STAT 131 - Intro to Probability Theory

Lecture 1: Probability and Counting

Suzana Șerboi, Ph.D.
Mathematics and Statistics, UCSC

Teaching Team

Instructor:

- Suzana Șerboi (smilea@ucsc.edu)
- Professor Șerboi or Dr. Șerboi, or simply Suzana.

Teaching Assistant:

- Peter Trubey (ptrubey@ucsc.edu)

LSS Tutor:

- Debi Majumdar (demajumd@ucsc.edu)

Materials

Textbook:

- Joe Blitzstein and Jessica Hwang (2014). Introduction to Probability. Second Edition. Chapman & Hall.
- A free online version of the book is available at http://probabilitybook.net

Additional Reference:

- M.H. DeGroot and M.J. Schervish (2002). Probability and Statistics. Fourth Edition. Addison Wesley.

Technology:

- R and Rstudio installed on your computer recommended:

https://rstudio-education.github.io/hopr/starting.html

- You can also write R code at: https://www.mycompiler.io/new/r

Tips to succeed in this class

- Read the syllabus and understand the evaluation of the class.
- Read the suggested chapters from the textbook every week.
- Attend class, and be an active participant during lectures and discussion sessions.
- Start working on the homework problems early (don't wait until the due date!); this will make studying for the quizzes and exams a lot less stressful.
- Find a study group and commit to it; this will make the work easier. Learning is always better with a learning community.
- Work on your formula sheet every week. Include theorems, definitions, formulas, and properties discussed in each class.
- Don't leave questions about the material unanswered; ask your questions on our discussion forum Ed Discussion or in office hours.

Learning Outcomes

After completing this course, successful students will be able to:

- Be familiar with the basic approaches to the definition of probability.
- Understand basic theory to construct probability models for both discrete and continuous random variables.
- Be able to use distribution functions.
- Be able to apply the meaning and the applications of joint probability and joint distribution functions.
- Be able to apply the concepts and expectations with respect to a given probability function.
- Understand the meaning and be able to apply the concept of conditional and marginal probability functions.
- Understand and be able to apply the Central Limit Theorem, the Law of Large Numbers, and the concept of Markov Chains.

Why study Probability?

Mathematics is the logic of certainty; probability is the logic of uncertainty.

- The field of probability originated from analyzing gambling and chance-based games.
- It underwent centuries of development before achieving a fully rigorous mathematical foundation.
- Nowadays, probability is used in fields such as medicine, meteorology, photography from satellites, marketing, earthquake prediction, human behavior, the design of computer systems, finance, genetics, law, etc.
- Mastering probability is key to navigating our uncertain world.

Experiment, sample space and events

An **experiment** is any process, real or hypothetical, in which the possible outcomes can be identified ahead of time.

For example, flipping two coins is an experiment.

The **sample space** is the collection of all the possible outcomes of an experiment. The sample space can be finite or infinite.

For example, when flipping two coins, write Heads as H and Tails as T. Then the corresponding sample space is $S = \{HH, HT, TH, TT\}$.

Any subset E of the sample space is known as an **event**. In other words, an event is a set consisting of possible outcomes of the experiment.

Let E be the even that the two coins come up different. Then $E = \{HT, TH\}$.

Naive Definition of Probability

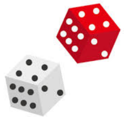
If S is a finite nonempty sample space of equally likely outcomes, and A is an event, that is, a subset of S, then the probability of A is

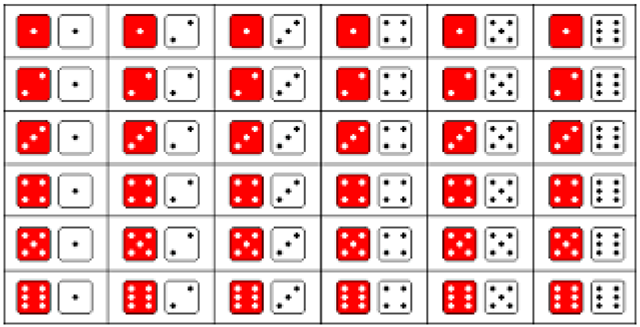
$$p(A) = \frac{|A|}{|S|} = \frac{\text{number of outcomes favorable to A}}{\text{total number of outcomes in S}},$$

where $|A|$ is the size (cardinality) of set A.

