Probability Theory I Notes

Dean Quach

Course Info

Probability Theory I

Janurary 17, 2023 to May 9, 2023
MWF — 1:00 pm - 1:50 pm — Wachman 407
Prof. Mr. Wei-Shih Yang
CRN 37229

3 credits $\Longrightarrow$ 6 hours of homework

Text

Introduction to Probability Theory

- David F. Anderson
- Timo Sappalainen
- Benedek Valko

Edition 1

Other Notes

2 midterms $\Longrightarrow$20% each Final $\Longrightarrow$35%

Quizzes $\Longrightarrow$12% each (every other Friday)

HW $\Longrightarrow$3% each (due every Friday)

Office Hours: Wachman 534
MWF 3:00 pm - 4:00 pm
Phone for texting: (215) 983 - 3816

Spring 2023
Math 3031
Section 001

Contents

Chapter 1

Experiments with random outcomes

1.1 Sample Spaces & Probability

Definition. (Sample Space) The set of all possible outcomes of an expirement is called a sample space, denoted by Ω.

Definition. A subset of a sample space Ω is called an "Event".

1.1.1 The Sample Spaces

We first look at the notion of Sample Spaces through some examples. **E.g. 1** - Tossing a coin.

- $\Omega = \{H, T\}$.
- $A = \{H\} \subseteq \Omega$.
- $B = \{T\} \subseteq \Omega$.

- Also by definition, $\emptyset \subseteq \Omega$.
  - This is by definition of a set, but in real life: "It balances on the edge"[1], "lost the coin", "coin rolled away".
- $\Omega \subseteq \Omega$.
  - With $P(\Omega) = 1$, or the probability that "heads or tails" shows up.

E.g. 2 - Rolling a die. (assume 6-sided) Some examples of events include,

- $A = \{even\ \#'s\} = \{2, 4, 6\}$
- $B = \{1, 5\}$
- $\emptyset \Longrightarrow P(\emptyset) = 0$
- $\Omega \Longrightarrow P(\Omega) = 1$

and all events are subsets within a sample space,
i.e. $\{A, B, \emptyset, \Omega\} \subseteq \Omega$.

Notice, $P(\emptyset) = 0$ always.
Also notice, " $\subseteq$ "$\Longrightarrow$ proper subset. So this is why Ω is an event itself.

Question: How many posibilities?
Well you can think of each option within a simple sample

[1] Although you may say, but theres a "chance" it lands on it. The main idea is we weren't measuring/counting that possibility. As in with measure theory. But if you may, think of a table, we were only filling out H or T, and *on the edge* isn't even an option. Therefore implying 0 probability.

space as being on or off, a binary choice $1 \vee 0$. In the mind-set of "Measure Theory", you can treat all these posibilities/probabilities as measures of a sample space, i.e. asking for "all the posibilities" $\implies$ "all the subsets" $\implies$ "the cardinality of the (power set)":= $\mathbb{P}$[2] of the sample space" $= 2^{|\Omega|} = 2^6 = 64$, where in this instance "$||$" asks for the cardinality of a set. So we have a new idea,

Idea (Event Space) The *class* (set) of *all events* of a sample space Ω is denoted by $\mathcal{F}$.

This in itself can be a sample space or an event space, but this is more or less just an informal definition of what an event space should be.

But note, this is for a sample space of $\Omega = \{1, 2, 3, 4, ...n\}$, noting in real life, rolling a die you can only have **1 face up**, and therefore $\mathcal{F} = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$.

E.g. 2 continued - Now for Rolling a die. (assume 6-sided), what is $\mathcal{F}$?

By hand/written out, it would be

$$\mathcal{F} = \Big\{ \emptyset,$$

$$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\},$$

[2]The reason why I will use $\mathbb{P}$ for power set is because in probability theory, $P()$ is already used to describe probability. But for the rest of my notes, I will just specify each time I use it (just in case). I may end up keeping this notation from now on; I will link back to this footnote if need be. Since sometimes its easier to just type $P()$ and also $\mathbb{P}$ is used for polynomial vector space in Theoretical Linear Algebra, but I'll just stop the rant and say "If in this course 3031 notes, and I didn't link back to this, you know what I meant". OH!, also I forgot, but another way is $\mathbb{P}() \Longrightarrow 2^{\Omega}$, without cardinality symbol, thats the best work around I'll probably also use.

$$
\begin{gathered}
\{1,2\},\{1,3\},\{1,4\},\{1,5\},\{1,6\},\\
\{2,3\},\{2,4\},\{2,5\},\{2,6\},\\
\{3,4\},\{3,5\},\{3,6\},\\
\{4,5\},\{4,6\},\\
\{5,6\},
\end{gathered}
$$

$$
\begin{gathered}
\{1,2,3\},\{1,2,4\},\{1,2,5\},\{1,2,6\},\\
\{1,3,4\},\{1,3,5\},\{1,3,6\},\\
\{1,4,5\},\{1,4,6\},\\
\{1,5,6\},\\
\{2,3,4\},\{2,3,5\},\{2,3,6\},\\
\{2,4,5\},\{2,4,6\},\\
\{2,4,6\},\\
\{3,4,5\},\{3,4,6\},\\
\{3,5,6\},\\
\{4,5,6\},
\end{gathered}
$$

$$
\begin{gathered}
\{1,2,3,4\},\{1,2,3,5\},\{1,2,3,6\}\\
\{1,2,4,5\},\{1,2,4,6\},\\
\{1,2,5,6\},\\
\{1,3,4,5\},\{1,3,4,6\},\\
\{1,3,5,6\},\\
\{1,4,5,6\},\\
\{2,3,4,5\},\{2,3,4,6\},
\end{gathered}
$$

$$\{2,3,5,6\},$$
$$\{2,4,5,6\},$$
$$\{3,4,5,6\},$$

$$\{1,2,3,4,5\},\{1,2,3,4,6\},$$
$$\{1,2,3,5,6\},$$
$$\{1,2,4,5,6\},$$
$$\{1,3,4,5,6\},$$
$$\{2,3,4,5,6\},$$

$$\Omega \Big\}.$$

Which has $|\mathcal{F}| = 2^{|\Omega|} = 2^{36}$ events (subsets).

Although we can find how many events are in the event space, we don't have a practical way to list them all out. Also This is not generalizable, because sometimes we are looking at only a few of the events, such as drawing 3 dice out of an urn with 6 die. For this we turn to...

1.1.2 The Event Space

Definition. (Event Space) The collection of all events is called the "Event Space", denoted by $\mathcal{F}$ (or any calligraphic letter such as $\mathcal{A}$)

Now with the example earlier, we may write more efficiently

$$\mathcal{F} = \Big\{ \{i_1, i_2, i_3, ..., i_k\} \;\Big|\; i_1, i_2, i_3, ..., i_k \in \Omega,$$

$$\forall\ i_j \neq i_k \text{ "distinct"},\ for\ k = 0,1,2,3,4,5,6 \Big\}.$$

Remark. This definition of $\mathcal{F}$, is far more concrete since it considers when were not looking for all subsets. i.e. If you tossed one die, then there are 6 sample points, but also only 6 events (the event that you toss a die and obtain 4 numbers on the top shouldn't be considered). Also look at Ex 1.1 from the Homework as an example of this, noticing $\mathcal{F}$ is obviously not $2^{|\Omega|} = 2^{36}$, rather it is just $6 * \dot{6}$ since there are 2 dice with only 1 face from each as an event.[3]

Now, this efficient definition of $\mathcal{F}$ yields the idea of a probability measure P or $\mathbb{P}$.

