

Probability Theory I

Notes

Dean Quach

Course Info

Probability Theory I

January 17, 2023 to May 9, 2023

MWF — 1:00 pm - 1:50 pm — Wachman 407

Prof. Mr. Wei-Shih Yang

CRN 37229

3 credits \implies 6 hours of homework

Text

Introduction to Probability Theory

- David F. Anderson
- Timo Sappalainen
- Benedek Valko

Edition 1

Other Notes

2 Midterms \Rightarrow 20% each

Final \Rightarrow 35%

Quizzes \Rightarrow 12% each
(every other Friday)

HW \Rightarrow 3% each
(due every Friday)

Office Hours: Wachman 534

MWF 3:00 pm - 4:00 pm

Phone for texting: *removed, because I'm not stupid*

Spring 2023
Math 3031
Section 001

Preface

Those of you who aren't familiar with me, I am just a below average math major. Struggling as always; in his 4th semester at Temple University (Spring 2023). I had originally wrote this for some close friends that I had met along the way, mainly one which had introduced me to a bunch of his other friends which had introduced me to more, like induction or whatever :)

First I'd like to thank you for looking at my first book.
Use it for whatever you need.

Whether this be an aid to if your sick and you can “borrow” my notes like I'm your friend¹, or this may be a way for you to see what an undergraduate math course looks like at a (what I would call) normal university. Worst case for me, but best case for you, this could be something that is hilarious to you, not only because of my footnotes, but because you might be; a graduate student, an actual adult (with a full time job), one of my professors (oh dear god), someone I dislike (please no)...a relative or friend (it can't get worse)...or my future self...
No! God!...No God, please no, No!...NO!...NOOOOO!!!
— Michael Scott, The Office.

¹its why I typed them up, since my handwritting is cursive. I.E. It's beautiful, but illegible to most.

Now, those friends know who they are, and have helped me feel comfortable adopting the phrase I say a lot, which is “it’s like ‘math major’ is my whole fucking personality”. They did not directly sit down with me, hold a weapon to my head, and made me believe that by force... Rather they had asked me many (and I mean many) questions over 2 years (at this point). Along with doing my research project on the Perfect Cuboid (since my GPA at the time isn’t great due to my laziness, so of course I could not get other academic jobs), I feel as if everyone close around me is aware that they are finally helping me break out of my imposter syndrome of being a mathematician. I certainly believe I will never break out of the imposter syndrome, I definitely do not believe I deserve to be here, let alone have the “right” to call myself a mathematician (this feels very wrong for me to say, even type). To me I still feel like I am just a teenager (even though I’m 20 as of writing this), just doing math problems for fun, and equally butchering my quizzes and tests².

I hope this helps as another way someone relates to you, but enough about me and more about the book. These were meant to be my personal notes. I had originally started typing them up in Obsidian, which is a great program with “atomic” style linking. But even though it had \LaTeX (actually Math-Jax) and other coding implements, I found myself just linking in a string all the lectures together. Sometimes the atomic style worked, but it more or less became a thing I had to organize, and although I could have set it up with a tutorial on how to make it future proof, I believe it made more sense for me to go back to \LaTeX and just comment out what lecture # it is, and basically write my notes like a book.

²Ironically, these close friends I am talking about, have no idea what I am talking about. I am referring to other math classes that kicked my ass. But in the classes I took with them (such as this class and Calc 3), they always saw me get good grades with 0 sleep and 80% caffeinated blood. Hilariously stupid on my end.

I don't know if it is obvious by know, but these, as I said, were meant to be my personal notes. I had actually only meant to share it to those close friends I talked about on my GitHub. Anyways, much of this is literally based off of the textbook my course used. Also it wasn't meant to be a book, but I had realized about page 73 (the week of my midterm and me procrastinating as always), that I am basically just writing a book. At this point I had the cool idea that I wanted to print this out for my personal collection, at least I think it'll be pretty cool to have a bunch of books made with "no additional effort", that are technically mine. And then I could possibly go back and rewrite my old notes for other courses, and the new courses will be an easy case. And then I realized I could probably just publish this. So I hope, like in paragraph 2, that you find this useful or funny, hopefully both. ³ Thank you for reading!

Disclaimer: Some things I write may be completely wrong. Usually the arithmetic. This is a personal journal, NOT an official study guide/aid. If you fail because of me, (ha!) but I'm not responsible. Go study for real instead of reading this book. Actually no, please read this book to, you might find some interesting things. Also please contact me at deanquach.com if theres any problems. I do want to fix any problems and I'll submit new editions when enough are caught. On the mean time, you should be looking at this from my GitHub. You can message message me there instead.

³Also I hope that this is the first book you read out of all of my (hopefully) upcoming books. Since this is my first book, and therefore, this is also the first preface (which should give a good insight to who I am and how I write).

Contents

1	Experiments with random outcomes	1
1.1	Sample Spaces & Probability	1
1.1.1	The Sample Spaces	1
1.1.2	The Event Space	5
1.1.3	The Probability Space	6
1.1.4	Introducing Cartesian Products	7
1.1.5	Set Theory Examples	9
1.2	Random Sampling	13
1.2.1	Population and Sample Size	13
1.2.2	Sampling with replacement, order matters	15
1.2.3	Sampling without replacement, order matters	17
1.2.4	Sampling without replacement, order doesn't matter	20
1.2.5	Sampling with replacement, order doesn't matter	22
1.2.6	Random Sampling Examples	26
1.3	Infinite Many Outcomes	31
1.3.1	Discrete Infinite Outcomes	31

1.3.2	Continuous Infinite Outcomes	32
1.4	Consequences of the Rules of Probability	33
1.4.1	Decomposing an event	34
1.4.2	Events and Complements	39
1.4.3	Monotonicity of Probability	39
1.4.4	More Examples of these Consequences	41
1.4.5	Inclusion-Exclusion Principle	42
1.4.6	Some Examples	46
1.5	Random Variables	47
1.5.1	Graphing and Recording Table of Random Variables	47
1.5.2	Discrete Random Variable	49
1.5.3	Continuous Random Variable	49
1.5.4	Probability Mass Function	50
1.5.5	Properties of PMF's	51
1.5.6	Random Variables and PMF Examples	52
2	Conditional Probability and Independence	55
2.1	Conditional Probability	55
2.1.1	Multiplication Rule	56
2.1.2	Law of Total Possibility	58
2.1.3	Examples of Conditional Probability	58
2.2	Bayes' Formula	60
2.2.1	Weird Observations of Bayes' Theorem	63
2.3	Independence	71

Chapter 1

Experiments with random outcomes

1.1 Sample Spaces & Probability

Definition (Sample Space). The set of all possible outcomes of an experiment 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,

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

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 ” \implies proper subset. So this is why Ω is an event itself.

Question: How many possibilities?

Well you can think of each option within a simple sample

¹Although 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 possibilities/probabilities as measures of a sample space, i.e. asking for “all the possibilities” \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 $\Omega = \{1, 2, 3, 4, \dots, n\}$, 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

$$\mathcal{F} = \left\{ \emptyset, \right. \\ \left. \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \right.$$

²The 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 $\mathbb{P}() \implies 2^\Omega$, without cardinality symbol, thats the best work around I'll probably also use.