- This definition assumes that experiments have finitely many, "equally likely" outcomes.
- It works well for simple scenarios like fair coin tosses or dice rolls.
- However, it becomes problematic when applied to more complex situations. For instance, it's not logical to claim that the probability of extraterrestrial life on the moon is 1/2 simply because there are two possible outcomes (existence or non-existence).

Example (Rolling Two Dice)

Experiment: Rolling two dice. 



Sample space:

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

An event: Let E denote the event that the sum of the rolls is 4. Then
$E = \{(1,3),(2,2),(3,1)\}$ and $p(E) = \frac{|E|}{|S|} = \frac{3}{36} = \frac{1}{12}$.

Another event: Let F be the event that the two rolls are the same. Then
$F = \{(i,i) : i = 1 \ldots 6\}$ and $p(F) = \frac{|F|}{|S|} = \frac{6}{36} = \frac{1}{6}$.

Set Theory Notation

The sample space of an experiment can be thought of as a set, or collection, of different possible outcomes; and each outcome can be thought of as an element in the sample space. Similarly, events can be thought of as subsets of the sample space.

English	Sets
Events and occurrences	
sample space	S
s is a possible outcome	$s \in S$
A is an event	$A \subseteq S$
New events from old events	
A or B (inclusive)	$A \cup B$
A and B	$A \cap B$
not A	A^c (the complement of A)
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	$A_1 \cup \cdots \cup A_n = S, A_i \cap A_j = \emptyset \text{ for } i \neq j$

De Morgan's laws

Saying that it is not the case that at least one of A and B occur is the same as saying that A does not occur and B does not occur:

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

Saying that it is not the case that both occur is the same as saying that at least one does not occur:

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

Analogous results hold for unions and intersections of more than two events. Let A_i's be events *(e.g., Person i has an umbrella)*. Then

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

Not true that someone has an umbrella. = Everyone doesn't have an umbrella.

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

Not true that everyone has an umbrella. = Someone doesn't have an umbrella.

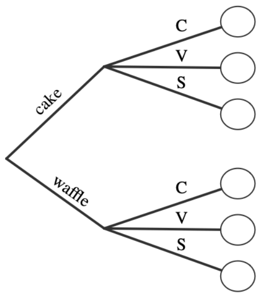
Counting

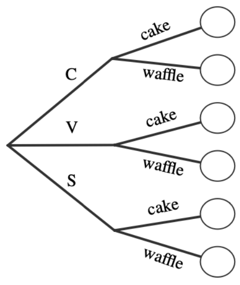
Many basic probability problems involve counting.

Multiplication Principle Consider a compound experiment consisting of two sub-experiments: experiment A and experiment B. Suppose that experiment A has a possible outcomes, and for each of those outcomes experiment B has b possible outcomes. Then the compound experiment has ab possible outcomes.

Example (Ice cream cones) Suppose you are buying an ice cream cone. You can choose whether to have a cake cone or a waffle cone, and whether to have chocolate, vanilla, or strawberry as your flavor.

This decision process can be visualized with a tree diagram. Regardless of whether the type of cone or the flavor is chosen first, there are $2 \cdot 3 = 3 \cdot 2 = 6$ possibilities.





Permutations and factorials Suppose that n positions are to be filled with n different objects. There are n choices for filling the first position, $n-1$ for the second, ..., and 1 choice for the last position. So, by the multiplication principle, there are

$$n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1 = n!$$

possible arrangements. The symbol $n!$ is read **"n factorial"**. We define $0! = 1$; that is, we say that zero positions can be filled with zero objects in one way.