Definition. (Probability Measure) Let Ω be a sample space and $\mathcal{F}$ be the set of all events of Ω. A "probability measure" is a function $P : \mathcal{F} \to \mathbb{R},\ s.t.$

(i) $0 \leq P(A) \leq 1,\ \forall\ A \subseteq \Omega$

(ii) $P(\Omega) = 1$ and $P(\emptyset) = 0$

(iii) Let $A_1, A_2, ...$ be a sequence of pairwise disjoint events. Then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Where A is an event in the event space $\mathcal{F}$, or just $A \in \mathcal{F}$.

1.1.3 The Probability Space

Important!
Definition (Probability Space) The triple $(\Omega, \mathcal{F}, P)$ is called

[3] In fact later we will look at how to count the number of possible events.

a probability space. [4]

Remark. This is for infinite measures (since we are definind a σ-algebra). But for finite sequence we have the following theorem.

Theorem 1.2 *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $A_1, A_2, ...A_n$ be a finite sequence of pair pairwise disjoint events. Then*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

Basically, this is saying, $\dot{\bigcup} A_i$ for ∞ many $\implies \dot{\bigcup} A_n$ works for finitely many. Seems doable and reasonable, so...

Proof: Let $A_1, A_2, ..., A_n$ be pairwise disjoint events. Now, we need to set the "rest" of the events to have probability 0, so we are setting ∞ many of them. Set $A_{n+1} = A_{n+1} = \cdots = \emptyset$. Then $A_1 \bigcup \cdots \bigcup A_n = A_1 \dot{\bigcup} \cdots \dot{\bigcup} A_n \; \dot{\bigcup} A_{n+1} \dot{\bigcup} A_{n+2} \dot{\bigcup} \cdots$

By axiom *(iii)*, we have $P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ $= \left(\sum_{i=1}^{\infty} P(A_i)\right) + 0 + 0 + 0 + \cdots = \sum_{i=1}^{n} P(A_i)$.

$$\therefore P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)\ .$$

q.e.d.

[4]Note, $A_1, A_2, ...$ are said to be pairwise disjoint if $A_i \cap A_j = \emptyset$, $\forall\, i \neq j$.

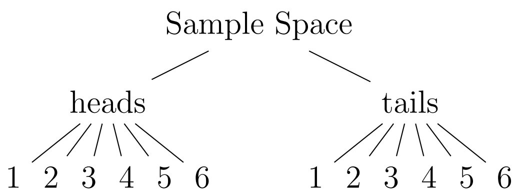
1.1.4 Introducing Cartesian Products

Let A, B be sets. Then

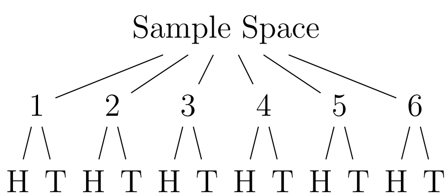
$$A \times B = \left\{ (a, b) \;\middle|\; a \in A,\ b \in B \right\}.$$

E.g. Toss a coin & Roll a die. What is the sample space Ω?

Using what we call a "Tree Diagram"...



or alternatively



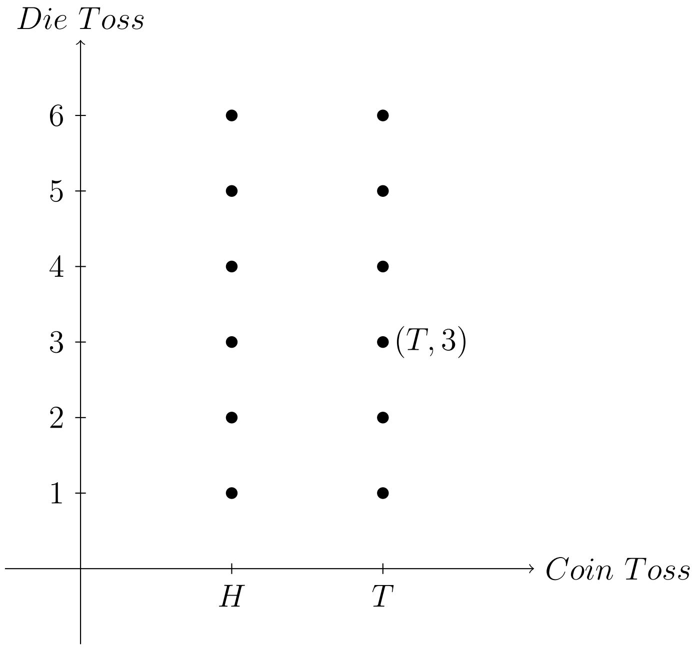
either way, they both imply

$$\Omega = \Big\{(H,1),(H,2),(H,3),(H,4),(H,5),(H,6),$$

$$(T,1),(T,2),(T,3),(T,4),(T,5),(T,6),\Big\}$$

Or by definition of cartesian products, we can say, $\Omega = \{H,T\} \times \{1,2,3,4,5,6\}$.[5] Which we know we can graph, at least for cartisian products of 3 or less sets (usually only for 2).

[5] Recall: The definition of cartisian product is that its an ordered pair, not an unordered set. So our graph must reflect that.



1.1.5 Set Theory Examples

Exercise B.4. Show that identity the "de Morgan" Law,

$$\left(\bigcap_i A_i\right)^C = \bigcup_i A_i^C$$

is true.

Recall: We need to show that the LHS and RHS are subsets of eachother.
Similar to "if and only if", however instead of **saying**

- "if $\rightarrow$ then...
- Conversely $\leftarrow$ then...

- Therefore $\Longleftrightarrow$..."

For set theory, we would **check** (for proper subsets btw),

- $\subseteq$: "Let $a \in A$, check that $a \in B$... yes! So since a started in A, but is also in B, hence $A \subseteq B$"

- $\supseteq$: "**Now** (not conversely), let $b \in B$, check that $b \in B$... yes! So since b started in B, but is also in A, hence $B \subseteq A$"

- $=$: Therefore since $\subseteq$ and $\supseteq$, then $A = B$.

Now that we remember this,

Proof by Professor: "$\subseteq$" direction first.
Let $x \in \left(\bigcap_i A_i\right)^C$.

$$\Longrightarrow x \notin A_i = \{y \mid y \in A_i,\ \forall i\}.$$

Then $\exists i$ *s.t.* $x \notin A_i$.
So $x \in \bigcup_i A_i^C$
Hence, $\left(\bigcap_i A_i\right)^C \subseteq \bigcup_i A_i^C$.

"$\subseteq$" direction now.

Now let $x \in \bigcup_i A_i^C$.
Then $x \in A_i^C$, for some $i \in \mathbb{Z}$.
Then $x \notin A_i$, for some $i \in \mathbb{Z}$.
Then $x \notin \bigcap_i A_i \Longrightarrow x \in \left(\bigcap_i A_i\right)^C$,

Hence, $\bigcup_i A_i^C \subseteq \left(\bigcap_i A_i\right)^C$.

Therefore,

$$\left(\bigcap_i A_i\right)^C = \bigcup_i A_i^C$$

q.e.d.

Exercise B.5. If A and B are sets then let $AtriangleB$ denote their symmetric difference: the set of elements that are in exactly one of the two sets. (In words this is "A or B, but not both.") Prove that the symmetric difference is associative.

Solution not Proof :)

We then go over B.5. however he goes over it with a proof by venn diagram. Which is somewhat trivial, and does not require much explanation/studying (which is why I dont really consider it a proof). *no figure here* However this is similar to doing proofs by table and I dislike this. But this is one of the few times I'll say, the proof is left as an exercise to the reader.[6]

Actually, since I'm stubborn + I'm a nerd = Hard way Both are direct proofs.

Proof by Internet
Similar to the venn diagram, we know
$(A\triangle B)\triangle C = (A\cup B\cup C)\cap(A\cup B^C\cup C^C)\cap(A^C\cup B\cup C^C)\cap(A^C\cup B^C\cup C)$
$= (B\cup C\cup A)\cap(B\cup C^C\cup A^C)\cap(B^C\cup C\cup A^C)\cap(B^C\cup C^C\cup A)$
$= (B\triangle C)\triangle A.$

Now, $\triangle$ is commutative since by def^n, obviously

[6] I am aware (and not naive...I think), that prooving this properly would take 10 pages and 5 hours (not actually).

