$\binom{9}{5}$

Solution

Imagine a row of six (empty) urns labeled "1's", "2's",..., "6's" and 10 identical balls. If there are a_1 "1's" among the 10 indistinguishable dice we toss, then we will put a_1 balls into the "1's" urn, if there are a_2 "2's" among the 10 indistinguishable dice we toss, then we will put a_2 balls into the "2's" urn, etc.

We distribute identical balls because while we want to keep track of *how many* of the 10 dice are "1's", *how many* of the 10 dice are "2's", … but we do not want to distinguish *which* of the 10 rolls of the dice are the "1's", the "2's", … because the 10 dice are indistinguishable.

Thus, the problem is equivalent to putting 10 identical balls into 6 distinguishable boxes with at least one ball per box.

Hence, this is just the special case with $r = 10$ and $n = 6$ of Problem 7b.

$$\binom{r-1}{n-1} = \binom{10-1}{6-1} = \binom{9}{5}.$$

■

Problem 8a

There are 8 people in a room. Everybody shook hands with everybody else exactly once. How many handshakes were there in all?

Answer

$\binom{8}{2}$

Solution

We can construct all possible handshakes by finding all possible selections of 2 people from a group of 8 distinct people when sampling without replacement (nobody is shaking their own hand) and when order is not important (Bob and Sally shaking hands is not distinct from Sally and Bob shaking hands).

So, this is just **combinations – type I**.

$$\binom{8}{2}$$

■

Problem 8b (MSHSML, 2011-12 State Tournament, Invitational Event)

A total of 28 handshakes were exchanged at the conclusion of a party. Assuming that each person shook hands exactly once with each of the others, how many people were present?

Solution

Generalizing the idea of part (a) of this problem there are

$$\binom{n}{2} = \frac{n!}{2!\,(n-2)!} = \frac{n(n-1)}{2}$$

possible handshakes between n people at a party. So,

$$\binom{n}{2} = \frac{n(n-1)}{2} = 28.$$

This gives us the quadratic equation $n^2 - n - 48 = 0$ which has solutions $n = 8$ and $n = -7$. The number of people at the party cannot be -7 so the only possibility is $n = 8$.

■

Problem 9a

How many different ways can 52 distinguishable objects (for example the 52 cards in a standard deck of cards) be distributed among 4 people if each person gets exactly 13 objects (for example, 13 cards)?

Answer

$$\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} = \frac{52!}{13!\,13!\,13!\,13!}$$

Solution

Method 1

The standard way to deal out a deck of cards is to go around in a circle giving each person one card at a time and continuing to the next person in the circle over and over until everybody had their 13 cards. But it would come to the same thing if you were to deal the first person all 13 of their cards at once and then went to the next person and gave them their 13 cards, etc.

The first person is selecting 13 cards from a deck (urn) with 52 distinguishable cards (balls). They are selecting without replacement (they cannot be given the same card twice) and where order is not important (in the context of hands of cards they order of the cards in your hand is irrelevant). This is **combinations, type I**. So, there are

$$\binom{52}{13}$$

ways for the first person to get their hand of 13 cards. At this point we move on to the next person (*i.e.* Step 2). There would only be 39 cards left so there are

$$\binom{39}{13}$$

ways for the second person to get their hand of 13 cards.

Clearly, what cards the second person gets does depend on what cards the first person got. But the number of different ways to give the second person their 13 cards does not depend on what cards the first person got.

That is, Step 1 and Step 2 are count independent steps and so the Rule of Product will apply here.

Continuing in this way, there would be

$$\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}$$

different ways to deal out the 52 cards. However, this product simplifies. We note that

$$\begin{aligned}\binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13} &= \frac{52!}{13!\,39!}\cdot\frac{39!}{13!\,26!}\cdot\frac{26!}{13!\,13!}\cdot\frac{13!}{13!\,0!} \\ &= \frac{52!}{13!\,13!\,13!\,13!}.\end{aligned}$$

Method 2

This is just an example of distributing $r = 52$ distinct objects (cards) into $R = 4$ distinguishable boxes (people) such that $t_{13} = 4$ of these boxes get 13 objects (cards) and $t_j = 0$ for all $j \neq 13$.

Plugging these numbers into our general count formula for distributing distinct objects into distinguishable boxes with given sizes we have

$$\frac{r!}{(1!)^{t_1}\cdots(r!)^{t_r}\cdot(t_1!\,t_2!\cdots t_r!)}\cdot\frac{R!}{t_0!} = \frac{52!}{(13!)^4\cdot 4!}\cdot\frac{4!}{0!} = \frac{52!}{13!\,13!\,13!\,13!}.$$

■

Problem 9b

Suppose you have a "Lucky Charms" deck of 52 playing cards consisting of 13 identical pink heart cards, 13 identical yellow moon cards, 13 identical orange star cards and 13 identical green clover cards. How many different decks can make?

Answer

$$\frac{52!}{13!\,13!\,13!\,13!}$$

Solution

Arranging the cards as a deck is just another way of describing the cards arranged in a long row. Thus, we can model this problem as a special case of the general problem of permutations with repeated letters. In this case we have

$a_1 = 13$ H's (hearts)

$a_2 = 13$ M's (moons)

$a_3 = 13$ S's (stars)

$a_4 = 13$ C's (clovers).

So, from our general formula

$$\frac{(a_1 + a_2 + \cdots + a_r)!}{a_1!\, a_2! \cdots !\, a_r!}$$

there are

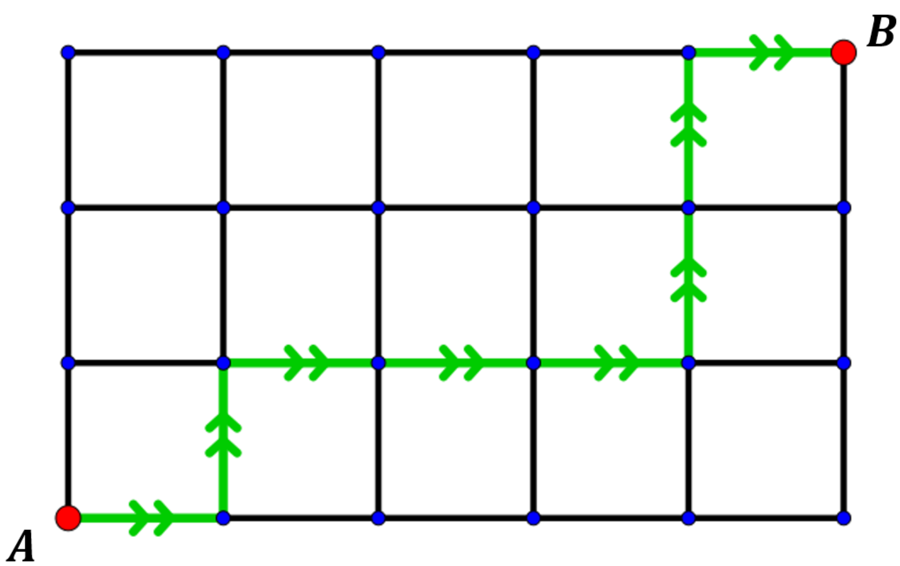
$$\frac{(13 + 13 + 13 + 13)!}{13!\, 13!\, 13!\, 13!} = \frac{52!}{13!\, 13!\, 13!\, 13!}$$

different ways can you arrange these 52 cards in a line (deck).

■

Problem 10

How many paths are there from point A located at grid point $(0,0)$ to the grid point B located at $(5,3)$ if you are only allowed to move up or to the right? One of the many possible paths is shown below in green.



Answer

$\binom{8}{5}$

Solution

We can denote a rightward direction step by R and an upward step by U. So the above path, which is (Right, Up, Right, Right, Right, Up, Up, Right) would be denoted by RURRRUUR. But to get from A to B (under the rules that you cannot come back left or come back down) your only option is to go Right five times and Up three times in some order.

That is, every path from A to B can be matched with some arrangement of 5 R's and 3 U's. But from our work in Problem 9b that there would be

$$\frac{8!}{5!\,3!} = \binom{8}{5}$$

ways to arrange 5 R's and 3 U's.

∎

Problem 11a

How many ways "full-houses" can be formed using a standard deck of 52 cards?

<u>Discussion</u>

A "full house" is a term used in the card game of poker. For those unfamiliar with a "standard" deck of cards and/or the game of poker, let me define my terms.

Each card in a deck of 52 cards has a "value" and a "suit". There are thirteen "values" in a deck which have the names Aces, 2's, 3's, 4's, … , 10's, Jacks, Queens, and Kings. Each value has one of four suits which have the names Spades, Hearts, Diamonds and Clubs. These suits are also identified by symbols which are ♠, ♥, ♦, and ♣ respectively. This gives us our $13 \times 4 = 52$ cards. Any of the 26 cards with either a space ♠ or club ♣ suit is referred to as a black card and any of the 26 cards with either a heart ♥ or diamond ♦ suit is referred to as a red card

A "full house" is a set of 5 cards from a single standard deck that is made up of exactly two values, where one value occurs two times and the other value occurs three times. Notice we used the term "set" of 5 cards on purpose. The re-ordering the same 5 cards does not make for different "full houses".

The named "hands" (a set of 5 cards) in poker are

Bust Any hand which is not one of the other nine named hands described below.	A♦ 7♥ 5♣ 3♦ 2♠
Pair This hand (a set of five cards) contains exactly four different values.	A♣ A♦ 9♥ 6♠ 4♦
Two Pair This hand contains exactly three different values with two of the values occurring twice.	K♥ K♠ J♣ J♦ 9♦
Three of a Kind This hand contains exactly three different values with one value occurring three times.	Q♠ Q♥ Q♦ 5♠ 9♣

Straight The five cards in a straight must be sequential and not *all* of the same suit. The sequence of values are A, 2,3,4, … ,10,J,Q,K,A. Notice the exception that an Ace ("A") can be at either end of the sequence. Also, a straight cannot "wrap around". For example, the five cards Q♠, K♠, A♥, 2♦, 3♦ is <u>not</u> a straight.	
Flush The five cards in a flush all have the same suit but are not sequential.	
Full House This hand contains exactly two values with one of the values occurring exactly three times.	
4 of a Kind This hand contains exactly two values with one of the values occurring four times.	
Straight Flush The five cards have the same suit and are sequential but the highest card in the sequence is not an Ace.	
Royal Flush The five cards have the same suit and are sequential and the highest card in the sequence is an Ace.	

These ten "named" hands are listed in decreasing order of how many there are. That is, there are more "busts" than "pairs" and more "pairs" than "two pairs", etc. For this reason, when playing poker we say that a pair "beats" a bust, two pair "beats" a pair, three of a kind "beats" two pair, etc.

Now you have all the terms and information you need to solve the problem of counting the number of full houses.

<u>Solution</u>

Full House This hand contains exactly two values with one of the values occurring exactly three times.	

We can construct a full house in five steps.

First do

Step 1. Pick 2 of the 13 values

and then do

Step 2. Pick 1 of these 2 values and triple it

and then do

Step 3. Pick 1 of the remaining 1 unpicked values and double it

and then do

Step 4. Pick 3 of the 4 suits to attach to the value which was tripled

and then do

Step 5. Pick 2 of the 4 suits to attach to the value which was doubled.

Why are these five steps count independent? Can you work this out clearly in your head yet?

Why are each of these steps modeled by **Combinations-Type I** (sampling without replacement, order not important)? Is this obvious to you by this point?

$$\binom{13}{2}\binom{2}{1}\binom{1}{1}\binom{4}{3}\binom{4}{2} = 78 \times 2 \times 1 \times 4 \times 6 = 3744.$$

■

Problem 11b

How many ways “pairs” can be formed using a standard deck of 52 cards?

<u>Solution</u>

Pair This hand (a set of five cards) contains exactly four different values.	

First do

Step 1. Pick 3 of the 13 values

and then do

Step 2. Pick 1 of these 3 values and double it

and then do

Step 3. Pick 2 of the 4 suits to attach to the value which was doubled

and then do

Step 4. Pick 1 of the 4 suits to attach to a value which occurs once

and then do

Step 5. Pick 1 of the 4 suits to attach to a value (other than the value in Step 4) which occurs once

and then do

Step 6. Pick 1 of the 4 suits to attach to a value (other than the value in Step 4 or Step 5) which occurs once.

$$\binom{13}{3}\binom{3}{1}\binom{4}{2}\binom{4}{1}\binom{4}{1}\binom{4}{1} = 286 \times 3 \times 6 \times 4^3 = 329472.$$

■

Problem 12 (Brilliant - Math and science done right, https://brilliant.org/)

Four distinct numbers are randomly chosen from the set $\{1,2,3,\dots,14\}$. What is the probability that there exists some pair of the selected numbers that are consecutive?

Solution

Let S be the set of all ways to select 4 distinct numbers from the set $\{1,2,3,\dots,14\}$. Let $A \subset S$ be the event (subset of S) where some pair of the 4 chosen numbers are consecutive. It is tempting to take the combinatorial probability formula

$$P(A) = \frac{N(A)}{N(S)}$$

and start counting. But first check the necessary condition – is there some reason to know that every element in S has an equal probability of occurring as an outcome?

Yes! The phrase "randomly chosen". This is the standard shorthand for "assume all outcomes in the sample space are equally likely (probable)".

So, are we ready to start counting? Well, sort of. Something seems to be missing. We are told we are sampling without replacement, but should we assume that order matters (permutations-type I) or that order does not matter (combinations-type I)? The statement of the problem does not say one way or the other.

This was not an oversight by the problem poser. Remember that when proving the formulas for permutations and combinations we saw that

$$\text{\# permutations} = \text{\# combinations} \quad \times \quad r!.$$

So, ***in this problem*** if you solve the problem under the assumption that order matters then both your numerator and denominator are going to include a factor of $r! = 4!$ which will cancel out and leave you with the solution you would get if you solved the problem under that assumption that order does not matter.

So, you end up with the same final answer under both options!

Contest Trick of the Trade – Missing Vital Information

I've come across other contest problems, like this one, where some "vital" piece of information is not provided. If we assume the problem poser has not made a mistake, then it must be the case that the final answer will not depend on which option you pick or what value you take for that vital piece of information.

In such cases, ***always*** pick the option or pick a value for the variable that makes solving the problem as easy as possible.

But this trick of the trade comes with a **big caveat**. What if you overlooked how this vital piece of information can in fact be determined from the given information? Then just picking a convenient option or parameter value is not a valid approach and will most likely lead to the wrong final answer.

Not that it matters, but I will give the solution in terms of **combinations-type I**. That is, sampling without replacement and assuming that order does not matter.

It follows this assumption that

$$N(S) = \binom{14}{4} = 1001.$$

We've taken $A \subset S$ as the set of selections that *do* contain *some* pair of numbers that are consecutive.

It follows that $\overline{A} \subset S$ would be the set of selections where *none* of the selected numbers are consecutive. Now "Counting Rule for Not B" from Chapter 1 tells us that

$$N(A) = N(S) - N(\overline{A}).$$

So, we have two routes for finishing this problem

$$P(A) = \frac{N(A)}{\binom{14}{4}} \quad \text{or} \quad P(A) = \frac{\binom{14}{4} - N(\overline{A})}{\binom{14}{4}}.$$

Which route will be easier, finding $N(A)$ or finding $N(\overline{A})$? It can be critical on a contest where the time crunch is huge that you think this through because in many problems (including this one) one route is much easier than the other.

Knowing ahead of time during a contest which route will be easier often comes down to having gone through the agony of being bogged down on the "wrong" route when doing practice problems! "Practice make perfect" is more than another glib remark.

There are no absolutes when it comes to picking the easiest route, only some “rule of thumbs”. Look for that route where the rules more tightly control what can happen. In this problem, let $a < b < c < d$ be the four numbers we pick and consider the possibilities for the vector

(Are a and b consecutive?, Are b and c consecutive?, Are c and d consecutive?).

If we go the route of finding $N(A)$ then we have to account for each of the seven possible answers

(Yes, Yes, Yes), (Yes, Yes, No), (Yes, No, Yes), (No, Yes, Yes)

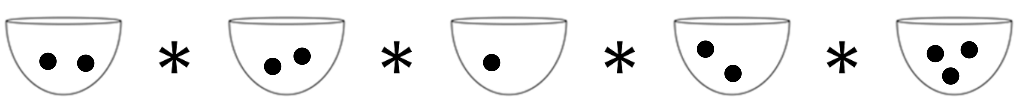
(Yes, No, No), (No, Yes, No), (No, No, Yes).

But if we go the route of finding $N(\overline{A})$ then there is only one possible answer (No, No, No) we have to consider. So, this appears to be the better route to take.

You will get much more practice on having to decide “what’s the best route” when we get to Chapter 4 Advanced Inclusion-Exclusion.

The key to this problem is figure out a way to translate it into a Type II (balls into urns) problem.

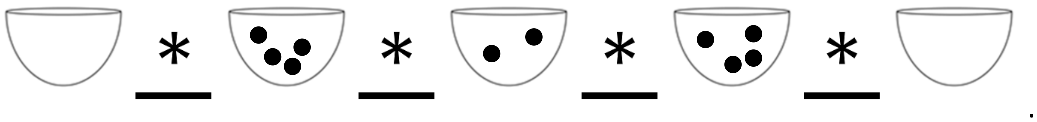
In the diagram below let a $*$ represent a number that we *do* pick and let a ● represent a number that we *do not* pick. Using just these symbols we could (unambiguously) represent any pick of four numbers. For example, consider the pick $\{3,6,8,11\}$. It becomes



i.e.

1,2	**3**	4,5	**6**	7	**8**	9,10	**11**	12,13, 14

And the pick $\{1,6,9,14\}$ is represented as



i.e.

	1	2,3, 4,5	**6**	7,8	**9**	9,10, 11,12 13	**14**	

How would you represent the pick $\{2,5,10,13\}$?

Can you see how to make a bijection between a set of four numbers chosen without replacement from $\{1,2,3,\ldots,14\}$ such that no two numbers are consecutive and a distribution of four $*$'s and ten ●'s? That is, does each set of four numbers map to one and only one urn model representation and vice versa? Why?

What rule(s) do we need on the distribution of the ten ●'s into the five urns to guarantee that the four chosen numbers represented by the $*$'s are not adjacent?

We must require that each of the three "interior" urns has at least one ●.

Here is the three-step construction we need.

- Step 1. Put a single $*$ in each of the four slots
- Step 2. Put a single ● in each of the three interior urns
- Step 3. Distribute the remaining $10 - 3 = 7$ ●'s in any of the five urns without any restrictions on the number of ●'s per urn.

There is only one distinguishable way to put a single $*$ in the four slots and only one distinguishable way to put a single ● in each of the three interior urns.

This just leaves Step 3. How can we model this? Clearly it is a Type II situation of putting balls into urns.

Are the balls distinguishable or indistinguishable? Indistinguishable.

Are the urns distinguishable or indistinguishable? Distinguishable. (Leaving the first urn empty represents a different situation than leaving the second urn empty, for example.)

Are you limited to at most one ball per urn or can any number of balls go into any urn? Any number.

We have just described **Multisets-Type II**. So, there are

$$\binom{n+r-1}{r} = \binom{5+7-1}{7} = \binom{11}{7} = 330$$

distinct ways to accomplish Step 3.

It is also clear that these steps are count independent by default because no matter which of the 1 way to do Steps 1 and 2 you do there are the still 330 ways to do Step 3.

Hence,

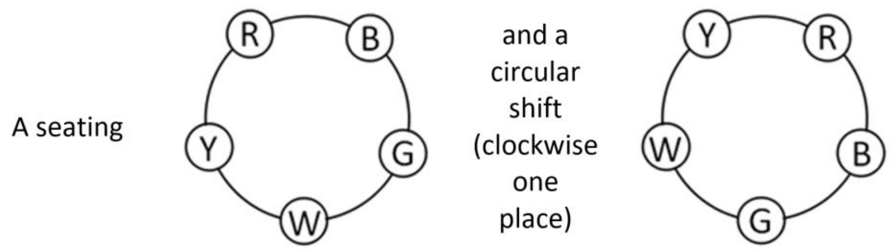
$$P(A) = \frac{\binom{14}{4} - \binom{11}{7}}{\binom{14}{4}} = \frac{1001 - 330}{1001} = \frac{61}{91}.$$

■

Permutations on a Circle, All Objects Distinct

Problem 13a

How many ways can n people be seated at a round table with n chairs? Assume that circular shifts are NOT distinguishable (and hence should not be part of the count more than once).