Each of the $n!$ arrangements (in a row) of n different objects is called a **permutation** of the n objects.

For example, $3, 5, 1, 2, 4$ is a permutation of $1, 2, 3, 4, 5$.

Example The number of permutations of the four letters a, b, c, and d is

$$4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24.$$

However, the number of possible four-letter code words using the four letters a, b, c, and d if letters may be repeated is $4^4 = 256$, because in this case each selection can be performed in four ways.

Sampling with replacement Consider n objects and making k choices from them, one at a time with replacement (i.e., choosing a certain object does not preclude it from being chosen again). Then there are n^k possible outcomes (where order matters, in the sense that, e.g., choosing object 3 and then object 7 is counted as a different outcome than choosing object 7 and then object 3.)

For example, imagine a jar with n balls, labeled from 1 to n. We sample balls one at a time with replacement, meaning that each time a ball is chosen, it is returned to the jar. Each sampled ball is a sub-experiment with n possible outcomes, and there are k sub-experiments. Thus, by the multiplication rule there are

$$\underbrace{n \cdot n \cdot \ldots \cdot n}_{k \text{ times}} = n^k$$

ways to obtain a sample of size k.

Example A die is rolled seven times. Note that rolling a die is equivalent to sampling with replacement from the set $\{1, 2, 3, 4, 5, 6\}$. The number of possible ordered samples is $6^7 = 279,936$.

Sampling without replacement Consider n objects and making k choices from them, one at a time without replacement (i.e., choosing a certain object precludes it from being chosen again). Then there are

$$n(n-1)\dots(n-k+1)$$

possible outcomes for $1 \leq k \leq n$, and 0 possibilities for $k > n$ (where order matters). It is equivalent to the number of permutations of n objects taken k at a time and is denoted by $P(n,k)$ or ${}_nP_k$.

In terms of factorials, we have

$$P(n,k) = \frac{n(n-1)\dots(n-k+1)(n-k)\cdot\ldots\cdot 2\cdot 1}{(n-k)\cdot\ldots\cdot 2\cdot 1} = \frac{n!}{(n-k)!}$$

Again, imagine a jar with n balls, labeled from 1 to n. We sample balls one at a time without replacement, meaning that each time a ball is chosen, it is NOT returned to the jar. The number of possible choices decreases by 1 each time. The above result follows directly from the multiplication rule.

Example The number of possible four-letter code words, selecting from the 26 letters in the alphabet, in which all four letters are different is

$$26\cdot 25\cdot 24\cdot 23 = 358,800.$$

Binomial coefficient For any nonnegative integers k and n, the binomial coefficient $\binom{n}{k}$, read as "n choose k", is the number of subsets of size k from a set of n elements. Algebraically, binomial coefficients can be computed as follows. For $k > n$, $\binom{n}{k} = 0$. For $k \leq n$, we have

$$\binom{n}{k} = \frac{n \cdot (n-1) \cdot \ldots \cdot (n-k+1)}{k!} = \frac{n!}{(n-k)!k!}$$

Sets and subsets are by definition **unordered**, e.g., $\{3, 1, 4\} = \{4, 1, 3\}$, so we are counting the number of ways to choose k objects out of n, without replacement and without distinguishing between the different orders.

There are $n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$ 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 can get the correct count by dividing by $k!$.

