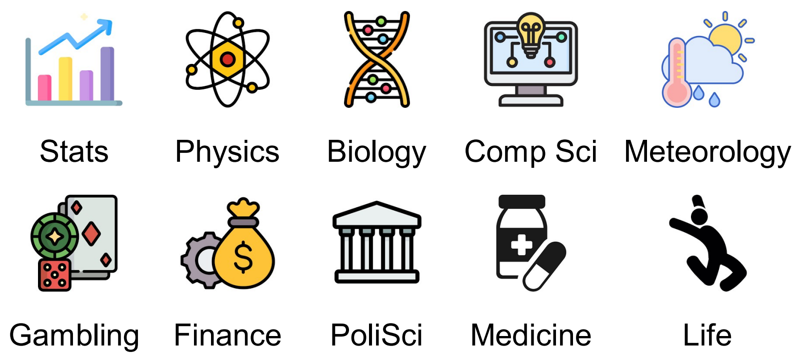
Probability Theory

Textbook: *Introduction to Probability* by Blitzstein and Hwang

§1.1 - Motivation

- Probability (prob) is the logic of uncertainty. We build prob models to help us understand:



- Intuition fails us (as it did for Newton/Leibniz w.r.t. probability).
  Prob teaches us procedures to shore-up our intuition, avoid fallacies.

- Real life is messy, requiring prob, not deductive logic.

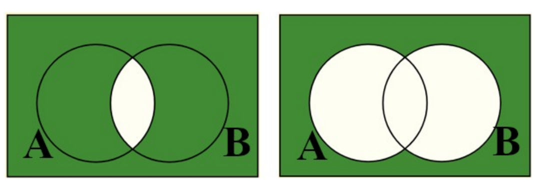
- Resolving philosophical arguments often requires prob.

Harvard Video: youtube.com/watch?v=KbB0FjPg0mw&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=2

§1.2 - Sample Spaces

Prob is built on *sets*: review Appendix A.1.1 - A.1.4.

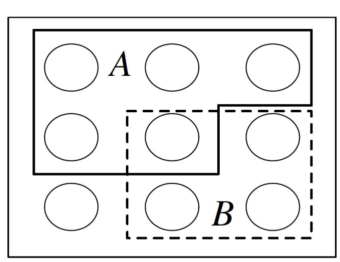
Useful - **De Morgan's Law**:



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

Prob often imagines an *experiment* is conducted. Before the experiment, the *outcome* is unknown.

Def (**Sample Space**): Sample space S of an experiment is set of all possible *outcomes* (could be finite or infinite).
Event A is a subset of S. We say A *occurred* if the actual outcome is *in* A.



Performing the experiment amounts to randomly selecting one outcome.

Example (Ex) - (Coin Flips). A coin is flipped 10 times (heads H, tails T).

A possible outcome is *HHHTHHTTHT*. Sample space?

Let's code H, T as $1, 0$, so that an outcome is a sequence $(s_1, \dots, s_{10})$ w/ $s_j \in \{0,1\}$.

We can define some events:

- ♦ A_1 - first flip is heads:

  $A_1 = \left\{(1, s_2, \dots, s_{10}) \,:\, s_j \in \{0,1\} \text{ for } 2 \le j \le 10\right\}$. Or similarly A_j.

- ♦ B - at least one flip was heads:

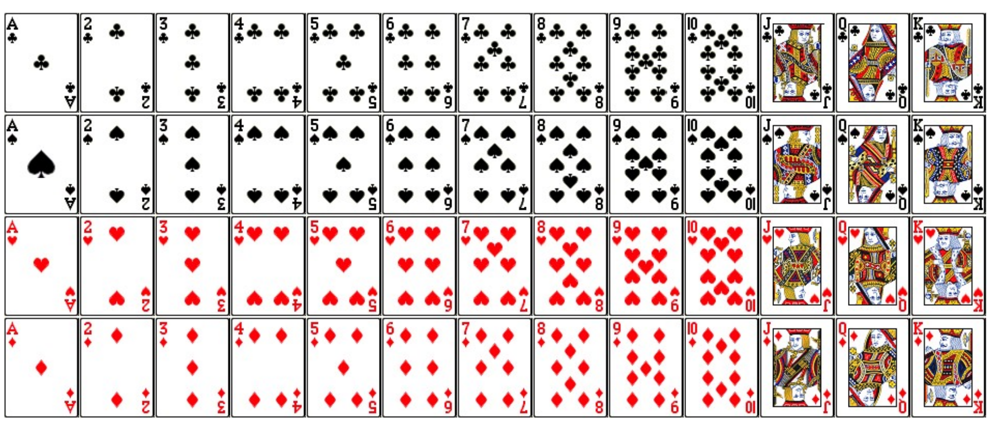
  $B = \bigcup_{j=1}^{10} A_j$. (this means $A_1 \cup A_2 \cup \dots \cup A_{10}$)

