

Standard deviation

- The variance and standard deviation are measures of spread that indicate how far values in the data set are from the mean, on average.
- Consider the observations $x_1, x_2, x_3, \dots, x_n$.
- The deviations $(x_i - \bar{x})$ display the spread of x_i about their mean \bar{x} .
- The sum of the deviations is **always** 0, as some of the deviations are positive and others are negative.
- Squaring the deviations makes them all positive. Observations far from the mean will have large positive squared deviations.
- The **variance** is the 'average' squared deviation.

Standard deviation

- If we have n observations $x_1, x_2, x_3, \dots, x_n$. The **variance** is defined as

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

- The **standard deviation**, s , is the square root of the variance.
 1. s is a measure of spread about the mean and should be used when the mean is used as the measure of center.
 2. If $s=0$, then all the values in the data set are exactly the same (no spread). **Why?**
 3. The more spread out the data, the greater the standard deviation.
 4. s is always positive.
 5. s has the same unit of measurement as the original data

A short cut formula for s^2

Theorem

An alternative expression for variance s^2 is

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n(\bar{x})^2 \right)$$

Proof.

Do some algebra on the numerator.

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2\bar{x} \cdot x_i + (\bar{x})^2) \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + \sum_{i=1}^n (\bar{x})^2 \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \cdot n\bar{x} + n(\bar{x})^2 \\ &= \sum_{i=1}^n x_i^2 - n(\bar{x})^2 \end{aligned}$$



Probability

What is randomness?

- The world is full of random events that we seek to understand.
- An event is **random** if we know what outcomes could occur, but not the particular values that will happen.
- The outcome of these events is uncertain, but they follow a regular pattern.
- Deterministic models vs. Random models.
- **Probability theory** is the mathematical representation of random phenomena.

Notation

- An **experiment** is any action or process whose outcome is subject to uncertainty. e.g. tossing a coin once or several times; selecting a card or cards from a deck; weighing a loaf of bread; etc.
- The **sample space** of an experiment, denoted by S , is the set of all possible outcomes of that experiment.

Ex. Flip a coin. Two possible outcomes: Heads (H) or Tails (T). $S=\{H,T\}$.

Ex. Battery life. $S=\{x: 0 \leq x < \infty\}$.

Notation

- An **event** is any collection of possible outcomes, that is, any subset of S (including S itself). An event is **simple** if it consists of exactly one outcome and **compound** if it consists of more than one outcome.
- If the outcome of a random phenomenon is contained in an event A , then we say that **A has occurred**.

Ex. Flip a coin twice. Four possible outcomes, $S=\{HH, HT, TH, TT\}$. Let A be the event that we obtain at least one H in the two flips. $A=\{HH, HT, TH\}$. Let B be the event that we obtain two H's in the two flips. $B=\{HH\}$.

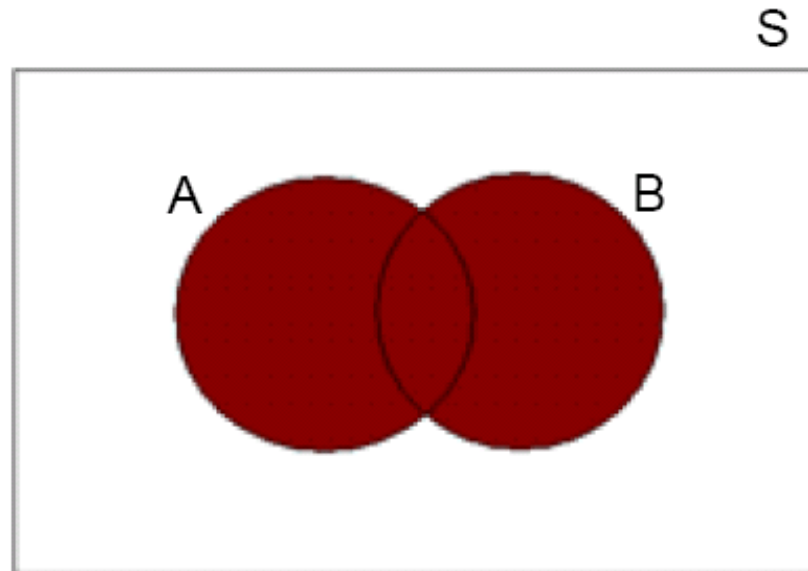
Ex. Battery life example. The event that the battery lasts less than 3 hours is denoted as $A=\{x: 0 \leq x < 3\}$.

