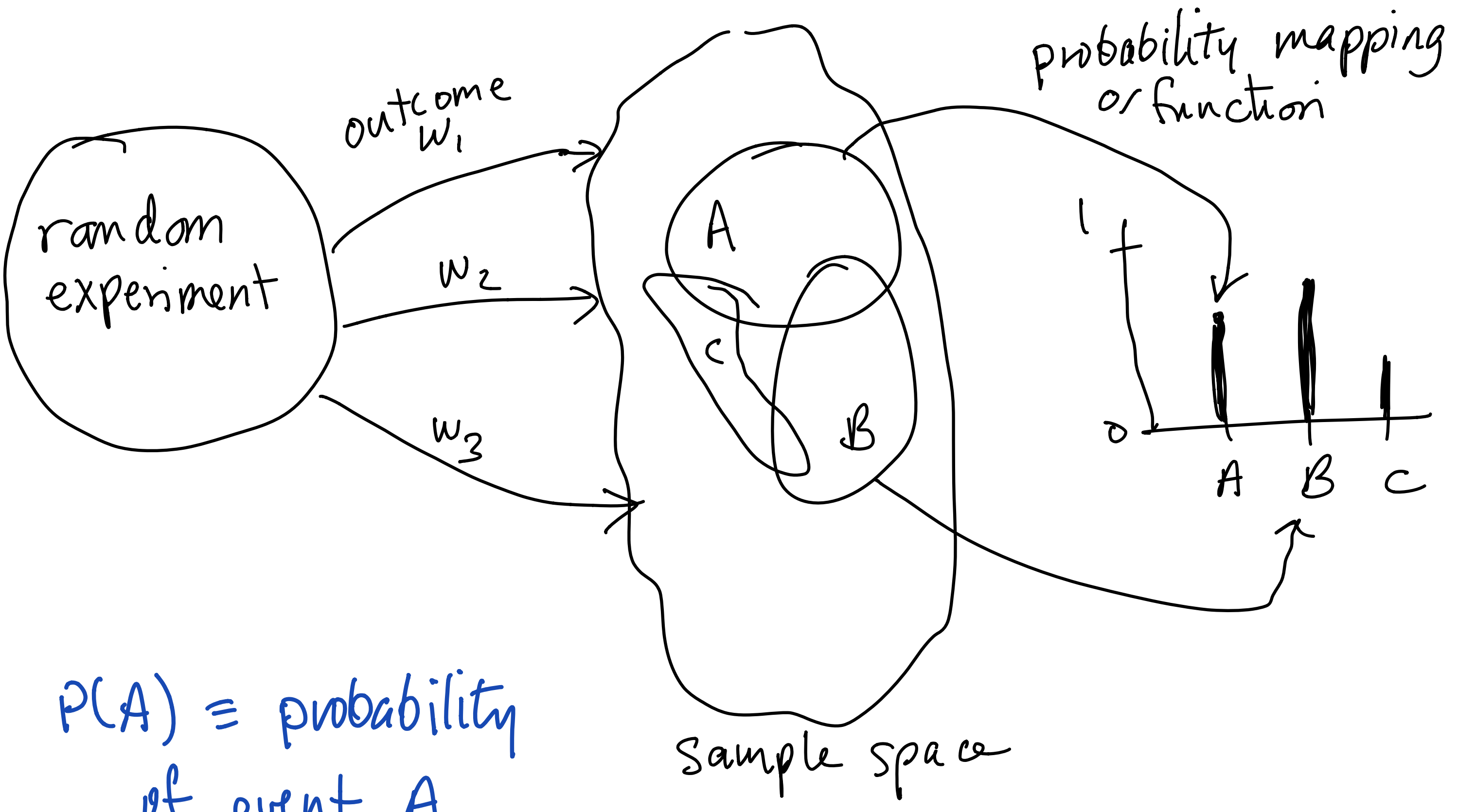
Probability models and the Axioms of Probability (Chapter 2.2)

3 components of a probability model

1. Sample space (set of all possible outcomes)
2. Events (subsets of the sample space)
3. Probability mapping $P(\cdot)$ a function



$P(A) \equiv$ probability of event A

P: event $\longrightarrow [0,1]$

The probability function takes an event and maps it into a real number between 0 and 1 inclusive

Several interpretations of probability

- how likely an event is
- relative frequency, based on repeated experimental trials

ex: if you perform the same experiment N times, and observe event A N_A times, then

$$r_A = N_A / N \equiv \text{relative frequency of event } A$$

and as N gets large,

$$P(A) \approx r_A$$

But! not all experiments are repeatable!