- ♦ C - all the flips were heads:

  $C = \bigcap_{j=1}^{10} A_j$. (this means $A_1 \cap A_2 \cap \dots \cap A_{10}$)

- ♦ D - there were at least two consecutive heads:

  $D = \bigcup_{j=1}^{9} (A_j \cap A_{j+1})$.



Ex (**Deck of Cards**). Pick a card from a standard deck of 52 cards. The sample space S is set of 52.

Consider events:

- ▸ A - is an ace
- ▸ B - has a black suit.
- ▸ D - is a diamond ♦.
- ▸ H - is a heart ♡.

What are the following events:

- $A \cap H$
- $A \cap B$
- $A \cup D \cup H$
- $(A \cup B)^c$

Some Basics: Let A, B be events.
$P(A) = 1 - P(A^c)$.   (complement rule)

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ (prob of a union)

$P(B) = P(B \cap A) + (B \cap A^c)$ (breaking up an event using a partition)

$P(A \cup A^c) = P(A) + P(A^c)$ (prob of disjoint unions can be summed)

Activity 2

Set Dictionary.

English	**Sets**
Events and occurrences	
sample space	S
s is a possible outcome	$s \in S$
A is an event	$A \subseteq S$
A occurred	$s_{\text{actual}} \in A$
something must happen	$s_{\text{actual}} \in S$
New events from old events	
A or B (inclusive)	$A \cup B$
A and B	$A \cap B$
not A	A^c
A or B, but not both	$(A \cap B^c) \cup (A^c \cap B)$
at least one of $A_1, \ldots, A_n$	$A_1 \cup \cdots \cup A_n$
all of $A_1, \ldots, A_n$	$A_1 \cap \cdots \cap A_n$

Relationships between events	
A implies B	$A \subseteq B$
A and B are mutually exclusive	$A \cap B = \emptyset$
$A_1, \ldots, A_n$ are a partition of S	$A_1 \cup \cdots \cup A_n = S, A_i \cap A_j = \emptyset$ for $i \neq j$

§1.3 - Naive Definition of Probability

Count # of ways an **event** could happen, then divide by **total** # of possible outcomes for the experiment.

Prob of rolling 6? There are 6 possible outcomes, 1 way in which "6" happens: $\frac{1}{6}$.

Def (**Naive Probability**): Let A be an event for an experiment w/finite S.

Naive prob of A: $P_{\text{naive}}(A) = \frac{|A|}{|S|} = \frac{\text{\# of outcomes favorable to } A}{\text{total \# of outcomes in } S}$.

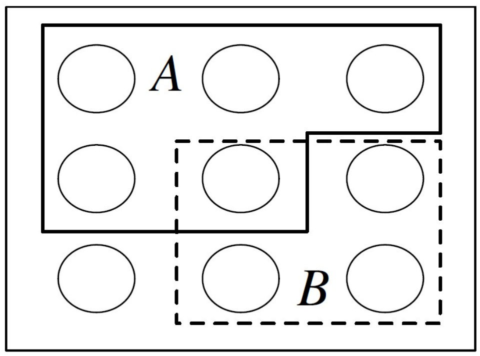


Figure 1: Depiction of a sample space

From figure 1, calculate these: $P_{\text{naive}}(A) = ?$, $P_{\text{naive}}(B) = ?$, $P_{\text{naive}}(A \cup B) = ?$, $P_{\text{naive}}(A \cap B) = ?$

$$P_{\text{naive}}(A) = \tfrac{5}{9}, \quad P_{\text{naive}}(B) = \tfrac{4}{9}, \quad P_{\text{naive}}(A \cup B) = \tfrac{8}{9}, \quad P_{\text{naive}}(A \cap B) = \tfrac{1}{9}.$$