Set Operations

- Given any two events (or sets) A and B , we have the following elementary set operations:
 - The union
 - The intersection
 - The complement
- Venn diagrams are often used to illustrate relationships between sets.

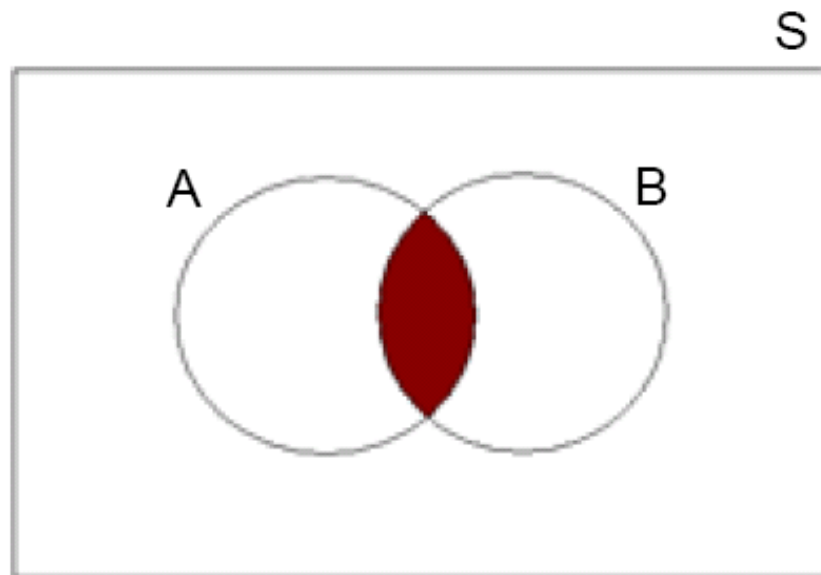
Union

- The **union** of A and B, written as $A \cup B$ and read “A or B”, is the set of outcomes that belong to either A or B or both.



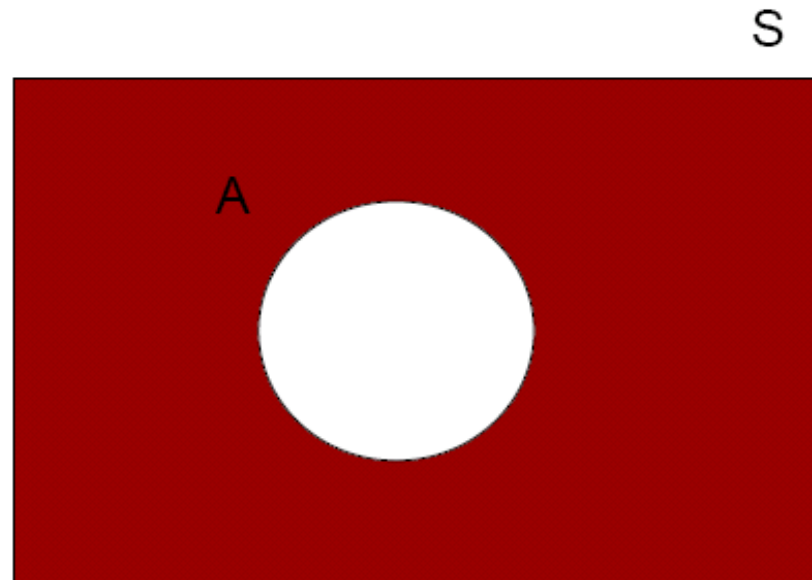
Intersection

- The **intersection** of A and B, written as $A \cap B$, read “A and B”, is the set of outcomes that belong to both A and B.