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

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

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

$$\begin{aligned}
& \{2, 3, 5, 6\}, \\
& \{2, 4, 5, 6\}, \\
& \{3, 4, 5, 6\}, \\
& \{1, 2, 3, 4, 5\}, \{1, 2, 3, 4, 6\}, \\
& \{1, 2, 3, 5, 6\}, \\
& \{1, 2, 4, 5, 6\}, \\
& \{1, 3, 4, 5, 6\}, \\
& \{2, 3, 4, 5, 6\}, \\
& \Omega \}.
\end{aligned}$$

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

$$\begin{aligned}
\mathcal{F} = & \left\{ \{i_1, i_2, i_3, \dots, i_k\} \mid i_1, i_2, i_3, \dots, i_k \in \Omega, \right. \\
& \left. \forall i_j \neq i_k \text{ “distinct”, for } k = 0, 1, 2, 3, 4, 5, 6 \right\}.
\end{aligned}$$

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 $6 * 6$ 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 $P : \mathcal{F} \rightarrow \mathbb{R}$, 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, \dots 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 (Ω, \mathcal{F}, P) is called a probability space.⁴

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

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

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

Theorem 1.2. *Let (Ω, \mathcal{F}, P) be a probability space. Let A_1, A_2, \dots, 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, $\bigcup A_i$ for ∞ many $\implies \bigcup A_n$ works for finitely many. Seems doable and reasonable, so...

Proof: Let A_1, A_2, \dots, 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 $A_{n+1} = A_{n+2} = \dots = \emptyset$. Then $A_1 \cup \dots \cup A_n = A_1 \cup \dots \cup A_n \cup A_{n+1} \cup A_{n+2} \cup \dots$

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

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

q.e.d.

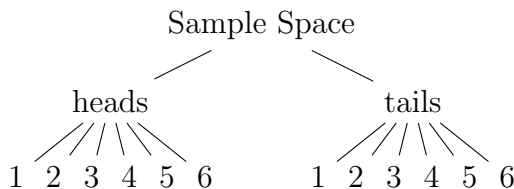
1.1.4 Introducing Cartesian Products

Let A, B be sets. Then

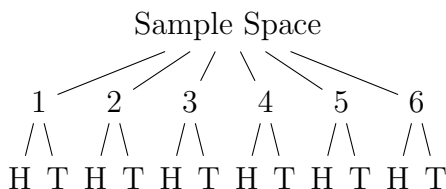
$$A \times B = \left\{ (a, b) \mid a \in A, b \in B \right\}.$$

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

Using what we call a “Tree Diagram” ...



or alternatively

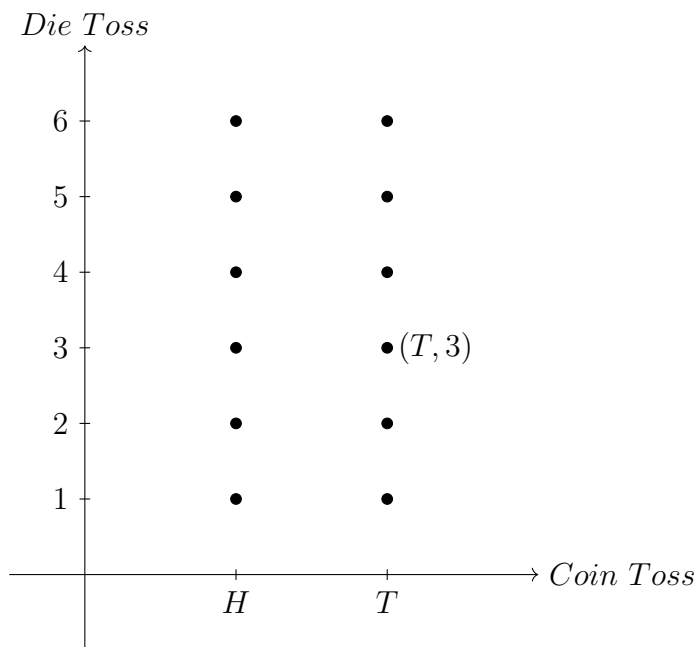


either way, they both imply

$$\Omega = \left\{ (H, 1), (H, 2), (H, 3), (H, 4), (H, 5), (H, 6), \right. \\ \left. (T, 1), (T, 2), (T, 3), (T, 4), (T, 5), (T, 6), \right\}$$

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 cartesian products of 3 or less sets
 (usually only for 2).

⁵Recall: The definition of cartesian 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,

$$\left(\bigcap_i A_i \right)^C = \bigcup_i A_i^C, \quad \text{is true.}$$

Recall: We need to show that the LHS and RHS are subsets of each other.

Similar to “if and only if”, however instead of **saying**

- “if \rightarrow then...”
- Conversely \leftarrow then...
- Therefore \iff ...”

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 $b \in A$... 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 $x \in \left(\bigcap_i A_i \right)^C$.

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

Then $\exists i$ s.t. $x \notin A_i$.

So $x \in \bigcup_i A_i^C$.

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

“ \supseteq ” 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 $x \notin \bigcap_i A_i \implies x \in \left(\bigcap_i A_i \right)^C$,

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 $A \triangle B$ 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. Refer to Fig. 1. Which is somewhat trivial, and does not require much explanation/studying (which is why I dont really consider it a proof). 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

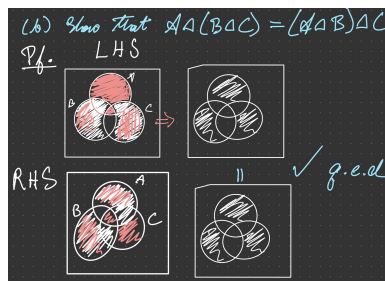


Fig. 1

⁶I am aware (and not naive...I think), that proving this properly would take 10 pages and 5 hours (not actually). Also yess thats my handwriting, see its not that bad, kind of...

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

$$\begin{aligned}(A \triangle B) \triangle C &= (A \cup B \cup C) \cap (A \cup B^C \cup C^C) \cap (A^C \cup B \cup C^C) \cap \\ &\quad (A^C \cup B^C \cup C) \\ &= (B \cup C \cup A) \cap (B \cup C^C \cup A^C) \cap (B^C \cup C \cup A^C) \cap (B^C \cup C^C \cup A) \\ &= (B \triangle C) \triangle A.\end{aligned}$$

Now, \triangle is commutative since by *defⁿ*, obviously
 $A \triangle B = (A \cap B^C) \cup (B \cap A^C) = (B \cap A^C) \cup (A \cap B^C) = B \triangle A$.
So $(B \triangle C) \triangle A = A \triangle (B \triangle C)$. *q.e.d.*

Proof by Stubbornness

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

Then

$$\begin{aligned}(A \triangle B) \triangle C &= ((A \triangle B) \cap C^C) \cup (C \cap (A \triangle B)^C) \\ &= (((A \cap B^C) \cup (B \cap A^C)) \cap C^C) \cup (C \cap ((A \cap B^C) \cup (B \cap A^C))^C)\end{aligned}$$

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,

$$\begin{aligned}&(((A \cap B^C) \cup (B \cap A^C)) \cap C^C) \cup (C \cap ((A \cap B^C) \cup (B \cap A^C))^C) \\ &= (((A \cap B^C) \cup (B \cap A^C)) \cap C^C) \cup (C \cap (A \cap B^C)^C \cap (B \cap A^C)^C) \\ &= (((A \cap B^C) \cup (B \cap A^C)) \cap C^C) \cup (C \cap A^f C \cup B \cap B^C \cup A)\end{aligned}$$

... 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... □

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 elements.

