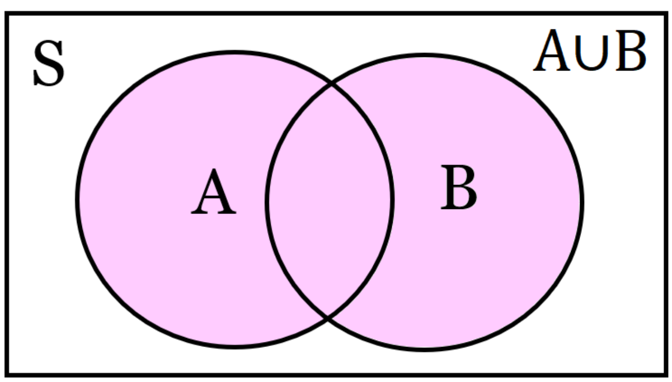
LECTURE: PROBABILITY AXIOMS

1. Basic Set Operations

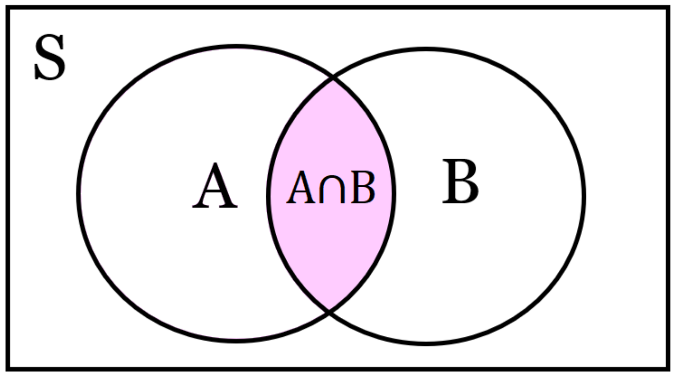
Let S be any set (sample space) and A and B be two subsets of S (events), then we can define the following:

Definition: The **union** $A \cup B$ of A and B is the set of all elements which are either in A or B (or both)

In other words, all elements that are in at least one of the sets

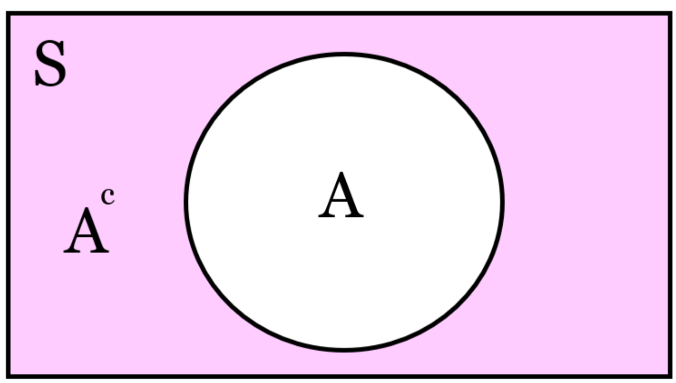


Definition: The **intersection** of $A \cap B$ of A and B is the set of all elements which are in both A and B



Definition: Two events A and B are **disjoint** or **mutually exclusive** if they have no elements in common, ie $A \cap B = \emptyset$

Definition: The **complement** A^c is the set of all points in S that are not in A



Distributive Laws:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

This is much easier understood with examples:

The first one says: If you order a Burger and a (Coke or Sprite), you either order a (Burger and Coke) or a (Burger and Sprite)

The second one says: If you take the Final or (Midterm 1 and Midterm 2), you take the (Final or Midterm 1) and (Final or Midterm 2).

De Morgan's Laws:

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

In other words, whenever you apply ^{c} you flip the $\cap$ with the $\cup$

In terms of examples, if you don't eat an (Apple or Banana), you didn't eat an Apple **and** you didn't eat a Banana. And if you don't eat both an apple and a banana, you didn't eat at least one of them

2. Basic Probability Axioms

We are now ready to define probability in a rigorous way, using axioms

Let S be our sample space and A an event in S

Definition: The Probability P is a function with the following property:

(1) For any A, $0 \leq P(A) \leq 1$

(2) $P(\emptyset) = 0$

(3) $P(S) = 1$

(4) If $A \subseteq B$ then $P(A) \leq P(B)$

(5) If $A_1, A_2, A_3, \cdots$ are pairwise disjoint (see below) then

$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots$$

Remarks:

(1) says that the probability is always between 0 and 1; 0 means that the event will never occur and 1 means the event will always occur.

(2) says that the probability that nothing happens is 0, i.e. something must happen

(4) says that if we make a set bigger, its probability can only increase (or stay the same)

(5) needs more explanation: It's a generalization of the fact that if A and B are disjoint then $P(A \cup B) = P(A) + P(B)$

Definition: Sets $A_1, A_2, A_3, \cdots$ are **pairwise disjoint** if $A_i \cap A_j = \emptyset$ for $i \neq j$

For example, for three sets A, B, C this would mean that $A \cap B = \emptyset, A \cap C = \emptyset$, and $B \cap C = \emptyset$. This is stronger than saying $A \cap B \cap C = \emptyset$ (see HW)

And (5) says that the probability of the union of disjoint sets is the sum of the probabilities.