$$P_{\text{naive}}(A^c) = ?, \quad P_{\text{naive}}(B^c) = ?, \quad P_{\text{naive}}((A \cup B)^c) = ?, \quad P_{\text{naive}}((A \cap B)^c) = ?.$$

$$P_{\text{naive}}(A^c) = \tfrac{4}{9}, \quad P_{\text{naive}}(B^c) = \tfrac{5}{9}, \quad P_{\text{naive}}((A \cup B)^c) = \tfrac{1}{9}, \quad P_{\text{naive}}((A \cap B)^c) = \tfrac{8}{9}.$$

Note that $P_{\text{naive}}(A^c) = $??

$$= \frac{|A^c|}{|S|} = \frac{|S|-|A|}{|S|} = 1 - \frac{|A|}{|S|} = 1 - P_{\text{naive}}(A).$$

Restrictions: S must be finite. Outcomes must be equally likely.

Later (non-naive) definition won't have these restrictions.

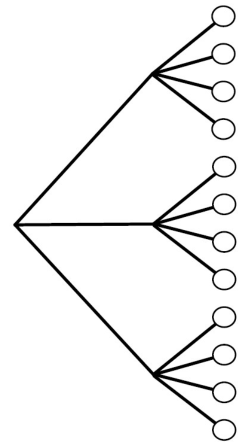
So it seems we'll have to do a lot of counting...

§1.4 - How to Count

How to count the # of outcomes in A and S?

It can get complicated, so we need better tools!

Thm (**Multiplication Rule**): Consider an experiment consisting of two sub-experiments:
Experiment A and Experiment B. Suppose A has a possible outcomes,
and for each of those outcomes B has b possible outcomes.
Then the compound experiment has ab possible outcomes.

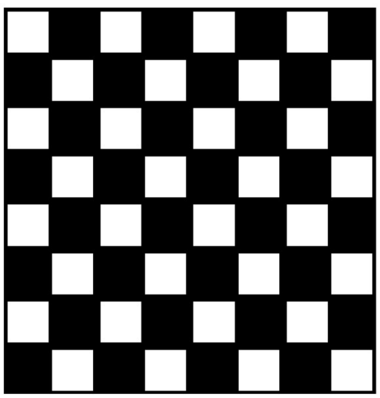


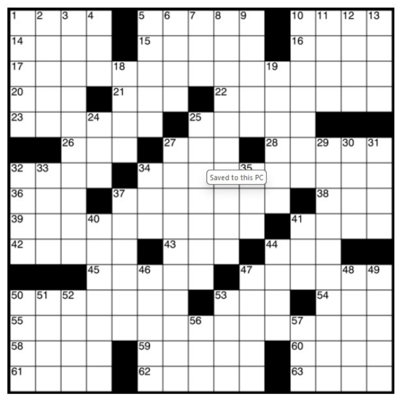
Tree Diagram

A has 3 outcomes,

B has 4 outcomes.

❗ Experiment A need not occur *before* Experiment B (they could occur simultaneously!).





Ex: How many squares are there in an 8×8 chessboard?

Solution: To specify a square, is to say which row (experiment A) and column (experiment B) it's in.

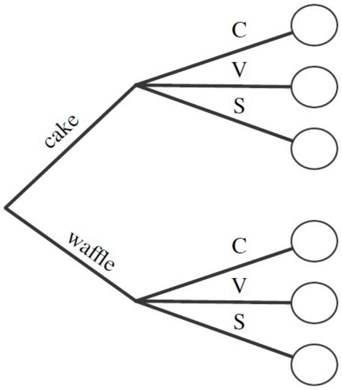
8 choices of row, each of which has 8 choices of column. (64)

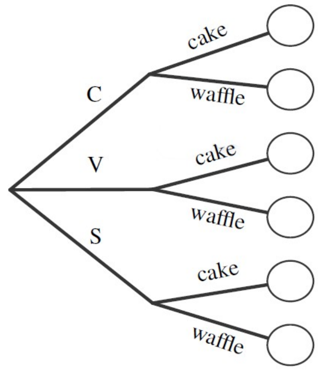
How many white squares?

How many white squares in the crossword puzzle? (symmetry is useful!)