$A\triangle B = (A\cap B^C)\cup(B\cap A^C) = (B\cap A^C)\cup(A\cap B^C) = B\triangle A$.
So $(B\triangle C)\triangle A = A\triangle(B\triangle C)$. *q.e.d.*

Proof by Stubborness
Notice $A\triangle B = (A \cap B^C) \cup (B \cap A^C)$.

Then
$(A\triangle B)\triangle C = ((A\triangle B) \cap C^C) \cup (C \cap (A\triangle B)^C)$
$= (((A\cap B^C)\cup(B\cap A^C))\cap C^C)\cup(C\cap((A\cap B^C)\cup(B\cap A^C))^C)$

Recall the De Morgan's law, $(A \cap B)^C = A^C \cup B^C$ and $(A \cup B)^C = A^C \cap B^C$

So,
$(((A\cap B^C)\cup(B\cap A^C))\cap C^C)\cup(C\cap((A\cap B^C)\cup(B\cap A^C))^C)$
$= (((A\cap B^C)\cup(B\cap A^C))\cap C^C)\cup(C\cap(A\cap B^C)^C\cap(B\cap A^C)^C)$
$= (((A\cap B^C)\cup(B\cap A^C))\cap C^C)\cup(C\cap A^f C\cup B\cap B^C\cup A)$

... yeah so I give up on this, I'll link a note on this when I find a clever but complete solution. Now that I have officially wasted 30 minutes of my saturday... $\square$

1.2 Random Sampling

1.2.1 Population and Sample Size

Some notation. Simply put,

- Population Set: What we've been calling the sample space.
  - "Population Size" $\#\Omega = n$.

- Sample Set: What we've been calling an event.
  - "Sample Size" $\#A = k$, where A is an event from the "event space".

I believe our professor is changing the terminology since it's more used or more practical, and less mathematical with its measure theory roots.

Main takeaways being, sample space $\neq$ sample set. They're different things, not related.

Also, we may still use Ω for both "Sample Space" and "Population Set". As well as $\mathcal{F}$ for the "Event Space" and "Sample Size". But more efficiently: I've been using the cardinality of these sets to denote the number of elements. But to step away from set theory, the textbook and other places use $\#$ instead.

Quick example.
Let the "Population Set" $\Omega = \{a, b, c, d\}$,
then the "Population Size" $\#\Omega = n = 4$.

Let our "sample" be this event $A = \{b, d\}$,
then the "sample size" $\#A = k = 2$.

How Many Samples are there?
This is our next important question.
— How many samples are there given a population set,
or
— rather how many events are there from a sample space.

Remark. This question is kind of ambiguous, as I am not asking amount of all events in the "event space" which would be the full set of all events (i.e. the cardinality of the power set). I am talking about what is the number of all the events if we have restrictions on *how* we sample.

So what are the restrictions/"procedures" we can do when making an event from a sample space. Well we can,

1. Sample with replacement, order matters $:= n^k$
2. Sampling without replacement, order matters $:= nPk = \frac{n!}{(n-k)!}$
3. Sample without replacement, order doesn't matter $:= nCk = \binom{n}{k} = \frac{n!}{k!(n-k)!}$
4. Sample with replacement, order doesn't matter $:= \binom{k+n-1}{n-1} = \binom{k+n-1}{k}$

All of these are somewhat explanatory, increasing in complexity.

Also similarly, finding out (the amount of events) or (the different ways), we can sample goes from simple to complex as well.

But I'll try to walk through each logically, as well as proof wise, and I'll derive the formulas for finding out the number of possible samples/events.

1.2.2 Sampling with replacement, order matters

Suppose we have a sample space with n elemtents.
— I have a bag of $n = 15$ marbles. They are colored.
Now suppose we want to take a sample of size k.
— I want to pick $k = 4$ marbles.

Let's choose:
We want the order to matter, so we will pick up one at a time, and record it.

- I grab 1 marble, its red. (then I mark it down)

We want the to sample **with** replacement, so after we pick one, we put it back in the bag.
— I'll put this back in the bag (so theres a chance I could pick it again).

Then we repeat until we get our full sample (of size k).

- I grab 1 marble, its blue. (I mark it down) Ok, I'll put it back.
- I grab 1 marble, its red. (I mark it down) Ok, I'll put it back.
- I grab 1 marble, its clear. (I mark it down) Ok, I'll put it back.

Thats what sampling with replacement looks like.
— Overall: I have a bag of 15 marbles, and my sample was "I chose 4 marbles".

In this scenario, Imma put back the marbles each time (so I could get the same one each time). Also in this scenario, I care about the order.
— So this time I got $\omega =$(red, blue, red, clear).

How many outcomes are there? Logically.
Well, simply put:
The first marble has n posibilities.
The second marble has n posibilities.
The third marble has n posibilities.
The... ...posibilities.
$\downarrow$ $\downarrow$
The kth marble nas n posibilities.

This is because we're sample with replacement,
i.e. each time I choose a marble, I put it back in the bag,
so I always have 15 marbles in the bag, so 15 posibilities.

So we have

$$n * n * n * n * \cdots * n = n^k$$

$$k \quad times$$

And the number of outcomes is

$$\#\Omega = n^k \ . \quad \square$$

How many outcomes are there? Proof wise.
Proof. Let $S = \{1, 2, ..., n\}$ $for\ some\ n \in \mathbb{N}$.
Set an element of our sample space $\omega \in \Omega$ to be an ordered k-tuple of elements $s_i \in S$.
So $\omega = (s_1, s_2, s_3, ..., s_k),\ for\ some\ k \in \mathbb{N}$.
Then

$$\Omega = \left\{ (s_1, s_2, ..., s_k)_{k \in \mathbb{N}} \ \middle|\ s_i \in S,\ i = 1, 2, ...k \right\}$$

but this implies

$$\Omega = S \times S \times \cdots \times S,\ k\ times.$$

So therefore

$$\#\Omega = \prod_{i=1}^{k} S_i = n^k,\ for\ some\ k \in \mathbb{N}.^{7}$$

q.e.d.

[7] Also quickly note, the probability measure (probability), if each event/outcome is equally likely implies

$$P(\omega) = \frac{1}{\#\Omega} = n^{-k} \ . \quad \square$$

1.2.3 Sampling without replacement, order matters

Suppose we have a sample space with n elemtents.
— I have a bag of $n = 15$ marbles. They are colored.
Now suppose we want to take a sample of size k.
— I want to pick $k = 4$ marbles.

Let's choose:
We want the order to matter, so we will pick up one at a time, and record it.

- I grab 1 marble, its red.

This time, we sample **without** replacement, so after we pick one, we keep it and don't put it in the bag.
— I put this on a rack, now the bag has $n - 1 = 15 - 1 = 14$ marbles left.

Then we repeat until we get our full sample (of size k).

- I grab 1 marble, its blue. Ok, I'll put it on the 2nd position on my rack. Now theres $n - 2 = 15 - 2 = 13$ marbles left.

- I grab 1 marble, its red. Ok, I'll put it on the 3nd position on my rack. Now theres $n - 3 = 15 - 3 = 12$ marbles left.

- I grab 1 marble, its clear. Ok, I'll put it on the 4nd position on my rack. Now theres $n - 4 = 15 - 4 = 11$ marbles left.

Thats what sampling with replacement looks like.
— Overall: I have a bag of 15 marbles, and my sample was "I chose 4 marbles".

In this scenario, keep the marbles in a rack, in order. (each time, theres 1 less marble, noting that i can only do this a max of $n = 15$ times).
— So this time I got $\omega =$(red, blue, red, clear).