Answer

$$\frac{n!}{n} = (n-1)!$$

Solution

Suppose for a moment that we did want to count circular seatings as distinct. Then the problem would reduce to just counting the total number of ways of seating n people into n distinct chairs. And we know there would be $n!$ ways of doing this.

But we could create all possible seatings at a round table (continuing to assume for the moment that circular seatings are distinct) in a two step procedure.

Step 1. Construct the distinct seatings when circular shifts are not distinct.

Step 2. Construct all possible circular shifts of each seating from Step 1.

Clearly these two steps are count independent (the number of possible circular shifts does not depend on how the people where seated in the first step).

So $n! = N(\text{Step 1}) \cdot N(\text{Step 2})$. Now there are n possible circular shifts at a table with n seats.

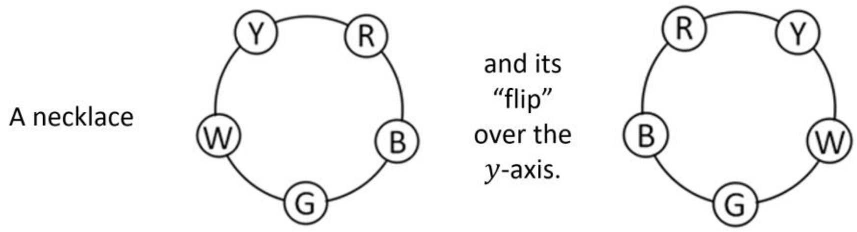
Therefore,

$$\begin{aligned} N(\text{Step 1}) &= N(\text{all distinct seatings when circular shifts are not counted}) \\ &= \frac{n!}{N(\text{Step 2})} = \frac{n!}{n} = (n-1)!. \end{aligned}$$

■

Problem 13b

How many ways are there to arrange n distinct beads on a necklace if we assume that circular shifts and “flips” are NOT distinguishable.



<u>Answer</u>

$$\frac{(n-1)!}{2}$$

<u>Solution</u>

To understand what is meant by a “flip”, imagine a necklace of five colored beads in the clockwise order red→blue→green→white→yellow→red. Then the five beads of this necklace would have counterclockwise order red←blue←green←white←yellow←red.

The “flip” switches the clockwise and counterclockwise orderings. So for this necklace, the flip would have clockwise order red→yellow→white→green→blue→red and counterclockwise order red←yellow←white←green←blue←red.

So, the above count for seating people at a round table would be off by a factor of 2. Therefore, there are $(n-1)!/2$ ways to arrange n distinct beads on a necklace if we assume that circular shifts (*e.g.* shifting each object clockwise one position) and “flips” are NOT distinguishable.

■

3 Binomial and Multinomial Theorems

3.1 Binomial Theorem

Here is a common problem to see on MSHSML exams:

"What is the coefficient of x^4y^5 if you expand $(x+y)^9$ into all its terms?"

The obvious approach is to expand the binomial $(x+y)^9$ and read off the requested coefficient. But how?

The last thing you want to do is to find

$$(x+y)(x+y)(x+y)(x+y)(x+y)(x+y)(x+y)(x+y)(x+y)$$

by actually going through all the multiplication. Clearly it would be a big mess and would take forever.

What is the shortcut? The process of expanding

$$(x+y)(x+y)(x+y)(x+y)(x+y)(x+y)(x+y)(x+y)(x+y)$$

comes down to going through all ways of selecting a term from the first factor, a term from the second factor, a term from the third factor, etc. and multiplying them together.

So x^4y^5 results every time we select the x from 4 of the factors and the y from 5 of the factors.

Selecting x from the first four factors and y from the last five factors gives us

$$x\,x\,x\,x\,y\,y\,y\,y\,y = x^4y^5$$

while

$$x\,y\,x\,y\,y\,y\,x\,y\,x = x^4y^5$$

results from selecting x from the $1st, 3rd, 7th$ and $9th$ factors and y from the $2nd$, $4th$, $5th$, $6th$ and $8th$ factors.

We can see there is a bijection (one to one and onto correspondence) between every arrangement of four x's and five y's and each way of getting x^4y^5 when we expand out $(x+y)^9$.

That is, the coefficient of x^4y^5 in $(x+y)^9$ is just the number of ways of arranging four x's and five y's.

But we verified in the previous section "Arrangements with Repeated Letters" in Chapter 2 that there are

$$\frac{9!}{4!\,5!} = \binom{9}{4}$$

such arrangements. So, the coefficient of x^4y^5 in $(x+y)^9$ is

$$\frac{9!}{4!\,5!} = \binom{9}{4}.$$

Following in this manner for all coefficient of $(x+y)^9$ we find

$$(x+y)^9 = \binom{9}{0}x^0y^9 + \binom{9}{1}x^1y^8 + \cdots + \binom{9}{4}x^4y^5 + \cdots + \binom{9}{9}x^9y^0.$$

We get the binomial theorem by generalizing this result to an arbitrary non-negative integer n.

Binomial Theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

for all real number x and y and all non-negative integers n.

Historical Aside: Around the year 1665, Sir Isaac Newton generalized the binomial theorem to allow for all real n. So, for example, Newton was able to find the coefficients a_k such that

$$(1+x)^{1/2} = \sum_{k=0}^{\infty} a_k x^k.$$

This turned out to be a key step in Newton's development of calculus!

■

3.2 Multinomial Theorem

What is the pattern if we consider the expansion of $(x+y+z)^9$?

Just as we noted in the binomial expansion, the process of expanding

$$(x+y+z)(x+y+z)(x+y+z)\cdots(x+y+z)$$

comes down to going through all ways of selecting a term from the first factor, a term from the second factor, a term from the third factor, etc. and multiplying them together.

So $x^2y^3z^4$ results every time we select the x from 2 of the factors, the y from 3 of the factors and the z from 4 of the factors.

As in the binomial expansion, there is a bijection between every arrangement of two x's, three y's and four z's and each way of getting $x^2y^3z^4$ when we expand out $(x+y+z)^9$.

That is, the coefficient of $x^2y^3z^4$ in $(x+y+z)^9$ is just the number of ways of arranging 2 x's, three y's and four z's in a line.

But from our work on "Arrangements with Repeated Letters" in Chapter 2 we know that there are

$$\frac{9!}{2!\,3!\,4!}$$

such arrangements. So, the coefficient of $x^2y^3z^4$ in $(x+y+z)^9$ is

$$\frac{9!}{2!\,3!\,4!}.$$

Going through this same argument for an arbitrary non-negative integer value for n gives us the multinomial theorem.

Multinomial Theorem

$$(x_1+x_2+\cdots+x_r)^n = \sum_{\substack{(n_1,\ldots,n_r):\\ n_1+\cdots+n_r=n\\ n_j\in\{0,1,2,\ldots\}\\ j=1,2,\ldots,r}} \frac{n!}{n_1!\,n_2!\cdots n_r!}x_1^{n_1}x_2^{n_2}\cdots x_r^{n_r}$$

Notation:

$$\binom{n}{n_1,n_2,\ldots,n_r} = \frac{n!}{n_1!\,n_2!\cdots n_r!}$$

is often used for multinomial coefficients.

Counting Distinct Terms

We see that the sum is over all nonnegative integer-valued vectors, $(n_1,n_2,\ldots,n_r)$, such that $n_1+n_2+\cdots+n_r=n$.

Now suppose you have n identical balls to distribute among r distinct urns without any restrictions on the number of balls per urn. Then each value of the vector $(n_1, n_2, \dots, n_r)$ is a different way of making this distribution of the n balls.

That is, there is a bijection between the number of terms in this expansion and the number of distinct ways to distribute n identical balls into r distinct urns with any number of balls per urn. But we recognize this as the **multisets-type II model**.

> Be careful with how the parameters n and r are used here! When we defined multisets-type II in Chapter 2 we took n as the number of distinct *urns* and r as the number of identical *balls* which is backwards from how they are defined in this problem. To avoid such confusion in the future it is best to remember the multisets-type II formula as
>
> $$\binom{\text{\# urns} + \text{\# balls} - 1}{\text{\# balls}}.$$

Hence there are

$$\binom{r+n-1}{n}$$

distinct terms in the multinomial expansion of $(x_1 + x_2 + \cdots + x_r)^n$ *provided no pair of variables among* $x_1, x_2, \dots, x_r$ *have a common factor*.

3.3 Combinatorial Identities

We are just providing a list without proofs of the most commonly needed combinatorial identities for AMC and AIME problems. You should familiarize yourself with these identities.

(1)	$\binom{n}{k} = \binom{n}{n-k}$
(2)	$\binom{n-1}{k} + \binom{n-1}{k+1} = \binom{n}{k+1}$

(3)	$\sum_{k=0}^{n} \binom{n}{k} = 2^n$
(4)	$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0$
(5)	$\sum_{k=m}^{n} \binom{k}{m} = \binom{n+1}{m+1}$
(6)	$\sum_{k=0}^{n} \binom{m+k}{k} = \binom{n+m+1}{n}$

3.4 Practice Problems with Solutions and Discussion

Problem 1a. (MSHSML, Event 5C, 2006-07)

In the expansion of $(x+y)^{15}$, what is the coefficient of x^9y^6?

Answer

$$\binom{15}{9}$$

Solution

By the Binomial Theorem

$$(x+y)^{15} = \sum_{j=0}^{15} \binom{15}{j} x^j y^{15-j}.$$

So, the coefficient of x^9y^6 in $(x+y)^{15}$ is

$$\binom{15}{9} = \frac{15!}{9!\,(15-9)!} = 5005.$$

■

Problem 1b. (MSHSML, Event 5C, 2015-16, State Tournament, Event C)

Calculate the coefficient of the x^2 term in $(5x-1)^7$.

Answer

$$\binom{7}{2} 5^2(-1)^5 = 21 \cdot 25 \cdot (-1) = -525$$

Solution

$$(5x-1)^7 = \sum_{j=0}^{7} \binom{7}{j} (5x)^j(-1)^{7-j}$$

The only term in the above with x^2 is when $j = 2$. In this case we have

$$\left(\binom{7}{2} 5^2(-1)^5\right) \cdot x^2.$$

So, the coefficient of x^2 in $(5x-1)^7$ is

$$\binom{7}{2} 5^2(-1)^5 = 21 \cdot 25 \cdot (-1) = -525.$$

■

Problem 1c.

Find the coefficient of $x^2y^3z^5$ in $(x + 4y + z)^{10}$.

Solution

By the multinomial theorem

$$(x+4y+z)^{10} = \sum \cdots \sum \frac{10!}{j_1!\,j_2!\,j_3!} x^{j_1}(4y)^{j_2}z^{j_3}$$
$$= \sum \cdots \sum \left(\frac{10!}{j_1!\,j_2!\,j_3!} 4^{j_2}\right) x^{j_1}y^{j_2}z^{j_3}$$

where are sum is over all vectors (j_1, j_2, j_3) such that $j_1 + j_2 + +j_3 = 10$ and $j_1 \in \{0,1,2,\dots,10\}, j_2 \in \{0,1,2,\dots,10\}, j_3 \in \{0,1,2,\dots,10\}$.

So, the coefficient of $x^2y^3z^5$ in $(x+4y+z)^{10}$ is

$$\frac{10!}{2!\,3!\,5!} 4^3.$$

■

Problem 1d.

Find the coefficient of $x^6y^6z^6$ in $(x^2+y^3+z)^{11}$.

<u>Solution</u>

By the multinomial theorem

$$(x^2+y^3+z)^{11} = \sum \cdots \sum \frac{11!}{j_1!\,j_2!\,j_3!} (x^2)^{j_1}(y^3)^{j_2}z^{j_3}$$
$$= \sum \cdots \sum \frac{11!}{j_1!\,j_2!\,j_3!} x^{2j_1}y^{3j_2}z^{j_3}$$

where the sum is over all vectors (j_1, j_2, j_3) such that $j_1 + j_2 + +j_3 = 11$ and $j_1 \in \{0,1,2,\dots,10\}, j_2 \in \{0,1,2,\dots,10\}, j_3 \in \{0,1,2,\dots,10\}$.

The term in the above sum with $j_1 = 3, j_2 = 2$ and $j_3 = 6$ (note that $j_1 + j_2 + j_3 = 11$) gives us

$$\frac{11!}{3!\,2!\,6!} x^{2(3)} y^{3(2)} z^6 = \frac{11!}{3!\,2!\,6!} x^6 y^6 z^6.$$

So, the coefficient of $x^6 y^6 z^6$ in $(x^2 + y^3 + z)^{11}$ equals

$$\frac{11!}{3!\,2!\,6!}.$$

■

Problem 1e.

What is the constant term in $\left(x^2 - \frac{1}{x}\right)^6$?

<u>Solution</u>

$$\begin{aligned}\left(x^2 - \frac{1}{x}\right)^6 &= \sum_{n=0}^{6} \binom{6}{n} (-1)^{6-n} (x^2)^n \left(\frac{1}{x}\right)^{6-n} \\ &= \sum_{n=0}^{6} \binom{6}{n} (-1)^{6-n} x^{2n-(6-n)} = \sum_{n=0}^{6} \binom{6}{n} (-1)^{6-n} x^{3n-6}\end{aligned}$$

We get a constant term (*i.e.* no x's) whenever we have x^0. That is, when $3n - 6 = 0$. This only happens for $n = 2$. The coefficient of x^0 is therefore,

$$\binom{6}{2} (-1)^{6-2} = 15.$$

■

4 Method of Inclusion-Exclusion

4.1 Union of Three Sets

Previously we established the inclusion-exclusion formula for two sets.

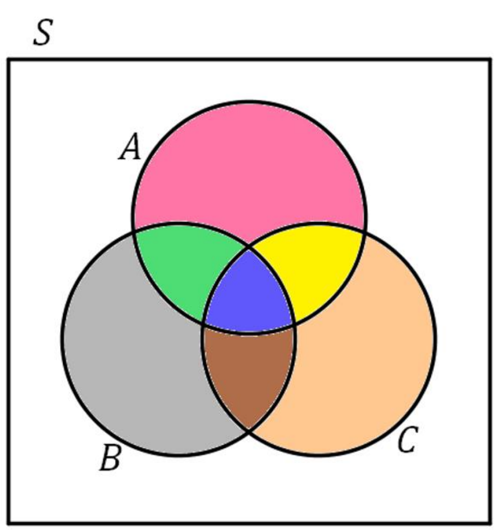
$$N(A \cup B) = N(A) + N(B) - N(A \cap B).$$

This result might have left you wondering if there is a similar result for $N(A \cup B \cup C)$. It turns out there is. In particular, it turns out that

$$\begin{aligned} N(A \cup B \cup C) = \quad & + \quad \big(N(A) + N(B) + N(C)\big) \\ & - \quad \big(N(A \cap B) + N(A \cap C) + N(B \cap C)\big) \\ & + \quad N(A \cap B \cap C). \end{aligned}$$

Notice that in the middle row where we have "two-way" intersections this formula does *not* include $N(A \cap A)$ and it does *not* both $N(A \cap B)$ and $N(B \cap A)$. The middle row of "two-way" intersections includes the counts of all possible intersections of two sets taken without replacement and when order is not important from A, B, C, D and E.

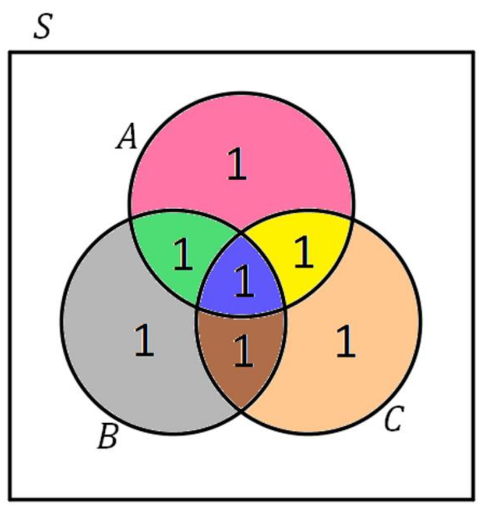
A formal algebraic proof is straightforward but not very enlightening. The following pictorial justification is more revealing of what is really going on.



$N(A \cup B \cup C)$

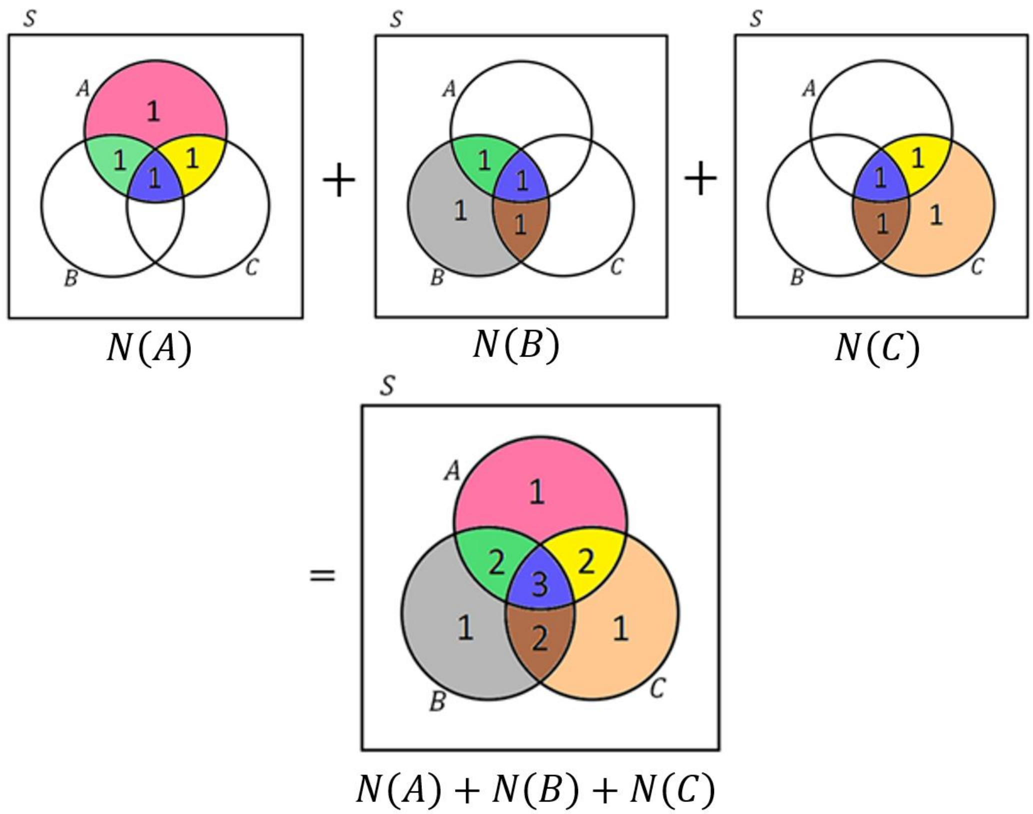
The cardinality of the union of sets A, B, and C, that is $N(A \cup B \cup C)$, is the sum of the cardinalities of the seven disjoint and different colored regions (the grey, green, pink, blue, brown, yellow and tan) shown above. Remember from our Chapter 1 notes that by the **Rule of Sum** the cardinality of the *union* of *disjoint* regions is the sum of the cardinalities of each region. But what if sets, such as sets A, B and C above, are *not* disjoint?

In order to avoid double counting or skipping, we need to make sure we add the cardinalities of each of these seven regions once and only once.



$N(A \cup B \cup C)$

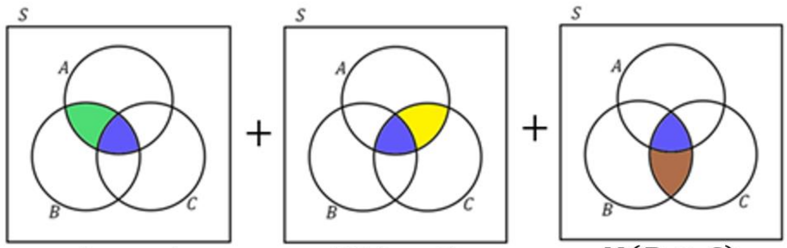
Suppose we consider just the top row, namely $N(A) + N(B) + N(C)$, of the formula given above for $N(A \cup B \cup C)$. Pictorially, $N(A) + N(B) + N(C)$ looks like



In the above, we can see that the count of the pink region got included just once (in $N(A)$). Similarly, that the counts of the grey and tan regions are included just once. However, the green region is included twice (in $N(A)$ and in $N(B)$). Similarly, the yellow and brown regions are included twice. Finally, the cardinality of the blue region is included three times (in $N(A)$ and in $N(B)$ and in $N(C)$). The number of times the cardinality of each region has been included is shown in the second row of the Venn diagrams above.