- Probability as belief
  - probability expresses how confident you are that an event will happen

Kolmogorov proposed a unified, axiomatic theory of probability in 1930s

Axioms of probability (Section 2.2)

An **axiom** is a statement or proposition that is regarded as accepted, or self-evidently true

Axiom I : $P(A) \geq 0$

Axiom II : $P(S) = 1$

Axiom III : If $A \cap B = \emptyset$ (A and B are disjoint)
then $P(A \cup B) = P(A) + P(B)$

These are the basics. Everything else can be derived from these!

Axiom III$'$: if $A_1, A_2, \ldots$ is an infinite sequence of events such that A_i and A_j are disjoint for all $i \neq j$ $(A_i \cap A_j = \emptyset)$
then

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

We will build useful facts about probability from these axioms

Corollaries to the axioms

Corollary 1: $\boxed{P(A^c) = 1 - P(A)}$

why? recall $S = A \cup A^c$ and $A \cap A^c = \phi$

so $P(A \cup A^c) = P(S) = 1$ (by Axiom II)

$= P(A) + P(A^c)$ (by Axiom III)

so $P(A^c) = 1 - P(A)$

Corollary 2: $\boxed{P(A) \leq 1}$

why? $P(A) \geq 0$ and $P(A^c) \geq 0$

and $P(A^c) = 1 - P(A) \geq 0 \Rightarrow P(A) \leq 1$

Corollary 3: $\boxed{P(\phi) = 0}$

why? $\phi = S^c$

$P(\phi) = P(S^c) = 1 - P(S) = 1 - 1 = 0$

Corollary 4: Suppose we have an event A that is a bit complicated. We can break A into disjoint components: $A_1, A_2, \ldots, A_n$. Such that $A = \bigcup_{i=1}^{n} A_i$

Then $P(A) = \sum_{i=1}^{n} P(A_i)$

Corollary 5: $\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$

Note: Axiom III can only be used when A and B are disjoint. This can be used for any A and B.

Why? We'll use a standard approach: manipulate the set inside $P(\cdot)$ to call it something else, <u>without</u> changing its contents

	B	B^c
A	$A \cap B$	$A \cap B^c$
A^c	$A^c \cap B$	$A^c \cap B^c$

write

$$A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$$

All 3 pieces are disjoint

Apply axiom III

$$P(A \cup B) = P(A \cap B) + P(A \cap B^c) + P(A^c \cap B)$$

and

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

$$P(B) = P(A \cap B) + P(A^c \cap B)$$

combining

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

We will use this corollary extensively!

Corollary 6: extends this to more than 2 events

ex:

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

(add all individual, subtract all pairwise, add all three way, subtract all 4 way, etc)

Corollary 7 $\boxed{\text{If } A \subset B, \text{ then } P(A) \leq P(B)}$

(basically, can write $B = A + C$ where $C = B - A$ so $A \cap C = \phi$. $P(C) \geq 0$, so $P(A) \leq P(B)$)

This can be useful if events get complicated. We may not know all events in A but we could find a simpler set B that contains A.

Union Bound: $\boxed{P(A \cup B) \leq P(A) + P(B)}$

This can be useful when we don't know how to characterize $A \cap B$, or if $P(A)$ and $P(B)$ are both small (i.e., A and B are both rare)

Breaking an event into disjoint components and applying Corollary 4 is a key strategy for solving probability problems.

1) We can apply it to the sample space

Suppose the collection of sets $A_1, A_2, \dots, A_n$ partition the sample space.

Then $P(S) = 1 = P(A_1 \cup A_2 \cup \dots A_n)$

$= P(A_1) + P(A_2) + \dots P(A_n)$

$$1 = \sum_{i=1}^{n} P(A_i)$$

2) If we have some other set, B, we can break it into pieces based on the partition of A_i's.

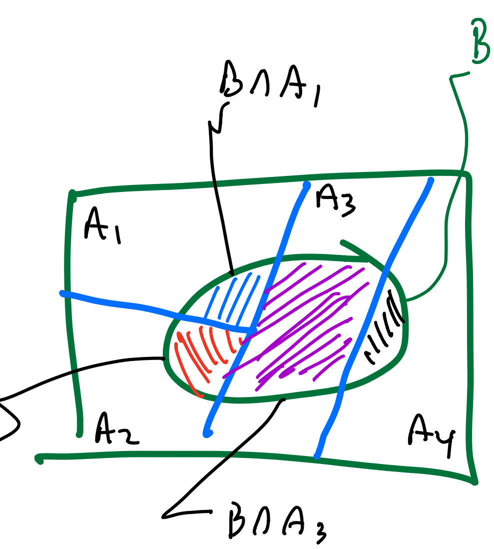
$B = B \cap S = B \cap (A_1 \cup A_2 \cup \dots \cup A_n)$