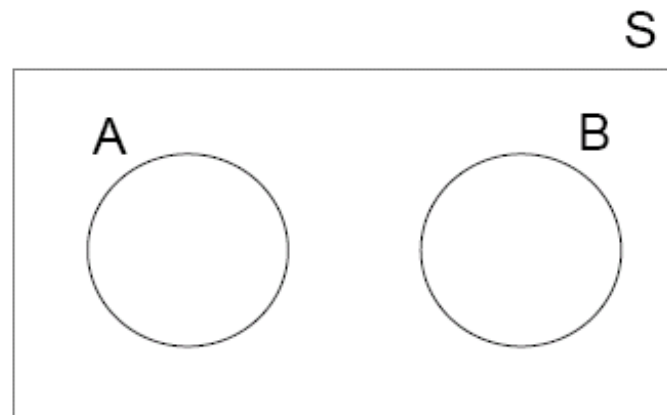
Complement

- The complement of A , written as A' or A^c , is the set of all outcomes in S that are not in A .



Disjoint events

- If $A \cap B = \emptyset$ then A and B are said to be **mutually exclusive** or **disjoint** events.



- Any event and its complement are disjoint!

Probability models

- A **probability model** consists of a **sample space** and the **assignment of probabilities** to each possible outcome.
- Probability that event A occurs is written as $P(A)$, which will give a precise **measure** of the chance that A will occur.
- To ensure the probability assignments will be consistent with our intuitive notions of probability, all assignments should satisfy the following axioms (basic properties) of probability.
 1. For any event A, $P(A) \geq 0$.
 2. $P(S) = 1$.
 3. If A_1, A_2, A_3, \dots is an infinite (finite) collection of disjoint events, then
$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = \sum P(A_i)$$

Propositions

- ▶ For any event A , $0 \leq P(A) \leq 1$.
- ▶ $P(A) + P(A^c) = 1$.
- ▶ If event A is contained in event B , in the sense that every outcome in A is also in B , then

$$P(A) \leq P(B)$$

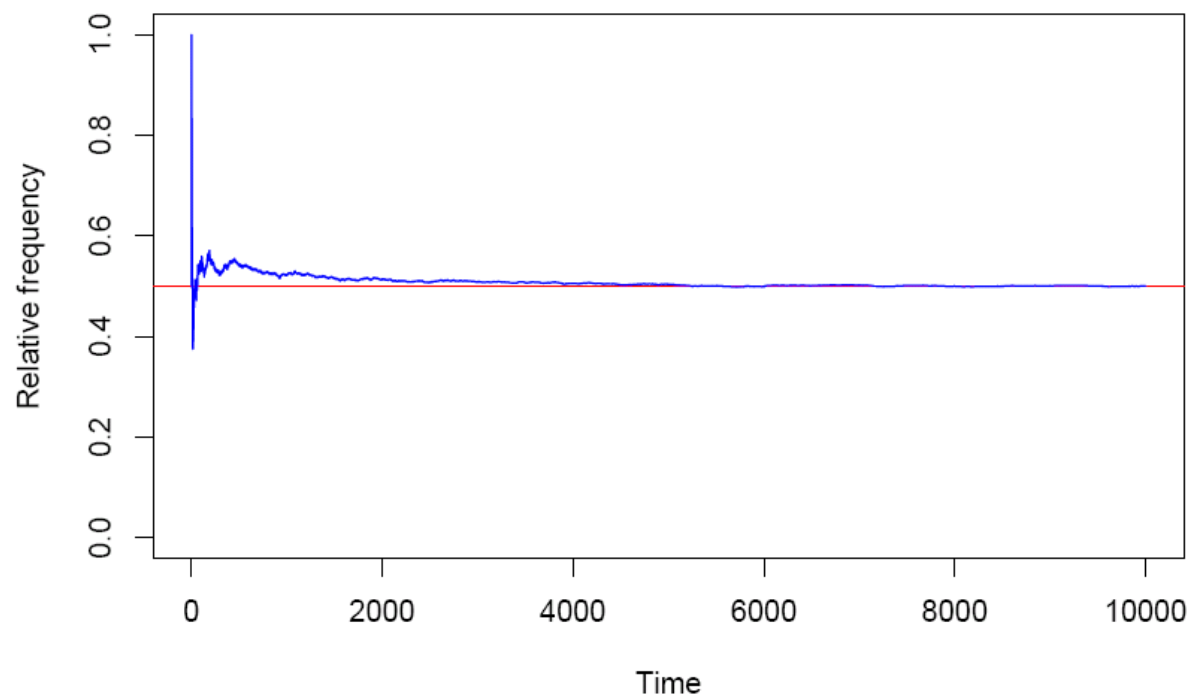
- ▶ $P(\emptyset) = 0$.