— 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 $\omega = (\text{red}, \text{blue}, \text{red}, \text{clear})$.

How many outcomes are there? Logically.

Well, simply put:

The first marble has n possibilities.

The second marble has n possibilities.

The third marble has n possibilities.

The... ..possibilities.

↓

↓

The k th marble has n possibilities.

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 possibilities.

So we have

$$n * n * n * n * \dots * n = n^k$$

k times

And the number of outcomes is

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

How many outcomes are there? Proof wise.

Proof. Let $S = \{1, 2, \dots, n\}$ 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 $\omega = (s_1, s_2, s_3, \dots, s_k)$, for some $k \in \mathbb{N}$.

Then

$$\Omega = \{(s_1, s_2, \dots, s_k)_{k \in \mathbb{N}} \mid s_i \in S, i = 1, 2, \dots, k\}$$

but this implies

$$\Omega = S \times S \times \dots \times S, \text{ } k \text{ times.}$$

So therefore

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

q.e.d.

1.2.3 Sampling without replacement, order matters

Suppose we have a sample space with n elements.

— 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 there's $n - 2 = 15 - 2 = 13$ marbles left.

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

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

- I grab 1 marble, its red. Ok, I'll put it on the 3rd 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 $\omega = (\text{red}, \text{blue}, \text{red}, \text{clear})$.

How many outcomes are there? Logically.

Well, simply put:

The first marble has n possibilities.

The second marble has $n - 1$ possibilities.

The third marble has $n - 2$ possibilities.

The... ...possibilities.

↓

↓

The k th marble has $n - k$ possibilities, noting that $k \not\geq n$ since you can't pick up more than n marbles.

This does imply that the k th marble has only 1 possibility if instead $k \geq n$, so we should say it has $(n - k) + 1$ possibility.

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

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

How many outcomes are there? Proof wise.

Proof. Let $S = \{1, 2, \dots, n\}$ 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 $\omega = (s_1, s_2, s_3, \dots, s_k)$, 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

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

but this implies

$$\begin{aligned} & n * (n-1) * (n-2) * \cdots * (n-k+1) = nPk \\ &= n * (n-1) * (n-2) * \cdots * (n-k+1) * \frac{(n-k)(n-k-1)(n-k-2) \cdots 3 * 2 * 1}{(n-k)(n-k-1)(n-k-2) \cdots 3 * 2 * 1} \\ &= \frac{n * (n-1) * (n-2) \cdots * (n-k+1)(n-k)(n-k-1)(n-k-2) \cdots 3 * 2 * 1}{(n-k)(n-k-1)(n-k-2) \cdots 3 * 2 * 1} \\ &= \frac{n!}{(n-k)!}. \end{aligned}$$

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 elements.

— 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.

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, I must 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** $\omega = \{2 \text{ red's}, \text{blue}, \text{clear}\}$ in my hand.

How many outcomes are there? Logically.

Well, simply put:

The first marble has n possibilities.

The second marble has $n - 1$ possibilities.

The third marble has $n - 2$ possibilities.

The... ..possibilities.

↓

↓

The k th marble has $n - k$ possibilities, noting that $k \not\geq n$ since you can't pick up more than n marbles.

This does imply that the k th marble has only 1 possibility if instead $k \geq n$, so we should say it has $(n - k) + 1$ possibility.

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

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

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 of the $n!$ possibilities, the $(n - k)!k! \implies (n - k)!k!$.

And so,

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

which makes sense since $nPa \geq nCk$. \square

— Read: “ n choose k ”

How many outcomes are there? Proof wise.

Proof. Let $S = \{1, 2, \dots, n\}$ 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 implies that all ω_i are proper subsets of S .

So

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

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

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

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 elements.

— 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** $\omega = \{2 \text{ red's}, \text{blue}, \text{clear}\}$.

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 $\omega = (x_1, x_2, x_3 \dots x_i)$).

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 fashion, 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. $x_1 = 2 \text{ red} = 1 \text{ red} + 1 \text{ red} = 0 + 2 \text{ red} = 2 \text{ red} + 0$, 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

$x_i = \{3 \text{ chocolate}, 4 \text{ strawberry}, 5 \text{ glazed}\}$ into

$x_i = \{1 \text{ chocolate}, 1 \text{ chocolate}, 1 \text{ chocolate}, 1 \text{ strawberry}, 1 \text{ strawberry}, 1 \text{ strawberry}, 1 \text{ glazed}, 1 \text{ glazed}, 1 \text{ glazed}\}$

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, \dots, 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 wherever 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

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

How many outcomes are there? Proof wise.

Proof. Let $S = \{1, 2, \dots, n\}$ 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 implies that all ω_i are proper subsets of S .

So

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

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

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

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 $n_1, \dots, n_i, \dots, n_k$, 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. $\dots \dot{\cup} \dot{\cup} \dots = \dots \dot{\cup} \emptyset \dot{\cup} \dots$.

So then we have

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

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 \implies putting marbles back into the bag
2. Sampling without replacement \implies keeping the marbles out of the bag
3. Order matters \implies put it in on a rack in order, or mark down the tuple (which came first) \implies permutations
4. Order doesn't matter \implies hold it in your hand, or mark down the set \implies 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 $C = \{2 \text{ E}, 1 \text{ A}, 1 \text{ N}\}$. (Note sets are unordered)
Suppose you have

$$\begin{aligned}
 & P(1st \text{ E}, 2nd \text{ E}, 3rd \text{ A}, 4th \text{ N}) \\
 & + P(1st \text{ E}, 2nd \text{ A}, 3rd \text{ E}, 4th \text{ N}) \\
 & + P(1st \text{ N}, 2nd \text{ A}, 3rd \text{ E}, 4th \text{ E}) + \dots \\
 = & \frac{(5 * 4) * (4) * (3)}{14 * 13 * 12 * 11} + \frac{(5 * 4) * (4) * (3)}{14 * 13 * 12 * 11} + \frac{(3) * (4) * (5) * (4)}{14 * 13 * 12 * 11} + \dots \\
 = & \frac{5 * 4 * 4 * 3}{14 * 13 * 12 * 11} \binom{4}{2} \binom{2}{1} \binom{1}{1} \\
 = & \frac{5 * 4 * 4 * 3}{14 * 13 * 12 * 11} \left(\frac{4 * 3}{2 * 1} \right) \left(\frac{2 * 1}{1} \right) \left(\frac{1}{1} \right) \approx 0.0599. \quad \square
 \end{aligned}$$

- (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 $S = \{1, 2, 3, 4, \dots, 50\}$, then

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

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 separate this into the probability of 2 days, and probability of 3 days (which includes 2).

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

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

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

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

Solution.