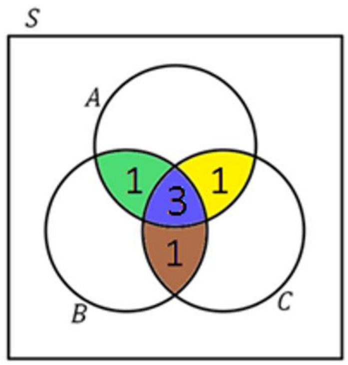
Now let's consider the middle row, $N(A \cap B) + N(A \cap C) + N(B \cap C)$, of the formula for $N(A \cup B \cup C)$, *i.e.* all of the two-way intersections.

Pictorially, $N(A \cap B) + N(A \cap C) + N(B \cap C)$ looks like



$N(A \cap B)$ $N(A \cap C)$ $N(B \cap C)$

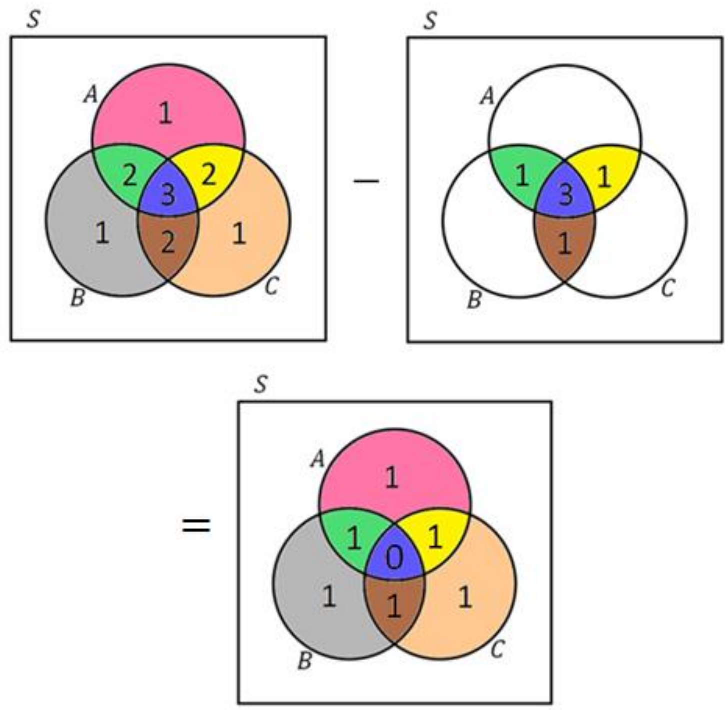
$=$



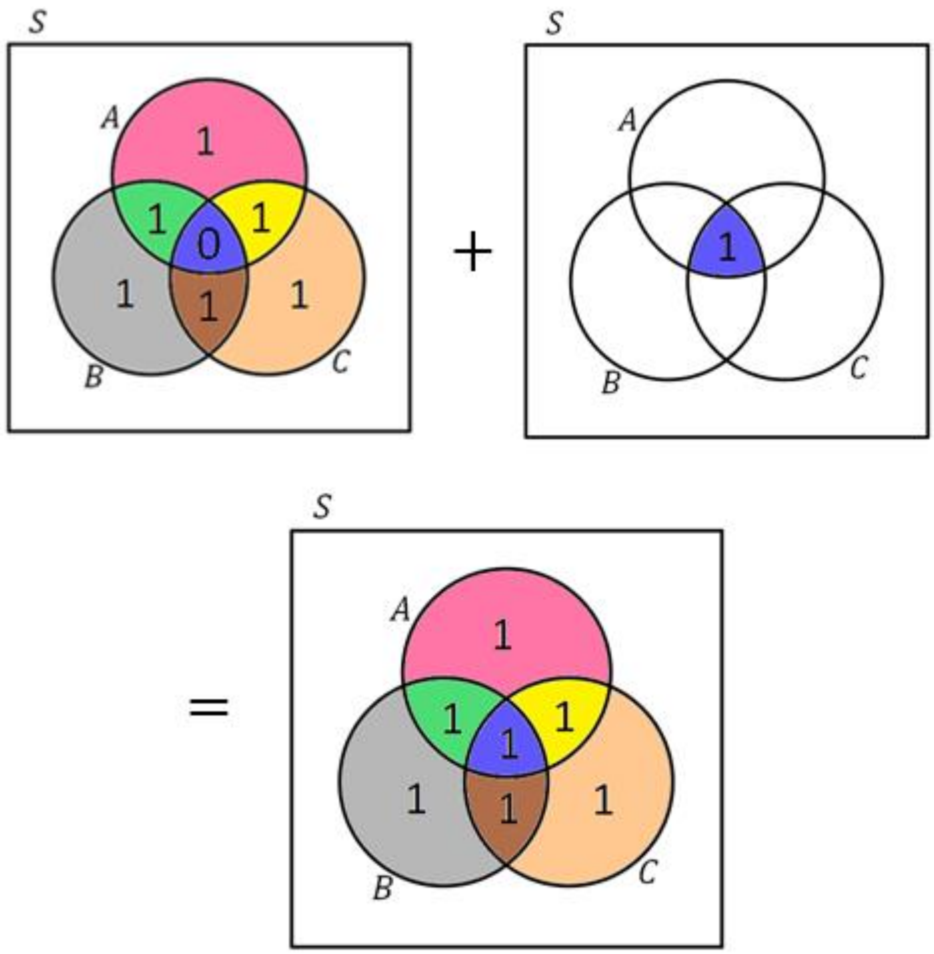
$$N(A \cap B) + N(A \cap C) + N(B \cap C)$$

Now consider the difference in the first two rows of the formula for $N(A \cup B \cup C)$.

$$\Big(N(A) + N(B) + N(C)\Big) - \Big(N(A \cap B) + N(A \cap C) + N(B \cap C)\Big)$$



The count of each section, except the blue section, is included exactly once. The final correction we need to do is to add back the count of this blue section. But this blue section is $A \cap B \cap C$. So, we need to add in $N(A \cap B \cap C)$. This would give us



which we recognize as $N(A \cup B \cup C)$.

But what if we move up to the union of four, five or more sets? The inclusion-exclusion formula for five sets given below illustrates the general idea.

4.2 Union of Four or More Sets

Method of Inclusion – Exclusion

$$\begin{aligned} N(A \cup B \cup C \cup D \cup E) = \quad & + \quad (N(A) + N(B) + N(C) + N(D) + N(E)) \\ & - \quad (N(A \cap B) + \cdots + N(D \cap E)) \\ & + \quad (N(A \cap B \cap C) + \cdots + N(C \cap D \cap E)) \end{aligned}$$

$$- \quad (N(A \cap B \cap C \cap D) + \cdots + N(B \cap C \cap D \cap E))$$

$$+ \quad N(A \cap B \cap C \cap D \cap E)$$

where, for example, the row of "three-way" intersections includes the counts of all possible intersections of three sets taken without replacement and when order is not important from A, B, C, D and E.

Note the pattern: Include (add) all "one way" intersections, then exclude (subtract) all "two way" intersections, then include (add) all "three way" intersections, etc.

Critical Point:

In an Inclusion-Exclusion problem when, for example, you are finding $N(A \cap B)$ as part of $N(A \cup B \cup C \cup D \cup E)$, you are making no assumptions on whether conditions C, D and/or E apply or not. In other words, when finding $N(A \cap B)$, we ignore C, D and E altogether.

4.3 Inclusion-Exclusion in the Presence of Symmetry

In many applications there is a symmetry inherent in the problem that makes the count of each "one-way" intersection equal (*i.e.* $N(A) = N(B) = N(C) = \cdots$) and makes the count of each "two-way" intersection equal (*i.e.* $N(A \cap B) = N(A \cap C) = N(A \cap D) = \cdots$) and makes the count of each "three-way" intersection equal (*i.e.* $N(A \cap B \cap C) = N(A \cap B \cap D) = N(A \cap B \cap E) = \cdots$) etc.

In such cases the Inclusion-Exclusion formula simplifies a great deal because we can combine all the "one-way" intersections into a single term and we can combine all the "two-way" intersections into a single term, etc.

When we consider $N(A \cup B \cup C \cup D \cup E)$ in the presence of symmetry, how many "one-way" intersections will we be combining? How many "two-way" intersections are there? "three-way"?

Remember that the row of "three-way" intersections included the counts of all possible intersections of 3 sets taken without replacement and when order is not important from the 5 distinct events A, B, C, D and E. But sampling without replacement and when order is not important is the **combinations-type I** problem. Hence there are $\binom{5}{3}$ "three-way" intersections. And more generally, there would be $\binom{n}{j}$ possible "j-way" intersections among the n events $A_1, A_2, \dots, A_n$.

Hence, when this kind of symmetry is present and we combine all like terms, the inclusion-exclusion formula collapses to the following more manageable form.

$$N(A_1 \cup A_2 \cup \cdots \cup A_n)$$
$$= \binom{n}{1} N(A_1) - \binom{n}{2} N(A_1 \cap A_2) + \binom{n}{3} N(A_1 \cap A_2 \cap A_3) - \binom{n}{4} N(A_1 \cap A_2 \cap A_3 \cap A_4)$$
$$+ \cdots + (-1)^{n-1} \binom{n}{n-1} N(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) + (-1)^n \binom{n}{n} N(A_1 \cap A_2 \cap \cdots \cap A_n).$$

4.4 Inclusion-Exclusion Principle Applies for *Any* Additive Set Function

Up to this point we have only applied the principle of inclusion-exclusion to the counting function $N(\ )$ for sets. But if you go back to where we established both $N(A \cup B)$ and $N(A \cup B \cup C)$ and look closely you will see that the ideas and justifications would still hold at every step if we replaced the counting set function $N(\ )$ with *any* other set function $W(\ )$ where $W(A \cup B) = W(A) + W(B)$ whenever $A \cap B = \emptyset$.

A set function $W(\ )$ that has the property where $W(A \cup B) = W(A) + W(B)$ whenever $A \cap B = \emptyset$ is called an ***additive set function***.

4.4.1 Method of Inclusion - Exclusion for Probabilities and Percentages

Fortunately, you don't have to learn a new set of formulas here. The inclusion-exclusion formulas for counting apply to probabilities exactly as they are with the simple switch of a P or Pe for probability and percentages respectively, instead of the N for counting.

$$P(A \cup B \cup C \cup D \cup E) = \quad + \quad (P(A) + P(B) + P(C) + P(D) + P(E))$$

$$
\begin{aligned}
&- \quad (P(A\cap B)+\cdots+P(D\cap E)) \\
&+ \quad (P(A\cap B\cap C)+\cdots+P(C\cap D\cap E)) \\
&- \quad (P(A\cap B\cap C\cap D)+\cdots+P(B\cap C\cap D\cap E)) \\
&+ \quad P(A\cap B\cap C\cap D\cap E).
\end{aligned}
$$

And if there is a symmetry in the problem that makes the count of each "one-way" intersection equal (*i.e.* $P(A)=P(B)=P(C)=\cdots$) and makes the count of each "two-way" intersection equal (*i.e.* $P(A\cap B)=P(A\cap C)=P(A\cap D)=\cdots$) and makes the count of each "three-way" intersection equal (*i.e.* $P(A\cap B\cap C)=P(A\cap B\cap D)=P(A\cap B\cap E)=\cdots$) etc., then

$$
\begin{aligned}
&P(A_1\cup A_2\cup\cdots\cup A_n) \\
&\quad=\binom{n}{1}P(A_1)-\binom{n}{2}P(A_1\cap A_2)+\binom{n}{3}P(A_1\cap A_2\cap A_3)-\binom{n}{4}P(A_1\cap A_2\cap A_3\cap A_4) \\
&\qquad+\cdots+\;(-1)^{n-1}\binom{n}{n-1}P(A_1\cap A_2\cap\cdots\cap A_{n-1})+\;(-1)^{n}\binom{n}{n}P(A_1\cap A_2\cap\cdots\cap A_n).
\end{aligned}
$$

4.5 Practice Problems with Solutions and Discussion

Problem 1.

Three newspapers, *The Daily News*, *The Post Gazette* and *The Times* are published in Gotham City. The following results were obtained in a survey of the adult population of the city: 20 percent read *The Daily News*, 16 percent read *The Post Gazette*, 14 percent read the *The Times*, 8 percent read both the *Daily* and the *Post*, 5 percent read both the *Daily* and the *Times*, 4 percent read both the *Post* and the *Times*, and 2 percent read all three? What percent of the adult population of the city reads none of the papers?

<u>**Solution**</u>

Define the following events:

A: reads *The Daily News*

B: reads *The Post Gazette*

C: reads *The Times*

The problem is asking for $Pe(A' \cap B' \cap C')$, the percentage of the adult population in Gotham City that don't read any of the three newspapers.

$Pe(A' \cap B' \cap C')$

$$= 100 - Pe(A \cup B \cup C)$$

$$= 100 - \Big[\big(Pe(A) + Pe(B) + Pe(C)\big)$$
$$- \big(Pe(A \cap B) + Pe(A \cap C) + Pe(B \cap C)\big) + Pe(A \cap B \cap C)\Big]$$

$$= 100 - \big[(20 + 16 + 14) - (8 + 5 + 4) + 2\big]$$

$$= 100 - (50 - 17 + 2) = 65.$$

So 65 percent of the adult population of Gotham City don't read any of that city's three newspapers.

■

Problem 2 (MSHSML, 2014-15, Meet 5, Team Event)

4 women each store a distinct hat in the same box. If all 4 women reach into the box randomly and independently, what is the probability that no woman picks her own hat?

<u>Solution</u>

Define the events

A_1: first woman to pick does get her own hat

A_2: second woman to pick does get her own hat

A_3: third woman to pick does get her own hat

A_4: fourth woman to pick does get her own hat

We want

$$P(A_1' \cap A_2' \cap A_3' \cap A_4') = P(S) - P(A_1 \cup A_2 \cup A_3 \cup A_4)$$

$$= P(S) - (P_1^4 - P_2^4 + P_3^4 - P_4^4)$$

$$= P(S) - \left(\binom{4}{1} P(A_1) - \binom{4}{2} P(A_1 \cap A_2)\right.$$

$$\left. + \binom{4}{3} P(A_1 \cap A_2 \cap A_3) - \binom{4}{4} P(A_1 \cap A_2 \cap A_3 \cap A_4)\right)$$

$$= 1 - \binom{4}{1}\left(\frac{3!}{4!}\right) + \binom{4}{2}\left(\frac{2!}{4!}\right) - \binom{4}{3}\left(\frac{1!}{4!}\right) + \binom{4}{4}\left(\frac{0!}{4!}\right)$$

$$= \frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} = \frac{3}{8}.$$

■

Problem 3 (MSHSML, 2010-11, Meet 5, Event C)

In how many ways can 10 distinct math team members be assigned to ride to a meet in 3 different school vehicles (a maxi-van, a mini-bus, and a full-sized school bus), if each vehicle must have at least one math team member riding in it?

<u>Solution</u>

Define the events

S : all possible ways (unrestricted) to assign the 10 students to a vehicle

A_1: seating assignments where the maxi-van is left empty

A_2: seating assignments where the mini-bus is left empty

A_3: seating assignments where the full-sized school bus is left empty

We want

$$N(A_1' \cap A_2' \cap A_3') = N(S) - N(A_1 \cup A_2 \cup A_3)$$

$$= N(S) - \Big(\big(N(A_1) + N(A_2) + N(A_3)\big)$$

$$-\big(N(A_1 \cap A_2) + N(A_1 \cap A_3) + N(A_2 \cap A_3)\big) + N(A_1 \cap A_2 \cap A_3)\Big)$$

$$= N(S) - \left(\binom{3}{1} N(A_1) - \binom{3}{2} N(A_1 \cap A_2) + \binom{3}{3} N(A_1 \cap A_2 \cap A_3)\right)$$

$$= 3^{10} - \left(\binom{3}{1} \cdot 2^{10} - \binom{3}{2} \cdot 1^{10} + \binom{3}{3} \cdot 0\right)$$

$$= \sum_{j=0}^{3} \binom{3}{j} (-1)^j (3-j)^{10} = 55980.$$

■

Problem 4

How many positive integers less than or equal to 2018 are not divisible by 2 or 5 or 6?

<u>Solution</u>

Let $S = \{1,2,3, \dots ,2018\}$ and define the sets

A_1: all integers in S which are divisible by 2

A_2: all integers in S which are divisible by 5

A_3: all integers in S which are divisible by 6.

In terms of these events A_1, A_2 and A_3 the problem is asking for $N(A_1' \cap A_2' \cap A_3')$.

Using DeMorgan's Laws and the Principle of Inclusion-Exclusion together we have

$$N(A_1' \cap A_2' \cap A_3') = N(S) - N(A_1 \cup A_2 \cup A_3)$$

$$= N(S) - \sum_{j=1}^{3} (-1)^{j-1} N_j^3.$$

This is a problem where we do not have symmetry among the events A_1, A_2, and A_3 which means we have to find each of the terms in N_j^3 separately.

We've already used the key result from we need from number theory several times now.

> The number of integers in $\{1,2,\ldots,n\}$ which are divisible by the positive integers a and b, both less than or equal to n, equals
>
> $$\text{floor}\left(\frac{n}{\text{lcm}(a,b)}\right)$$
>
> where the floor(x) function equals the largest integer less than or equal to x and where lcm(a,b) is the least common multiple of the two integers a and b.

It follows from this result that

$$N(A_1) = \text{floor}\left(\frac{2018}{2}\right) = 1009$$

$$N(A_2) = \text{floor}\left(\frac{2018}{5}\right) = \text{floor}(403.6) = 403$$

$$N(A_3) = \text{floor}\left(\frac{2018}{6}\right) = \text{floor}(336.\overline{3}) = 336$$

$$N(A_1 \cap A_2) = \text{floor}\left(\frac{2018}{\text{lcm}(2,5)}\right) = \text{floor}\left(\frac{2018}{10}\right) = \text{floor}(201.8) = 201$$

$$N(A_1 \cap A_3) = \text{floor}\left(\frac{2018}{\text{lcm}(2,6)}\right) = \text{floor}\left(\frac{2018}{6}\right) = 336$$

$$N(A_2 \cap A_3) = \text{floor}\left(\frac{2018}{\text{lcm}(5,6)}\right) = \text{floor}\left(\frac{2018}{30}\right) = \text{floor}\big(67.2\overline{6}\big) = 67$$

$$N(A_1 \cap A_2 \cap A_3) = \text{floor}\left(\frac{2018}{\text{lcm}(2,5,6)}\right) = \text{floor}\left(\frac{2018}{30}\right) = 67.$$

From this data we see that

$$N_1^3 = N(A_1) + N(A_2) + N(A_3) = 1009 + 403 + 336 = 1748$$

$$N_2^3 = N(A_1 \cap A_2) + N(A_1 \cap A_3) + N(A_2 \cap A_3) = 201 + 336 + 67 = 604$$

$$N_3^3 = N(A_1 \cap A_2 \cap A_3) = 67.$$

Hence,

$N(\text{numbers in } S \text{ not divisible by 2 or 5 or 6})$

$$= N(A_1' \cap A_2' \cap A_3') = N(S) - \sum_{j=1}^{3} (-1)^{j-1} N_j^3$$

$$= 2018 - (1748 - 604 + 67) = 807.$$

■

5 Integer Solutions of the Equation

This section is about counting the number of solutions $(x_1, x_2, \dots, x_n)$ to the equation $x_1 + x_2 + \cdots + x_n = r$ if $x_1 \in A_1, x_2 \in A_2, \dots, x_n \in A_n$ where each A_j is a set of integers.

Motivating Example

The Winona City Parks Board has purchased 20 new park benches to be distributed to their 5 city parks. The board considered many factors (land size, usage data, etc.) in trying to decide how many of these 20 benches should go to each park. But the board realized that aesthetics was a critical factor and this could best be decided by the facilities staff at the time of installation. So, the park board only gave the staff the following general information for the number of benches per park.

$$3 \leq \text{Park } 1 \leq 6$$
$$1 \leq \text{Park } 2 \leq 3$$
$$2 \leq \text{Park } 3 \leq 9$$
$$0 \leq \text{Park } 4 \leq 3$$
$$5 \leq \text{Park } 5 \leq 12.$$

To be certain, if we let x_j equal the number of benches that get placed in Park J, then $x_1 + x_2 + x_3 + x_4 + x_5 = 20$ and the restriction $3 \leq \text{Park } 1 \leq 6$, for example, would tell us that $x_1 \in \{3,4,5,6\}$. That is, $A_1 = \{3,4,5,6\}$. Similarly, we have $A_2 = \{1,2,3\}$, $A_3 = \{2,3,4, \dots, 9\}$, $A_4 = \{0,1,2,3\}$ and $A_5 = \{5,6, \dots 12\}$.