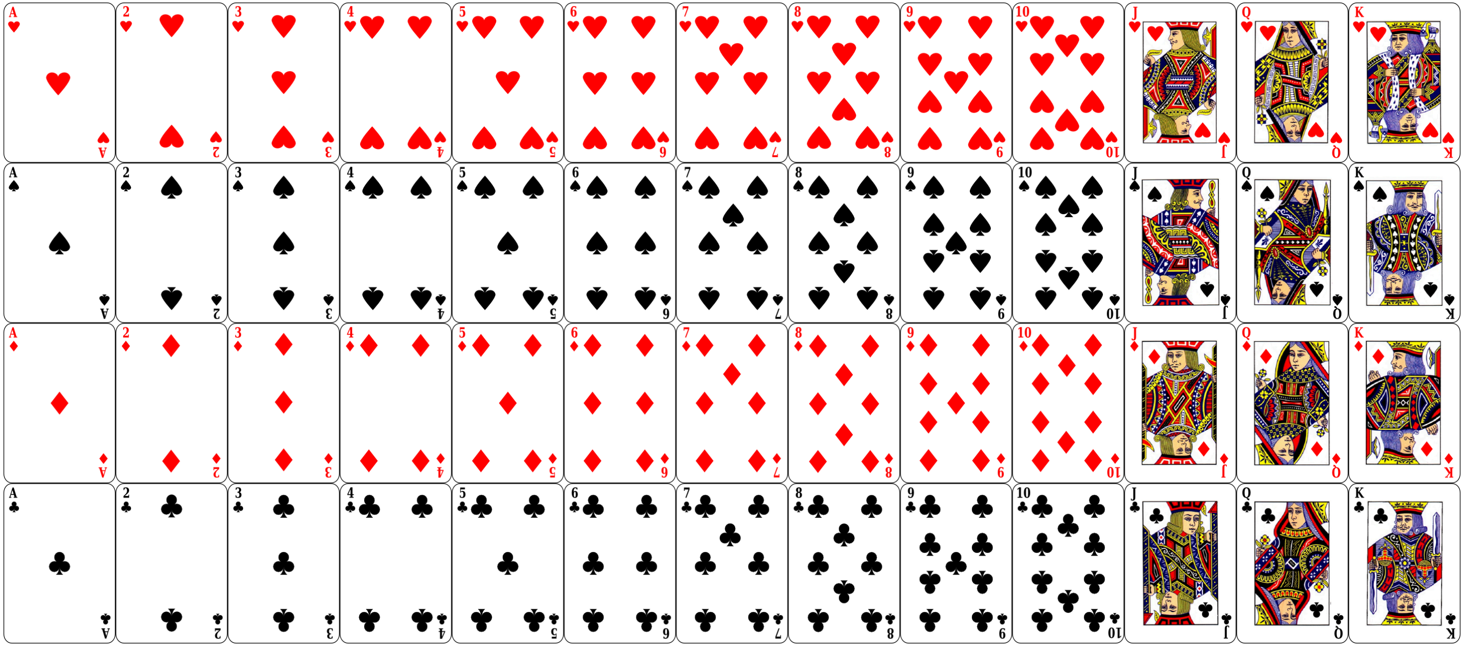
Example The number of possible five-card hands (in five-card poker) drawn from a deck of 52 playing cards is $\binom{52}{5}$.

```
choose(52,5)
```

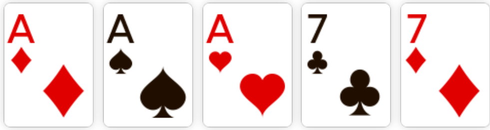
```
## [1] 2598960
```

Deck of Cards

A "standard" deck of playing cards consists of 52 Cards in each of the 4 suits of Hearts, Spades, Diamonds, and Clubs. Each suit contains 13 cards: Ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, King.



Example (Full House) What is the probability that a poker hand contains a full house, that is, three of one kind and two of another kind?



Solution: Note that the order of the two kinds matters, because, for instance, three queens and two aces is different from three aces and two queens.

- There are 13 choices for what kind we have three of.
- There are $\binom{4}{3}$ ways to choose three cards out of four of a given kind.
- Then there are 12 choices for what kind we have two of, and
- $\binom{4}{2}$ ways to choose two of that kind.