Ex: You're buying an ice cream cone. You choose whether to have a cake or a waffle cone,
and whether to have chocolate, vanilla, or strawberry ice cream.

Visualized with a tree diagram:





Doesn’t matter whether you choose type of cone first ("waffle w/chocolate ice cream, plz.") or flavor first ("chocolate ice cream on a waffle.").

Suppose you buy two ice cream cones on a certain day, one in the afternoon and the other in the evening. How many possibilities?

$6 \times 6 = 6^2 = 36$ possibilities.

How many if you’re only interested in what **kinds** of ice cream cones you had that day, not the order in which you had them, so you don’t want to distinguish between (cakeC, waffleV) and (waffleV, cakeC)?

Are there now $36/2 = 18$ possibilities?

No, since possibilities like (cakeC, cakeC) were already only listed once each in the 6^2.

So, there are only $6 \times 5 = 30$ ordered possibilities (x, y) with $x \neq y$, which turn into 15 possibilities if we treat (x, y) as equivalent to (y, x).

Adding back in the 6 w/the form (x, x), gives 21 possibilities. ■

Counting w/different kinds of sampling

In experiments, we often sample from a larger population.

Multiplication Rule lets us count when sampling with, or without, *replacement*.

Thm (**Sampling w/Replacement**): Consider n objects from which we choose k of them, one at a time w/replacement (choosing an object doesn't preclude it from being chosen again). There are n^k possible outcomes (where order matters $(3,7) \neq (7,3)$).

Concretely: Imagine a jar w/n balls, labeled 1 to n. Sample k balls, one at a time w/replacement, (each time a ball is chosen, return it to the jar). Each sampled ball is a sub-experiment w/n possible outcomes. There are k sub-experiments. Thus, by multiplication rule there are n^k ways to obtain a sample of size k.

Thm (**Sampling without Replacement**): Consider n objects from which we choose k of them, one at a time **without** replacement (choosing an object precludes it from being chosen again). Then there are $n(n-1)\ldots(n-k+1)$ possible outcomes for $1 \leq k \leq n$, and 0 possibilities for $k > n$ (where order matters).

Proof: This follows from multiplication rule: each sampled ball is a sub-experiment.

The # of outcomes decreases by 1 each time.

Ex (**Permutations and Factorials**). A permutation of $1,2,\ldots,n$ is an arrangement of them in some order, e.g., $3,5,1,2,4$ is a permutation of $1,2,3,4,5$. By previous thm w/$k = n$, there are $n!$ permutations of $1,2,\ldots,n$.

Label objects! Or else you'll wrongly believe that  is same outcome as . $(3,4) \neq (4,3)$.

Adjusting for Overcounting

It's very easy to overcount (like we did with the ice cream).

Ex (**Teams of Two**): Consider a group of four people $\{1,2,3,4\}$. How many ways are there to choose a two-person committee?

Here, order doesn't matter. That is: $(1,2)$ *is* the same as $(2,1)$, DON'T OVERCOUNT! (It's not $4 \times 3 = 12$)

The possibilities are: $(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)$ $\left(\frac{4\times 3}{2} = 6 \text{ ways}\right)$

So how, in general, do we count the (unordered) # of ways to choose k objects out of n, without replacement (unordered means $(3,1,4) = (4,1,3)$)?

Def (**Binomial Coefficient**): For any nonnegative integers k and n, the binomial coefficient $\binom{n}{k}$ (read as "n choose k") is # of subsets of size k from a set of size n. (this works since sets are unordered)

Thm (**Binomial Coefficient Formula**): For $k \leq n$, we have: $\binom{n}{k} = \frac{n(n-1)\ldots(n-k+1)}{k!} = \frac{n!}{(n-k)!k!}$.

For $k > n$, we have $\binom{n}{k} = 0$.

Proof. Let A be a set with $|A| = n$. Any subset of A has size at most n, so $\binom{n}{k} = 0$ for $k > n$.
So let $k \leq n$. By Sampling w/Replacement Thm, there are $n(n-1)\ldots(n-k+1)$ ways to make an ordered choice of k elements without replacement. This overcounts each subset of interest by a factor of $k!$ (since we don't care how these elements are ordered), so we get the correct count by dividing by $k!$. ■

Ex (**Club Officers**): In a club w/n people, how many ways are there to choose a president, vice president, and treasurer?