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

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,

$$P(\omega) = P(y, g, g) + P(g, y, g) + P(g, g, y)$$

$$\begin{aligned}
&= \frac{3 * 4 * 2}{7 * 6 * 5} + \frac{3 * 2 * 4}{7 * 6 * 5} + \frac{4 * 3 * 2}{7 * 6 * 5} \\
&= \frac{3 * 4 * 2}{7 * 6 * 5} * \binom{3}{1} = \frac{3 * 4 * 2}{7 * 6 * 5} * 3 = \frac{\binom{3}{2} \binom{4}{1}}{\binom{7}{3}}. \quad \square
\end{aligned}$$

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

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

Also our probability measure should be for a finite amount of trials, set this amount to be “ k -trials”. Then⁸,

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

(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

$$\begin{aligned}
P(\infty) &= 1 - \sum_{k=1}^{\infty} P(k) = 1 - \sum_{k=1}^{\infty} \frac{5^{k-1}}{6^k} \\
&= 1 - \frac{a}{1-r} = 1 - \frac{\frac{1}{6}}{1 - \frac{5}{6}}
\end{aligned}$$

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

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

So,⁹

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

1.3 Infinite Many Outcomes

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

1.3.1 Discrete Infinite Outcomes

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

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

(b) Find $P(k)$.

Solution. Well, there are $n = 2$ possibilities for each sample point.

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

The k th event can be described as the following,

$$k = 5 \implies H * H * H * H * H * T.$$

So

$$P(k) = \frac{1 * 1 * 1 * \dots * 1}{2^k}, \text{ for } k = 1, 2, 3, \dots$$

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 $k = 1$,

$$\begin{aligned} 1 - \sum_{k=1}^{\infty} \frac{1}{2^k} &= 1 - \frac{a}{1 - r} = 1 - \frac{\frac{1}{2}}{1 - \frac{1}{2}} \\ &= 1 - \frac{\frac{1}{2}}{\frac{1}{2}} = 1 - 1 = 0. \end{aligned}$$

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

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

Remark. Discrete Infinite Outcomes \implies an Infinite Series.

1.3.2 Continuous Infinite Outcomes

E.g. 1.17- We pick a real number uniformly at random from 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.

$$\begin{aligned} P(\text{hitting the bullseye}) &= \frac{\text{Area}(\text{bullseye})}{\text{Total area}} \\ &= \frac{\pi\left(\frac{1}{9}\right)^2}{\pi(9)^2} = \frac{\frac{1}{16}}{81} = \frac{1}{(16)(81)}. \quad \square \end{aligned}$$

Remark. Continuous Infinite Outcomes \implies an Area/Length (Depending on the Sample Space, usually Euclidean space \mathbb{R}^n , also usually $n = 1, 2$).

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, \dots 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, \dots 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, $P(\text{exactly 1 red or exactly 1 yellow})$.

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.

$P(\text{exactly 1 red or exactly 1 yellow})$

$$= P(\{1r, 1y\} \cup \{1r, 1g\} \cup \{1y, 1g\})$$

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

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

Recall If A_1, A_2, \dots are pairwise disjoint events, then

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

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 $A_i = \{\text{Peter wins at his } i^{\text{th}} \text{ roll}\}$

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”

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* : $T : \frac{4}{6}$

Then

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

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

$$\begin{aligned}
 &= \sum_{i=1}^4 \left(\frac{4}{6} * \frac{2}{6} \right)^{i-1} \left(\frac{2}{6} \right) \\
 &= \frac{1}{3} \sum_{i=1}^4 \left(\frac{2}{9} \right)^{i-1} \\
 &= \frac{1}{3} \left[\left(\frac{2}{9} \right)^0 + \left(\frac{2}{9} \right)^1 + \left(\frac{2}{9} \right)^2 + \left(\frac{2}{9} \right)^3 \right] \\
 &= \frac{1}{3} \left[\frac{1}{3} + \frac{2}{9} + \left(\frac{2}{9} \right)^2 + \left(\frac{2}{9} \right)^3 \right] \\
 &= \frac{1}{3} \frac{1 - \left(\frac{2}{9} \right)^4}{1 - \left(\frac{2}{9} \right)} \approx 0.427 \quad \square
 \end{aligned}$$

(b) What is the probability that May wins?

Solution. Let $B_i = \{\text{Mary wins at her } i^{\text{th}} \text{ roll}\}$
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* : $F : \frac{2}{6}$
5. *Peter* : $F : \frac{4}{6}$
6. *Mary* : $F : \frac{2}{6}$
- .
- .
- .

Then

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

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.

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

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

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

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

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 $B = A \dot{\cup} (B \setminus A)$

($\dot{\cup}$ meaning disjoint union).

Since this is a disjoint union, we can take the individual probabilities.

So,

$$P(B) = P(A) + P(B \setminus A) \geq P(A) + 0 = P(A) \quad q.e.d.$$

E.g. 1.22 Suppose we toss a fair coin repeatedly. What is the probability that tails never occurs?

Recall. We know $P(\text{tails never occur}) = 0$ already. This is by using a summation generalized to k , and as it approaches ∞ , we find $P(A) \rightarrow 0$.

However let's prove this using the new property “Monotonicity of Probability” and the squeeze theorem.

Solution. We know

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

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 k th flip has a probability measure of $\frac{1}{2}$.

So now

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

or alternatively

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

And by the squeeze theorem (from Calculus)

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

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

$P(H) = 0.99 \implies P(T) = 0.01$.

But then, like the previous example, we get the same results, since by squeeze theorem

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

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 $W = \{\text{white ball did not appear}\}$,

$G = \{\text{green ball did not appear}\}$, and

$R = \{\text{red ball did not appear}\}$.

Then

$$\begin{aligned} &P\{\text{we did see all three colors}\} \\ &= P(W \cup G \cup R) \\ &= P(W) + P(G) + P(R) \\ &\quad - P(W \cap G) - P(W \cap R) - P(G \cap R) \\ &\quad + P(W \cap G \cap R) \end{aligned}$$

$$\begin{aligned}
&= \frac{3 * 3 * 3}{4 * 4 * 4} + \frac{3 * 3 * 3}{4 * 4 * 4} + \frac{2 * 2 * 2}{4 * 4 * 4} \\
&\quad - \frac{2 * 2 * 2}{4 * 4 * 4} - \frac{1 * 1 * 1}{4 * 4 * 4} - \frac{1 * 1 * 1}{4 * 4 * 4} \\
&\quad + \frac{0 * 0 * 0}{4 * 4 * 4} \\
&= \frac{3^3 + 3^3 - 1 - 1}{4^3} = \frac{52}{64} = \frac{26}{32} = \frac{13}{16}. \quad \square
\end{aligned}$$

(b) Solve using the complement.

Solution.

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

1.4.5 Inclusion-Exclusion 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

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

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

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

Or more concisely (3)

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

¹¹I 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 $A_i = i^{\text{th}}$ 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

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

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, $A \cap B \subseteq A \implies P(A \cap B) \leq P(A) = 0.4$

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.

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

E.g. Toss a coin. Let X be the number of heads.

$$\Omega = \{H, T\}$$

Then,

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

$\implies X$ 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 $k \in X$.

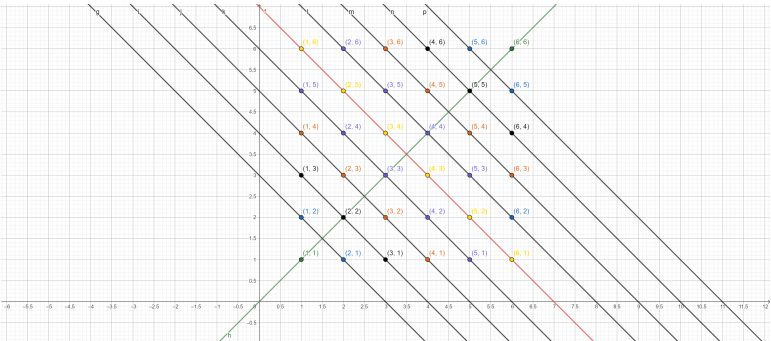
Solution.

