Probabilities

$\def\E{\operatorname{E}}$ $\def\Var{\operatorname{Var}}$ $\def\Cov{\operatorname{Cov}}$ $\def\dd{\mathrm{d}}$ $\def\ee{\mathrm{e}}$ $\def\Norm{\mathcal{N}}$ $\def\Uniform{\mathcal{U}}$

$$\def\E{\operatorname{E}}$$ $$\def\Var{\operatorname{Var}}$$ $$\def\Cov{\operatorname{Cov}}$$ $$\def\dd{\mathrm{d}}$$ $$\def\ee{\mathrm{e}}$$ $$\def\Norm{\mathcal{N}}$$ $$\def\Uniform{\mathcal{U}}$$

Outline

- What are probabilities?
- Notation and basic concepts
    - Sets
    - Outcomes, events
    - Probabilities
        - Addition and multiplication
        - Independence, conditional
        - Bayes' theorem
- Exercises
    - Birthday problem
    - Monty Hall problem

What are probabilities?

- **Probability as frequency of outcome of events**:
    - In this way of thinking about probabilities we look at the number of times a given event happens over a large number of tries.
    - It is difficult to define consistently however, without running into circular reasoning.

- **Probability as degree of belief**:
    - Subjective probability is associated with personal judgements about how likely something is to happen.
    - For example, 'I believe that team X will beat team Y, because teams Y's star player has an injury, while team X has been training really hard.' Such statements can be made even if teams X and Y have never played each other.
    - By requiring that two people arrive at the same conclusion if given the same assumptions and data, this definition of probability can be formalised into a mathematical system equivalent to the other definitions.

- **Probably derived from axioms**:
    - Probability is a measure that satisfies a set of axioms derived from logic and set theory, such as the Kolmogorov or Cox axioms.
    - This sidesteps the frequentist vs Bayesian interpretation by sticking to purely mathematical concepts.

In this course we start out with this definition. In general we will follow the Bayesian degree-of-believe way of thinking about probability.

Notation and basic concepts

Set notation

- A set is a collection of elements, e.g.: $A = \{1, 2, 3\}$

- $e \in A$ means e is a member of the set A, e.g.: $1 \in \{1, 2, 3\}$

- A set can also be represented by a rule: $\{x|x ~{\rm follows ~a ~rule} \}$

  For example, the set E of even integers: $E = \{x| x=2y,\, y \in \mathbb{Z} \}$

- Set inclusion ($\subseteq$). A is included in B (or is a subset of B) if all the elements of A are also elements of B.

  For example: $\{ 1, 2 \} \subseteq \{1, 2, 3\}$

Set operators

Let $A = \{1, 3, 5\}$; $B = \{ 2, 3, 4\}$

- **Union $\cup$** All elements of A and all elements of B

  $A \cup B = \{1, 2, 3, 4, 5\}$

- **Intersection $\cap$**: Elements that are in both A and B

  $A \cap B = \{3\} $

- **Difference $\setminus$** Elements that are in A but not in B

  $A \setminus B = \{1, 5\}$

  $B \setminus A = \{2, 4 \}$

- **Complement**:

  The complement of A in reference to Ω includes all elements in Ω that are not in A. For the die example $\Omega = \{1, 2, 3, 4, 5, 6\}$,

  $A^c = \{ 2, 4, 6 \}$ or

  $A^c = \{\omega:\omega \in \Omega ~{\rm and}~ \omega\not\in A\}$

- **Empty set $\varnothing$**

  The empty set, $\varnothing$, is the complement of the universal set:

  $\Omega^c = \varnothing$ and $\varnothing^c = \Omega$.

  This means, $A\cup \varnothing = A$ and $A\cap\varnothing = \varnothing$.

- **Power set**

  Collection of all possible sets of a given set

  $A = \{1, 3, 5\}$

  $\mathcal P(A) = \left\{\varnothing,\{ 1\},\{3 \},\{5 \},\{1, 3 \},\{1, 5 \},\{3, 5 \}, \{1, 3, 5\} \right\}$

Outcomes, events, probability

Outcomes and sample space

The outcomes ω of an experiment are elements of the set of all possible outcomes, called the sample space Ω.