$= (B \cap A_1) \cup (B \cap A_2) \cup \dots \cup (B \cap A_n)$

$= \bigcup_{i=1}^{n} B \cap A_i$ and so

$$\boxed{P(B) = \sum_{i=1}^{n} P(B \cap A_i)}$$

one version of the theorem of total probability



A problem-solving tool: The probability table

A table of disjoint events, each labeled with a probability

	A_1	A_2	A_3	A_4	
B_1					$P(B_1)$
B_2					$P(B_2)$
	$P(A_1)$	$P(A_2)$	$P(A_3)$	$P(A_4)$	1

The A's form a partition

The B's form another partition

The sum of the probabilities across rows are $P(B_i)$

The sum of the probabilities down the columns are $P(A_j)$

Together, since the A's form a partition,
the sum along the bottom is 1

and since the B's form a partition,
the sum down the right column is also 1

Note that together, the events associated with each block also form a partition

Exercise 3.

Chips are produced by two machines, and each chip has a probabiity of being bad. The probability a chip is made from machine 1 is 0.4, and the probability that a chip is good is 0.7. Also, suppose we know that the probability of a chip being either bad or made from machine 1 is 0.7.

(a) What is the probability a chip is made from machine 2.

(b) What is the probability a chip is made from machine 1 and is bad?

$A = \{\text{chip made on machine } 1\}$

$B = \{\text{chip is bad}\}$

We can turn this word problem into math. Define A as the event that the chip is made from machine 1, and define B as the event that the chip is Bad. Then $P(A) = 0.4$, $P(B^c) = 0.7$, and $P(A \cup B) = 0.7$. The questions are then to determine

(a) $P(A^c)$.

(b) $P(A \cap B)$.

$P(A) = 0.4$

$P(B^c) = 0.7$

$P(A \cup B) = 0.7$

a) $P(A^c) = 1 - P(A) = 1 - 0.4 = 0.6$

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

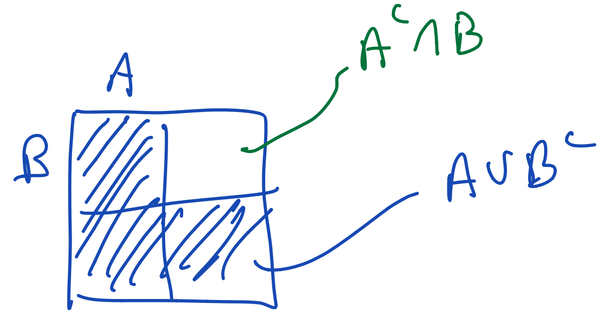
$= 0.4 + [1 - P(B^c)] - 0.7$

$= 0.4 + 0.3 - 0.7$

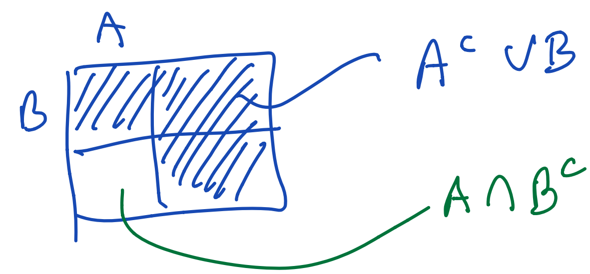
$= 0$

$\Rightarrow$ machine 1 does not make bad chips!

Exercise 6.
Events A and B are defined on a sample space S. if $P(A) = 0.6$, $P(B) = 0.5$, and $P(A \cup B^c) = 0.9$, what is $P(A^c \cup B)$?



Given: $P(A \cup B^c) = 0.9$.



Want: $P(A^c \cup B)$.

Break $A \cup B^c$ into disjoint events.

$$P(A \cup B^c) = P(A \cap B) + P(A \cap B^c) + P(A^c \cap B^c)$$
$$= 1 - P(A^c \cap B) = 0.9$$

so $P(A^c \cap B) = 0.1$

And $P(A^c \cup B) = P(A^c) + P(B) - P(A^c \cap B) = (1 - 0.6) + 0.5 - 0.1$
$= 0.8$

In fact, there are many ways to solve this.
The completed probability table looks like:

	A	A^c	
B	0.4	0.1	0.5
B^c	0.2	0.3	0.5
	0.6	0.4	