How many outcomes are there? Logically.
Well, simply put:
The first marble has n posibilities.
The second marble has $n-1$ posibilities.
The third marble has $n-2$ posibilities.
The... ...posibilities.
$\downarrow$ $\downarrow$
The kth marble nas $n-k$ posibilities, noting that $k \not> n$ since you can't pick up more than n marbles.
This does imply that the kth marble has only 1 posibility if instead $k \geq n$, so we should say it has $(n-k)+1$ posibility.

So we have

$$n * (n-1) * (n-2) * \cdots * (n-k+1) = nPk.$$

Where we say nPk is the number of k permutations of n.

The number of outcomes is easier to calculate if we rewrite this formula as...

$$\#\Omega = nPk = \frac{n!}{(n-k)!}$$

since

$$n * (n-1) * (n-2) * \cdots * (n-k+1)$$
$$= \frac{n!}{(n-k)(n-k-1)(n-k-2)\cdots 3 * 2 * 1}. \quad \square$$

How many outcomes are there? Proof wise.
Proof. Let $S = \{1, 2, ..., n\}$ $for\ some$ $n \in \mathbb{N}$.
Set an element of our sample space $\omega \in \Omega$ to be an ordered k-tuple of elements $s_i \in S$.

So $\omega = (s_1, s_2, s_3, ..., s_k),\ for\ some\ k \in \mathbb{N}$.
But this time this tuple has elements $s_i \in S$, that are distinct.
So, $s_i \neq s_j,\ \forall i \neq j$.
Then

$$\Omega = \left\{(s_1, s_2, ..., s_k)_{k\in\mathbb{N}} \,\middle|\, s_i \in S, s_i \neq s_j,\ \forall i \neq j.\right\}$$

but this implies

$$n * (n-1) * (n-2) * \cdots * (n-k+1) = nPk$$

$$= n*(n-1)*(n-2)*\cdots*(n-k+1)*\tfrac{(n-k)(n-k-1)(n-k-2)*\cdots*3*2*1}{(n-k)(n-k-1)(n-k-2)*\cdots*3*2*1}$$

$$= \tfrac{n*(n-1)*(n-2)*\cdots*(n-k+1)(n-k)(n-k-1)(n-k-2)*\cdots*3*2*1}{(n-k)(n-k-1)(n-k-2)*\cdots*3*2*1}$$

$$= \frac{n!}{(n-k)!}.$$

So therefore

$$\#\Omega = nPk = \frac{n!}{(n-k)!}.$$

q.e.d.

1.2.4 Sampling without replacement, order doesn't matter

Suppose we have a sample space with n elemtents.
— I have a bag of $n = 15$ marbles. They are colored.
Now suppose we want to take a sample of size k.
— I want to pick $k = 4$ marbles.

Let's choose:
We don't care about order, so we will just hold it in our hand.

- I grab 1 marble, its red. It is in my hand.

We are not sampling with replacement, so after we pick one I pick up another one and keep it in my hand.

- I grab 1 marble, its blue. Ok, I'll keep this in my hand.

Then we repeat until we get our full sample (of size k).

- I grab 1 marble, its red. Ok, I keep this in my hand.

- I grab 1 marble, its clear. Ok, I keep this in my hand.

Thats what sampling with replacement looks like.
— Overall: I have a bag of 15 marbles, and my sample was "I chose 4 marbles".

In this scenario, Imma put back the marbles each time (so I could get the same one each time). Also in this scenario, I care about the order.
— So this time I got the **set** $\omega = \left\{2\ red's, blue, clear\right\}$ in my hand.

How many outcomes are there? Logically.
Well, simply put:
The first marble has n posibilities.
The second marble has $n-1$ posibilities.
The third marble has $n-2$ posibilities.
The... ...posibilities.
$\downarrow$ $\downarrow$
The kth marble nas $n-k$ posibilities, noting that $k \not> n$ since you can't pick up more than n marbles.
This does imply that the kth marble has only 1 posibility if instead $k \geq n$, so we should say it has $(n-k)+1$ posibility.

So we have

$$n * (n-1) * (n-2) * \cdots * (n-k+1).$$

The number of outcomes is easier to calculate if we rewrite this formula as...

$$\#\Omega = nPk = \frac{n!}{(n-k)!}$$

since

$$n * (n-1) * (n-2) * \cdots * (n-k+1)$$
$$= \frac{n!}{(n-k)(n-k-1)(n-k-2)\cdots 3 * 2 * 1}.$$

Now lastly note that for each event ω, it is a proper subset of S. Also this is not an ordered k-tuple. So each ω can be rearranged in many ways. I.E. The permutations do not matter, what we care about is the **combinations**.

We can kind of cheat and make this simpler, since we are only picking up k amount, we don't have to consider all of the possible unordered subsets, just the k amount of subsets we pick. Each subset can be rearranged in $k!$ ways. So out ot the $n!$ posibilities, the $(n-k)!k! \Longrightarrow (n-k)!k!$.

And so,

$$\#\Omega = \frac{n!}{(n-k)!k!} = nCk = \binom{n}{k},$$

which makes sense since $nPa \geq nCk$. □
— Read: "n choose k"

How many outcomes are there? Proof wise.
Proof. Let $S = \{1, 2, ..., n\}$ *for some* $n \in \mathbb{N}$.
Set an element of our sample space $\omega \in \Omega$ to be an unordered set with elements $s_i \in S$. Since we are going without replacement, the amount of elements in each subset have to add up to the total amount of elements in our sample space, i.e. $\sum(\#s_i \in \omega_i) = \#\Omega$. But this also imples that all ω_i are proper subsets of S.
So

$$\Omega = \Big\{\omega \subset S \;\Big|\; \# = k,\ 1 \leq k \leq n\Big\}.$$

Now we know for the amount of permutations of Ω is,

$$\#\Omega = nPk = \frac{n!}{(n-k)!}.$$

And the total amount of subsets ω, has been counted $k!$ times. (first k has been counted k times $= k$ permutations).
So we need to delete this amount from our number of permutations formula
but this implies

$$nCk \Longrightarrow \frac{nPk}{k!} = \frac{n*(n-1)*(n-2)*\cdots*(n-k+1)}{k!}$$

which we know we can simplify to

$$nCk = \frac{n!}{k!(n-k)!}.$$

q.e.d.

1.2.5 Sampling with replacement, order doesn't matter

Suppose we have a sample space with n elemtents.
— I have a bag of $n = 15$ marbles. They are colored.
Now suppose we want to take a sample of size k.
— I want to pick $k = 4$ marbles.

Let's choose:
We don't care about order, so we will just mark it down as a set/combination.

- I grab 1 marble, its red. Marked down 1 red.

We want the to sample with replacement, so after we pick one, we put it back in the bag.

- I got one blue, I'll put this back in the bag (so theres a chance I could pick it again).

Then we repeat until we get our full sample (of size k).

- I grab 1 marble, its red. Ok, I'll put it back.
- I grab 1 marble, its clear. Ok, I'll put it back.

That's what sampling with replacement looks like.
— Overall: I have a bag of 15 marbles, and my sample was "I chose 4 marbles".

In this scenario, Imma put back the marbles each time (so I could get the same one each time). Also in this scenario, I care dont care about the order, just the combination.
— So this time I got the **set** $\omega = \left\{2\ red's, blue, clear\right\}$.

How many outcomes are there? Logically.
Well, simply put:
We want combinations of these, so we will use

$$\Omega = \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

But this time, we can take any combination of each combination. So consider each combination's elements, $x_i \in \omega$, (where $\omega = (x_1, x_2, x_3 ... x_i)$).

Then what I am saying is that we can rearrange x_i in many different ways. For simplicity sake (and by our professor's idea), consider using "dividers". In this fasion, there are k distinct samples from the nCk equation. And for each sample we can then take each x_i and split it into its components.
E.g. $x_1 = 2\ red = 1\ red + 1\ red = 0 + 2\ red = 2\ red + 0$, where in this case, we treat $+$ as the divider.