How many different ways could the facilities staff distribute the 20 benches to the 5 parks under these instructions?

For the example given above you might well be thinking, "Isn't this just another way of phrasing urn models? Couldn't we think of Park J as "Urn J" and ask for the number of ways to distribute 20 identical balls into a row of 5 urns with the same restrictions on the number of balls per urn as we had on the number of benches per park?"

The short answer is, "Yes". We can think of an "Integer Solutions of the Equation" problem as an urn model problem disguised in algebra terminology.

So why bother? It will turn out that an algebraic formulation is a convenient context for many applications, is a natural way to include negative numbers in the the range of variables and provides for an intuitive way to express and generalize solutions.

There is no "clean" answer for *general* sets of integers $A_1, A_2, \ldots, A_n$. Fortunately, there are simple solutions to this type of problem for certain useful special cases of the A_j's.

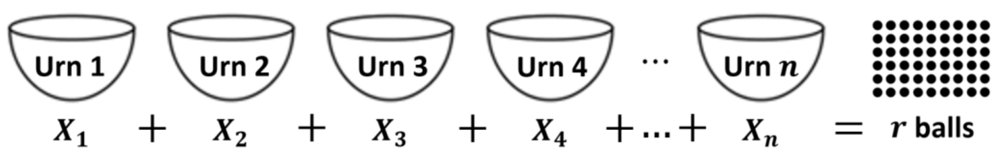
5.1 Nonnegative Integer Solutions of the Equation

How many solutions are there to the equation $X_1 + X_2 + \cdots + X_n = r$ with $X_j \in \{0,1,2,\ldots\}$ for $j = 1,2,\ldots,n$?

Answer

$\binom{n+r-1}{r}$

Explanation



Every vector $(X_1, X_2, \ldots, X_n)$ which is a solution of

$$X_1 + X_2 + \cdots + X_n = r \text{ with } X_i \in \{0,1,2,\ldots\}, \; i = 1,2,\ldots,n$$

represents a unique way of distributing r identical balls into n distinguishable urns with any number of balls per urn. (Just let X_j be the number of balls distributed into the j^{th} urn.). This is true in the other direction as well. That is, every way of distributing r identical balls into n distinguishable urns with any number of ball per urn represents a unique solution to the above equation if we again let X_j equal the number of balls in the j^{th} urn.

That is, these two problems have the same count. But we recognize that the above problem of distributing balls into urns is just the **multisets-type II** situation and has

$$\binom{\#\text{urns} + \#\text{balls} - 1}{\#\text{balls}} = \binom{n + r - 1}{r}$$

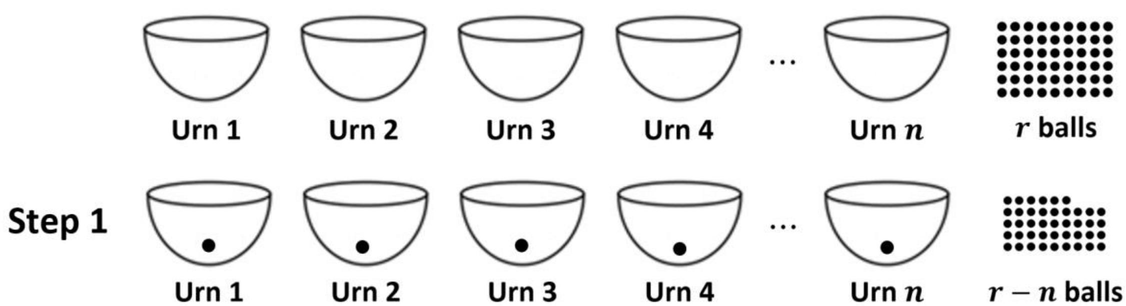
solutions.

■

5.2 Positive Integer Solutions of the Equation

How many solutions are there to the equation $X_1 + X_2 + \cdots + X_n = r$ with $X_j \in \{1,2,\ldots\}$ for $j = 1,2,\ldots,n$?

Solution



Answer

$$\binom{r-1}{r-n}$$

Explanation

Step 1. Start by putting a single ball in each urn. Because the balls are identical there is only 1 distinguishable way to do this.

Step 2. Distribute the remaining $r - n$ identical balls into the n urns without restriction. But as before this is just the problem of Multisets-Type II.

$$1 \cdot \binom{\#\text{urns} + \#\text{balls} - 1}{\#\text{balls}} = \binom{n + (r-n) - 1}{r-n} = \binom{r-1}{r-n}$$

■

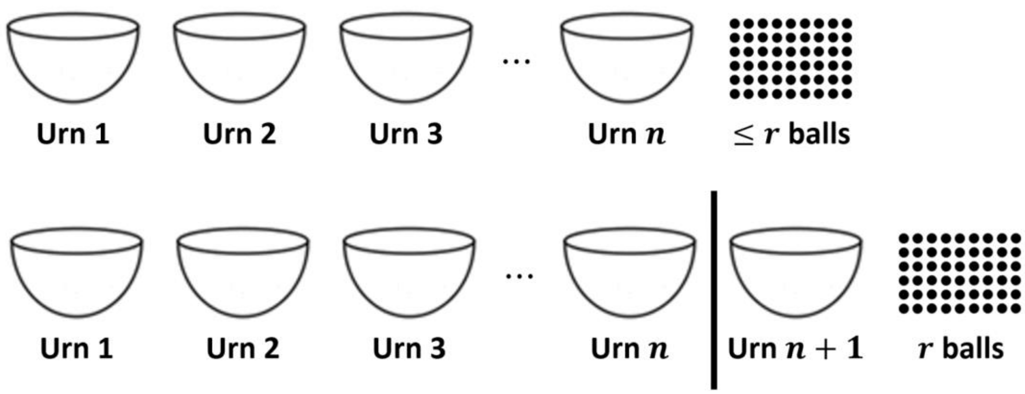
5.3 Nonnegative Integer Solutions of an Inequality

How many solutions are there to the equation $X_1 + X_2 + \cdots + X_n \leq r$ with $X_j \in \{0,1,2,\ldots\}$ for $j = 1,2,\ldots,n$?

Answer

$\binom{n+r}{r}$

Explanation



To make the initially understandable let's consider the special case $n = 3$ and $r = 8$. That is, let's show that the number of solutions to the equation

$$X_1 + X_2 + X_3 \leq 8 \text{ with } X_i \in \{0,1,2,\ldots\} \text{ for } i = 1,2,3$$

equals

$$\binom{3+8}{8} = \binom{11}{8}.$$

For the purpose of illustration let's write down several solutions (X_1, X_2, X_3) to the inequality $X_1 + X_2 + X_3 \leq 8$. That is, our set of solutions contains each of the following (and more):

$$(0,4,1), (2,2,4), (0,0,0), (8,0,0), (1,1,0), (3,3,1), \ldots$$

For each of these solutions (and all other solutions not listed above) we can indicate how much short of the maximum value of 8 we are. That is, we could write the above in the form

$$(0,4,1|3), (2,2,4|0), (0,0,0|8), (8,0,0|0), (1,1,0|6), (3,3,1|1), \ldots$$

where this last number equals $8-(X_1+X_2+X_3)$.

We note that tacking on this last number to each solution in our solution set will not change the number of solutions in our solution set. It just changes how the solutions look.

Because the X's are distinct there is no ambiguity to write these solutions in the form

$$(0,4,1,3),(2,2,4,0),(0,0,0,8),(8,0,0,0),(1,1,0,6),(3,3,1,1),\ldots$$

Now let's change gears and consider a new problem. (Our purpose is to ultimately show this new problem will have the same solution and hence the same count as the problem we started with.)

How many solutions to the equation

$$X_1+X_2+X_3+X_4=8 \text{ with } X_i\in\{0,1,2,\ldots\} \text{ for } i=1,2,3,4$$

are there?

For the purpose of illustration lets write down several solutions (X_1,X_2,X_3,X_4) to the equality $X_1+X_2+X_3+X_4=8$. Our set of solutions to this new problem contains each of the following (and more):

$$(0,4,1,3),(2,2,4,0),(0,0,0,8),(8,0,0,0),(1,1,0,6),(3,3,1,1),\ldots$$

That is every solution to this new problem is a solution to the original problem and vice versa. This means that the number of ways to do each problem must be the same.

However, we have already solved this new problem. This is just Model Problem 1 of this set with $n=4$ and $r=8$. Hence there are

$$\binom{4+8-1}{8}=\binom{11}{8}$$

solutions to the equality

$$X_1+X_2+X_3+X_4=8 \text{ with } X_i\in\{0,1,2,\ldots\} \text{ for } i=1,2,3,4$$

and hence also

$$\binom{4+8-1}{8}=\binom{11}{8}$$

solutions to the inequality

$$X_1+X_2+X_3\le 8 \text{ with } X_i\in\{0,1,2,\ldots\} \text{ for } i=1,2,3.$$

Now let's jump back to the general form $X_1 + X_2 + \cdots + X_n \le r$. We need only mimic the argument used in the above special case to see that this inequality problem has the same count as the problem of counting the solutions to the equality

$$X_1 + X_2 + \cdots + X_n + X_{n+1} = r \text{ with } X_i \in \{0,1,2,\ldots\} \text{ for each } i.$$

From our solution to the part (a) of this problem, this equals

$$\binom{(n+1)+r-1}{r} = \binom{n+r}{r}.$$

(Note: This idea of adding an extra variable (X_{n+1} in our case) to switch an inequality problem to an equality problem occurs in various branches of mathematics and is particularly important in the field of *Linear Programming*. It is generally known as the method of "adding a *slack variable*".)

■

6 The Basics of Probability

The starting point for probability is an ***experiment*** with a given set $S = \{o_1, o_2, \ldots\}$ of possible ***outcomes***. The set S of all possible outcomes is called the ***sample space*** of that experiment.

For example, our experiment could be to roll a standard six-sided die. In this case our sample space of all possible outcomes to this experiment is $S = \{1,2,3,4,5,6\}$. In this example, our experiment has a sample space S made up of six outcomes.

Subsets of the sample space S of an experiment are called ***events***. Events are typically denoted by capital letters from the beginning of the English alphabet, such as events A or B, etc.

The distinction between outcomes and events is that ***outcomes cannot be further refined*** into smaller pieces.

Continuing with our above example, let $E = \{2,4,6\}$. E is an event of Sbecause $E \subset S = \{1,2,3,4,5,6\}$. $\{2\}, \{4\}, \{6\}$ are each outcomes of S because they cannot be broken apart into more refined parts. On the other hand, $E = \{2,4,6\}$ is not an outcome because it can be broken apart into three more refined pieces.

Our intuitive notions of probability stem from chance experiments which we can perform over and over – think again in terms of rolling a die or flipping a coin.

If we focus on repeatable experiments, we can define the ***probability of an outcome*** o_j as the long run percentage[*] of times this outcome occurs.

[*]Long Run Percentage – the theoretical value of the ratio $a_j(m)/m$ as m increases to infinity, where $a_j(m)$ is the number of times outcome o_j occurs in m replications of the experiment.

Events can be written in set notation as the **union** of outcomes.

This allows us to define $P(A)$, the ***probability of the event*** $\boldsymbol{A}$, as the sum of the probabilities of the distinct outcomes making up event A.

It should be understood that when we write $P(A)$ or when we say "the probability of event A", that is actually a shortcut for the more precise statement "the probability that the outcome of the experiment belongs to the set A".

Connecting probability to long run percentages (real or thought experiments) is called the empirical or frequentist approach to probability. Throughout these notes we will adopt the frequentist model.

The Language of Probability

"Or" and "And" in Probability

In set theory the ***union*** symbol ∪ translates to the word "**or**" while the **Intersection** symbol ∩ translates to the word "**and**".

This is also true in the language of probability theory. So, for example, we would read $P(A \cup B)$ as the probability that event A occurs "or" the event B occurs and we would read $P(A \cap B)$ as the probability that event A occurs "and" the event B occurs.

"Or" Can Be Ambiguous Outside of Mathematics

The phrase "A or B" in some contexts carries the meaning ***one but not both*** of A and B are true. For example, "This coupon entitles you to a free hamburger or fish sandwich."

In some contexts, "A or B" is used to indicate ***not just one but both*** of A and B are true. For example, "I did not see you in the classroom or the hallway."

In some contexts, "A or B" is meant to be understood as ***one***, ***both*** or ***neither*** of A and B are true. For example, "Would you like cream or sugar in your coffee?"

Finally, in some contexts "A or B" means ***at least one*** of A and B is true. For example, "You will earn an "A" in the class if you have at least a 93 average on all the tests or if you make at least a 98 on the final exam."

In probability, and in mathematics more broadly, the word "***or***" is by definition ***always*** to be interpreted in the "at least one" sense. So $P(A \cup B)$ is necessarily to be interpreted as "the probability that the outcome of the experiment belongs to ***at least one*** of the sets A and B".

Fundamental Properties

Theorem 1. If S is the sample space of an experiment, then $P(S) = 1$.

By definition, the sample space $S = \{o_1, o_2, \dots, o_n\}$ of an experiment is the union of all outcomes of that experiment. That is,

$$S = \{o_1\} \cup \{o_2\} \cup \cdots \cup \{o_n\}.$$

Furthermore, S is by definition an event in the sample space S.

Therefore, by the definition of the probability of an event as the sum of the probability of the outcomes making up that event, it is necessarily true that

$$P(S) = P(o_1) + P(o_2) + \cdots + P(o_n) = 1.$$

■

Two sets A and B are ***disjoint*** if their intersection (set of shared elements) is the empty set. So the two sets $A = \{2,3\}$ and $B = \{7,11,12\}$ are disjoint because $A \cap B = \emptyset$.

If the set of outcomes making up two events are disjoint, we say those two events are ***mutually exclusive***. Imagine now that our experiment is to roll a pair of dice and to calculate the sum. This is another example of a stochastic repeatable experiment and in this case the sample space would be $S = \{2,3,4, \dots ,12\}$. Clearly, $A = \{2,3\}$ and $B = \{7,11,12\}$ are both events of this experiment and clearly these two sets are disjoint. Therefore, we would say that these two events are *mutually exclusive*.

In the case of more than two events, we say the t events $A_1, A_2, \dots, A_t$ are mutually exclusive if the set of outcomes defining these events are ***pairwise disjoint***. That is if $A_i \cap A_j = \emptyset$ for all $i \neq j$.

So, if we define events $A = \{2,3\}$, $B = \{1,3\}$ and $C = \{4,5,6\}$ of the sample space $S = \{2,3,4,\dots,12\}$ then $A \cap B \cap C = \emptyset$ because there is no outcome shared by *all three events*. However, these events are **not** mutually exclusive because these events are not pairwise disjoint. In particular, $A \cap B \neq \emptyset$.

Theorem 2. If $A_1, A_2, \dots, A_k$ are mutually exclusive events of the sample space S, then

$$P(A_1 \cup A_2 \cup \dots \cup A_k) = P(A_1) + P(A_2) + \dots + P(A_k).$$

Consider the special case of events $A = \{2,3\}$, $B = \{4,5,6\}$ and $C = \{7\}$ from the sample space $S = \{2,3,4,\dots,11,12\}$.

In this case, the events A, B and C are mutually exclusive* because they are pairwise disjoint. That is, $(A \cap B) = (A \cap C) = (B \cap C) = \emptyset$. Also, we can see that $A \cup B \cup C$ is an event (it is a subset of S). To be clear,

$$A \cup B \cup C = \{2,3,4,5,6,7\} \subset \{2,3,4,5,6,7,8,9,10,11,12\}.$$

But (by definition) the *probability of an event* is the sum of the probabilities of the distinct outcomes making up that event. Therefore,

$$P(A \cup B \cup C) = P(\{2\}) + P(\{3\}) + P(\{4\}) + P(\{5\}) + P(\{6\}) + P(\{7\}).$$

But A, B and C are also events. Therefore,

$$\begin{aligned} P(A) &= P(\{2\}) + P(\{3\}) \\ P(B) &= P(\{4\}) + P(\{5\}) + P(\{6\}) \\ P(C) &= P(\{7\}). \end{aligned}$$

And in this case, it is clear by inspection that $P(A \cup B \cup C) = P(A) + P(B) + P(C)$.

*Note: The assumption that $A_1, A_2, \dots, A_k$ are mutually exclusive is the **key** to this argument. By definition all elements in a ***set*** are distinct – no repetitions allowed. So even if these events were not mutually exclusive, the probability of shared outcomes would only be counted once in the left-hand side term $P(A_1 \cup A_2 \cup \dots \cup A_k)$. However, any shared outcomes would be doubly (or multiply) counted in the right-hand side terms $P(A_1) + P(A_2) + \dots + P(A_k)$. Hence the two sides could ***not*** be equal.

■

The ***complement*** of a set A, denoted as A' (alternately denoted as $\overline{A}$ or A^c) is defined as the set of all outcomes in S that are not in S. Therefore, by definition $A \cap A' = \emptyset$ and $A \cup A' = S$.

Accordingly, if A is an event in the sample space S, then by definition A' is also an event in S and furthermore the **events A and A' are mutually exclusive**.

Theorem 3. For any event A,

$$P(A') = 1 - P(A).$$

Because the events A and A' are mutually exclusive, it follows from Theorem 2 that $P(A \cup A') = P(A) + P(A')$. But $P(A \cup A') = P(S) = 1$ by Theorem 1, therefore,

$$P(A) + P(A') = 1 \text{ and } P(A') = 1 - P(A).$$

■

Corollary of Theorem 3. For any experiment, $P(\emptyset) = 0$.

By definition the complement of the sample space of an experiment is the empty set. Therefore, by Theorems 1 and 3, $P(\emptyset) = 1 - P(S) = 1 - 1 = 0$.

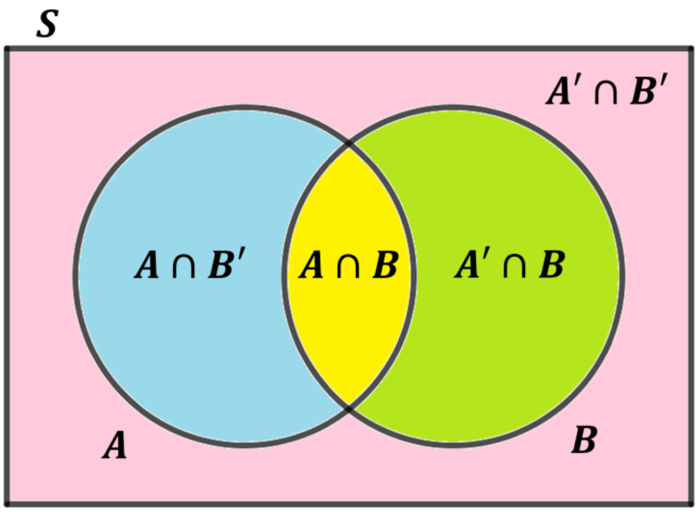
■

Theorem 4. If A and B are any* two events in the sample space S, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

*Events A and B are not necessarily mutually exclusive in Theorem 4. If A and B are mutually exclusive, then $P(A \cap B) = P(\emptyset) = 0$ and this result is just a special case of Theorem 2.

The essence of Theorem 4 is most clearly seen from a Venn diagram.



Notice that

(i) $(A \cap B')$, $(A \cap B)$ and $(A' \cap B)$ are mutually exclusive events in S.