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”)



With the table of values being

x	2	3	4	5	6	7	8	9	10	11	12
$P_x(k)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Note the symmetry for $P_X(7)$, also note the non-uniformity, where there is less probabilities for the “rarer” cases.

We say $P_X(k)$ is called the probability mass function of X , commonly abbreviated and pronounced “PMF”.

Quick E.g. Using the table. $P(x = 2, 5, 7) = ?$

Solution. $P(x = 2, 5, 7)$
 $= P(x = 2) + P(x = 5) + P(x = 7)$
 $= P_x(2) + P_x(5) + P_x(7)$
 $= \frac{1}{36} + \frac{4}{36} + \frac{6}{36} = \frac{11}{36}. \quad \square$

1.5.2 Discrete Random Variable

Definition. A random variable X is said to be “discrete” if the set of possible values of X is either finite or countable.

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 $k \in X$.

This example has discrete random variables since there are finite values of x .

1.5.3 Continuous Random Variable

Conversely, consider a continuous interval such as $[0, 1] \in \mathbb{R}$. We know about the cardinality of \mathbb{R} compared to \mathbb{N} , as \mathbb{R} is uncountable.

So $x \in [0, 1] \implies x$ is continuous, and x has uncountable possible values.

E.g. 1.38 We have a dart board of radius 9 inches. The board is divided into four parts by three concentric circles of radii 1, 3, 6 inches. If our dart hits the smallest disk, we get 10 points, if it hits the next region then we get 5 points, and we get 2 and 1 points for the other 2 regions. Let x denote the number of points we get when we throw a dart randomly at the board. Find the probability of x .

Solution. Note: we are throwing 1 dart with discrete random variables, since for each zone, we get a single number (that isn't on a continuous range).

So we should make a table of values (or plot the points if wanted).

Our dart board looks something like this

We'll have circles with:

- $r = 1$
 - inside of this, $0 < r < 1$ is 10 points
- $r = 3$
 - $1 < r < 3$ is 5 points
- $r = 6$
 - $3 < r < 6$ is 2 points
- $r = 9$ which is the edge of the board
 - $6 < x < 9$ is 1 points

And our table of values I will use to calculate,

x	1	2	5	10
$P_x(k)$	$\frac{\pi(9)^2 - \pi(6)^2}{\pi(9)^2}$	$\frac{\pi(6)^2 - \pi(3)^2}{\pi(9)^2}$	$\frac{\pi(3)^2 - \pi(1)^2}{\pi(9)^2}$	$\frac{\pi(1)^2}{\pi(9)^2}$
\parallel	$\frac{45\pi}{81\pi}$	$\frac{27\pi}{81\pi}$	$\frac{8\pi}{81\pi}$	$\frac{\pi}{81\pi}$
\parallel	$\frac{45}{81}$	$\frac{27}{81}$	$\frac{8}{81}$	$\frac{1}{81}$
$=$	$\frac{5}{9}$	$\frac{1}{3}$	$\frac{8}{81}$	$\frac{1}{81}$

Which are the probabilities of x .¹² \square

1.5.4 Probability Mass Function

We probably should finally define the notion of Probability Mass Functions, as we have already observed in different scenarios.

¹²Notice the sum of all probabilities is 1 ($\sum P_x(k) = 1$).

The idea that the sum of all probabilities being 1 is always true for all PMF's. This is due to our axiomatic definition of a probability measure, which is a function $P : \mathcal{F} \rightarrow \mathbb{R}$, s.t. P has the properties of a σ -algebra with total mass 1.

That is, **Definition.** Suppose $X = x_1, x_2, \dots$. Then we say $P_x(x_k) = P(x = x_k)$, for $k = 1, 2, 3, \dots$ is called the probability mass function of X . And we know (from paragraph 2) $P_x(x_k) \geq 0$

$$\implies \sum_{k=1}^{\infty} P_x(x_k) = 1.$$

1.5.5 Properties of PMF's

Now suppose we are given a PMF.

Then we should be able to calculate $P_x(k)$ for any x .

But also we can find properties of these PMF's using the values of $P_x(k)$ we find.

For example

The Weighted Sum

We say

$$\text{Mean of } X = \sum_k k P_x(k)$$

For the previous example **1.38**.

$$\begin{aligned} \text{Mean of } X &= \sum_k k P_x(k) \\ &= 1 * \frac{45}{81} + 2 * \frac{27}{81} + 5 * \frac{8}{81} + 10 * \frac{1}{81} \\ &= \frac{45 + 54 + 40 + 10}{81} \end{aligned}$$

$$= \frac{109}{81} = 1.8395. \quad \square$$

But what does this “mean”¹³?

This is another form of what you may have learned for the mean in highschool, that being

$$\frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

But in probability theory, we have a more sophisticated form of

$$\frac{x_1 + x_2 + \cdots + x_n}{n} = \sum_k k P_x(k)$$

since the probability measure of a random variable has n in the denominator.

But overall, to answer what does this “mean”, we can say that $x_1 + x_2 + \cdots + x_n$ is finite, but large (like the a series) so in a limit process, i.e. over time, your over all average score will be (with arbitrarily many n throws)

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \xrightarrow{x \rightarrow \infty} 1.8395.$$

1.5.6 Random Variables and PMF Examples

E.g. 1.17 An urn contains 4 red balls and 3 green balls. Two balls are drawn randomly.

- (a) Let Z denote the number of green balls in the sample when the draws are done without replacement. Find the probability mass function of Z .

¹³HAHAHAHahaa haa haahaahaa ha... sorry

Solution by Professor.

Z	0	1	2
$P_Z(k)$	$\frac{4 * 3}{7 * 7}$	$\frac{\binom{4}{1}\binom{3}{1}}{\binom{7}{2}}$	$\frac{\binom{4}{0}\binom{3}{2}}{\binom{7}{2}}$

- (b) Let W denote the number of green balls in the sample when the draws are done with replacement. Find the probability mass function of W .

Solution by Professor.

Z	0	1	2
$P_W(k)$	$\frac{4 * 4}{7 * 7}$	$\frac{4 * 3 + 3 * 4}{7 * 7}$	$\frac{3 * 3}{7 * 7}$

E.g. 1.16 We flip a fair coin five times. For every heads you pay me \$1 and for every tails I pay you \$1. Let X denote my net winnings at the end of five flips. Find the possible values and the probability mass function of X .

Solution. ...can be found in the last hw, homework 1.5. \square

Exercise 1.19 You throw a dart and it lands uniformly a random on a circular dartboard of radius 6 inches. If your dart gets to within 2 inches of the center I will reward you with 5 dollars. But if your dart lands farther than 2 inches away from the center I will give you only 1 dollar. Let X denote the amount of your reward in dollars. Find the possible values and the probability mass function of X .