Thus, by the multiplication rule, the number of hands containing a full house is $13 \cdot \binom{4}{3} \cdot 12 \cdot \binom{4}{2} = 3,744.$

Because there are $\binom{52}{5} = 2,598,960$ poker hands, the probability is

$$p(\text{full house}) = \frac{3,744}{2,598,960} \approx 0.0014.$$

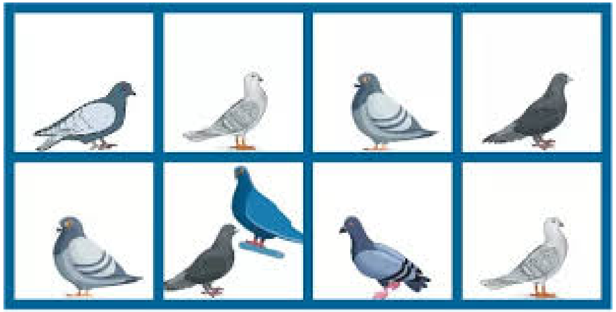
The birthday problem

The birthday problem, a classic probability puzzle, illustrates both sampling with and without replacement in its solution.

Given k people, find the probability that two have the same birthday.

Assumptions: exclude February 29th, assume the other 365 days are equally likely and assume birthdays are independent (no twins).

Pigeonehole Principle If k objects are placed into n boxes, where $k > n$, then there must be at least one box that contains more than one object.



In the birthday problem (assuming there are 365 days in a year), with 366 or more people there is guaranteed to be at least one birthday match. So, when $k > 365$, the probability is 1.

The birthday problem (cont.)

Let $k \leq 365$. A good strategy when trying to find the probability of an event is to start by thinking about whether it will be easier to find the probability of the event or the probability of its complement. In this case, it is easier to find the probability of the complement, i.e. no two people have the same birthday. Then the probability of having a birthday match will be

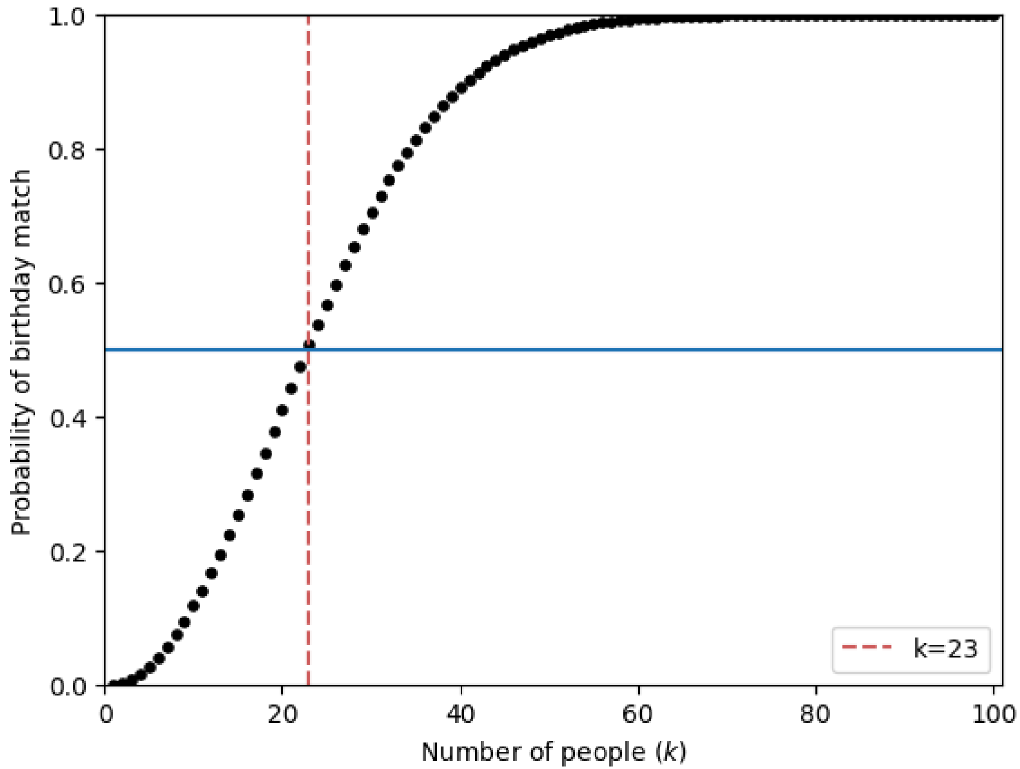
$$1 - P(\text{``no match''}) = 1 - \frac{365 \cdot 364 \cdot \ldots \cdot [365 - (k-1)]}{365^k}$$