(ii) $A = (A \cap B') \cup (A \cap B)$

(iii) $B = (A \cap B) \cup (A' \cap B)$

(iv) $A \cup B = (A \cap B') \cup (A \cap B) \cup (A' \cap B)$,

Because the blue, yellow and green define mutually exclusive events in S, it follows from Theorem 2 that

$$P(A) = P(A \cap B') + P(A \cap B)$$
$$P(B) = P(A \cap B) + P(A' \cap B)$$
$$P(A \cup B) = P(A \cap B') + P(A \cap B) + P(A' \cap B).$$

Hence,

$$\begin{aligned} P(A \cup B) &= P(A \cap B') + P(A \cap B) + P(A' \cap B) \\ &= P(A \cap B') + P(A \cap B) + P(A' \cap B) + \big(\boldsymbol{P(A \cap B) - P(A \cap B)}\big) \\ &= \big(P(A \cap B') + P(A \cap B)\big) + \big(P(A' \cap B) + P(A \cap B)\big) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B). \end{aligned}$$

■

Having established this formula for $P(A \cup B)$ which applies whether the events A and B are mutually exclusive or not, it is natural to consider if there is a parallel formula for $P(A \cup B \cup C)$ or $P(A \cup B \cup \cdots \cup Z)$ which holds whether the events are mutually exclusive or not. This leads us to the "**method of inclusion-exclusion**". We illustrate the method for the case of five events.

$$
\begin{aligned}
P(A \cup B \cup C \cup D \cup E) \;=\; & + \quad \big(P(A) + P(B) + P(C) + P(D) + P(E)\big) \\
& - \quad \big(P(A \cap B) + \cdots + P(D \cap E)\big) \\
& + \quad \big(P(A \cap B \cap C) + \cdots + P(C \cap D \cap E)\big) \\
& - \quad \big(P(A \cap B \cap C \cap D) + \cdots + P(B \cap C \cap D \cap E)\big) \\
& + \quad P(A \cap B \cap C \cap D \cap E).
\end{aligned}
$$

To be absolutely clear about this notation, the abbreviated notation used in third row of "three-way" intersections would in its entirety be

$$
\begin{aligned}
-\Big(& P(A \cap B \cap C) + P(A \cap B \cap D) + P(A \cap B \cap E) + P(A \cap C \cap D) + P(A \cap C \cap E) \\
& + P(A \cap D \cap E) + P(B \cap C \cap D) + P(B \cap C \cap E) + P(B \cap D \cap E) \\
& + P(C \cap D \cap E)\Big).
\end{aligned}
$$

The pattern of first adding (including) all "one way" intersections, then subtracting (excluding) all "two way" intersections, then adding (including) all "three way" intersections, etc. continues for cases with any number of events.

Problems requiring the method of inclusion-exclusion come up occasionally on MSHSML tests and we will consider a few of these in our final chapter where we take up "Miscellaneous" problems.

The two results known as **DeMorgan's Laws** are often part of solving inclusion-exclusion problems so we mention them now. Suppose $A_1, A_2, \ldots, A_n$ are events in a sample space S. Then

$$P\big((A_1 \cup A_2 \cup \cdots \cup A_n)'\big) = P\big(A_1' \cap A_2' \cap \cdots \cap A_n'\big)$$

and

$$P\big((A_1 \cap A_2 \cap \cdots \cap A_n)'\big) = P\big(A_1' \cup A_2' \cup \cdots \cup A_n'\big).$$

The verification of DeMorgan's Laws are an exercise in what is called "set chasing" in contest problem solver blogs. In set theory, we say the sets $A = B$ if for all x it is true that $x \in A \Leftrightarrow x \in B$.

We note that

$$
\begin{aligned}
& x \in (A_1 \cup A_2 \cup \cdots \cup A_n)' \\
& \quad \Leftrightarrow x \notin (A_1 \cup A_2 \cup \cdots \cup A_n) \\
& \quad \Leftrightarrow (x \notin A_1) \text{ and } (x \notin A_2) \text{ and } \cdots \text{ and } (x \notin A_n) \\
& \quad \Leftrightarrow \big(x \in (A_1)'\big) \text{ and } \big(x \in (A_2)'\big) \text{ and } \cdots \text{ and } \big(x \in (A_n)'\big) \\
& \quad \Leftrightarrow x \in \big((A_1)' \cap (A_2)' \cap \cdots \cap (A_n)'\big).
\end{aligned}
$$

This establishes that $(A_1 \cup A_2 \cup \cdots \cup A_n)' = A_1' \cap A_2' \cap \cdots \cap A_n'$. From here it follows that

$$P\big((A_1 \cup A_2 \cup \cdots \cup A_n)'\big) = P\big(A_1' \cap A_2' \cap \cdots \cap A_n'\big).$$

The proof of the other DeMorgan's Law follows similarly.

■

Theorem 5. Problems with Equally Likely Outcomes

If the sample space S of an experiment consists of n equally likely outcomes $o_1, o_2, \ldots, o_n$ and if the event A of S consists of k of those n outcomes, then $\boldsymbol{P(A) = k/n}$.

By Theorem 1, $P(S) = P(o_1) + P(o_2) + \cdots + P(o_n) = 1$. In the case of all equally likely outcomes, it follows that $P(o_j) = 1/n$ for all $1 \le j \le n$. Furthermore, by the definition of the probability of an event, if $A = \{o_{i_1}\} \cup \{o_{i_2}\} \cup \cdots \cup \{o_{i_k}\}$ then

$$
\begin{aligned}
P(A) &= P\big(\{o_{i_1}\}\big) + P\big(\{o_{i_1}\}\big) + \cdots + P\big(\{o_{i_k}\}\big) \\
&= \frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n} = \frac{k}{n}.
\end{aligned}
$$

■

Theorem 5 is sometimes called the **counting definition** of probability because it can be restated as,

$$P(A) = \frac{\text{number of outcomes in } A}{\text{number of outcomes in } S}$$

for all events A in S whenever all outcomes in S are equally likely.

Conditional (Rescaled) Probabilities

A familiar way to think about conditional probability is to make an analogy to test grades. Suppose you have a class where the tests together count for 60% of your overall grade, where the final exam counts for 20% of your grade, your homework counts for 10% and a project counts for 10% of your overall grade.

Now suppose that midway through the semester your teacher announces that he/she is cancelling the required project and will reset the percentages for the tests, final and homework so that the remaining components keep their same relative importance. What would the new percentages have to be?

For grades to keep their "same relative importance" means that if the original plan gave weights w_1, w_2, w_3, w_4 to the test, final, homework and project, then the rescaled grades become kw_1, kw_2 and kw_3 after the project has been scrapped. The constant k is that number necessary to make the updated weights again sum of 100%. That is,

$$k(60\%) + k(20\%) + k(10\%) = 100\% \Leftrightarrow k = 100/90.$$

So the new grading plan would be to count the tests for

$$(60\%) \cdot \frac{100}{90} = 66.\bar{6}\,\%$$

$$(20\%) \cdot \frac{100}{90} = 22.\bar{2}\,\%$$

and

$$(10\%) \cdot \frac{100}{90} = 11.\bar{1}\,\%.$$

Notice that the relative weights have not changed. The tests are still weighted three times as heavily as the final exam and the final exam is still worth twice as much as the homework.

Now let's keep the same numbers but change the storyline to one where John has agreed to a foot race with his good friends Mary, Max and Min. Max is the track star of the group and his probability of winning the race has been accessed to be 0.60 (*i.e.* 60%). John, Mary and Min come in close together with winning probabilities of 20%, 10% and 10% respectively. On the afternoon of the race Mary finds out she has chores and cannot be there for the race after all. What are the new probabilities of winning for Max, John and Min now that Mary has dropped out?

By direct analogy to the question with rescaling test grades they will become $66.\overline{6}\ \%$, $22.\overline{2}\ \%$ and $11.\overline{1}\ \%$ respectively.

Let's generalize this. Suppose we have an experiment with sample space $S = \{o_1, o_2, \ldots, o_n\}$. Suppose something causes all the outcomes <u>not</u> in $F = \{o_{f_1}, \ldots, o_{f_r}\} \subset S$ to be ruled out as possible outcomes of the experiment. (In analogy with removing the project from the set of graded elements for the class and Mary dropping out of the race.)

What are the conditional (rescaled) probabilities for the remaining outcomes $\{o_{f_1}, \ldots, o_{f_r}\}$? These conditional probabilities are denoted by $P\left(o_{f_j} \middle| F\right), j = 1,2, \ldots, r$.

Theorem 6. Conditional Probability of Outcomes

$$P\left(o_{f_j} \middle| F\right) = \frac{P\left(o_{f_j}\right)}{P(F)}$$

for all outcomes in $F = \{o_{f_1}, \ldots, o_{f_r}\} \subset S$.

First note that this theorem shows how the conditional probability $P\left(o_{f_j} \middle| F\right)$ can be computed as the ratio of two unconditional probabilities $P\left(o_{f_j}\right)$ and $P(F)$.

The fundamental principle of conditioning on the event F is to declare that all outcomes in S but not in F (*i.e.* the outcomes in F') are no longer possible (*i.e.* have dropped out of the race) and to keep the <u>relative conditional probability</u> of all outcomes in F the same as they were in the original (unconditional) sample space S.

i.e.

$$\frac{P\left(o_{f_i} \middle| F\right)}{P\left(o_{f_j} \middle| F\right)} = \frac{P\left(o_{f_i}\right)}{P\left(o_{f_j}\right)}.$$

From this principle

$$P(o_{f_i}|F) = \left(\frac{P\left(o_{f_j}|F\right)}{P\left(o_{f_j}\right)}\right) P(o_{f_i}) = k \cdot P(o_{f_i}).$$

The constant k can be determined by the requirement that the sum of these rescaled probabilities equals 1.

$$1 = \sum_{i=1}^{r} P(o_{f_i}|F) = \sum_{i=1}^{r} k \cdot P(o_{f_i}) = k \sum_{i=1}^{r} P(o_{f_i}) = kP(F).$$

Therefore,

$$k = \frac{1}{P(F)}$$

and

$$P(o_{f_i}|F) = k \cdot P(o_{f_i}) = \frac{P(o_{f_i})}{P(F)}.$$

■

Conditional Probability of an Event

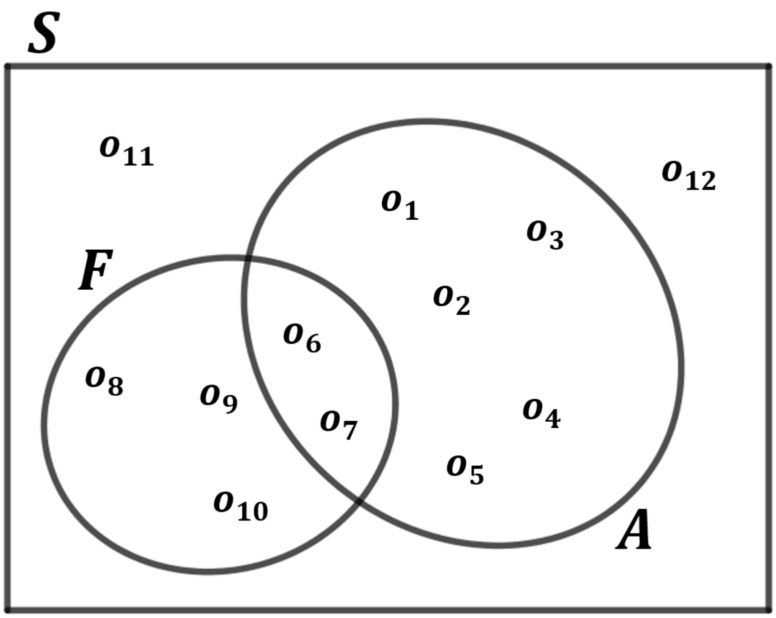
In alignment with how we previously defined (unconditional) probability we can now make the following definition.

> We define $P(A|F)$, the ***probability of the event A <u>conditional</u> on the event F***, as the sum of the rescaled probabilities <u>conditional</u> on F of the distinct outcomes making up event A.

The following result is a consequence of this definition.

> **Theorem 7.**
>
> If A and F are events in S then
>
> $$P(A|F) = \frac{P(A \cap F)}{P(F)}.$$

Once again we can use a Venn diagram to illustrate the central idea behind this result.



The two core ideas of conditioning on the event F are

(i) to remove all outcomes not in F from S, the sample space of possible outcomes

(ii) to rescale the probabilities of all outcomes in F to make $P(F|F) = 1$.

So, the first step in solving for $P(A|F)$ is set $P(o_j|F) = 0$ for all outcomes $o_j \notin F$. In particular

$$P(o_1|F) = P(o_2|F) = P(o_3|F) = P(o_4|F) = P(o_5|F) = P(o_{11}|F) = P(o_{12}|F) = 0.$$

The second step (as established by Theorem 6) is to set

$$P(o_j|F) = \frac{P(o_j)}{P(F)} \text{ for all outcomes } o_j \in F.$$

We note that in this example we have $A = \{o_1, o_2, o_3, o_4, o_5, o_6, o_7\}$. Hence, by the definition of the condition probability of an event, we have

$$P(A|F) = P(o_1|F) + P(o_2|F) + P(o_3|F) + P(o_4|F) + P(o_5|F) + P(o_6|F) + P(o_7|F)$$
$$= 0 + 0 + 0 + 0 + 0 + P(o_6|F) + P(o_7|F)$$

$$= \frac{P(o_6)}{P(F)} + \frac{P(o_7)}{P(F)} = \frac{P(o_6) + P(o_7)}{P(F)}.$$

But we also know that in this example $A \cap F = \{o_6, o_7\}$. Therefore,

$$P(A \cap F) = P(o_6) + P(o_7).$$

and

$$P(A|F) = \frac{P(o_6) + P(o_7)}{P(F)} = \frac{P(A \cap F)}{P(F)}.$$

■

Conditional Probability Versions of Theorems 1 – 5 and DeMorgan's Laws

Theorems 1 – 5 and DeMorgan's Laws remain valid for conditional probabilities. We can rewrite them as:

Theorem 1′. If C is any event in the sample space S of an experiment, then $P(C|C) = 1$.

Theorem 2′. If $A_1, A_2, \dots, A_k$ are mutually exclusive events of the sample space S and if C is any event in the sample space S, then

$$P\big((A_1 \cup A_2 \cup \dots \cup A_k)\big|C\big) = P(A_1|C) + P(A_2|C) + \dots + P(A_k|C).$$

Theorem 3′. For any events A and C in the sample space S

$$P\big(A'\big|C\big) = 1 - P(A|C).$$

Theorem 4′. If A, B and C are any events in the sample space S, then

$$P\big((A \cup B)\big|C\big) = P(A|C) + P(B|C) - P\big((A \cap B)\big|C\big).$$

(**DeMorgan's Laws**)′

If $A_1, A_2, \ldots, A_n$ and C are events in the sample space S, then

$$P\big((A_1 \cup A_2 \cup \cdots \cup A_n)'\big|C\big) = P\left(\left(A_1' \cap A_2' \cap \cdots \cap A_n'\right)\middle|C\right)$$

and

$$P\big((A_1 \cap A_2 \cap \cdots \cap A_n)'\big|C\big) = P\left(\left(A_1' \cup A_2' \cup \cdots \cup A_n'\right)\middle|C\right).$$

Theorem 5′. Problems with Equally Likely Outcomes

If A and C are events in the sample space S such that

(i) S consists of n equally likely outcomes
(ii) C consists of r of the n outcomes in S
(iii) $A \cap C$ consists of k of the n outcomes in S

then $P(A|C) = k/r$.

Chain Rule (or General Product Rule) for Probabilities

Theorem 8.

If $A_1, A_2, A_3 \ldots, A_n$ are events in S then

$$P(A_1 \cap A_2 \cap \cdots \cap A_n)$$
$$= P(A_1) \cdot P(A_2|A_1) \cdot P\big(A_3\big|(A_1 \cap A_2)\big) \cdot \ \cdots \ \cdot P\big(A_n\big|(A_1 \cap A_2 \cap \cdots \cap A_{n-1})\big)$$

The case of $n = 2$ follows immediately from Theorem 7. Let's step through the logic for the case of $n = 3$. The general proof follows the same pattern with induction.

$$P(A_1) \cdot P(A_2|A_1) \cdot P\big(A_3\big|(A_1 \cap A_2)\big)$$

$$= \Big(P(A_1) \cdot P(A_2|A_1)\Big) \cdot \ P\big(A_3\big|(A_1 \cap A_2)\big)$$

$$= P(A_1 \cap A_2) \cdot \ P\big(A_3\big|(A_1 \cap A_2)\big) \qquad \text{by Theorem 7}$$

$$= P(A_1 \cap A_2 \cap A_3) \qquad \text{by Theorem 7.}$$

■

Example

You might well have used Theorem 8 many times without explicit thinking about it as a separate result.

Suppose you draw at random and without replacement from an urn containing 3 white balls and 3 black balls. What is the probability that you get a white ball on the first draw, a white ball on the second draw and a black ball on the third draw?

Solution

Define the events A:white ball on the first draw, B:white ball on the second draw and C:black ball on the third draw. Applying Theorem 8,

$$P(A \cap B \cap C) = P(A)P(B|A)P\big(C\big|(A \cap B)\big).$$

There are initially 3 white and 3 black balls so $P(A) = 3/6$. After event A there are 2 white and 3 black balls left in the urn so $P(B|A) = 2/5$. After events A and B there are 1 white and 3 black balls left in the urn so $P\big(C\big|(A \cap B)\big) = 3/4$.

So,

$$P(A \cap B \cap C) = P(A)P(B|A)P\big(C\big|(A \cap B)\big) = \left(\frac{3}{6}\right)\left(\frac{2}{5}\right)\left(\frac{3}{4}\right) = \frac{3}{20}.$$

■

Independent Events

We define events $A_1, A_2, \dots, A_k$ of sample space S to be **mutually independent**★ if

$$P\big(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}\big) = P\big(A_{i_1}\big)P\big(A_{i_2}\big)\cdots P\big(A_{i_r}\big)$$

for *all* subsets $\{i_1, i_2, \dots, i_r\}$ of $\{1,2,\dots,k\}$ with two or more elements.

★the term "mutually independent" is usually shortened to just "independent" in contest problems

Example

To verify that events A, B and C of sample space S are mutually independent requires a separate★★ proof of each of the 4 equalities shown below.

$P(A \cap B) = P(A)P(B)$	$P(A \cap C) = P(A)P(C)$	$P(B \cap C) = P(B)P(C)$
$P(A \cap B \cap C) = P(A)P(B)P(C)$.		

To verify that events A, B, C and D of sample space S are mutually independent requires a separate** proof of each of the 11 equalities shown below.

$P(A \cap B) = P(A)P(B)$	$P(A \cap C) = P(A)P(C)$	$P(A \cap D) = P(A)P(D)$
$P(B \cap C) = P(B)P(C)$	$P(B \cap D) = P(B)P(D)$	$P(C \cap D) = P(C)P(D)$
$P(A \cap B \cap C) = P(A)P(B)P(C)$	$P(A \cap B \cap D) = P(A)P(B)P(D)$	
$P(A \cap C \cap D) = P(A)P(C)P(D)$	$P(B \cap C \cap D) = P(B)P(C)P(D)$	
$P(A \cap B \cap C \cap D) = P(A)P(B)P(C)P(D)$.		

**No one of these equalities will imply all the others. In particular,

$$P(A \cap B \cap C \cap D) = P(A)P(B)P(C)P(D)$$

does <u>not</u> imply all the other equalities.

The following two theorems are direct consequences of the definition of (mutually) independent events but are useful enough to be highlighted on its own.

Theorem 9.

If events A and F of sample space S are mutually independent then

$$P(A|F) = P(A)$$

and

$$P(F|A) = P(F).$$

This result follows immediately from Theorem 7 and the definition of independence.

$$P(A|F) = \frac{P(A \cap F)}{P(F)} = \frac{P(A)P(F)}{P(F)} = P(A)$$

and

$$P(F|A) = \frac{P(F \cap A)}{P(A)} = \frac{P(F)P(A)}{P(A)} = P(F).$$

■

Theorem 10.

If the events $A_1, A_2, \dots, A_k$ of sample space S are mutually independent then set functions of pairwise disjoint subsets of the events $A_1, A_2, \dots, A_k$ are mutually independent.

An example of this idea will help to clarify. Suppose A, B, C, D and E are mutually independent events of a sample space S. Then by Theorem 10 the three set functions

(i) $(A \cup B)'$
(ii) C
(iii) $(D' \cap E)$

are mutually independent because no two of these set functions involve the same events from A, B, C, D and E.

Corollary of Theorem 10

If events A and B of sample space S are mutually independent then

$\overline{A}$ and B are mutually independent events,

A and $\overline{B}$ are mutually independent events,

$\overline{A}$ and $\overline{B}$ are mutually independent events.

Combining Theorems 9 and 10 we can get a sort of corollary to both which we will simply illustrate through examples.

If events A, B, C, D and E are mutually independent then

$P\left((D' \cap E)\middle|(A \cup B)'\right) = P\left((D' \cap E)\right).$

Independence and Intuition

Independence is a term used in probability and in everyday conversation. You might agree that whether it rains tonight is in no way influenced by (*i.e.* is independent of) whether you left your car windows open. But that intuitive understanding of how events influence each other is *for the most part* not sufficient to determine whether events meet the probability definition for independence.

There are a few notable exceptions. In a contest setting, you are expected to assume that the results of successive tosses of a coin are independent events ***whether it is stated or not***. The same goes for the successive rolls of a die and the outcome when drawing balls from an urn if the sampling is done with replacement.