Similarly, we can have a box of donuts, and there are "∞-many" choices (w/ replacement i.e. the same flavors), as well

as the order doesn't matter, "we got 3 chocolate, 4 strawberry, and 5 glazed". Then we can "break down" each
$x_i = \{3\ chocolate,\ 4\ strawberry,\ 5\ glazed\}$ into
$x_i = \{1\ chocolate,\ 1\ chocolate,\ 1\ chocolate,\ 1\ strawberry,\ 1\ strawberry$
and then use dividers to split this into as many partitions as we would like.

Now how many dividers are there again?
Well each solution can be represented by k elements, $x_1, ..., x_k$. And then there should be $n-1$ lines since there are only n options (colors, flavors of donuts, etc.)

And the main idea is we can "permute" these dividers whereever we like. To cheat though, we can just consider **all** the combinations of the sample points (colors/flavors) + all the dividers, compared to how many we pick (k amount (which can be larger than the $\#\Omega$).

So finally our equation is

$$\binom{n}{k} \Longrightarrow \binom{[n+(k-1)]}{k} = \binom{n+k-1}{n-1}. \quad \square$$

How many outcomes are there? Proof wise.
Proof. Let $S = \{1, 2, ..., n\}\ for\ some\ n \in \mathbb{N}$.
Set an element of our sample space $\omega \in \Omega$ to be an unordered set with elements $s_i \in S$. Since we are going without replacement, the amount of elements in each subset have to add up to the total amount of elements in our sample space, i.e. $\sum(\#s_i \in \omega_i) = \#\Omega$. But this also imples that all ω_i are proper subsets of S.
So

$$\Omega = \left\{\omega \subset S \;\middle|\; \# = k,\ 1 \le k \le n\right\}.$$

Now we know for the amount of permutations of Ω is,

$$\#\Omega = nPk = \frac{n!}{(n-k)!}.$$

And the total amount of subsets ω, has been counted $k!$ times. (first k has been counted k times = k permutations).
So we need to delete this amount from our number of permutations formula
but this implies

$$nCk \Longrightarrow \frac{nPa}{k!} = \frac{n*(n-1)*(n-2)*\cdots*(n-k+1)}{k!}$$

which we know we can simplify to

$$nCk = \frac{n!}{k!(n-k)!}.$$

Now, consider the elements of our sample space $\omega_i \in \Omega$.
Note that since $\omega \in \Omega$, then for $\omega \ni n_i \in S$.
But now we can partition ω and its k elements $n_1, ..., n_i, ...n_k$, with $n-1$ disjoint unions.
This implies there are now $k + (n-1)$ "objects", which we can take any combination of. Where for each instance of two adjacent disjoint unions, we take the union of the empty set, i.e. $\cdots \dot{\cup}\, \dot{\cup} \cdots = \cdots \dot{\cup}\, \emptyset\, \dot{\cup} \cdots$.
So then we have

$$\binom{n}{k} \Longrightarrow \binom{[n+(k-1)]}{k} \Longrightarrow \binom{n+k-1}{k} = \binom{n+k-1}{n-1}. \qquad \blacksquare$$

Summary: Just Memorize

1. Sample with replacement, order matters := n^k
2. Sampling without replacement, order matters := $nPk = \frac{n!}{(n-k)!}$
3. Sample without replacement, order doesn't matter := $nCk = \binom{n}{k} = \frac{n!}{k!(n-k)!}$
4. Sample with replacement, order doesn't matter := $\binom{k+n-1}{n-1} = \binom{k+n-1}{k}$

As well as, for grabbing marbles one by one...

1. Sampling with replacement $\Longrightarrow$ putting marbles back into the bag
2. Sampling without replacement $\Longrightarrow$ keeping the marbles out of the bag
3. Order matters $\Longrightarrow$ put it in on a rack in order, or mark down the tuple (which came first) $\Longrightarrow$ permutations
4. Order doesn't matter $\Longrightarrow$ hold it in your hand, or mark down the set $\Longrightarrow$ combinations

1.2.6 Random Sampling Examples

E.g. 1.8 Suppose a bag of scrabble tiles contains 5 $E's, 4\ A's$, 3 $N's$, and 2 $B's$. Draw 4 tiles without replacement.

(a) Find $P(C)$ by imagining that the tiles are drawn one by one, as an ordered sample.

Solution. Let $C = \{2\ E,\ 1\ A,\ 1\ N\}$. (Note sets are unordered) Suppose you have

$$\begin{aligned} &P(1st\ E,\ 2nd\ E,\ 3rd\ A,\ 4th\ N) \\ &+ P(1st\ E,\ 2nd\ A,\ 3rd\ E,\ 4th\ N) \\ &+ P(1st\ N,\ 2nd\ A,\ 3rd\ E,\ 4th\ E) + \cdots \end{aligned}$$

$$= \frac{(5*4)*(4)*(3)}{14*13*12*11} + \frac{(5*4)*(4)*(3)}{14*13*12*11} + \frac{(3)*(4)*(5)*(4)}{14*13*12*11} + \cdots$$

$$= \frac{5*4*4*3}{14*13*12*11}\binom{4}{2}\binom{2}{1}\binom{1}{1}$$

$$= \frac{5*4*4*3}{14*13*12*11}\left(\frac{4*3}{2*1}\right)\left(\frac{2*1}{1}\right)\left(\frac{1}{1}\right) \approx 0.0599. \quad \square$$

(b) Find $P(C)$ by imaging that the tiles are drawn all at once as an unordered sample.

Or more simply,

$$\frac{\binom{5}{2}\binom{3}{1}\binom{4}{1}}{\binom{14}{4}}.$$

which gives the same value, ≈ 0.059. $\square$

E.g. 1.8 Everyday a kindergarten class chooses randomly one of the 50 state flags to hang on the wall, without regard to previous choices.

(a) Find the sample space.

Solution. Let $S = \{1, 2, 3, 4, ..., 50\}$, then

$$\Omega = S \times S \times S = \Big\{(i, j, k) \;\Big|\; i, j, k \in S\Big\}.$$

and then

$$\#\Omega = 50 * 50 * 50 = 50^3. \quad \square$$

(c) Find the probability that the Wisconsin flag will be hung at least two of the three days, Monday, Tuesday, and Wednesday.

Solution.
Lets seperate this into the probability of 2 days, and probability of 3 days (which includes 2).

$$P(W\ W\ else) + P(W\ else\ W) + P(else\ W\ W) + P(W\ W\ W)$$
$$= \frac{1 * 1 * 49}{50 * 50 * 50} * 3 + \frac{1 * 1 * 1}{50^3}. \quad \square$$

E.g. 1.13 Suppose we have a class of 24 students.

(a) A team of three children is chosen at random. Find the probability that the team consists of Shane, Heather, and Laura.

Solution.
So we need to find

$$P\big(\{Shane,\ Heather,\ Laura\}\big) = \frac{1}{\binom{24}{3}}. \quad \square$$

(c) Find the probability that Mary is on the team.

Solution.

$$P\Big(\big\{\{Shane,\ Heather,\ Laura\} \wedge \{Mary\}\big\}\Big) = \frac{\binom{23}{2}}{\binom{24}{3}}. \quad \square$$

E.g. ?.? Three students are chosen to be class president, vice president, and treasurer. No student can hold more than one office.

(a) Find the probability that mary is president, cory is vice president, and matt is treasurer.

Solution.
Let event A be the event in part (a) holds. Then

$$P(A) = \frac{1 * 1 * 1}{24 * 23 * 22}. \quad \square$$

(b) What is the probability that ben is either president or vice president.