The first value of k for which the probability exceeds 0.5 or 50% is $k = 23$.In R, the following code uses `prod` (which gives the product of a vector) to calculate the probability of at least one birthday match in a group of 23 people:

```
k <- 23
1-prod((365-k+1):365)/365^k
```

```
## [1] 0.5072972
```

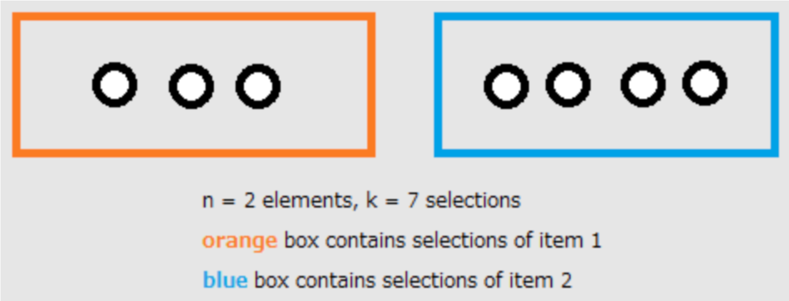
Go to https://www.mycompiler.io/new/r and replace `k <- 23` with `k <- 57`.



In the above plot, we see the first value of k for which the probability of a match exceeds 0.5 is $k = 23$. Thus, in a group of 23 people, there is a better than 50% chance that two or more of them will have the same birthday. For a quick intuition into why it should not be so surprising, note that with 23 people there are $\binom{23}{2} = 253$ pairs of people, any of which could be a birthday match.

Sampling with replacement, unordered

Although not needed as often in the study of probability, it is interesting to count the number of possible samples of size k that can be selected out of n objects when **the order is irrelevant** and when sampling **with replacement**.



To count the number of possible outcomes, notice that what we are really doing here is placing $n-1=2-1=1$ dividers between k elements.

Or in other words, we are choosing k slots for the elements out of $(n-1)+k$ slots in total.

The number of ways to select k items out of n, unordered and with replacement, is: $\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$.

*This counting method is sometimes called the stars and bars argument, but here we used circles in place of stars. See some examples at https://discrete.openmathbooks.org/dmoi2/sec_stars-and-bars.html

General Definition of Probability

A probability space consists of a sample space S and a probability function p which takes an event $A \subseteq S$ as input and returns $p(A)$, a real number between 0 and 1, as output. The function p must satisfy the following axioms:

Axiom 1. $p(\emptyset) = 0, p(S) = 1.$

Axiom 2. If $A_1, A_2, \ldots$ are disjoint events, then

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

(Saying that these events are disjoint means that they are mutually exclusive: $A_i \cap A_j = \emptyset$ for $i \neq j$.)

With this definition, we can model experiments in which outcomes are either equally likely or not equally likely by choosing the appropriate probability function p.

Example For a fair coin, the probability that heads comes up when the coin is flipped equals the probability that tails comes up, so the outcomes are equally likely. Consequently, we assign the probability 1/2 to each of the two possible outcomes. That is, $p(H) = p(T) = 1 / 2$, where H is the event that heads comes up and T is the event that tails comes up.
What probabilities should be assigned to these outcomes when the coin is biased so that heads comes up twice as often as tails?

Solution: For the biased coin we have $p(H) = 2p(T)$.
Because $p(S) = p(H \cup T) = p(H) + p(T) = 1$, it follows that $2p(T) + p(T) = 3p(T) = 1$.
We conclude that $p(T) = \frac{1}{3}$ and $p(H) = \frac{2}{3}$.