However, apart from such notable exceptions, you should work under the assumption that events are not necessarily independent unless it is explicitly stated in the problem.

Here are a few examples where independence is *implicitly* assumed in some MSHSML problems:

(5C141) … probability that a fair coin will land tails three times in a row

(5D102) … a and b chosen with replacement from $\{1,2, \dots ,8,9\}$

(5C182) … a and b are the results from rolling an 8-sided die twice

(5C091) … a, b and c are the results from rolling a 6-sided die three times

(5C122) … Kathy flips a fair coin until she get three heads in a row

And here are a few examples where independence is *explicitly* stated in some MSHSML problems:

(5C164) … Assuming that Ashley, Ben, and Charlie chose their ice cream flavors independently, what is the probability no flavor is chosen more than once

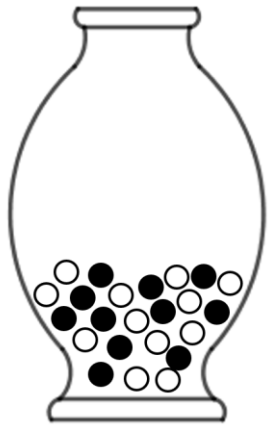
(5D104) … a frog choses the direction of its next jump independently from previous jumps

(5T104) … a candy cane is broken at two points chosen independently and at random.

7 Random Variables with Repeated Independent Trials

7.1 Binomial Random Variable

I. Suppose an urn contains w white balls and b black balls. You reach into this urn and randomly select a ball, note its color and ***then return the ball to the urn***. You repeat this process for a total of n draws. Let X equal the number of times you select a white ball and let $P(X = x)$ represent the probability that you get exactly x white balls in these n draws.



Then

$$P(X = x) = \begin{cases} \binom{n}{x}\left(\frac{w}{w+b}\right)^{x}\left(\frac{b}{w+b}\right)^{n-x} & x \in \{0,1,\dots,n\} \\ 0 & x \notin \{0,1,\dots,n\}. \end{cases}$$

II. Suppose you roll a fair die a total of n times. Let X equal the number of times you roll either a 3 or 6. Then

$$P(X = x) = \begin{cases} \binom{n}{x}\left(\frac{1}{3}\right)^{x}\left(\frac{2}{3}\right)^{n-x} & x \in \{0,1,\dots,n\} \\ 0 & x \notin \{0,1,\dots,n\}. \end{cases}$$

III. You have a penny that has been altered in such a way that it lands heads up with probability $5/8$ when flipped. Suppose you flip this penny n times and let X equal the number of times it lands heads in these n flips. Then

$$P(X=x)=\begin{cases}\binom{n}{x}\left(\frac{5}{8}\right)^x\left(\frac{3}{8}\right)^{n-x} & x\in\{0,1,\dots,n\}\\ 0 & x\notin\{0,1,\dots,n\}.\end{cases}$$

Now consider inventing a generic experiment that incorporates the common aspects of these three problems above.

Suppose our generic experiment consists of n repeated trials where these trials meet the following conditions:

(1) the trials are independent, *i.e*. the outcome in one trial will have no impact on the outcome of a different trial

(2) there are only two possible outcomes for a trial. We will refer to these two possible outcomes as "success" or "failure"

(3) $P(\text{success})$ is the same for every trial. We will use the letter p to represent $P(\text{success})$ for notational simplicity

(4) the total number of trials performed is fixed (not random)

(5) our interest is in the total number of successes that occur in these fixed number of trials

Trials that meet these conditions are called **Bernoulli trials** in honor of Jacob Bernoulli (1654-1705) who wrote about such trials in his seminal paper *Ara Conjectandi* (The Art of Conjecturing).

Let $X=$ number of successes that occur in n Bernoulli trials. Then

$$P(X=x)=\begin{cases}\binom{n}{x}p^x(1-p)^{n-x} & x\in\{0,1,\dots,n\}\\ 0 & x\notin\{0,1,\dots,n\}\end{cases}$$

where $p=P(\text{success})$.

In the language of probability and statistics, whenever $P(X=x)$ is determined by the above formula we say that X is a **binomial random variable** or that X follows the **binomial distribution**.

To understand where this result comes from, let's examine the case where $X=2$ heads in $n=3$ flips of a penny with $p=P(\text{heads})=5/8$. Let $\underline{H}$ represent heads and $\underline{T}$ represent tails. Then

$$P(X=2\text{ heads})=P\left(\left(\underline{H}\text{ and }\underline{H}\text{ and }\underline{T}\right)\text{ or }\left(\underline{H}\text{ and }\underline{T}\text{ and }\underline{H}\right)\text{ or }\left(\underline{T}\text{ and }\underline{H}\text{ and }\underline{H}\right)\right)$$

$$= P(\underline{H} \text{ and } \underline{H} \text{ and } \underline{T}) + P(\underline{H} \text{ and } \underline{H} \text{ and } \underline{T}) + P(\underline{H} \text{ and } \underline{H} \text{ and } \underline{T})$$
$$= \left(P(\underline{H}) \cdot P(\underline{H}) \cdot P(\underline{T})\right) + \left(P(\underline{H}) \cdot P(\underline{T}) \cdot P(\underline{H})\right) + \left(P(\underline{T}) \cdot P(\underline{H}) \cdot P(\underline{H})\right)$$
$$= \left(\frac{5}{8}\right)^2 \left(\frac{3}{8}\right) + \left(\frac{5}{8}\right)^2 \left(\frac{3}{8}\right) + \left(\frac{5}{8}\right)^2 \left(\frac{3}{8}\right)$$
$$= 3 \cdot \left(\frac{5}{8}\right)^2 \left(\frac{3}{8}\right).$$

As a reminder, $P(A \text{ or } B) = P(A) + P(B)$ when events A and B are mutually exclusive (*i.e.* they cannot both be true at the same time) and $P(A \text{ and } B) = P(A) \cdot P(B)$ when events A and B are independent (*i.e.* one event occurring or not does not influence whether the other event occurs or not).

It is straightforward to extend this to the general case where we want to find the probability of x heads in n flips with $p = P(\text{heads})$.

Look back at each of the parts in the above answer

$$P(X = 2) = 3 \cdot \left(\frac{5}{8}\right)^2 \left(\frac{3}{8}\right).$$

The leading coefficient 3 was just the number of possible ways to arrange two $\underline{H}$'s and one $\underline{T}$. In general, there are $\binom{n}{x}$ possible ways to arrange x $\underline{H}$'s and $(n - x)$ $\underline{T}$'s.

The factor $p = P(\text{heads}) = 5/8$ was raised to the second power and the factor $(1 - p) = P(\text{tails}) = 3/8$ was raised to the first power because there two $\underline{H}$'s and one $\underline{T}$. In the general case with x $\underline{H}$'s and $(n - x)$ $\underline{T}$'s these factors become $p^x(1 - p)^{n-x}$.

This establishes the general result. If we let X equal the number of successes that occur in n Bernoulli trials then

$$P(X = x) = \begin{cases} \binom{n}{x} p^x (1 - p)^{n-x} & x \in \{0,1, \dots, n\} \\ 0 & x \notin \{0,1, \dots, n\} \end{cases}$$

where $p = P(\text{success})$.

7.2 Negative Binomial Random Variable

The starting point for the *negative* binomial model is nearly identical to our starting point the binomial model.

We consider a generic experiment that consists of repeated trials where these trials meet the following conditions:

(1) the trials are independent, *i.e*. the outcome in one trial will have no impact on the outcome of a different trial

(2) there are only two possible outcomes for a trial. We will refer to these two possible outcomes as "success" or "failure"

(3) $P(\text{success})$ is the same for every trial. We will use the letter p to represent $P(\text{success})$ for notational simplicity

These are the same first three assumptions about the repeated trials in the binomial model. It is in the fourth and fifth assumptions where the difference between the binomial and the negative binomial shows up.

In the <u>binomial</u> model we took

(4) the total number of trials performed is n (n is fixed, not random)

But in the <u>negative</u> binomial model the

(4) trials continue until we get observe the r^{th} success (r is fixed, not random)

In the <u>binomial</u> model

(5) our interest is in the total number of successes that occur in these n trials

But in the <u>negative</u> binomial model

(5) our interest is in the total number of trials that required to observe r successes

The following chart makes clear the essential distinction between the "Binomial" and "Negative Binomial" models.

	Total Number of Trials Performed	**Total Number of Successes Observed**
Binomial	Fixed (predetermined)	Random
Negative Binomial	Random	Fixed (predetermined)

Theorem

Let X equal the number of (Bernoulli) trials required in order to get a total of r successes. Then

$$P(X=k)=\begin{cases}\binom{k-1}{r-1}p^r(1-p)^{k-r} & k\in\{r,r+1,r+2,\dots\}\\ 0 & k\notin\{r,r+1,r+2,\dots\}.\end{cases}$$

The above formula is called the ***negative binomial distribution***. Note that the range of k (the possible values of k) start at r because it certainly takes at least r trials to reach the r^{th} success. But there is no upper limit on the number of trials it might take you to reach the r success. It is unlikely, but you might be there forever waiting for the r^{th} success to occur.

To justify this formula we start by defining the events:

A: get the r^{th} success on the k^{th} Bernoulli trial

B: get exactly $(r-1)$ successes in the first $(k-1)$ Bernoulli trials

C: get a success on the k^{th} Bernoulli trial.

Then

$$P(\text{get the } r^{th} \text{ success on the } k^{th} \text{ Bernoulli trial}) = P(A).$$

Remember that the trials stop right after observing the r^{th} success. This means that the **last trial had to be a success** (in fact, the r^{th} success). Hence saying that we got the r^{th} success on the k^{th} trial is equivalent to saying that we got $r-1$ successes in the first $k-1$ trials and then got a success on the k^{th} trial.

That is,

$$P(\text{get the } r^{th} \text{ success on the } k^{th} \text{ Bernoulli trial})$$

$$= P(A) = P(B\cap C).$$

But from Chapter 1,

$$P(B\cap C) = P(C|B)P(B).$$

Now look again at the definition for the event B: getting exactly $(r-1)$ successes in the first $(k-1)$ Bernoulli trials. We can write this as

$$P(B) = P(X = r-1 \text{ successes in } n = k-1 \text{ trials})$$

to help bring home the point. We are asking for the probability of a given number of successes $(r-1)$ in a given number of trials $(k-1)$. But that is what the Binomial probability gives us. That is, B follows the *binomial* distribution. So,

$$\begin{aligned} P(B) &= \binom{k-1}{r-1} p^{r-1}(1-p)^{(k-1)-(r-1)} \\ &= \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r}. \end{aligned}$$

Now look at the factor $P(C|B)$, which is asking for the probability of getting a success on the k^{th} trial conditional on the event that we got exactly $(r-1)$ successes in the first $(k-1)$ trials. But remember that all trials are independent. The probability of getting a success on ***any*** trial is p. That is, $P(C|B) = p$.

So, we have established that

$$\begin{aligned} P(X = k) &= P(\text{get the } r^{th} \text{ success on the } k^{th} \text{ Bernoulli trial}) \\ &= P(C|B) \cdot P(B) \\ &= p \cdot \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r} \\ &= \binom{k-1}{r-1} p^{r}(1-p)^{k-r}. \end{aligned}$$

■

7.3 Multinomial Probability Distribution

The ***multinomial distribution*** is an extension of the binomial distribution to the case where there are ***more than two possible outcomes per trial***.

Suppose an experiment consists of repeated trials where these trials meet the following conditions:

(1) the trials are independent, *i.e*. the outcome in one trial will have no impact on the outcome of a different trial

(2) there are k possible outcomes on any trial

(3) $p_j = P(\text{type } j \text{ outcome})$ is the same for every trial, for each $j = 1,2,\dots,k$

Let X_j equal the number of Type j outcomes that occur in n repeated trials. Then

$$P(X_1 = x_1, \dots, X_k = x_k) = \begin{cases} \dfrac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} & \begin{array}{c} x_j \in \{0,1,2,\dots,n\}, j = 1,\dots,k \\ x_1 + x_2 + \cdots + x_k = n \end{array} \\ 0 & \text{else.} \end{cases}$$

7.4 Exercises for Problems Involving Repeated Independent Trials

Exercises 1 to 24 were all taken from old MSHSML tests. Before trying to solve them, read them several times. As you are reading them, consider how each of them depends in various ways on a binomial, negative binomial or multinomial random variable.

1. (5C191) Jamie flips four fair coins. Determine exactly the probability that she gets more heads than tails.

2. (5C194) Amy and Ben each have biased coins but they are unfair in different ways: Amy's coin comes up <u>heads</u> only $1/3$ of the time and Ben's coin comes up <u>tails</u> only $1/3$ of the times. They both flip their respective coins twice. Determine exactly the probability that they get the same number of heads.

3. (5C174) A server noted that when diners have cherry pie for dessert, 3 out of 5 will leave a big tip. If there are 6 diners who have cherry pie, what is the probability that a big tip is left by exactly 4 of them?

4. (5C153) Three quarters and three dimes are tossed in the air. Determine exactly the probability that the same number of quarters and dimes turn up heads.

5. (TT156) An unfair coin lands tails with a probability of $1/5$. When tossed n times, the probability of exactly three tails is the same as the probability of exactly 4 heads. What is the value of n?

6. (5C141) What is the probability that a fair coin, flipped three times, will land all tails?

7. (5C142) Determine exactly the probability that a student, by randomly guessing, achieves a score of 3 out of 5 on a pop quiz whose questions are multiple choice with four choices each.

8. (MB0610) A die is rolled six times. What is the probability of getting either a 1 or a 6 on at least three rolls?

9. (5A054) (a) Amy, Beth, and Christine toss a coin 15,16 and 17 times respectively. Which girl is least likely to get more heads than tails?

10. (5A054) (b) Amy, Beth and Christine toss a coin 18,19 and 20 times respectively. Which girl is least likely to get more heads than tails?

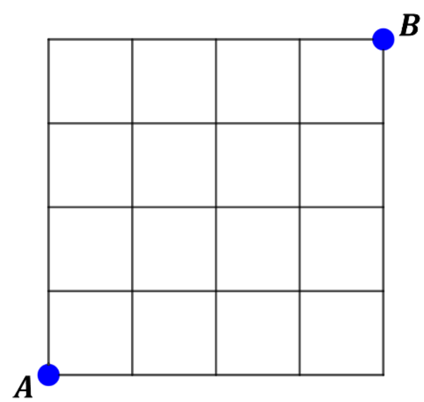
11. (5C054) Sarah was sent to get 8 cans of soda to have on hand for the study session. When she got to the machine, she found that she had six choices. Remembering that she had a die in her purse for some homework for her probability class, she decided to roll it 8 times, choosing the first flavor if a 1 came up, etc. for all six possibilities. Using this scheme, what is the probability that she gets exactly four diet cokes?

12. (TI0512) John flips a fair coin 12 times. What is the probability that he gets more heads than tails?

13. (TI008) If the probability of A beating B is $3/5$ (so the probability of losing to B is $2/5$), find the probability of A winning exactly 8 of 12 games the teams play during a season.

14. (5C974) In the game of Zonk, one throws six dice on each turn. What is the probability that on a random throw, exactly three dice will be 2's?

15. (TT974) John starts at A and walks toward B (eight blocks), moving only to the right or up (east or north). Jane starts at B and walks toward A (eight blocks), moving only to the left or down (west or south). If they start at the same time and walk at the same rate, and they each choose their direction at intersections (when they have a choice) in a random manner, what is the probability they will meet along the way?



16. (TC942) A fair coin is flipped 8 times. What is the probability of obtaining exactly 3 heads?

17. (TI9213) A pitcher throws a sequence of pitches and a blind umpire calls each one a ball or strike at random, meaning that independent of previous calls, the probability that a

particular pitch will be called a strike is $1/2$. The following questions assume that the batter does not swing at any pitch (*i.e.* is taking).

(a) What is the probability that the first three pitches will be called strikes?

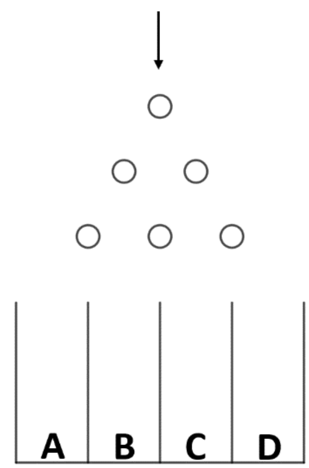
(b) What is the probability that after two pitches, the count will be 1 ball, 1 strike?

(c) What is the probability that the batter will be called out (have 3 strikes called) before a fifth pitch has been thrown? (No more pitches will be thrown to a batter once three strikes have been called.)

(d) What is the probability that the umpire will call 3 strikes before calling 4 balls.

18. (TI8912) It is rumored that Eric Vee, Math Leaguer from Moose Lake, beats his father Stan, Math League coach at Barnum, about $2/3$ of the time in chess. Assuming this to be true, what is the probability that Eric will beat his father in a best of three match (which, in the usual fashion, is over as soon as someone wins two games)?

19. (5D852) A ball in a certain pinball machine bounces to your left with probability $1/3$ and to your right with probability $2/3$ whenever it comes to a bumper. It falls through the configuration shown, hitting a bumper at each level, finally coming to rest in one of the slots A, B, C or D. What is the probability that the ball will come to rest in slot C?



20. (TI8414) Oddsmakers believe the probability of the Tigers beating the Cubs in a single game is 0.6. The Tigers and the Cubs are to play a series of up to seven games, the first team to win four games being declared the champion.

(i) What probability should they assign to the outcome of the Cubs winning the series in exactly 6 games?

(ii) What probability should they assign to the outcome that the Tigers will win the series in less than 7 games?

21. (3T836) Rodney Roller is down to his last turn in a Yahtzee game. He has two 6's on the table, but he needs one more. He gets to roll three dice. What is the probability that Rodney will get at least one 6 when he rolls the three dice?

22. (5C182) When tossing two 8-sided dice, the sides of each die being numbered 1 through 8, determine the probability of rolling two numbers a and b, such that $|a-b|<3$.

23. (TC894) A fair die is rolled three times. What is the probability that two of the three rolls, but not all three will be equal?

24. Every hospital has backup generators for critical systems should the electricity go out. Independent but identical backup generators are installed so that the probability that at least one system will operate correctly when called upon is no less than 0.99. Let n denote the number of backup generators in a hospital. How large must n be to achieve the specified probability of at least one generator operating, if the probability that any backup generator will work correctly is 0.95?

25. You flipped a coin 10 times and got 8 heads and this made you wonder if this was a "fair" coin (i.e. $50/50$ chance of heads/tails). To find out you decide to run an experiment consisting of 5 replications of flipping this coin 10 times (in each replication) in a controlled manner. If this really is a fair coin, what is the probability of observing at least one replication where you observe at least 8 heads?

26. A tire maker knows from past experience that 20% of their top-of-the-line brand tire will not satisfy the conditions of their warranty. The tire maker also knows from experience that only 10% of their customers of this top-of-the-line brand tire will bother to make a claim on their warranty when they could. What is the probability that a particular tire store selling this top-of the-line brand tire will not see any appropriate claims against the warranty from their next 30 customers buying this brand of tire?

27. Consider a multiple-choice examination with 10 questions, each of which has 4 possible answers. If a student knows the correct answer with probability 0.8 and guesses with probability 0.2, what is the probability that this student will score at least an 80? Assume each question is worth 10 points and no partial credit is given. Also assume that the questions are answered independently, that is, whether this student answers question #3 correct or incorrect will not influence whether they answer question #7 correct or incorrect?

28. A fair die is rolled n times. What is the smallest value of n such that the probability of getting at least one six in these n rolls is 0.95 or higher?

29. Suppose you know from experience that 1% of the parts coming off an assembly line at a local manufacturing plant are defective.

    (i) What is the probability that a lot of 500 will have less than 3 defectives in it?

(ii) Suppose you cannot tell by just looking whether a part is defective but rather have to subject the part to a test in order to tell. What is the probability that you will have to test more than 150 parts before you find a defective one?

30. An airline knows from experience that 10% of the people holding reservations on a given flight will not appear. The plane holds 90 people. If 95 reservations have been sold, what is the probability that the airline will be able to accommodate everyone appearing for the flight?

31. Suppose that a four-engine plan can fly if at least two engines work and suppose that a two-engine plan can fly if at least one engine works. Would you rather fly on a four-engine or two-engine plan?