Solution.
Let event B be the event in part (b) holds. Then

$$P(A) = \frac{1 * 23 * 22}{24 * 23 * 22} + \frac{23 * 1 * 22}{24 * 23 * 22} = \frac{1}{24} + \frac{1}{24} = \frac{1}{12}. \quad \square$$

E.g. 1.7 We have an urn with 3 green and 4 yellow balls. We draw 3 one by one without replacement.

(a) Find the probability that the colors we see in order are green, yellow, green.

Solution.
Note: order matters, without replacement. So,

$$P(\omega) = \frac{3 * 4 * 5}{7 * 6 * 5}. \quad \square$$

(b) Find the probability that our sample of 3 balls contain 2 green balls and 1 yellow ball.

Solution.
Note: order doesn't matter, without replacement. So,

$$\begin{aligned} P(\omega) &= P(y, g, g) + P(g, y, g) + P(g, g, y) \\ &= \frac{3 * 4 * 2}{7 * 6 * 5} + \frac{3 * 2 * 4}{7 * 6 * 5} + \frac{4 * 3 * 2}{7 * 6 * 5} \\ &= \frac{3 * 4 * 2}{7 * 6 * 5} * \binom{3}{1} = \frac{3 * 4 * 2}{7 * 6 * 5} * 3 = \frac{\binom{3}{2}\binom{4}{1}}{\binom{7}{3}}. \quad \square \end{aligned}$$

E.g. 1.10 We roll a fair die repeatedly until we see the number four appear and then stop. The outcome of the expirement is the number of rolls.

(a) Find the sample space Ω and the probability measure P.

Solution.
The sample space goes on until infinity with discrete steps, so

$$\Omega = \{1, 2, 3, 4, ...\} \cup \{\infty\}.$$

Also our probability measure should be for a finite amount of trials, set this amount to be "k-trials". Then[8],

$$P(k) = \frac{5^{k-1} * 1}{6^k}. \quad \blacksquare$$

(b) Calculate the probability that the number never appears.

Solution.
The probability that the number never appears implies the probability that there are infinite events. Which is

$$P(\infty) = 1 - \sum_{k=1}^{\infty} P(k) = 1 - \sum_{k=1}^{\infty} \frac{5^{k-1}}{6^k}$$

$$= 1 - \frac{a}{1-r} = 1 - \frac{\frac{1}{6}}{1 - \frac{5}{6}}$$

$$= 1 - \frac{\frac{1}{6}}{\frac{1}{6}} = 1 - 1 = 0.$$

So,[9]

$$P(\infty) = 0. \quad \square$$

[8] Technnically, we should define a probability measure using the definition. So for and event that is also $A \subseteq \Omega$, we set $P(A) = \sum_{k \in A} P(A)$.

[9] This such definition can be taken in a limit process as well, at least in this case. So you would set $\{\infty\} \subseteq \{F, F, F..., F\}$, of size k. Or better defined, $|\{F, F, F..., F\}| = k, \ \forall k \in \mathbb{N}$. Then $P(\infty) \leq P(F...F) = \frac{5^k}{6^k} = \left(\frac{5}{6}\right)^k \to 0$, as $k \to 0$. This is opposite to the probability I get a 100% in this class, so at least I have that going for me.

1.3 Infinite many outcomes

Key Idea: infinite/uncountable sets $\Longrightarrow$ probabilities are continuous over $\mathbb{R}$. ($\mathbb{R}$ is the new thing here). We shall look at some examples, it should be fairly simple.

E.g. 1.16 Flip a coin until the first tails comes up. Record the number of flips required as the outcome of the experiment.

(a) Find the sample space.

Solution. Then the sample space is

$$\Omega = \{1, 2, 3, ...\} \cup \{\infty\}. \quad \square$$

(b) Find $P(k)$.

Solution. Well, there are $n = 2$ posibilities for each sample point.
The kth event can be described as the following,
$k = 5 \Longrightarrow H * H * H * H * H * T.$
So

$$P(k) = \frac{1 * 1 * 1 * \cdots * 1}{2^k}, \; for \; k = 1, 2, 3, ...$$

And also

$$P(\infty) = 1 - \sum_{k=1}^{\infty} P(k) = 1 - \sum_{k=1}^{\infty} \frac{1}{2^k},$$

which is a geometric series, with $a_1 = \frac{1}{2}$, common ratio $r = \frac{1}{2}$.
So since this one starts at $k = 1$,

$$1 - \sum_{k=1}^{\infty} \frac{1}{2^k} = 1 - \frac{a}{1-r} = 1 - \frac{\frac{1}{2}}{1 - \frac{1}{2}}$$

$$= 1 - \frac{\frac{1}{2}}{\frac{1}{2}} = 1 - 1 = 0.$$

So the probability that you never get tails/you keep getting heads, in this limit process of a **fair** coin, is

$$P(\infty) = 0. \quad \square$$

E.g. 1.17- We pick a real number uniformly at random reom the closed interval $[0, 1]$.

(a) Find the sample space.

Solution.

$$\Omega = \{a \in \mathbb{R} \mid 0 \leq x \leq 1\} = [0, 1]. \quad \square$$

(b) Find $P\left(\frac{1}{2} \leq x \leq \frac{3}{4}\right)$

Solution.

$$P\left(\frac{1}{2} \leq x \leq \frac{3}{4}\right) = \frac{\frac{3}{4} - \frac{1}{2}}{1} = \frac{1}{4}. \quad \square$$

E.g. 1.18 - Consider a dartboard in the shape of a disk with radius of 9 inches. The bullseye is a disk of diameter $\frac{1}{2}$ inch in the middle of the board. What is the probability that a dart randomly thrown on the board hits the bullseye?
Solution.

$$P(\textit{hitting the bullseye}) = \frac{\textit{Area}(\textit{bullseye})}{\textit{Total area}}$$

$$= \frac{\pi\left(\frac{1}{9}\right)^2}{\pi(9)^2} = \frac{\frac{1}{16}}{81} = \frac{1}{(16)(81)}. \quad \square$$

1.4 Consequences of the Rules of Probability

These are the three main properties that come from the Rules of Probability (i.e. the way we had described and defined our σ-algebra "$P(A), A \in \mathcal{F} = |\Omega|$" and Probability Space)

1. Decomposition of an event.
   - If $A = \bigcup_{i=1}^{\infty} A_i$, where $A_1, A_2, ...$ are pairwise disjoint events, then $P(A) = \sum_{i=1}^{\infty} P(A_i)$.
2. Events and Complements
   - $P(A) + P(A^C) = 1$
   - $P(A^C) = 1 - P(A)$
3. Monotonicity of Probability
   - If $A \subseteq B$, then $P(A) \subseteq P(B)$.

Below we will give examples and prove some of these consequences.

1.4.1 Decomposing an event

If $A = \bigcup_{i=1}^{\infty} A_i$, where $A_1, A_2, ...$ are pairwise disjoint events, then $P(A) = \sum_{i=1}^{\infty} P(A_i)$

E.g. 1.19 Suppose an urn contains 30 red, 20 green and 10 yellow balls. Draw two without replacement. What is the probability that the sample contains exactly one red or exactly one yellow?