Solution. ...can be found in the last hw, homework 1.5. \square

Chapter 2

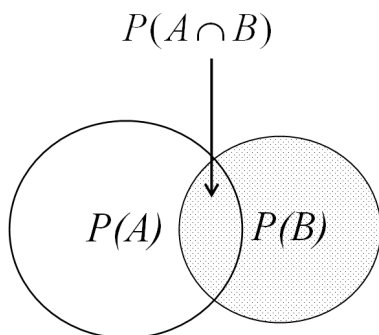
Conditional Probability and Independence

2.1 Conditional Probability

The Definition of Conditional Probability

Let A, B be events. The conditional probability of A given B has occurred is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ if } P(B) \neq 0.$$



(Visually with a venn diagram, think about what probabilities you are dividing by, and why.)

E.g. 2.5 You have 4 reds and 6 greens. Choose 3 balls without replacement. What is the probability you get 2 reds, given that you've have at least 1 red.

*Solution.*¹

$$\begin{aligned}
 & P(\text{exactly 2 red} \mid \text{at least 1 red}) \\
 &= \frac{P(\text{exactly 2 red} \cap \text{at least 1 red})}{\text{at least 1 red}} \\
 &= P \frac{\text{exactly 2 red}}{P(\text{at least 1 red})} \\
 &= \frac{\frac{\binom{4}{2} \binom{6}{1}}{\binom{4}{3}}}{1 - P(\text{no red})} \\
 &= \frac{\frac{\binom{4}{2} \binom{6}{1}}{\binom{4}{3}}}{1 - \frac{\binom{4}{1} \binom{6}{2}}{\binom{4}{3}}} \dots \quad \square
 \end{aligned}$$

2.1.1 Multiplication Rule

$$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A)$$

Proof. By the Definition of Conditional Probability, we have

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ if } P(B) \neq 0.$$

But of course,

$$P(B)P(A|B) = \frac{P(A \cap B)}{P(B)} P(B) \implies P(B)P(A|B) = P(A \cap B)$$

¹(I do not know if I copied this down correctly.)

Therefore,

$$P(A \cap B) = P(B)P(A|B)$$

q.e.d.

E.g. You have 8 reds and 4 greens. Choose 2 balls without replacement.

$$(a) \ P(1st \ red \mid 2nd \ red) = ?$$

Solution.

$$\begin{aligned} P(1st \ red \mid 2nd \ red) \\ &= P(1st \ red)P(2nd \ red \mid 1st \ red) \\ &= \frac{8}{12} * \frac{7}{11}. \quad \square \end{aligned}$$

$$(a) \ P(2nd \ ball \ is \ red) = ?$$

Solution.

$$\begin{aligned} P(2nd \ ball \ is \ red) \\ &= P(1st \ red \cap 2nd \ red) + P(1st \ white \cap 2nd \ red) \\ &= P(1st \ R)P(2nd \ R \mid 1st \ R) + P(1st \ W)P(2nd \ R \mid 1st \ W) \\ &= \frac{8}{12} * \frac{7}{11} + \frac{4}{12} * \frac{8}{11}. \quad \square \end{aligned}$$

2.1.2 Law of Total Possibility

Theorem

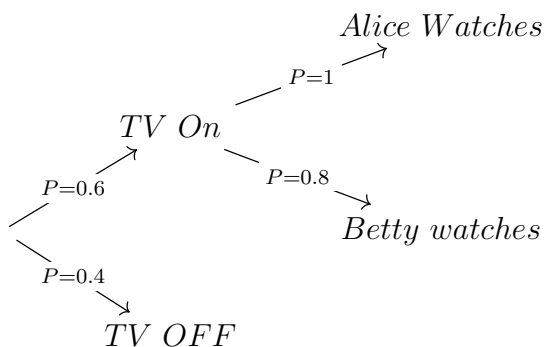
Let $\{B_1, B_2, B_3, \dots, B_n\}$ be a partition² of the sample space Ω . Then

$$P(A) = \sum_{i=1}^n P(B_i \cap A) = \sum_{i=1}^n P(B_i)P(A | B_i)$$

Where the first equality is the partition, while the second equality is generalizing the conditional probability.

2.1.3 Examples of Conditional Probability

Exercise 2.6 When Alice spends the day with the babysitter, there is a 0.6 probability that she turns on the TV and watches a show. Her little sister Betty cannot turn on the TV on by herself. But once the TV is on, Betty watches with probability 0.8. Tomorrow the girls spend the day with the babysitter.



²Recall: A Partition is a set of sets such that it is a disjoint union of a set. i.e. $\{B_1, B_2, B_3, \dots, B_n\}$ is called a partition of Ω if $\bigcup_{i=1}^n B_i = \Omega$ and $B_i \cap B_j = \emptyset, \forall i \neq j$.

- (a) What is the probability that both Alice and Betty watch TV tomorrow?

Solution.

$$\begin{aligned} &P(\text{Alice watches} \cap \text{Betty watches}) \\ &= P(\text{TV on})P(\text{Alice and Betty watches} \mid \text{TV on}) \\ &= (0.6)(1)(0.8) = (0.6)(0.8) = 0.48 . \quad \square \end{aligned}$$

- (b) What is the probability that only Betty watches TV tomorrow?

Solution. $(0.6)(0.8) = 0.48$. \square

- (c) What is the probability that only Alice watches TV tomorrow?

Solution. $(0.6)(0.2) = 0.12$. \square

Exercise 2.3

P (a randomly chosen number between 1 and 100 is divisible by 3 | the number has at least 1 digit equal to 5)

Solution.

$$P = P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$B = \{5, 15, 25, 35, 45, 55, 65, 75, 85, 95, 50, 51, 52, 53, 54, 56, 57, 58, 59\}$$

$$\text{and } A \cap B = \{15, 51, 45, 54, 57, 75\}$$

$$\frac{\frac{6}{100}}{\frac{19}{100}} = \frac{6}{19} . \quad \square$$

Exercise 2.3 From 52 cards deal three cards without replacement. Find $P(1st\ card\ is\ a\ queen,\ 2nd\ is\ a\ king,\ 3rd\ is\ an\ ace)$.

Solution 1.

Intuitively,

$$P(1st\ Q, 2nd\ K, 3rd\ A) = \frac{4}{52} * \frac{4}{51} * \frac{4}{50}. \quad \square$$

Solution 2.

More mathematically,

$$P(A \cap B \cap C) = P(A) P(B|A) P(C|A \cap B)$$

(using multiplication rule)

$$= \frac{4}{52} * \frac{4}{51} * \frac{4}{50}. \quad \square$$

2.2 Bayes' Formula

E.g. Say you have two boxes,

Box 1 := $\{2g, 1r\}$

Box 2 := $\{2g, 3r\}$

Suppose we choose a box randomly, then we draw a ball from the selected box.

If the sample ball is r , what is the conditional probability that the ball comes from box 1.

Solution.

The probability that you get a box 1 or 2 is both $\frac{1}{2}$.

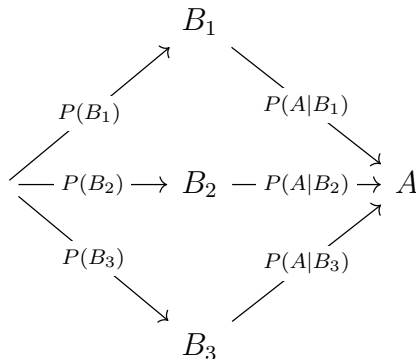
For box 1, getting red has probability $\frac{1}{3}$.