Example Suppose that a die is biased (or loaded) so that 3 appears twice as often as each other number but that the other five outcomes are equally likely. What is the probability that an odd number appears when we roll this die?

Solution: We want to find the probability of the event $E = \{1, 3, 5\}$.
We have $p(1) = p(2) = p(4) = p(5) = p(6) = \frac{1}{7}$ and $p(3) = \frac{2}{7}$.
It follows that $p(E) = p(1) + p(3) + p(5) = \frac{1}{7} + \frac{2}{7} + \frac{1}{7} = \frac{4}{7}$.

Frequentist vs Bayesian View of Probability

- The **frequentist** view of probability is that it represents a long-run frequency over a large number of repetitions of an experiment: if we say a coin has probability of Heads, that means the coin would land Heads 50% of the time if we tossed it over and over and over.
- The **Bayesian** view of probability is that it represents a degree of belief about the event in question, so we can assign probabilities to hypotheses like "candidate A will win the election" or "the defendant is guilty" even if it isn't possible to repeat the same election or the same crime over and over again.

The Bayesian and frequentist perspectives are complementary, and both will be helpful for developing intuition.

Regardless of how we choose to interpret probability, we can use the two axioms to derive other properties of probability, and these results will hold for any valid probability function.

Properties of probability

A **probability function** p has the following properties, for events A and B:

1. $p(A^c) = 1 - p(A)$.
2. If $A \subseteq B$, then $p(A) \leq p(B)$.
3. $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.

Proof of property 1: $p(A^c) = 1 - p(A)$

Since A and A^c are disjoint and their union is S, so the 2nd axiom gives:

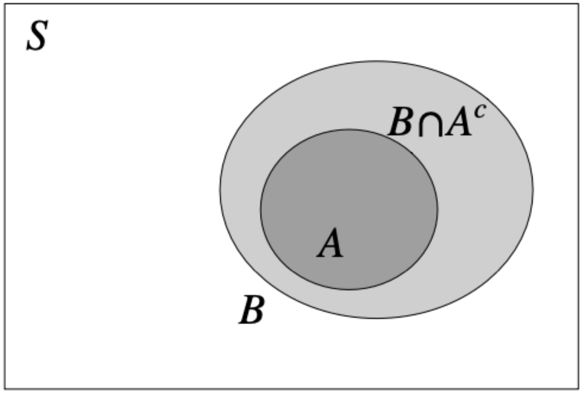
$$P(S) = P(A \cup A^c) = P(A) + P(A^c)$$

Also, $P(S) = 1$ by axiom 1.

Thus, $P(A) + P(A^c) = 1$.

Proof of property 2: If $A \subseteq B$, then $p(A) \leq p(B)$.

If $A \subseteq B$, then we can write B as the union of A and $B \cap A^c$, where this last event is the part of B not also in A.



Since A and $B \cap A^c$ are disjoint, we can apply the 2nd axiom:

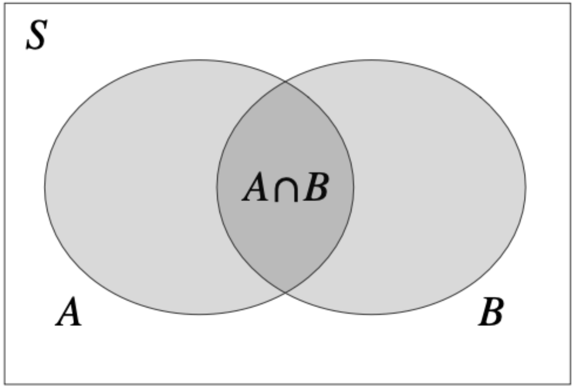
$$P(B) = P(A \cup (B \cap A^c)) = P(A) + \underbrace{P(B \cap A^c)}_{\geq 0}.$$

By definition, probability is non-negative, so $P(B \cap A^c) \geq 0$, thus

$$P(B) \geq P(A).$$

Proof of property 3: $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.