Solution
So the question is asking, $P(\textit{exactly } 1 \textit{ red or exactly } 1 \textit{ yellow})$.
Main question: Can you decompose this into other probabilities?
Without replacement $\implies$ means all the events can be broken up into disjoint unions. So therefore its probabilities can also be broken up.[10]
Ans: Yes you can! i.e.
$P(\textit{exactly } 1 \textit{ red or exactly } 1 \textit{ yellow})$

$$
\begin{aligned}
&= P\big(\ \{1r, 1y\} \cup \{1r, 1g\} \cup \{1y, 1g\}\ \big) \\
&= P\big(\{1r, 1y\}\big) + P\big(\{1r, 1g\}\big) + P\big(\{1y, 1g\}\big) \\
&= \frac{\binom{30}{1}\binom{10}{1}}{\binom{60}{2}} + \frac{\binom{10}{1}\binom{20}{1}}{\binom{60}{2}} + \frac{\binom{30}{1}\binom{20}{1}}{\binom{60}{2}} \\
&= \frac{\binom{30}{1}\binom{10}{1} + \binom{10}{1}\binom{20}{1} + \binom{30}{1}\binom{20}{1}}{\binom{60}{2}} \\
&= \frac{30 * 10 + 10 * 20 + 30 * 20}{\frac{60!}{2! * 58!}} \\
&= \frac{300 + 200 + 600}{\frac{\cancelto{30}{60} * 59 * \cancel{58!}}{\cancelto{1}{2} * \cancel{58!}}} \\
&= \frac{1100}{30 * 59} = \frac{550}{15 * 59} = \frac{110}{3 * 59} = \frac{110}{117}. \quad \blacksquare
\end{aligned}
$$

Recall If $A_1, A_2, ...$ are pairwise disjoint events, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

[10] You have to be very careful when you can or can't do this.

E.g. 1.20 Peter and Mary take turns rolling a fair die. If Peter rolls 1 or 2 he wins and the game stops. If Mary rolls 3, 4, 5, or 6, she wins and the game stops. They keep rolling in turn until one of them wins. Suppose Peter rolls first.

(a) What is the probability that Peter wins and rolls at most 4 times?

Solution. Let $A_i = \{Peter\ wins\ at\ his\ i^{th}\ roll\}$
Say the game looks like this...
I order it as "step number, the player, if they win or not, and probability that they would have won"

$$
\begin{aligned}
1.\ Peter : F : \frac{4}{6} \\
2.\ Mary : F : \frac{2}{6} \\
3.\ Peter : F : \frac{4}{6} \\
4.\ Mary : F : \frac{2}{6} \\
. \\
. \\
. \\
(i).\ Peter : T : \frac{4}{6}
\end{aligned}
$$

Then

$$P(Peter\ wins\ and\ rolls\ at\ most\ 4\ times)$$

$$= P(A_1) + P(A_2) + P(A_3) + P(A_4)$$

$$= \sum_{i=1}^{4} \left(\frac{4}{6} * \frac{2}{6} \right)^{i-1} \left(\frac{2}{6} \right)$$

$$= \frac{1}{3} \sum_{i=1}^{4} \left(\frac{2}{9} \right)^{i-1}$$

$$= \frac{1}{3} \left[\left(\frac{2}{9} \right)^{0} + \left(\frac{2}{9} \right)^{1} + \left(\frac{2}{9} \right)^{2} + \left(\frac{2}{9} \right)^{3} \right]$$

$$= \frac{1}{3} \left[\frac{1}{3} + \frac{2}{9} + \left(\frac{2}{9} \right)^{2} + \left(\frac{2}{9} \right)^{3} \right]$$

$$= \frac{1}{3} \frac{1 - \left(\frac{2}{9} \right)^{4}}{1 - \left(\frac{2}{9} \right)} \approx 0.427 \quad \square$$

(b) What is the probability that May wins?

Solution. Let $B_i = \{Mary\ wins\ at\ her\ i^{th}\ roll\}$
Say the game looks like this...

$$1.\ Peter : F : \frac{4}{6}$$
$$2.\ Mary : F : \frac{2}{6}$$

$$3.\ Peter : F : \frac{4}{6}$$
$$4.\ Mary : F : \frac{2}{6}$$
$$\vdots$$

Then

$$\begin{aligned}
P(Mary\ Wins) &= \sum_{i=1}^{\infty} P(B_i) \\
&= \sum_{i=1}^{\infty} \left[\left(\frac{4}{6}\right)\left(\frac{2}{6}\right)\right]^{i-1}\left(\frac{4}{6}\right) \\
&= \sum_{i=1}^{\infty} \left(\frac{2}{9}\right)^{i-1}\left(\frac{2}{3}\right) \\
&= \left(\frac{2}{3}\right)^{2}\frac{a}{1-r} \\
&= \left(\frac{2}{3}\right)^{2}\frac{1}{1-\frac{2}{9}} \\
&= \left(\frac{2}{3}\right)^{2}\left(\frac{9}{7}\right) = \frac{4}{7} \approx 0.571 \quad \square
\end{aligned}$$

1.4.2 Events and Complements

Both shall hold,

- $P(A) + P(A^C) = 1$
- $P(A^C) = 1 - P(A)$

E.g. 1.21 Roll a fair die 4 times. What is the probability that some numbers appears more than once?
Solution.

$$P(some\ \#'s\ appear\ more\ than\ once)$$

$$= 1 - P(all\ numbers\ are\ different)$$

$$= 1 - \frac{\cancel{6} * 5 * \cancelto{2}{4} * \cancel{3}}{\cancel{6} * 6 * \cancelto{3}{6} * \cancelto{2}{6}}$$

$$= 1 - \frac{5}{18} = \frac{13}{18} = .722 \quad \square$$

1.4.3 Monotonicity of Probability

Claim: If $A \subseteq B$, then $P(A) \subseteq P(B)$.
Proof. Let A, B be sets in Ω. Suppose $A \subset B$.
Then $B = A \dot{\cup} (B \backslash A)$
($\dot{\cup}$ meaning disjoint union).
Since this is a disjoint union, we can take the individual probabilities.
So,
$P(B) = P(A) + P(B \backslash A) \geq P(A) + 0 = P(A)$ *q.e.d.*

E.g. 1.22 Suppose we toss a fair coin repeatedly. What is the probability that tails never occurs?
Recall. We know $P(\textit{tails never occur}) = 0$ already. This is by using a a summation generalized to k, and as it approaches ∞, we find $P(A) \to 0$.
However lets prove this using the new property "Monotonicity of Probability" and the squeeze theorem.

Solution. We know

$$\begin{aligned} 0 &\leq P(\textit{tails never occur}) \\ &\leq P(\textit{tails never occur in the 1st } k \textit{ tosses}) \\ &= \left(\frac{1}{2}\right)^k, \ \forall k \in \mathbb{N}. \end{aligned}$$

Then by the right hand side inequality, you may see, how the squeeze theorem will be used. Also the ratio $\left(\frac{1}{2}\right)^k$ is because each kth flip has a probability measure of $\frac{1}{2}$.
So now

$$\lim_{k\to\infty} \left(\frac{1}{2}\right)^k = 0$$

or alternatively

$$P(\textit{tails never occur in the 1st } k \textit{ tosses}) = \left(\frac{1}{2}\right)^k \xrightarrow{k\to\infty} 0.$$

And by the squeeze theorem (from Calculus)

$$P(\textit{tails never occur}) = 0.$$

q.e.d.

Question: What if the probability measure was not fair?
Say a coin wasn't fair and had probability measures
$P(\omega) \neq P(H) = P(T)$, and we set
$P(H) = 0.99 \Longrightarrow P(T) = 0.01$.
But then, like the previous example, we get the same results, since by squeeze theorem

$$\begin{aligned} 0 \leq & P(\textit{tails never occur}) \\ \leq & P(\textit{tails never occur in the } 1st\ k \textit{ tosses}) \\ & = (0.99)^k \xrightarrow{k\to\infty} 0,\ \forall k \in \mathbb{N}. \end{aligned}$$

1.4.4 More Examples of these Consequences

E.g. 1.15 - An urn contains 4 balls: 1 white, 1 green, and 2 red. We draw 3 balls *with* replacement. Find the probability that we *did not* see all three colors.

(a) Solve with Inclusion-Exclusion Principle.