32. During the 1978 baseball season, Pete Rose of the Cincinnati Reds set a National League record by hitting safely in 44 consecutive games. Assume that Rose is a 300 hitter, (i.e. the probability he hits safely on any given time at bat is .300) and assume that he comes to bat four times each game. If each at bat is assumed to be an independent event, what is the probability of hitting safely in 44 consecutive games?

33. A coin is altered so that the probability that it lands on heads is less than $1/2$ and then the coin is flipped four times, the probability of an equal number of heads and tails is $1/6$. What is the probability that the coin lands on heads? (2010 AMC 12 A Problem 15)

34. Coin A is flipped three times and coin B is flipped four times. What is the probability that the number of heads obtained from flipping the two fair coins is the same? (2004 AMC 10a Problem 10)

35. A coin with an unknown probability of landing heads is tossed ten times and lands on heads exactly three times. Find the conditional probability that the first toss landed on heads. (Source: https://math.la.asu.edu/~jtaylor/teaching/Spring2017/STP421/problems/ps2-solutions.pdf)

Solution

Let p equal the unknown probability that this coin will on heads on any toss. Let E be the event of getting exactly 3 heads in 10 tosses. By the binomial distribution we have

$$P(E) = \binom{10}{3} p^3(1-p)^7.$$

Let H_1 be the event that the first toss landed on heads. Then

$$P(H_1|E) = \frac{P(E|H_1)P(H_1)}{P(E)}.$$

The distribution of E conditional on the information that the first toss landed on heads is the same as the unconditional probability of getting 2 heads in 9 tosses. But this is again modeled by the binomial distribution. That is,

$$P(E|H_1) = \binom{9}{2} p^2 (1-p)^7.$$

Also, we know that $P(H_1) = p$. Therefore,

$$P(H_1|E) = \frac{P(E|H_1)P(H_1)}{P(E)} = \frac{\binom{9}{2} p^2 (1-p)^7 \cdot p}{\binom{10}{3} p^3 (1-p)^7}$$

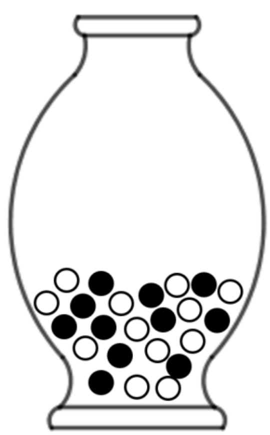
$$= \frac{\binom{9}{2}}{\binom{10}{3}} = \frac{9!}{10!} \cdot \frac{3!}{2!} = \frac{3}{10}.$$

■

8 Hypergeometric Random Variables

Suppose an urn contains w white balls and b black balls. You reach into this urn and randomly select a ball, note its color but ***then do not return the ball to the urn***. You repeat this process for a total of n draws. Let X equal the number of times you select a white ball and let $P(X = x)$ represent the probability that you get exactly x white balls in these n draws.

By the way, in some hypergeometric problems the phrasing might be to select n balls "one-by-one without replacement" and in others the phrasing might be to select n balls "all at once". From a probability point of view, these are *equivalent* if your only objective is to count the number of white balls among the n balls drawn.



Then

$$P(X = x) = \begin{cases} \dfrac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}} & x \in \{\max\{0, n-b\}, \dots, \min\{w, n\}\} \\ 0 & x \neq \{\max\{0, n-b\}, \dots, \min\{w, n\}\}. \end{cases}$$

In probability and statistics, whenever $P(X = x)$ is determined by the above formula we say that X is a **hypergeometric random variable** or that X follows the **hypergeometric distribution**.

First off, let's plug in some numbers to guide us through the messy looking range of X.

$$x \in \{\max\{0, n-b\}, \dots, \min\{w, n\}\}.$$

For now, let α (alpha) represent the smallest number and let ω (omega) represent the largest number of white balls we can get when drawing n times without replacement from an urn containing w white balls and b black balls.

What we want to verify is that in all situations $\alpha = \max\{0, n-b\}$, the max<u>imum</u> of the two numbers 0 and $n-b$ and that $\omega = \min\{w, n\}$, the min<u>imum</u> of the two numbers w and n.

Let's consider various cases:

<u>Case 1a</u>. $(n = 3, w = 4, b = 6)$

Suppose we draw $n = 3$ times without replacement from an urn containing $w = 4$ white balls, $b = 6$ black balls. In this scenario we could draw as few as 0 white balls.

<u>Case 1b</u>. $(n = 3, w = 4, b = 1)$

Suppose we draw $n = 3$ times without replacement from an urn containing $w = 4$ white balls, $b = 1$ black ball. Obviously, it is always true that we cannot draw less than 0 white balls but in this case, we can also see that it would be *impossible* to draw exactly 0 white balls. While we might get the $b = 1$ black ball on one of the $n = 3$ draws, we would have then run out of black balls and the other $n - b = 3 - 1 = 2$ draws would necessarily result in white balls.

What Cases 1a, 1b show us is that the number of white balls drawn cannot be less than <u>*either*</u> 0 or $n - b$. But this is logically equivalent to the single statement that the number of white balls drawn cannot be less than the <u>*maximum*</u> of 0 and $n - b$.

That is, $\alpha = \max\{0, n-b\}$.

<u>Case 2a</u>. $(n = 3, w = 4, b = 6)$

Here again we will draw $n = 3$ times without replacement from an urn containing $w = 4$ white balls, $b = 6$ black balls. By drawing a white ball each time draw we could draw as many as $n = 3$ white balls.

<u>Case 2b</u>. $(n = 3, w = 2, b = 6)$

In this case it would be *impossible* to get a white on each of the $n = 3$ draws because we are sampling without replacement and there are only $w = 2$ white balls in the urn.

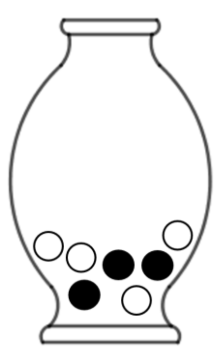
What Cases 2a, 2b show us is that the number of white balls drawn cannot be greater than ***either*** n or w. But this is logically equivalent to the single statement that the number of white balls drawn cannot be greater than the ***minimum*** of 0 and w.

That is, $\omega = \min\{w, n\}$. Hence, we have established that $x \in \{\max\{0, n - b\}, \dots, \min\{w, n\}\}$.

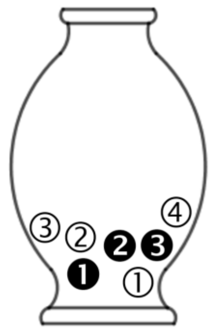
Derivation of the Hypergeometric Probability Distribution

$$P(X = x) = \begin{cases} \dfrac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}} & x \in \{\max\{0, n-b\}, \dots, \min\{w, n\}\} \\ 0 & x \neq \{\max\{0, n-b\}, \dots, \min\{w, n\}\}. \end{cases}$$

We will start our derivation by focusing on the special case of an urn with $w = 4$ identical white balls and $b = 3$ identical black balls. If we randomly take out $n = 3$ balls from this urn, what is the probability of getting $x = 2$ white balls (and 1 black ball)?

	What is the probability of getting 2 white balls and 1 black ball if we randomly select 3 balls from this urn with unlabeled balls?

But before we even being on solving this, let's introduce the related problem when the balls are labeled. What is the probability of getting $x = 2$ white balls (and 1 black ball) if we randomly take out $n = 3$ balls from the urn below with $w = 4$ labeled white balls and $b = 3$ labeled black balls?

	What is the probability of getting 2 white balls and 1 black ball if we randomly select 3 balls from this urn with labeled balls?

Practical experience tells us that the labels will not change how likely we are to get 2 white and 1 black if we randomly draw out 3 balls from this urn.

However, once we attach labels, we can view this as an urn where all the $w + b$ balls are *distinguishable*. How does it help us to how distinguishable balls in the urn?

The basic formula for *combinations* tells us there are $\binom{7}{3}$ distinct samples of size $n = 3$ from an urn of $w + b = 7$ **distinct** objects. The combinations formula does not apply unless all the objects are distinct.

Furthermore, because we are selecting balls at random, all $\binom{7}{3}$ samples are equally likely.

How many distinct ways can we select $x = 2$ white balls from this urn with $w = 4$ white balls? The same formula for combinations tells us there are $\binom{4}{2}$ distinct ways to select the $x = 2$ white balls from this urn. By the same reasoning there are $\binom{3}{1}$ ways to select $n - x = 3 - 2 = 1$ black balls from this urn.

Because we want to select $x = 2$ white balls **and** $n - x = 1$ black ball, the product rule for counting tells us there $\binom{4}{2} \cdot \binom{3}{1}$ ways to do both.

That is, there are a total of $\binom{4}{2}\binom{3}{1}$ ways to select 2 white balls and 1 black ball from this urn with 4 distinct white balls and 3 distinct black balls.

And because the $\binom{7}{3}$ total number of ways to select $n = 3$ balls (without any restriction on colors) are all equally likely, the ratio

$$\frac{\binom{4}{2}\binom{3}{1}}{\binom{7}{3}}$$

of counts gives the probability of getting 2 white and 1 black when we select 3 balls from the above urn (whether the labels are attached or not).

It is straightforward to extend this to the general case where we draw n times (without replacement) from an urn containing w white balls and b black balls and we want to find the probability of getting x white (and hence $n - x$ black) balls.

$$\frac{\binom{4}{2}\binom{3}{1}}{\binom{7}{3}} \longrightarrow \frac{\binom{w}{x}\binom{b}{n-x}}{\binom{w+b}{n}}.$$

8.1 Multivariate Hypergeometric Model

Suppose an experiment consists of drawing marbles at random and without replacement from an urn containing initially containing r_j Type j balls, $j = 1,2,\ldots,k$.

Let X_j equal the number of Type j balls selected in n draws. Then

$$P(X_1 = x_1, \ldots, X_k = x_k) = \begin{cases} \dfrac{\binom{r_1}{x_1}\binom{r_2}{x_2}\cdots\binom{r_k}{x_k}}{\binom{r_1+r_2+\cdots+r_k}{x_1+x_2+\cdots+x_k}} & \begin{array}{c} x_j \in \{0,1,2,\ldots,n\}, j = 1,\ldots,n \\ x_1 + x_2 + \cdots + x_k = n \end{array} \\ 0 & \text{else} \end{cases}$$

8.2 Exercises for Hypergeometric Random Variables

<u>Read</u> each of the following problems which were all taken from old MSHSML tests and consider how each of them depends in various ways on a hypergeometric random variable. (*i.e.* Identify what is playing the role of the white balls and the black walls, identify how is it made clear we are sampling *without* replacement and identify how many balls we are sampling in total).

1. (5T194) Johnny has a drawer containing two green socks, two red socks, two black socks, two white socks and two blue socks. Johnny likes to wear matching socks, but he is color blind and cannot distinguish red from green. Johnny randomly pulls our four socks from his drawer. Determine exactly the probability that Johnny does not get a pair of socks of the same color, but *thinks* he does.
2. (TI199) I have a bag with 21 beads of various colors, the majority being green. If I remove two beads (without replacement), the probability I will end up with one green and one non-green is $3/7$. How many green beads were in the bag originally?
3. (5C171) Sal's drawer contains 4 black socks and 2 red socks. If he chooses two socks at random without replacement, what is the probability he chooses two socks of the same color? Express your answer as a quotient of two relatively prime integers.
4. (5C102) An ordinary deck of 52 playing cards is shuffled and 2 cards are dealt face up. Calculate the probability that at least one of these is a spade. *[NYCC, Fall 1984]*
5. (MB097) Ten playing cards are laying face down on a table. Exactly three of them are aces. You get to turn over two cards. What is the probability that at least one of them is an ace?
6. (TI086) A hand of 5 cards is drawn from a standard deck of 52 cards. Find the probability of getting exactly one ace.
7. (TT082) A hand of 5 cards is drawn from a standard deck of 52 cards. Successful participants in this morning's Invitational found that the probability of getting exactly one ace is 0.299474. You get one point for finding the probabilities (again to six decimal places) of each of the other possibilities.
8. (TC063) A cooler contains 4 bottles of carbonated mineral water and 6 bottles of non-carbonated spring water. Not realizing there is a choice, Alicia and Beth each plunge their hands into the ice and grab a bottle. What is the probability that both girls get the same kind of water?
9. (5C052) If you draw three cards from an ordinary 52 card deck, what is the probability of drawing at least one of the twelve face cards (*i.e.* a Jack, Queen, or King)?

10. (5C012) Nine playing cards are lying face down on a table. Exactly two of them are Aces. You pick up two cards. What is the probability that at least one of them will be an Ace?

11. (5T003) A box contains eleven balls numbered $1,2,3,\ldots,11$. If six balls are drawn simultaneously at random, what is the probability that the sum of the numbers on the balls drawn is odd? [1984 AHSME, prob. 19]

12. (TI0010) Alicia grabs two bagels from a bag containing 10 whole wheat and 6 pumpernickel bagels. What is the probability that she gets 1 of each?

13. (5C991) Cards are turned over one at a time (without replacement) from an ordinary well-shuffled decks (52 cards). What is the probability that the first two cards are aces? (There are four aces in a deck.)

14. (TC994) A piggy bank contains 1 quarter, 3 dimes, 2 nickels, and 1 cent. Needing 46 cents, Aaron shakes out four coins. Assuming the coins are equally likely to be shaken out, what is the probability that he has shaken out exactly 46 cents?

15. (TI9911) A bag contains white marbles and black marbles. If you were to reach in without looking and pull out one marble, the probability that it would be white is $1/4$. However if you were to pull out two marbles, the probability that both would be white is $1/18$. How many white marbles are in the bag?

16. (5T985) Three cards are drawn at random from a standard deck of 52 playing cards. Calculate the probability (rounded to four decimal places) that at least one of these three cards is a face card. (Here, a face cards is a jack, queen, or king. The motivation for this problem is the question of whether or not one would give even odds that at least one of the three cards is a face card.)

17. (TT983) A bag contains 16 billiard balls, some black and the remainder white. Two balls are drawn at the same time. It is equally likely that the two balls will be the same color as different colors. How are the balls divided within the bag?

18. (5C972) In your drawer there is a pair of green gloves and a pair of brown gloves. You reach in with your eyes closed and pull out two gloves at random. What is the probability that you have a matched pair?

19. (5T913) I have n nuts in my hand, m of which will fit a bolt on my bicycle. If I select 2 of the nuts at random, there is an even chance (*i.e.* the probability is $1/2$) that both will fit the bolt. What is the smallest value of n?

20. (5C902) There are 4 black socks and 6 brown socks all mixed together in a drawer in a dark room. If Pat grabs two socks from the drawer and takes them into the light, what is the probability that they will match?

21. (5C903) There are 4 black socks and 6 brown socks all mixed together in a drawer in a dark room. If Pat grabs three socks from the drawer and takes them into the light,

    a) what is the probability of a match?

    b) what is the probability of at least two black socks?

22. (5T881) Amy, who puts only nickels and dimes in her bank, currently has an even number of nickels. When two coins are drawn at random from this bank, the probability that both are dimes is $1/2$. What is the smallest amount of money that Amy might have?

23. (TD883) A box contains nine balls numbered 1 through 9. If five balls are drawn at random and without replacement, what is the probability that the sum of the numbers on the balls will be odd?

24. Suppose that 100 cards marked $1, 2, \dots, 100$ are randomly arranged in a line. What is the probability that exactly 8 of the first 20 cards in this line are marked with an even number?

25. Suppose you play a lottery where you choose five <u>different</u> numbers from the integers $1, 2, \dots, 45$. Then you choose one number from $1, 2, \dots, 45$. This last number is often called the *powerball* and <u>can</u> (but does not have to) repeat one of the first five numbers you chose. Lottery officials choose five numbers and the powerball in the same way. What is the probability that you match 3 of the first 5 but do not match the powerball?

26. 15 women and 9 men were eligible for promotion at a university and from this group 2 women and 4 men were selected for promotion. Is there reason to suspect gender bias?

27. An urn contains 3 red, 5 blue and 7 yellow balls. You pull out 4 balls (without replacement). What is the probability you get exactly 2 yellow balls?

28. A Hypergeometric Problem Nested Inside a Binomial Problem

    Imagine a game that consists of two stages. In the first stage you have to draw 3 balls at random and without replacement from an urn containing 4 black and 8 white balls. According to the rules, in order to be able to proceed to the second stage of this game you have to get at least one black ball in this sample of size three during the first phase.

    (a) Out of ten players (all starting with the same urn of 4 black and 8 white balls), what is the probability that five or more will make it to the second stage?

    (b) How many out of these ten players do you expect to make it to the second stage?

29. Suppose there are two identical urns and that each urn contains N balls numbered from 1 to N. One of these urns is given to Mary and the other is given to Bob. Mary reaches into her urn and randomly selects n balls without replacement and makes a list of the numbers

she got. Bob reaches into his urn and randomly selects k balls without replacement and makes a list of the numbers he got.

Let X equal how many numbers are on both Mary's and Bob's list. Find $P(X = x)$.

30. In the game of Texas Hold'em, players are each dealt two private cards, and five community cards are dealt face-up on the table. Each player makes the best 5-card hand they can with their two private cards and the five community cards. What is the probability that a particular player can make a flush of spades (i.e. 5 spades)?

https://brilliant.org/wiki/hypergeometric-distribution/

9 Conditional Probability

The Phrase "Given that" is Ambiguous in Probability

Consider the following standard conditional probability problem:

> You flip a fair penny and a fair nickel. What is the probability that both coins landed heads ***given that*** at least one coin landed heads?

The phrase "given that" is a key identifier of conditional probability problems. Precisely what do we mean by the phrase "given that"?

Let's consider three related storylines. In each case you are given the information that at least one coin landed heads. As you read these storylines think about whether and/or how the difference in the storylines impacts the probability that both coins landed heads

> **Storyline 1.** Your teacher asks you to close your eyes and to flip a fair penny and a fair nickel. Your teacher sees the results but before you or anyone else gets the chance to see how the coins landed, your teacher covers each coin with a playing a card that completely hides the results. Your teacher tells you that the penny is on the left and the nickel is on the right.
>
> 
>
> Now you leave the room and your friend comes in. **Your friend looks under *both* cards and then covers them back up.**
>
> Finally, you come back into the room and ask your friend "Did I get at least one heads?" Your friend truthfully answers "Yes" and also tells you the exact process they used to learn this information.
>
> Given what you now know, what is the probability that both coins landed heads?

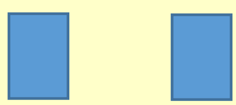
Storyline 2. Your teacher asks you to close your eyes and to flip a fair penny and a fair nickel. Your teacher sees the results but before you or anyone else gets the chance to see how the coins landed, your teacher covers each coin with a playing a card that completely hides the results. Your teacher tells you that the penny is on the left and the nickel is on the right.

Now you leave the room and your friend comes in. **Your friend has decided ahead of time that they will only look to see the outcome of the coin under the card on the left.**

Finally, you come back into the room and ask your friend "Did I get at least one heads?" Your friend truthfully answers "Yes" and also tells you the exact process they used to learn this information.

Given what you now know, what is the probability that both coins landed heads?

Storyline 3. Your teacher asks you to close your eyes and to flip a fair penny and a fair nickel. Your teacher sees the results but before you or anyone else gets the chance to see how the coins landed, your teacher covers each coin with a playing a card that completely hides the results. Your teacher tells you that the penny is on the left and the nickel is on the right.



Now you leave the room and your friend comes in. **Your friend looks under a randomly selected card (assume 50-50 chance they looked under the card on the left or the card on the right) and then puts the card back over the coin.**

Finally, you come back into the room and ask your friend "Did I get at least one heads?" Your friend truthfully answers "Yes" and also tells you the exact process they used to learn this information.

Given what you now know, what is the probability that both coins landed heads?

In each storyline you flipped two fair coins and you were ***given that*** at least one coin landed heads.

- Does it matter ***how*** your friend came to get this information they gave you? Ans: Yes.
- Can it change the probability that both coins landed heads? Ans: Yes.
- Assuming that it does, does it change ***how*** we go about solving each problem? Ans: Yes.

The key to sorting this all out depends on **establishing the appropriate sample space and the appropriate conditional sample space** for each storyline.

Best Advice: Even in a contest setting where time is critical, carefully enumerate and account for *all* random aspects of the story when building your sample space.

The sample space for the initial flips of the penny and nickel would just be

penny nickel

$$S = \{(H,T),(T,H),(H,H),(T,T)\}.$$

Note: We included (H,T) as well as (T,H) in the sample space because they are distinguishable.

In Storylines 1 and 2 there are no additional experiments performed with random outcomes. So, the above sample space S applies to both storyline 1 and storyline 2.