Interpreting Probability

- What does it mean when we say we have 50% chance of having a head when flipping a coin? Or what does it mean when we put $P(H)=0.5$?
- Probability is often treated as the *long-term relative frequency* or the *limiting relative frequency*.

Interpreting Probability

Ex. Flip a fair coin n times and calculate the proportion of heads.



- R demo. (Function: `sample(x, size); rbinom(x, size, prob)`)

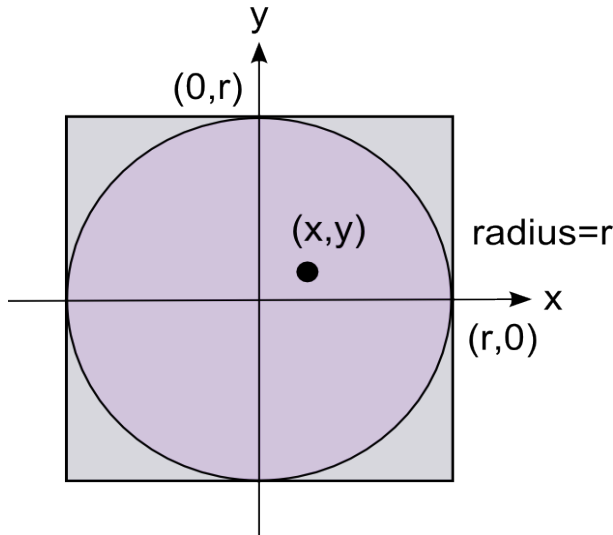
Law of Large Numbers

- The **law of large numbers** says that the long-run relative frequency of repeated independent events gets closer and closer to the true relative frequency as the number of trials increases.

$$\frac{\text{\# of occurrence of event A}}{\text{\# of trials } (n)} \xrightarrow[n \geq \infty]{} P(A)$$

How to calculate Pi

- ▶ An interesting application of Law of Large Numbers is to calculate Pi through simulations.
- ▶ If we spread a large quantity of seeds randomly but evenly on this square, what percentage of the seeds will lie inside the circle?



- ▶ R Demo.
- ▶ This type of simulation-based methods has a fancy name: Monte Carlo methods.

Assigning Probabilities

- The assignment of probabilities can often be derived from the physical set-up of an experiment.
- Suppose we have N outcomes in our sample space, each **equally likely to occur**. Each has a probability of $1/N$, and the probability of any event A is,

$$P(A) = \frac{\text{number of outcomes in } A}{N}$$

Ex. Roll a fair die. $S=\{1,2,3,4,5,6\}$. Our sample space consists of 6 points, each of which is equally likely to occur.

$P(\text{roll a } 1) = 1/6$.

Let $A = \text{roll a 4 or less} = \{1,2,3,4\}$. $P(A) = 4/6$.