Consequence:

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

Recall: An set is **simple** if it just has one element

Example: Consider once again tossing a single die. The sample space for this is $S = \{1, 2, 3, 4, 5, 6\}$. This sample space contains 6 simple events: $\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}$. Assuming we have a fair die, we can let $P\{n\} = \frac{1}{6}$ for $n = 1, 2, 3, 4, 5, 6$

But we don't have to do it that way. If we have a crooked die, which rolls a 6 half the time, we could assign probabilities: $P\{6\} = \frac{1}{2}$ and $P\{n\} = \frac{1}{10}$ for $n = 1, 2, \cdots, 5$

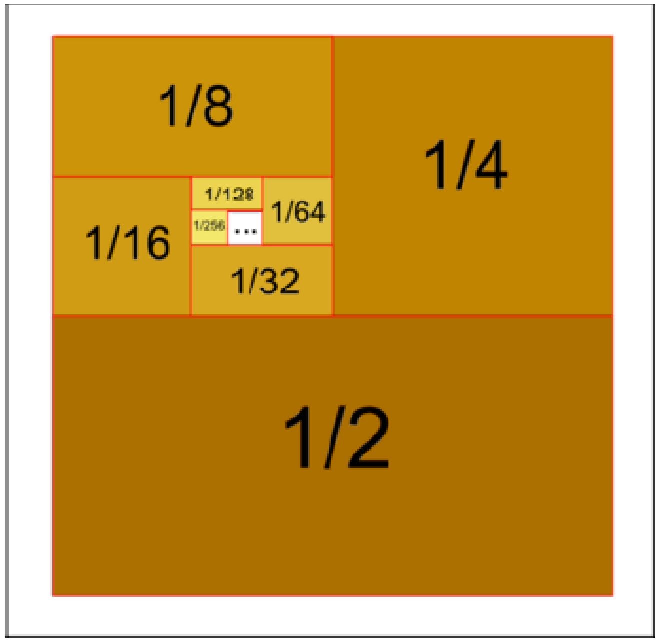
Example: Consider this time a countable sample space $S = \{1, 2, 3, \cdots\}$ One possibil- ity is to assign probabilities $P\{n\} = \frac{1}{2^n}$ for $n = 1, 2, 3, \cdots$

This works because

$$\sum_{n=1}^{\infty} P\{n\} = \sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 1$$

The last part follows because this is a geometric series with ratio $\frac{1}{2}$.

Here is a picture proof of this fact:



3. The Uniform Probability Distribution

The first probability distribution we will consider is the uniform distribution. In the discrete uniform distribution, every simple event is equally likely to occur.

Setting: Suppose we have a finite sample space S with n simple events. The discrete uniform distribution assigns each such event a probability of $\frac{1}{n}$

Why? Suppose our sample space S consists of n (disjoint) simple events $A_1, \cdots, A_n$, each of probability p, then by our axioms, we have:

$$\begin{aligned}
P(S) =& 1 \\
P(A_1 \cup \cdots A_n) =& 1 \\
P(A_1) + \cdots + P(A_n) =& 1 \\
\underbrace{p + p + \cdots + p}_{n \text{ times}} =& 1 \\
np =& 1 \\
p =& \frac{1}{n}
\end{aligned}$$

In general, if A is not simple, then you have

$$P(A) = \frac{\text{number of simple events in } A}{\text{number of simple events in } S}$$

Example: Suppose you're rolling 2 dice. What is the probability that the sum of the two dice is 7 ?

This sample space has 36 simple events, so each simple event has a probability of $\frac{1}{36}$

There are 6 simple events that give us a sum of 7:

$$(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$$

Thus the probability of a sum of 7 is $\frac{6}{36} = \frac{1}{6}$

Example: What is the probability that the sum is < 11?

Let A be the event "The sum is < 11"

In this case it's easier to calculate $P(A^c)$ which is the probability that the sum is ≥ 11 and then use $P(A^c) = 1 - P(A)$

A^c is composed of 3 simple events: $(5, 6), (6, 5), (6, 6)$

Thus $P(A^c) = \frac{3}{36} = \frac{1}{12}$

Thus we have $P(A) = 1 - P(A^c) = 1 - \frac{1}{12} = \frac{11}{12}$

In the previous example, it is relatively straightforward to draw the sample space, so we can essentially compute any probability we want simply by listing out all the elements

For more complicated problems, this is not as easy.

Example: A communication system consists of $n = 4$ antennas arranged in a line. Exactly $m = 2$ out of the n antennas are defective. The system is functional if no two consecutive antennas are defective. Assuming that each linear arrangement of the antennas is equally likely, what is the probability that the system will be functional?

For small values of n and m, we can write out all of the possible configurations. Representing a functional antenna by 1 and a defective antenna by 0, there are exactly six linear arrangements:

(1) 0011

(2) 0101 (functional)

(3) 0110 (functional)

(4) 1001

(5) 1010 (functional)

(6) 1100

In this case, the probability that the system is functional is $\frac{3}{6} = \frac{1}{2}$

For general n and m, it is not immediately obvious how to perform the requisite counting of configurations. Taking a cue from the Count on Sesame Street, we need to learn more about counting. The mathematical theory of counting is known as combinatorics.