Because both the penny and the nickel are assumed to be fair coins, all four of these outcomes are equally likely and hence each outcome has probability $1/4$.

However, in Storyline 3, your friend *randomly* selects whether they will look under the left or right card. We will designate these as L and R respectively. This additional random aspect of the story needs to be integrated into your sample space. So, for Storyline 3, our new sample space S_3 becomes:

penny nickel card lifted

$$S_3 = \Big\{(H,T,L),(H,T,R),(T,H,L),(T,H,R),(H,H,L),(H,H,R),(T,T,L),(T,T,R)\Big\}.$$

Because both the penny and the nickel are assumed to be fair coins and because your friend randomly selected whether to look under the left or right card, all eight of these outcomes are equally likely and hence each outcome has probability $1/8$.

In Storyline 1 you were given that at least one coin was a heads. So, C_1, the reduced (conditional) sample space for Storyline 1 is therefore,

penny nickel

$$C_1 = \Big\{(H,T),(T,H),(H,H)\Big\}.$$

As all outcomes in the unconditional sample space for Storyline 1 were equally likely, all outcomes in its conditional sample space C_1 must remain equally likely.

Only one of these three equally likely conditional outcomes has both coins land heads. Thus, for **Storyline 1, the probability that both coins landed heads equals $\mathbf{1/3}$**.

In Storyline 2 you were given that the penny landed heads (the coin under the left card). So, C_2, the reduced (conditional) sample space for Storyline 2 is therefore,

penny nickel

$$C_2 = \Big\{(H,T),(H,H)\Big\}.$$

As all outcomes in the unconditional sample space for Storyline 2 were equally likely, all outcomes in its conditional sample space C_2 must remain equally likely.

In just one of these two equally likely outcomes will both coins be heads. That is, for **Storyline 2, the probability that both coins landed heads equals $\mathbf{1/2}$**.

In Storyline 3 you were given that the coin under a randomly picked card was a heads. The outcomes $(H,T,R),(T,H,L),(T,T,L)$ an (T,T,R) because in each of these cases your friend the coin your friend would have seen was a T (tails).

So, C_3, the reduced (conditional) sample space for storyline three is therefore,

penny nickel card lifted

$$C_3 = \Big\{(H,T,L),(T,H,R),(H,H,L),(H,H,R)\Big\}.$$

As all outcomes in the unconditional sample space for Storyline 3 were equally likely, all four outcomes in its conditional sample space C_3 must remain equally likely.

In two of these four remaining equally likely outcomes both coins are heads. That is, for **Storyline 3, the probability that both coins landed heads equals $\mathbf{2/4 = 1/2}$**.

Summarizing, we have shown that in Storylines 2 and 3 the probability that both coins landed heads equaled $1/2$. However, in Storyline 1 the probability that both coins landed heads equaled $1/3$.

Is this what you expected?

Two Children Problem

9. (5C981) A woman goes to visit the house of some friends whom she has not seen in many years. She knows that, besides the two married adults in the household, there are two children of different ages. But she does not know their genders. When she knocks on the door of the house, a boy answers. What is the probability that the other child is a boy?

Solution

This problem is known as the "Two Children Problem" and shows up in many forms on puzzle blogs. This problem is notorious for tripping up problem solvers intuition and is also notorious for tripping up problem writers for creating ambiguity in the wording.

To make the solution provided here complete the underlying assumptions being used must be stated explicitly.

Assumptions:

(1) the gender determination of all children in a family are independent events

(2) $P(\text{boy}) = P(\text{girl}) = 1/2$ for all children in a family

(3) $P(\text{younger child answers the door}) = P(\text{older child answers the door}) = 1/2$.

The unconditional sample space must account for the gender of the younger child, the gender of the older child and which child (younger or older) opens the door.

Younger	Older	Opens Door	Probability
Male	Male	Younger	$1/8$
Male	Male	Older	$1/8$
Male	Female	Younger	$1/8$
Male	Female	Older	$1/8$

Female	Male	Younger	$1/8$
Female	Male	Older	$1/8$
Female	Female	Younger	$1/8$
Female	Female	Older	$1/8$

Which of these eight outcomes are **<u>not</u>** possible once we see that a boy answers the door?
Answer: (Male, Female, Older), (Female, Male, Younger), (Female, Female, Younger), (Female, Female, Older)

Removing these four outcomes and rescaling the remaining possibilities leaves us with our conditional sample space and their probabilities.

Younger	**Older**	**Opens Door**	**Conditional Probability**
Male	Male	Younger	$1/4$
Male	Male	Older	$1/4$
Male	Female	Younger	$1/4$
Female	Male	Older	$1/4$

Now we can easily find $P(\text{other child is boy}|\text{boy opens the door})$. We just need to sum the conditional probabilities where both children are boys.

$$P(\text{other child is boy}|\text{boy opens the door}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

■

The solution provided by MSHSML for this problem was $1/3$. This answer is correct if we amend the storyline as follows: "Suppose the visitor knows there are two children in the family they are visiting and when the visitor knocks the mother answers the door. The visitor then asks the mother if she has at least one boy and the mother informs her that she does have at least one boy but does not elaborate further. Based on this information what is the probability that the mother has two boys?"

Solution

Let A be the event that the younger child is a boy and let B be the event that the older child is a boy. Then the question in this amended form becomes

$$P(\text{two boys}|\text{at least one boy})$$

$$= P\big((A \cap B)\big|(A \cup B)\big)$$

$$= \frac{P\big((A \cap B) \cap (A \cup B)\big)}{P(A \cup B)}$$

$$= \frac{P(A \cap B)}{P(A \cup B)}$$

$$= \frac{P(A \cap B)}{P(A) + P(B) - P(A \cap B)}$$

$$= \frac{P(A)P(B)}{P(A) + P(B) - P(A)P(B)}$$

$$= \frac{\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}{\left(\frac{1}{2}\right) + \left(\frac{1}{2}\right) - \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)}$$

$$= \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}.$$

■

Notice that in both the original and the amended storylines the visitor learns that the family they are visiting has at least one son. The difference is in ***how*** this information is learned. In the original version this information is learned as the outcome of a random event (namely, which child answers the door). In contrast, in the amended version this information is learned as "given information" and not through any random event.

Multiplication Rule

Let $A_1, \dots, A_n$ be events such that $P(A_1 \cap \cdots \cap A_n) > 0$. Then

$$P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P\big(A_3\big|(A_1 \cap A_2)\big) \cdots P\big(A_n\big|(A_1 \cap \cdots \cap A_{n-1})\big)$$

Rule of Total Probability

Let $\{H_n\}, n = 1,2, \ldots$ be a sequence of events such that

$$H_i \cap H_j = \emptyset, \;\; i \neq j \;\text{ and }\; H_1 \cup \cdots \cup H_n = \Omega$$

Suppose that $P(H_n) > 0$ for all $n \geq 1$. Then for any event A

$$P(A) = \sum_{j=1}^{\infty} P(H_j)P(A|H_j)$$

Bayes's Rule

Let $\{H_n\}, n = 1,2, \ldots$ be a sequence of mutually exclusive events such that $P(H_j) > 0$, $j = 1,2, \ldots$ and $H_1 \cup \cdots \cup H_n = \Omega$. Let $A \subset \Omega$ be an event with $P(A) > 0$. Then for $j = 1,2, \ldots$

$$P(H_j|A) = \frac{P(H_j)P(A|H_j)}{\sum_{k=1}^{\infty} P\,(H_k)P(A|H_k)}$$

Extended Bayes' Rule

Let $\{H_k\}$ be a sequence of mutually exclusive events such that $(H_k) > 0$, $k = 1,2, \ldots$, and $\bigcup_{k=1}^{\infty} H_k = \Omega$. Let A be an event with $P(A \cap B) > 0$. Then for $j = 1,2, \ldots$

$$P(H_j|\,(A \cap B)) = \frac{P(H_j|B)P(A|(H_j \cap B))}{\sum_{k=1}^{\infty}\, P(H_k|B)\,P(A|(H_k \cap B))}.$$

9.1 Exercises for Conditional Probability

1. (5T184) Two teams, A and B, are playing in a tournament. They will play until one team wins four games (no ties allowed). The probability of either team winning the first game is 50%. For both teams the probability of winning the very next game after winning one is 60%, winning a third game after winning two in a row 70%, and winning a fourth game in a row is 75%. Determine the probability team A wins in exactly 5 games.

2. (5A164) Two people play a game, taking turns drawing a ball from an urn randomly, without replacing them. The urn starts out with 3 black balls and 4 white balls. If a player draws a black ball, their turn is complete. If a player draws a white ball, they must also (nonrandomly) take two black balls from the urn before they complete their turn. A player loses when they cannot complete their turn. If you draw first, determine the exact probability you will win the game.

3. (5C122) Kathy flips a fair coin until she gets three heads in a row. If it is known that she stopped after seven flips, what is the probability her <u>first</u> flip was heads?

4. (5C102) An ordinary deck of 52 playing cards is shuffled and 2 cards are dealt face up. Calculate the probability that at least one of these is a spade. ***[NYCC, FALL 1984]***

5. (5C083) At a carnival game, a black velvet pouch contains either a gold or silver coin, with equal probability of each. You win if you can correctly guess the color of the coin. While you are contemplating your choice, the game operator says, "Let me help you." He drops in two gold coins, shakes the pouch, then reaches in and draws out two gold coins. What is the probability the remaining coin is gold? (Variant of one of Lewis' Carroll's famous "*Pillow Problems*".)

6. (5C063) A bag contains 4 red and 8 white marbles, well mixed. One marble is removed and replaced by two marbles of the other color. Again, after thorough mixing, a marble is draw. What is the probability that this last marble to be removed is red? [original source AHSME 1995 #20]

7. (MB062) A bag contains ten balls numbered $1,2,\ldots,10$, that are thoroughly mixed. Alice draws a ball at random, looks at the number, and replaces the ball in the bag, which is then mixed again. Beth then draws a ball at random. What is the probability that the girls will have drawn the same number?

8. (5C991) Cards are turned over one at a time (without replacement) from an ordinary well-shuffled deck (52 cards). What is the probability that the first two cards are aces? (There are four aces in a deck.)

9. (5D942) A bag of popping corn contains $2/3$ white kernels and $1/3$ yellow kernels. Only $1/2$ of the white will pop, whereas $2/3$ of the yellow will pop. After the corn is popped, you select sight unseen a single popped piece from a well-mixed bag. What is the probability that it will be yellow?

10. (5C902) There are 4 black socks and 6 brown socks all mixed together in a drawer in a dark room. If Pat grabs two socks from the drawer and takes them into the light, what is the probability that they will match?

11. (5T894) Assume that a certain test for cancer is 98% accurate, by which we mean that if the test is administered to 100 people with cancer, it will detect the cancer in 98 of them; and if it is administered to 100 people without cancer, it will show 98 of them to be free of cancer. If this test is given to $10{,}000$ people $(1/2)\%$ of whom actually have cancer, and if Mr. Casetest is one of the people who is told that he has cancer, what is the probability that Mr. Casetest really does have cancer? [Adapted from Innumeracy by John Paulas.]

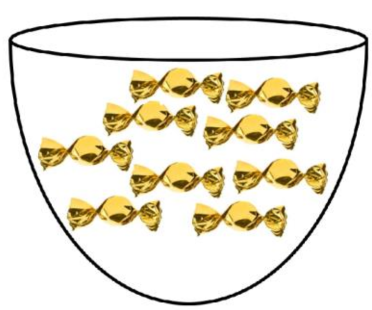
10 Exchangeable Random Variables

A box contains 3 red and 5 white balls. You reach into the box and randomly draw out a ball. What is the probability that it is red? We can all agree it is just $3/8$.

Now suppose we draw out a second ball, without replacing the first one. What is the probability that the second ball is red *taking into account that the first ball was red*. Obviously, there are now 7 balls left and 2 of them are red. So, the probability must be $2/7$.

But can we determine the probability that the second ball is red *without knowing the color of the first ball*? Is there a single answer?

Suppose that all eight balls are wrapped like a piece of candy so that the color of a ball cannot be determined just by inspection.



Now suppose we randomly draw out a wrapped ball from the initial distribution of 3 red and 5 balls and set it aside without unwrapping it. Then suppose we randomly draw out a second wrapped ball. Can we find the probability this second ball is red without knowing the color of the first ball?

To be certain, there is a unique answer to this question. If we repeated this experiment many, many times the percentage of times the second ball is red would begin to converge to a single number and that number is the probability the second ball is red *without taking into account the color of the first ball*.

How can we find this number using the rules of probability? Recall the formula

$$P(B) = P(B|A)P(A) + P(B|A')P(A')$$

which is valid for all generic events A and B. Let B represent the event that the second ball drawn is red and let A represent the event that the first ball drawn is red. In this case, A' represents the event that the first ball drawn is <u>not</u> red or equivalently that the first ball drawn is white.

Now we are in position to find $P(B)$, the probability that the second ball is red <u>unconditionally</u>. That is *without any knowledge of the color of the first ball*. Plugging into the above formula we have,

$$P(\text{second ball drawn out is red})$$

$$= P(\text{second ball red}|\text{first ball red})P(\text{first ball red})$$

$$+ P(\text{second ball red}|\text{first ball white})P(\text{first ball white})$$

$$= \left(\frac{2}{7}\right)\left(\frac{3}{8}\right) + \left(\frac{3}{7}\right)\left(\frac{5}{8}\right) = \frac{3(2+5)}{7 \cdot 8} = \frac{3}{8}.$$

Notice what we have just shown. When we do not take into account the color of the first ball,

$$P(\text{first ball drawn out is red}) = P(\text{second ball drawn out is red}).$$

Was this what you expected? Is it intuitive to you? Would this equality hold on the next draw as well?

In fact, it does. $P(i^{th}$ ball drawn out is red$) = 3/8$ for all $i = 1,2,\ldots,8$. Let's verify this mathematically for just one more step.

Let A_r (A_w) be the event that the first ball is red (white). Let B_r (B_w) be the event that the second ball is red (white) .

Let D be the event that the third ball drawn out is red and define the following four events.

$$C_1 = A_r \cap B_r$$
$$C_2 = A_r \cap B_w$$

$$C_3 = A_w \cap B_r$$
$$C_4 = A_w \cap B_w$$

Notice that $P(C_1 \cup C_2 \cup C_3 \cup C_4) = 1$ and $P(C_i \cap C_j) = 0$ for all $i \neq j$. Hence, it follows from the rule of total probability that

$$P(D) = P(D \cap C_1) + P(D \cap C_2) + P(D \cap C_3) + P(D \cap C_4)$$

and by substitution

$$P(D) = P(D \cap A_r \cap B_r) + P(D \cap A_r \cap B_w) + P(D \cap A_w \cap B_r) + P(D \cap A_w \cap B_w).$$

Now using the multiplication rule,

$$\begin{aligned} P(D) = {} & P\big(D\big|(B_r \cap A_r)\big)P(B_r|A_r)P(A_r) \\ & + P\big(D\big|(B_w \cap A_r)\big)P(B_w|A_r)P(A_r) \\ & + P\big(D\big|(B_r \cap A_w)\big)P(B_r|A_w)P(A_w) \\ & + P\big(D\big|(B_w \cap A_w)\big)P(B_w|A_w)P(A_w). \end{aligned}$$

From here we get

$$\begin{aligned} P(D) &= \left(\frac{1}{6}\right)\left(\frac{2}{7}\right)\left(\frac{3}{8}\right) + \left(\frac{2}{6}\right)\left(\frac{5}{7}\right)\left(\frac{3}{8}\right) + \left(\frac{2}{6}\right)\left(\frac{3}{7}\right)\left(\frac{5}{8}\right) + \left(\frac{3}{6}\right)\left(\frac{4}{7}\right)\left(\frac{5}{8}\right) \\ &= \frac{6 + 30 + 30 + 60}{6(7)(8)} = \frac{126}{6(7)(8)} = \frac{21}{7(8)} = \frac{3}{8}. \end{aligned}$$

As predicted, we have found that the unconditional probability that the third ball is red equals the unconditional probability that the second ball is red which also equals the unconditional probability that the first ball is red.

But what is the intuition behind this and why is this important? First, the intuition behind this result. Imagine we shook the balls around in the box, dumped them onto a table and lined them up.

Without any loss of generality, we can associate the first ball drawn with the ball on the far left. The second ball drawn is then associated with the ball immediately to its right. And so on. Is there anything in this process that would make a red ball more likely to land in one position in this line over another? No. But this means that for all $i = 1,2, \ldots, 8$

$$P(i^{th} \text{ ball drawn out is red}) = P(\text{a red ball lands in the } i^{th} \text{ position}) = 3/8.$$

We can extend this to unconditional problems involving multiple balls. For example, suppose we draw out all eight balls from this box of 3 red and 5 white balls one after the other and without replacement.

What is the probability that the first ball drawn is red and the last ball drawn is white?

In terms of the equivalent row model described above, this probability about balls in a box is equal to the probability that the first ball in the row is red and the last ball in the row is white.

But this must be the same as the probability the first ball in the row is red and the second ball in the row is white which in turn must be the same as the probability that the first ball drawn out of the urn is red and the second ball drawn out of the urn is white. In terms of our earlier notation this is just

$$P(A_r \cap B_w) = P(B_w|A_r)P(A_r) = \left(\frac{5}{7}\right)\left(\frac{3}{8}\right).$$

The formal word for this type of symmetry among events is *exchangeability*.

It is important to not get confused between *exchangeable* events and *independent* events.

Recall that events A and B are independent if $P(B|A) = P(A)$. But while the events A: the first ball is white and B: the second is red are exchangeable, they are not independent. More specifically,

$$P(\text{first ball is white and the second ball is red})$$

$$= P(A_w \cap B_r) = P(B_r|A_w)P(A_w) = \left(\frac{3}{7}\right)\left(\frac{5}{8}\right) = \frac{15}{56}$$

and

$$P(\text{first ball is red and the second ball is white})$$

$$= P(A_r \cap B_w) = P(B_w|A_r)P(A_r) = \left(\frac{5}{7}\right)\left(\frac{3}{8}\right) = \frac{15}{56}.$$

These probabilities agree so these events **are exchangeable**.

Now check for independence. We see that

$$P(\text{second ball is red}|\text{first ball is white}) = \frac{5}{7}$$

but as we determined above

$$P(\text{second ball is red}) = \frac{3}{8}.$$

These probabilities do not agree so these events **are <u>not</u> independent**.

Drawing Straws. Siblings Adam, Bob and Carrie all want to have the first chance to play a new video game they bought together. Their mom cuts a straw into three pieces of slightly different lengths and holds them together in such a way that all three pieces appear to be the same length. They agree that whoever draws the longest straw gets to go first. Adam, Bob and Carrie, in that order, each take a straw from their mom's hand. Is this a fair method for all three siblings?

It is. We could model this as a box containing one red and two white balls. Drawing the red ball would correspond to drawing the longest straw. But we know from exchangeability that for each of $i = 1,2,3$, the $P(i^{th}$ ball drawn is red$) = 1/3$.

That is, Adam, Bob and Carrie have an equal probability of drawing the longest straw.

10.1 Exercises for Exchangeable Random Variables

1. (5C013) A box contains exactly seven chips, four red and three white. Chips are randomly drawn one at a time without replacement until all the red chips are drawn or all the white chips are drawn. What is the probability that the last chip drawn is white?

   [Variation, AMC, 2001, No. 11]

2. (TC013) A bag contains 5 red balls and 4 white ones. They are to be drawn one at a time without replacement until all the red balls are drawn, or all the white ones are drawn. What is the probability that the last ball drawn is white?

3. A box contains 6 red balls, 7 green balls, and 9 yellow balls. Eleven balls are chosen at random one after the other without replacement. What is the probability that the 3rd ball chosen was yellow given that the 9th ball chosen was green?

4. A box contains a white balls and b black balls. Balls are drawn from the box at random and without replacement. What is the probability that all black balls are drawn before the last white ball?

5. You play a game with a friend where you each start with a shuffled deck of 52 cards turned face down. At each turn you both flip over the top card in your deck. If the cards are the same color, you win that turn. If the cards are not the same color, your friend wins that turn. On the next turn you repeat the process but with just the cards remaining in your

decks. What is the probability that you win the fifth turn (the cards are both red or both black on the fifth turn) of this game?

6. Four balls are drawn at random and without replacement from a box containing 10 red balls and 8 white balls and then discarded without noting the color of any of these four balls. At this point four more balls are drawn out at random and without replacement. What is the probability that exactly two of this second set of four balls are red?