Consider the experiment of tossing a (fair) coin twice:

- $\omega=\text{HH}$ ("two heads")
- $\omega \in \Omega=\{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$

Events and event space

An event F is a set of outcomes

- $F=\{\text{HH}, \text{HT}, \text{TH}\}$ ("at least one head")

Events are elements of the event space $\mathcal{F}$: the power set of the sample space (the set of all possible outcomes)

Note that Ω and $\mathcal{F}$ are not the same. The sample space contains the basic outcomes and event space contains sets of outcomes.

Probability

The probability function P assigns a probability (a number between 0 and 1) to events

- $F=\{\text{HH}, \text{HT}, \text{TH}\}$
- $\Pr(F) = \frac{3}{4}$

Kolmogorov's axioms of probability

- The probability measure of events is a real number equal or larger than 0:

  $0 \le \Pr(A)$

- The probability measure of the universal set is 1.

  $\Pr(\Omega) = 1$

- If the sets A_1, A_2, A_3 ... $\in \mathcal{F}$ are disjoint, then

  $\Pr(A_1 \cup A_2 \cup ...) = \Pr(A_1) + \Pr(A_2) + ...$

Consequences of the axioms of probability

- Numeric bound:

  $0 \leq \Pr(A) \leq 1$

- Monoticity:

  $A\subseteq B$ then $\Pr(A) \leq \Pr(B)$

- Complement rule:

  $\Pr(A^c) = 1 - \Pr(A)$

- Sum rule:

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

Example, a single fair die:

- $\Omega = \{1,2,3,4,5,6\}$
- $\Pr(\omega) = \frac{1}{6}\quad \forall \omega \in \Omega$
- Events $A = \{1, 3\}$ and $B = \{1,2,3,4\}$

- Monoticity:

  $A\subseteq B$, $\Pr(A)=\frac{1}{3} \leq \Pr(B)=\frac{2}{3}$

- Complement rule:

  $\Pr(A^c) = \Pr(\{2,4,5,6\}) = \frac{2}{3} = 1 - \Pr(A)$

- Sum rule:

  $\Pr(A \cup B) = \Pr(\{1,2,3,4\}) = \frac{2}{3}$

  $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) = \frac{1}{3} + \frac{2}{3} - \Pr(\{1, 3\}) = \frac{2}{3}$

Clicker

Two fair six-sided dice are rolled. What's the probability to obtain two sixes?

- 1/25
- 1/36
- 2/36
- 1/6

Conditional probabilities and independence

Conditional probabilities

The conditional probability of event A happening, given that event B happened, is

$\Pr(A|B) = \frac{\Pr(A\cap B)}{\Pr(B)}$

Instead of the Kolmogorov axioms, probability theory can also be defined in terms of conditional probabilities, using the Cox axioms.

Independence

If A is independent of B, $\Pr(A|B) = \Pr(A)$: the conditional probability of A given B does not depend on B. From this follows that

$\Pr(A\cap B) = \Pr(A)\Pr(B)$

Law of total probability

Let $\{H_1, H_2, ... \}$ be a countable collection of sets which is a partition of Ω, where

$H_i \cap H_j = \varnothing$ for $i \ne j$

$H_1 \cup H_2 \cup ... = \Omega$

The probability of an event D can be calculated as

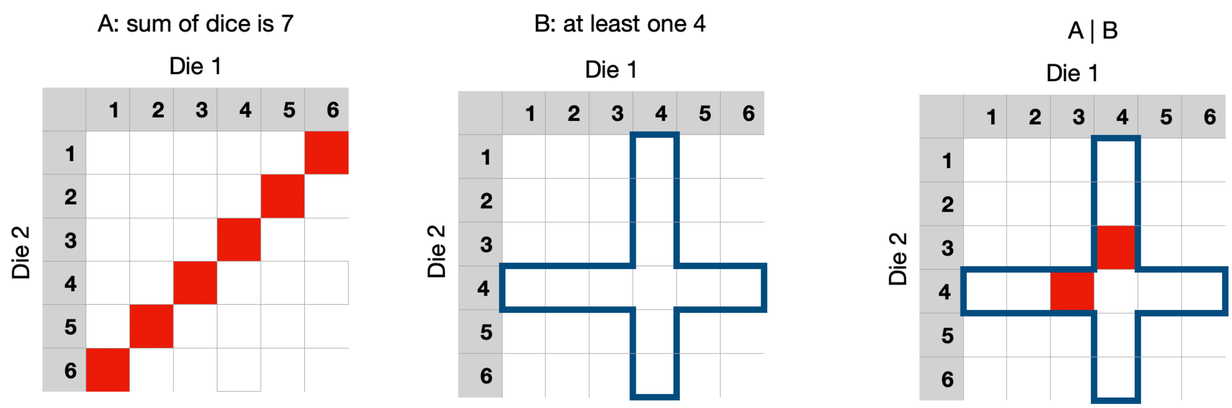
$\Pr(D) = \Pr(D \cap H_1) + \Pr(D \cap H_2) + \dots$