Solution.
Let $W = \{\textit{white ball did not appear}\}$,
$G = \{\textit{green ball did not appear}\}$, and
$R = \{\textit{red ball did not appear}\}$.
Then

$$P\{\textit{we did see all three colors}\}$$

$$= P(W \cup G \cup R)$$

$$\begin{aligned} &= P(W) + P(G) + P(R) \\ &P(W \cap G) - P(W \cap R) - P(G \cap R) \\ &+ P(W \cap G \cap R) \end{aligned}$$

$$
\begin{aligned}
&= \frac{3*3*3}{4*4*4} + \frac{3*3*3}{4*4*4} + \frac{2*2*2}{4*4*4} \\
&- \frac{2*2*2}{4*4*4} - \frac{1*1*1}{4*4*4} - \frac{1*1*1}{4*4*4} \\
&+ \frac{0*0*0}{4*4*4}
\end{aligned}
$$

$$
= \frac{3^3+3^3-1-1}{4^3} = \frac{52}{64} = \frac{26}{32} = \frac{13}{16}. \quad \square
$$

(b) Solve using the complement.

Solution.

$$
\begin{aligned}
&P\{we\ did\ see\ all\ three\ colors\} \\
&= 1 - P(we\ saw\ all\ three\ colors) \\
&= 1 - \frac{1*1*2*3!}{4*4*4} \\
&= 1 - \frac{12}{64} = 1 - \frac{3}{16} = \frac{13}{16}. \quad \square
\end{aligned}
$$

1.4.5 Inclusion-Excluion Principle

The goal is to rewrite unions as intersections. “This is because unions overlap and are difficult to calculate, but individual sets and intersections are simpler to calculate.”

For 2 sets:

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

The probability that one of the sets occur, implies probability of A and B, but then we are counting the intersection twice. This is why we subtract one of the intersections.

For 3 sets: Similarly,

$$\begin{aligned} P(A \cup B \cup C) =& P(A) + P(B) + P(C) \\ & - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ & + P(A \cap B \cap C) \end{aligned}$$

We add the last triple intersection, because when we delete the pairwise set intersections, we delete all events from the middle. One may draw out these examples, or I may find pictures to reference.

For n sets: Due to the parity (odd or even amount) of the number of sets,

- When its even, we have overlapping intersections we need to delete

- When its odd, we delete the even tuples, but then we completely cancel out the odd tuples. i.e. we add back the intersection

So we can think of this as (1)[11]

$$
\begin{aligned}
&|A \cup B \cup C \cup D \cup \cdots| = \\
&+|A| + |B| + |C| + |D| + \cdots (\textit{add all singletons}) \\
&-|A \cap B| - |A \cap C| - |A \cap D| - \cdots (\textit{minus all pairs}) \\
&+|A \cap B \cap C| + |A \cap B \cap D| + \cdots (\textit{plus all triples}) \\
&- \textit{quadruples} \\
&+ \textit{quintuples} \\
&\pm \textit{etc.}
\end{aligned}
$$

And we know we can take probabilities of these sets, so in general it would look like (2)

$$
\begin{aligned}
&P(A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n) \\
&= P(A_1) + P(A_1) + \cdots + P(A_n) \\
&- \sum_{i \leq i < j \leq n} P(A_i \cap A_j) \\
&+ \sum_{i \leq i < j \leq n} P(A_j \cap A_j \cap A_k) \\
&- \quad \cdots \\
&+ \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} (-1)^{k-1} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) \\
&\mp \quad \cdots \\
&+ (-1)^{n-1} P(A_1 \cap \cdots \cap A_n).
\end{aligned}
$$

Or more concisely (3)

$$
\boxed{P(A_1 \cup \cdots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} P(A_{i_1} \cap \cdots \cap A_{i_k})}
$$

[11] I would memorize (1), and know what it means well. Then memorize (3) for other computation.

E.g. 1.27 Suppose n people exchange their hats randomly. Find the probability that no one gets his/her own hat.

Solution.
Let the event $A_i = i^{th}$ *person gets his/her own hat.*
(Note this is the converse of what we want to find.)

Now we can apply what we know about complements and what we just learned about inclusion-exclusion principle. We find

$$
\begin{aligned}
&P(\textit{no one gets his/her own hat}) \\
&= P(A_1^C \cap A_2^C \cap \cdots \cap A_n^C) \\
&= 1 - P\Big((A_1^C \cap A_2^C \cap \cdots \cap A_n^C)^C\Big) \\
&= 1 - P(A_1 \cup \cdots A_n) \\
&= 1 - \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} P(A_{i_1} \cap \cdots \cap A_{i_k}) \\
&= 1 - \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \cdots < i_k \le n} \frac{(n-k)!}{n!} \\
&= 1 - \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!} \\
&= 1 - \sum_{k=1}^{n} (-1)^{k-1} \frac{\cancel{n}!}{\cancel{(n-k)}!k!} \frac{\cancel{(n-k)}!}{\cancel{n}!} \\
&= 1 + \sum_{k=1}^{n} (-1)^{k} \frac{1}{k!} \\
&= \sum_{k=0}^{n} \frac{1}{k!} \xrightarrow{n \to \infty} e^{-1} \approx \frac{1}{2.7}. \quad \square
\end{aligned}
$$

1.4.6 Some Examples

E.g. 1.14 Assume $P(A) = 0.4$ and $P(B) = 0.7$.
Show that $0.1 \leq P(AB) \leq 0.4$.

Proof by Professor:
Notice, $A \cap B \subseteq A \Longrightarrow P(A \cap B) \leq P(A) = 0.4$
Also notice, $P(A \cap B) = P(A) + P(B) - P(A \cup B)$, by the inclusion-exclusion principle.

So then

$$\begin{aligned} 1 \geq P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.4 + 0.7 - P(A \cap B) = 1.1 - P(A \cap B) \end{aligned}$$

Therefore $P(A \cap B) \geq 1.1 - 1 = 0.1$. *q.e.d.*

E.g. 1.9 We break a stick at a uniformly chosen random location. Find the probability that the shorter piece is less than 1/5 th of the original.

Solution by Professor:
Let $x \in [0, 1]$, i.e. x is the breaking point of a stick of length/total probability 1.
Then notice we can "break the stick" into two disjoint unions (i.e. 2 sticks). One of the sticks are x with the other $1 - x$.
Also note, $x \leq \frac{1}{2}$ or $x \geq \frac{1}{2}$.

So we want $x < \frac{1}{5}$, then there are two possible outcomes since $x \leq \frac{1}{2} \ \vee \ x \geq \frac{1}{2}$.

$$\begin{aligned}P(x < \frac{1}{2} \cap x \leq \frac{1}{2}) + P(1 - x < \frac{1}{5} \cap x > \frac{1}{2})\\ = P(x < \frac{1}{5}) + P(\frac{4}{5} < x < 1)\\ = \frac{1}{5} + (1 - \frac{4}{5}) = \boxed{\frac{2}{5}}. \quad \square\end{aligned}$$

1.5 Random Variables

Definition. (Random Variable) Let Ω be a sample space. A random variable is a function from Ω into the real numbers.

$$X : \Omega \to \mathbb{R}.$$

E.g. Toss a coin. Let X be the number of heads.
$\Omega = \{H, T\}$
Then,

- $X(H) = 1$
- $X(T) = 0$

$\Longrightarrow X$ is a random variable.

1.5.1 Graphing and Recording Table of Random Variables

E.g. 1.29 Roll a die 2 times. Let X be the sum of two die. Find $P_x(k) = P(x = k)$, for all $k \in X$.
Solution.
Notice $X = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

One can count or graph how many events obtain the sum of two die. i.e. for $x = 2$, only one possibility...die #1=1 and die #2=1. Similarly, for 3, one must be 2 and the other 1, so there are two combination.

Alternatively, by graphing (with die #1 on the "x-axis", and die #2 on the "y-axis")

Bibliography

[1] Anderson, David F., et al. *Introduction to Probability*, Cambridge University Press (1991)
ISBN: 978-1-108-41585-9