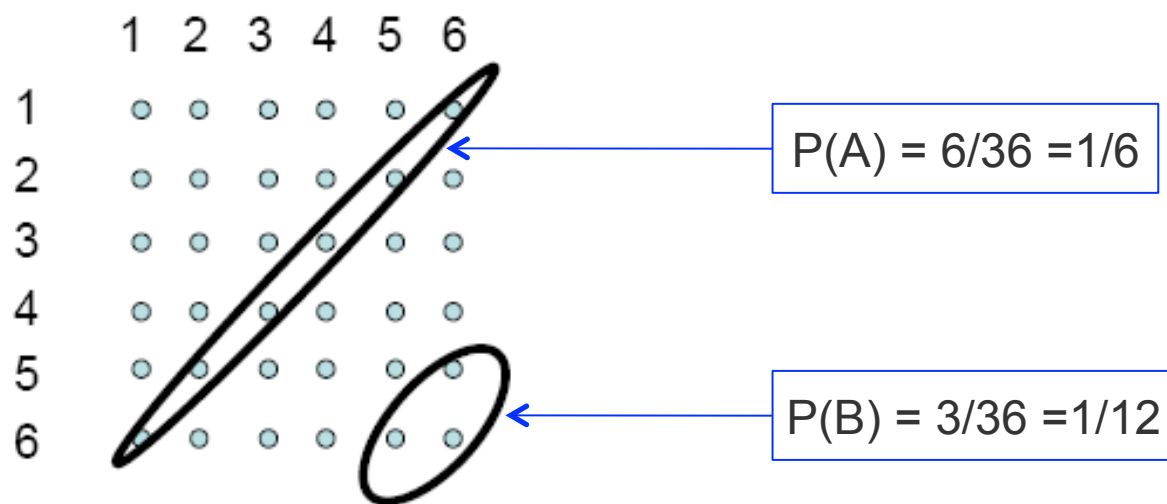
Let $B = \text{roll an even number} = \{2,4,6\}$. $P(B) = 3/6$.

Example

Ex. Roll two fair dice.

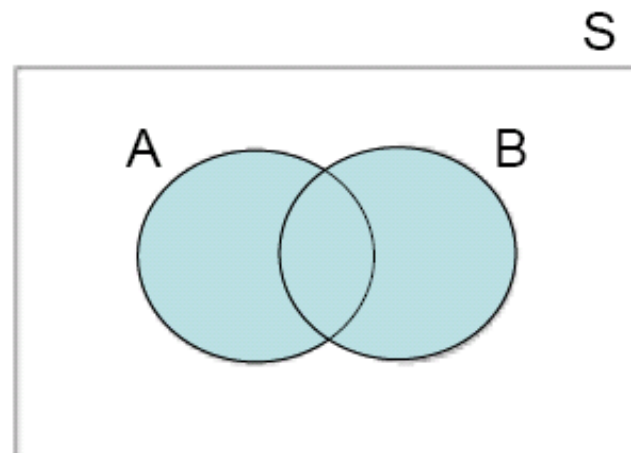
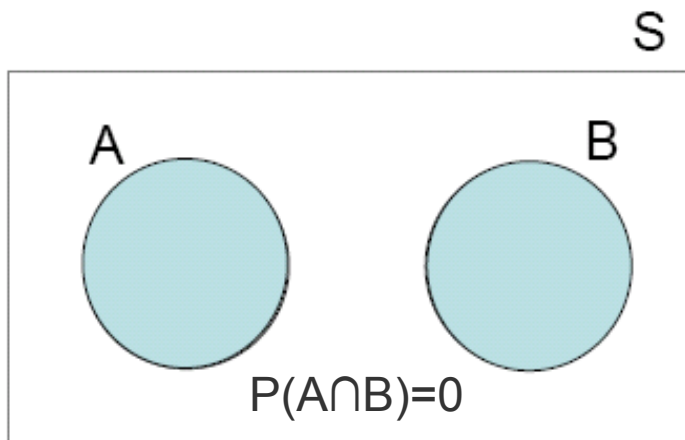
There are 36 possible outcomes: $\{(1,1),(1,2),(1,3),\dots,(6,5),(6,6)\}$.

Let A = sum of two rolls is 7; B = sum of two rolls is 11 or more. What are $P(A)$ and $P(B)$?



More Probability Properties

- Consider an experiment whose sample space is S . For each event A (B) in S , we assume that a number $P(A)$ is defined and satisfies the following rules:
 - $0 \leq P(A) \leq 1$.
 - $P(S)=1$.
 - $P(A^c)=1-P(A)$.
 - If A and B are disjoint, then $P(A \cup B)=P(A)+P(B)$.
 - For any two events A and B , $P(A \cup B)=P(A)+P(B)-P(A \cap B)$.