For box 2, getting red has probability $\frac{3}{5}$.

$$\begin{aligned}
 & P(\text{Box 1} \mid r) \\
 &= \frac{P(\text{Box 1} \cap r)}{P(r)} \\
 &= \frac{P(\text{Box 1} \cap r)}{P(\text{Box 1} \cap r) + P(\text{Box 2} \cap r)} \\
 &= \frac{P(\text{Box 1})P(r \mid \text{Box 1})}{P(\text{Box 1})P(r \mid \text{Box 1}) + P(\text{Box 2})P(r \mid \text{Box 2})} \\
 &= \frac{\frac{1}{2} * \frac{1}{3}}{\frac{1}{2} * \frac{1}{3} + \frac{1}{2} * \frac{3}{5}} \\
 &= \frac{\frac{1}{3}}{\frac{1}{3} + \frac{3}{5}} \\
 &= \frac{1}{\frac{5+9}{15}} = \frac{1}{3} * \frac{15}{14} = \frac{5}{14}. \quad \square
 \end{aligned}$$

Shall we try to generalize this?

Say we know,



Definition. 1 Where we call $P(B_1), P(B_2), \dots, P(B_n)$ the “Prior Probabilities”.

Definition. 2 Where we call $P(B_1), P(B_2), \dots, P(B_n)$ the “Posterior Probabilities”.

Of course, **Main Question:** $P(B_i|A) = ?$
i.e. whats the probability going backwards?

Theorem.

Let $B_1, B_2, B_3, \dots, B_n$ be a partition of the sample space Ω .
Then

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)}, \quad \forall i \in \mathbb{N}$$

General procedure:

Quickly remark, j is “all” the B_j ’s, while i is the one we want to find (“the specific B_i ”).

— Calculate $P(A)$. But notice that

— $P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \dots + P(B_j)P(A|B_j)$
by the sum of multiplication rules (instead of writting out all the intersection definitions³, for $j \in \mathbb{N}$.

— This is exactly what is on the bottom of the theorem, i.e.

$$P(A) = \sum_{j=1}^n P(B_j)P(A|B_j).$$

— Now all one needs to do is find the specific i th conditional probability, $P(B_i)P(A|B_i)$, where $1 < i < j$. Or you could just recalculate it.

— Then finally calculate $P(B_i) = \frac{P(B_i)P(A|B_i)}{P(A)}$. This is since,

³however, the professor may ask you to write/prove this for your question, in which case you have to write everything.

$$\begin{aligned}
 P(B_i|A) &= \frac{P(B_i)P(A|B_i)}{P(A)} \\
 &= \frac{P(B_i)P(A|B_i)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_j)P(A|B_j)} \\
 &= P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)}
 \end{aligned}$$

which is exactly our theorem, except simplified.

2.2.1 Weird Observations of Bayes' Theorem

Of course, this is not difficult⁴ after a few examples.

E.g. Say the probability of picking either box is equal.

- Box 1 := {2g, 1r}
- Box 2 := {3g, 5r}

Say we randomly choose one box.

Then we draw a ball randomly from the selected box.

If the sample ball is r , what is the conditional probability that the ball comes from box #2.

⁴seriously, I'm not lying, I think...

Proof Style. (like above)

$$\begin{aligned}
 P(\text{Box 2} \mid r) &= \frac{P(\text{Box 2} \cap r)}{P(r)} \\
 &= \frac{P(\text{Box 2} \cap r)}{P(\text{Box 2} \cap r) + P(\text{Box 1} \cap r)} \\
 &= \frac{P(\text{Box 2})P(r \mid \text{Box 2})}{P(\text{Box 1})P(r \mid \text{Box 1}) + P(\text{Box 2})P(r \mid \text{Box 2})} \\
 &= \frac{\frac{1}{2} * \frac{5}{8}}{\frac{1}{2} * \frac{1}{3} + \frac{1}{2} * \frac{5}{8}} \\
 &= \frac{\frac{5}{16}}{\frac{1}{6} \left(\frac{8}{3}\right) + \frac{5}{16} \left(\frac{3}{3}\right)} \quad (\text{just making common denominators}) \\
 &= \frac{\frac{5}{16}}{\frac{8+15}{48}} = \frac{5}{16} * \frac{48}{23} = \frac{15}{23}.
 \end{aligned}$$

q.e.d.

Better Workflow Solution.

We know:

- $P(B_1) = 1/2$
- $P(B_2) = 1/2$

We need to find:

1. $P(r) = ?$, and remember conditional of B_2
2. $\frac{\text{conditional of } B_2}{P(r)}$

Now,

$$\begin{aligned}
 P(r) &= P(B_1)P(r|B_1) + \boxed{P(B_2)P(r|B_2)} \quad (\boxed{\text{remember}}) \\
 &= \left(\frac{1}{2}\right) \frac{1}{3} + \boxed{\left(\frac{1}{2}\right) \frac{5}{8}} \\
 &= \frac{1}{6} + \boxed{\frac{5}{16}} = \frac{8}{48} + \boxed{\frac{15}{48}} = \frac{\mathbf{23}}{48}
 \end{aligned}$$

And finally

$$\begin{aligned}
 P(B_2|r) &= \frac{\boxed{P(B_2)P(r|B_2)}}{\mathbf{P(B_1)P(r|B_1)} + \mathbf{P(B_2)P(r|B_2)}} \\
 &= \frac{\left(\frac{1}{2}\right) \frac{5}{8}}{\left(\frac{1}{2}\right) \frac{1}{3} + \left(\frac{1}{2}\right) \frac{5}{8}} \\
 &= \frac{\frac{5}{16}}{\frac{23}{48}} = \boxed{\frac{5}{16}} \left(\frac{\mathbf{48}}{\mathbf{23}}\right) = \frac{15}{23}. \quad \square
 \end{aligned}$$

Remark: $\frac{15}{23} > \frac{1}{2}$. Or in english, this probability is **higher** than the prior probability. The probability that you chose the 2nd box if you have a red ball is higher than the probability you choose a random box. Which should make sense to you!⁵

E.g. 2.14 Suppose we have a medical test that detects a particular disease 96% of the time, but gives false positive 2% of the time. Assume that 0.5% of the population carries the disease. If a random person test positive for the disease, what is the probability that they actually carry the disease?

This is a typical example of using bayes' theorem. I will use the *Better Workflow Solution*. but even faster.

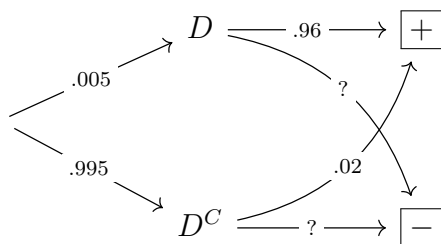
⁵I hope...

Better Workflow Solution (Faster).

Let D be the event “you are diseased”⁶, and the $\boxed{\pm}$ is tested positive or negative.

The question asks, $P(D) = ?$

A Quick Sketch:



$P(D|+) \implies \text{find } P(+|D)P(D) \wedge P(+)$
 thinking swap $D, +$

Now,

$$\begin{aligned}
 P(+) &= \boxed{P(D)P(+|D)} + P(D^C)P(+|D^C) \\
 &= \boxed{.005(.96)} + .995(.02) = \boxed{.0048} + .0199 \\
 &= \mathbf{.0247}
 \end{aligned}$$

And finally,

$$\begin{aligned}
 P(D|+) &= \frac{\boxed{P(D)P(+|D)}}{\mathbf{P(D)P(+|D)} + \mathbf{P(D^C)P(+|D^C)}} \\
 &= \frac{\boxed{.0048}}{\mathbf{0.247}} = .194 \quad \square
 \end{aligned}$$

⁶is this even a word?