or in terms of conditional probabilities

$\Pr(D) = \Pr(D | H_1)\Pr(H_1) + \Pr(D | H_2)\Pr(H_2) + \dots$

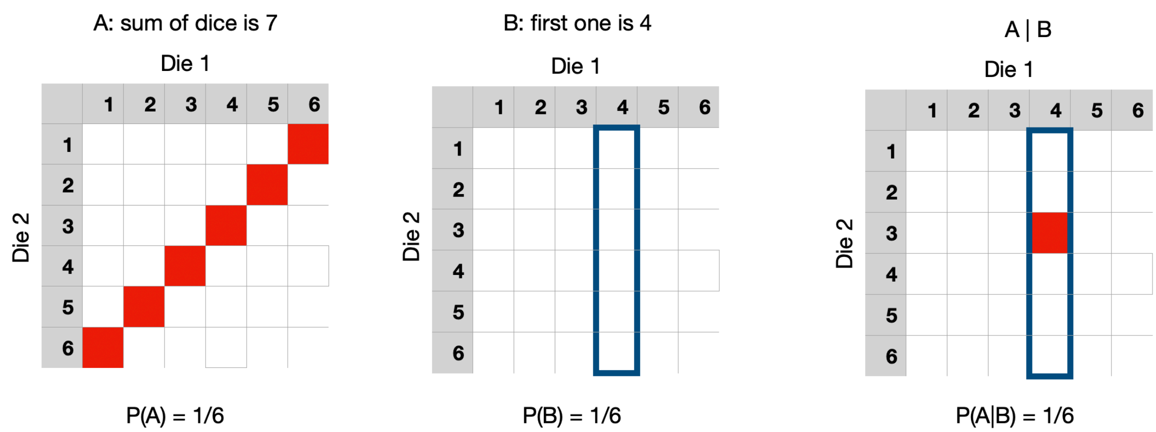
Clicker

Two fair six-sided dice are rolled. Let A be the event that the sum of the dice is 7, and let B be the event that at least one of the two dice is a 4. Are A and B independent?



P(A) = 1/6 P(B) = 11/36 P(A|B) = 2/11

What about when B is the event that the first die is a 4?



Bayes' theorem

Applying the definition of the conditional probability twice we get Bayes' theorem:

$\Pr(A|B) = \frac{\Pr(A\cap B)}{\Pr(B)} = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$

Named after Thomas Bayes, British clergyman, 1702-1761

Clicker: a test for rare events

Let us assume there is a rare disease that affects 0.1% of the population. There is a test that can detect this disease. It has a detection efficiency of 99% and a probability of error (false-positive) of 2%.

What is the probability $\Pr(D | +)$ of having the disease when receiving a positive test?

$\Pr(D | +) = \frac{\Pr(+ | D)\Pr(D)}{\Pr(+)} = \frac{\Pr(+ | D)\Pr(D)}{\Pr(+ | D)\Pr(D) + \Pr(+ | D^c)\Pr(D^c)}$ (law of total probability)

- $\Pr(+ | D) = 0.99$: probability of a postive test result, given the disease is present (detection efficiency of 99%)
- $\Pr(D) = 0.001$: the disease affects 0.1% of the population
- $\Pr(+ | D^c) = 0.02$: probability of a postive test result, given the disease is not present (false-positive rate of 2%)

$\Pr(D | +) = \frac{0.99 \cdot 0.001}{0.99 \cdot 0.001 + 0.02 \cdot 0.999} = 0.047$

The disease is only present in 5% of the cases where the test is positive!