Example

Ex. A store accepts either VISA or Mastercard. 50% of the stores customers have VISA, 30% have Mastercard and 10% have both. What is the probability that a customer has a credit card the store accepts?

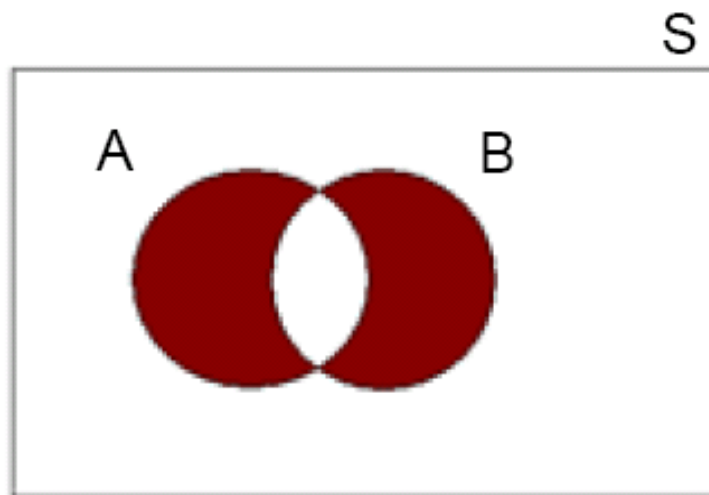
A = customers has VISA

B = customers has Mastercard

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.5 + 0.3 - 0.1 = 0.7 \end{aligned}$$

Example cont.

What is the probability that a customer has either a VISA or MC, but not both?



$$\begin{aligned} P(\text{A or B but not both}) &= P(A) + P(B) - 2P(A \cap B) \\ &= 0.5 + 0.3 - 0.2 = 0.6 \end{aligned}$$

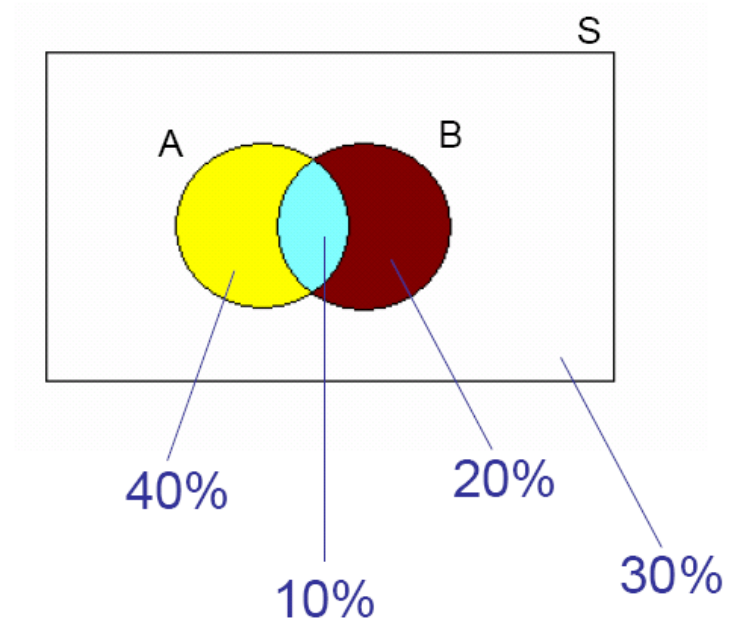
Example Cont.

What is the probability that a customer has a VISA but no MC?

$$\begin{aligned}P(\text{A but not both}) &= P(A) - P(A \cap B) \\ &= 0.5 - 0.1 = 0.4\end{aligned}$$

What is the probability that a customer has a MC but no VISA?

$$\begin{aligned}P(\text{B but not both}) &= P(B) - P(A \cap B) \\ &= 0.3 - 0.1 = 0.2\end{aligned}$$



Three Events

- For any three events A, B and C,

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