Some Remarks of E.g. 2.14

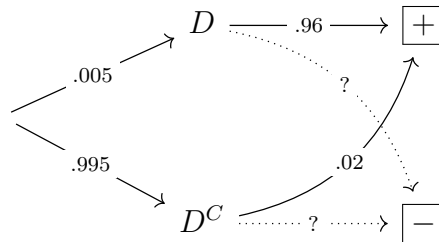
Observe the “prior probabilities” are

- $P(D) = .005$
- $P(D^C) = .995$

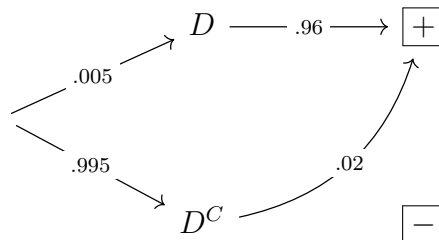
And the “posterior probabilities” are

- $P(+|D) = .96$
- $P(+|D^C) = .02$
- $P(-|D) = ?$
- $P(-|D^C) = ?$

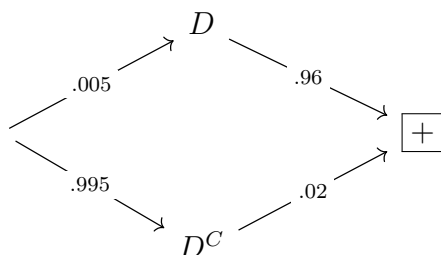
Notice that we were missing some of these, but it didn't matter since we were trying to find the conditional probability of the first, i.e. $P(D|+)$. Which is easy to see with the tree diagram.



In fact, here it is simplified:



which is the obviously just:



Also notice our final value, $P(D|+) = \boxed{.194}$.

This seems very low. *INSERT YOUR CONCERNED FACE.*

Anyways, this is pretty counterintuitive, but there's an interesting "philosophical" technicality. If you notice the prior probabilities, you can think of this (technically) as the probability that a random chosen person (from the side of the road or something) is sick. This implies, that the probability that they have a false positive is pretty high, since you don't know if they were sick or not. Or you can alternatively think the false positive rate is high since there is a very high chance they aren't actually sick.

Thinking like this should be similar to saying, there are more people that are sick that still test positive.

At least compared to those who are sick, even though they almost always test positive.

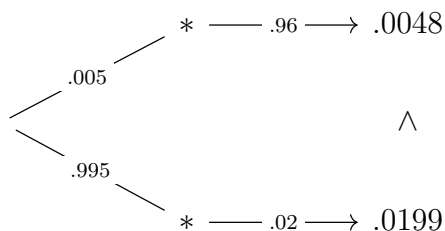
“Unraveling the Tree”

*I discovered this technique,*⁷ which is useful in checking if my answer makes sense.

⁷I didn't search this up, but it has probably been invented somewhere. If so, I do not take credit, but it be cool if I found new math more often...

Claim: *Depending on the products of each conditional probability, one of the “branches” have a higher probability than the other.*

Visually, one should think like this:



This does have its uses, since you're just computing the conditional probabilities ahead of time.

Let's actually modify **Ex. 2.14** to make the final value less concerning, I will also use this new method to calculate this even faster, then the faster version, of the *Better Workflow Solution*.⁸

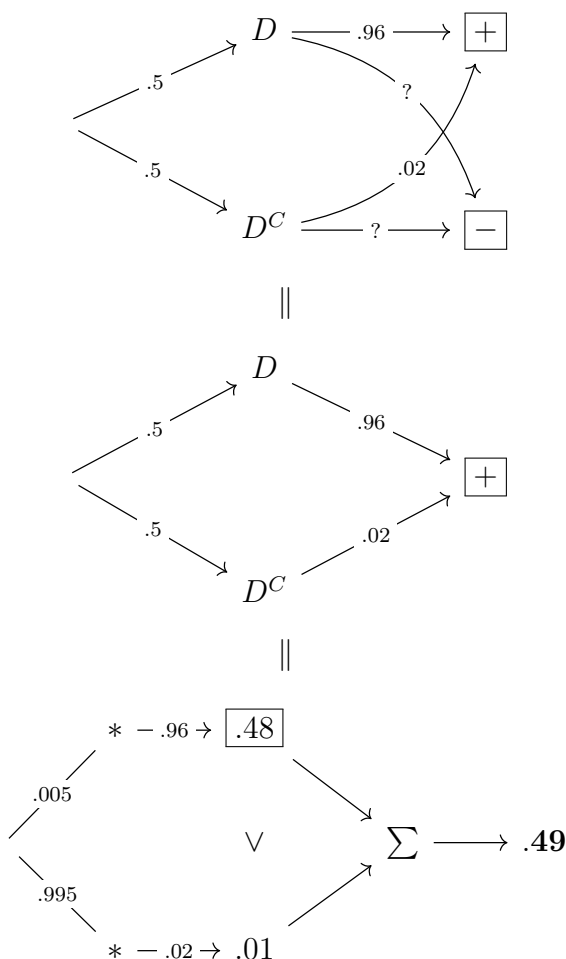
E.g. 2.14 (Modified) Let the posterior probabilities, i.e. the probability that you are sick or not, be equal.

Why? This number is out of thin air (that is $P(D) = P(D^C) = 1/2$, but the idea is that “if you are at the doctor, you are probably half as likely to be sick or not”. But to step away from making this too philosophical, (going back to math) this basically just means the posterior probabilities are normalized, so the only thing that changes final answer is the conditional probabilities.

(i.e. if you test positive, you are definitely sick (at least .96 of the time) and vice versa)

⁸you don't believe me, do you...

Enough Rambling Dean, Solution by rushing through Procrastination.



The $.48 > .01$ is not important, but good to know, and it makes sense. Now, $P(D|+) = \frac{\boxed{.48}}{.49} = .979 \quad \square^9$

⁹See told you so! But also you should probably do it *Better Workflow Solution (Faster)*. way for the professor to give most credit. Still, as I said, this useful to check work.

Final Remark Yes! This value makes sense in this context. The average person will most likely test positive, assuming that they went to the doctors and have a half chance of being sick in the first place...ok i'll shut up.

2.3 Independence

¹⁰ **Definition.** Events A and B are independent if

$$P(A \cap B) = P(A)P(B)$$

E.g. Suppose we have a bag with 2 green and 3 red. Choose 2 balls with replacement. $P(1st\ red \cap 2nd\ red)$

Solution

Since the sample size stays the same (with replacement), the events are independent.

$$\begin{aligned} P(1st\ red \cap 2nd\ red) &= P(1st\ red)P(2nd\ red) \\ &= \frac{3}{5} * \frac{3}{5} = \frac{9}{25} \end{aligned}$$

Good question from another student: What's the difference between this and disjoint events?

well disjoint events you can represent and draw (with venn diagram). You cannot represent independent events.

¹⁰ something I don't believe I have even though in 20 (at the time of writing this book)

Bibliography

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