Solution: Order matters (Pres, VP, TR) = (Sally, Bob, Julie) is not the same as
(Pres, VP, TR) = (Julie, Bob, Sally), so there are $n(n-1)(n-2)$ ways. (sampling without replacement)

How many ways are there to choose three generic administrators with equivalent titles/duties?

Order doesn't matter, so there are $\binom{n}{3} = \frac{n!}{(n-3)!3!}$ ways to choose.

Activity 3

Ex (**Permutations of a Word**). How many ways are there to permute the letters in LALALAAA?

We need to place each of the 8 letters into 8 locations in the word: _ _ _ _ _ _ _ _.

Note that we just need to choose where the 5 A's go (or, equivalently, just decide where the 3 L's go).

So there are $\binom{8}{5} = \binom{8}{3} = \frac{8 \cdot 7 \cdot 6}{3!} = 56$ permutations.

How many ways are there to permute the letters in STATISTICS?

We could choose where to put the 3 S's, then the 3 T's (from the remaining positions), then the 2 I's, then the one A (and then the C is determined).

$\binom{10}{3}\binom{7}{3}\binom{4}{2}\binom{2}{1}$

Alternatively, we can start with 10! and then adjust for overcounting, dividing by 3!3!2! to account for the fact that the S's can be permuted among themselves in any way, likewise for T's and I's.
This gives $\binom{10}{3}\binom{7}{3}\binom{4}{2}\binom{2}{1} = \frac{10!}{3!3!2!} = 50{,}400$ possibilities.

Thm (**Binomial**). $(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$, for any nonnegative integer n.

[Proof in Book]

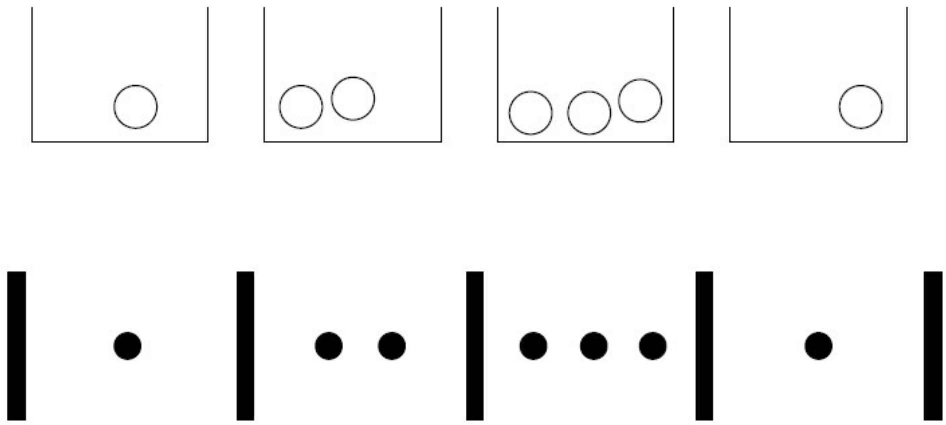
Ex (**Bose-Einstein**). How many ways are there to choose k times from a set of n objects *with replacement*, if *order doesn't matter*

(we only care about how many times each object was chosen, not the order in which they were chosen)?

With replacement, the multiplication rule gives n^k.

But order doesn't matter so, for instance: $\{1,2,2,3,3,3,4\} = \{4,1,2,3,2,3,3\}$. So we're overcounting.

Instead, visualize the n objects, not as objects, but as n buckets, and the k choices as dots (as below), then how many ways do we have of placing the dots?



Bose-Einstein with $k = 7$ and $n = 4$.

Solution: Convince yourself that the depictions above visualize the different possibilities when choosing.
Then, notice that between the outer walls, there are $n + k - 1$ objects (the | and •).
Once we place the k dots in these $n + k - 1$ locations, the locations for the "|"s are determined.
So the # of possibilities are: $\binom{n+k-1}{k}$.

In other words, from the $n + k - 1$ locations, we choose k of them to contain dots.

Harvard Video: youtube.com/watch?v=LZ5Wergp_PA&list=PL2SOU6wwxB0uwwH80KTQ6ht66KWxbzTIo&index=4

What did we learn?

- Sample Spaces
- Naive definition of probability
- How to count: multiplication rule
- Adjusting for over counting