We can write $A \cup B$ as the union of two disjoint events: A and $B \cap A^c$.



Then

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

Now we need to show that: $P(B \cap A^c) = P(B) - P(A \cap B)$.

For that, we can use the 2nd axiom:

$P(A \cap B) + P(B \cap A^c) = P(B)$, as $A \cap B$ and $B \cap A^c$ are disjoint and their union is B.

Thus $P(B \cap A^c) = P(B) - P(A \cap B)$.

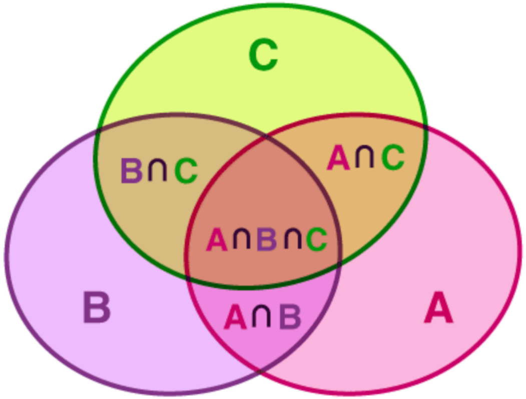
The third property is a special case of inclusion-exclusion, a formula for finding the probability of a union of events when the events are not necessarily disjoint:

Inclusion-exclusion. For any events $A_1, \ldots, A_n$, $p\left(\bigcup_{i=1}^{n} A_i\right) =$

$$\sum_i p(A_i) - \sum_{i<j} p(A_i \cap A_j) + \sum_{i<j<k} p(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} p\left(\bigcap_{i=1}^{n} A_i\right).$$

For three events, inclusion-exclusion says

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



Example A survey was taken of a group's viewing habits of sporting events on TV during the last year. Let A = {watched football}, B = {watched basketball}, and C = {watched baseball}.

The results indicate that if a person is selected at random from the surveyed group, then

$p(A) = 0.43, p(B) = 0.40, p(C) = 0.32,$

$p(A \cap B) = 0.29, p(A \cap C) = 0.22, P(B \cap C) = 0.20,$ and

$P(A \cap B \cap C) = 0.15.$

It then follows that $p(A \cup B \cup C) =$

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

$= 0.43 + 0.40 + 0.32 - 0.29 - 0.22 - 0.20 + 0.15 = 0.59$

is the probability that this person watched at least one of these sports.

Additional Practice Problems*

1. (a) How many 7-digit phone numbers are possible, assuming that the first digit can't be a 0 or a 1?

(b) How many 7-digit phone numbers are possible, assuming that the first digit can't be a 0 or a 1 and that the phone number is not allowed to start with 911?

2. (Leibniz's mistake) If we roll two fair dice, which is more likely: a sum of 11 or a sum of 12?

3. (a) How many ways are there to split 12 people into 3 teams, where one team has 2 people, and the other two teams have 5 people each?

(b) How many ways are there to split 12 people into 3 teams, where each team has 4 people?

4. (Four of a Kind Hand in Poker 

) Find the probability that a hand of five cards in poker contains four cards of one kind.

*Solutions will be discussed in lecture, if time permits.

5. A car repair can be performed either on time or late and either satisfactorily or unsatisfactorily. The probability of a repair being on time and satisfactory is 0.26. The probability of a repair being on time is 0.74. The probability of a repair being satisfactory is 0.41.
What is the probability of a repair being late and unsatisfactory?

6. In a certain city, three newspapers A, B, and C are published. Suppose that 60 percent of the families in the city subscribe to newspaper A, 40 percent of the families subscribe to newspaper B, and 30 percent subscribe to newspaper C. Suppose also that 20 percent of the families subscribe to both A and B, 10 percent subscribe to both A and C, 20 percent subscribe to both B and C, and 5 percent subscribe to all three newspapers A, B, and C. What percentage of the families in the city subscribe to at least one of the three newspapers?