Lecture Notes 1
Basic Probability

- Set Theory
- Elements of Probability
- Conditional probability
- Sequential Calculation of Probability
- Total Probability and Bayes Rule
- Independence
- Counting

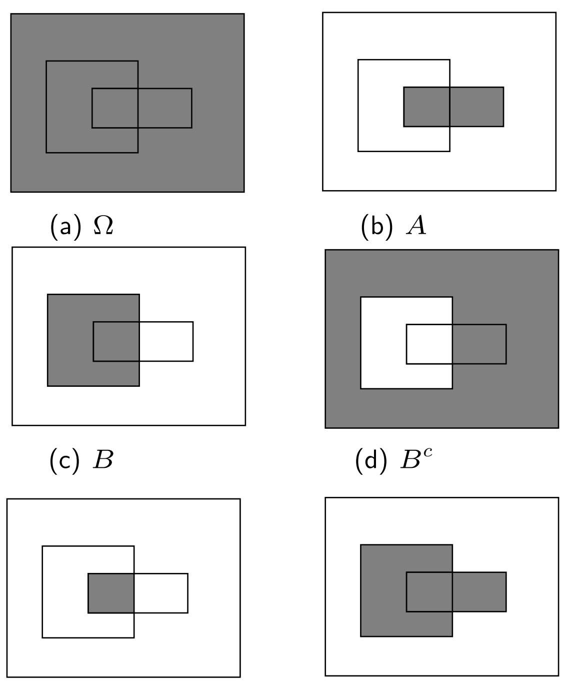
Set Theory Basics

- A set is a collection of objects, which are its *elements*
  - $\omega \in A$ means that ω is an element of the set A
  - A set with no elements is called the *empty set*, denoted by $\emptyset$
- Types of sets:
  - Finite: $A = \{\omega_1, \omega_2, \ldots, \omega_n\}$
  - Countably infinite: $A = \{\omega_1, \omega_2, \ldots\}$, e.g., the set of integers
  - Uncountable: A set that takes a continuous set of values, e.g., the $[0, 1]$ interval, the real line, etc.
- A set can be described by all ω having a certain property, e.g., $A = [0, 1]$ can be written as $A = \{\omega : 0 \leq \omega \leq 1\}$
- A set $B \subset A$ means that every element of B is an element of A
- A *universal set* Ω contains *all* objects of particular interest in a particular context, e.g., sample space for random experiment

Set Operations

- Assume a universal set Ω
- Three basic operations:
  - Complementation: A complement of a set A with respect to Ω is $A^c = \{\omega \in \Omega : \omega \notin A\}$, so $\Omega^c = \emptyset$
  - Intersection: $A \cap B = \{\omega : \omega \in A \text{ and } \omega \in B\}$
  - Union: $A \cup B = \{\omega : \omega \in A \text{ or } \omega \in B\}$
- Notation:
  - $\cup_{i=1}^n A_i = A_1 \cup A_2 \ldots \cup A_n$
  - $\cap_{i=1}^n A_i = A_1 \cap A_2 \ldots \cap A_n$
- A collection of sets $A_1, A_2, \ldots, A_n$ are *disjoint* or *mutually exclusive* if $A_i \cap A_j = \emptyset$ for all $i \neq j$, i.e., no two of them have a common element
- A collection of sets $A_1, A_2, \ldots, A_n$ *partition* Ω if they are disjoint and $\cup_{i=1}^n A_i = \Omega$

- Venn Diagrams



(a) Ω (b) A

(c) B (d) B^c

(e) $A \cap B$ (f) $A \cup B$

Algebra of Sets

- Basic relations:
  1. $S \cap \Omega = S$
  2. $(A^c)^c = A$
  3. $A \cap A^c = \emptyset$
  4. Commutative law: $A \cup B = B \cup A$
  5. Associative law: $A \cup (B \cup C) = (A \cup B) \cup C$
  6. Distributive law: $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
  7. DeMorgan's law: $(A \cap B)^c = A^c \cup B^c$
     DeMorgan's law can be generalized to n events:

$$(\cap_{i=1}^n A_i)^c = \cup_{i=1}^n A_i^c$$

- These can all be proven using the definition of set operations or visualized using Venn Diagrams

Elements of Probability

- Probability theory provides the mathematical rules for assigning probabilities to outcomes of random experiments, e.g., coin flips, packet arrivals, noise voltage
- Basic elements of probability:
  - *Sample space*: The set of all possible "elementary" or "finest grain" outcomes of the random experiment (also called *sample points*)
    - The sample points are all *disjoint*
    - The sample points are *collectively exhaustive*, i.e., together they make up the entire sample space
  - *Events*: Subsets of the sample space
  - *Probability law*: An assignment of probabilities to events in a mathematically consistent way

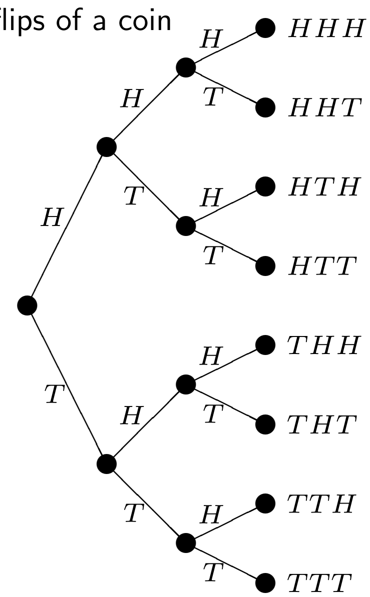
Discrete Sample Spaces

- Sample space is called *discrete* if it contains a countable number of sample points
- Examples:
  - Flip a coin once: $\Omega = \{H, T\}$
  - Flip a coin three times: $\Omega = \{HHH, HHT, HTH, \dots\} = \{H, T\}^3$
  - Flip a coin n times: $\Omega = \{H, T\}^n$ (set of sequences of H and T of length n)
  - Roll a die once: $\Omega = \{1, 2, 3, 4, 5, 6\}$
  - Roll a die twice: $\Omega = \{(1,1), (1,2), (2,1), \dots, (6,6)\} = \{1, 2, 3, 4, 5, 6\}^2$
  - Flip a coin until the first heads appears: $\Omega = \{H, TH, TTH, TTTH, \dots\}$
  - Number of packets arriving in time interval $(0, T]$ at a node in a communication network : $\Omega = \{0, 1, 2, 3, \dots\}$

  Note that the first five examples have *finite* Ω, whereas the last two have *countably infinite* Ω. Both types are called discrete

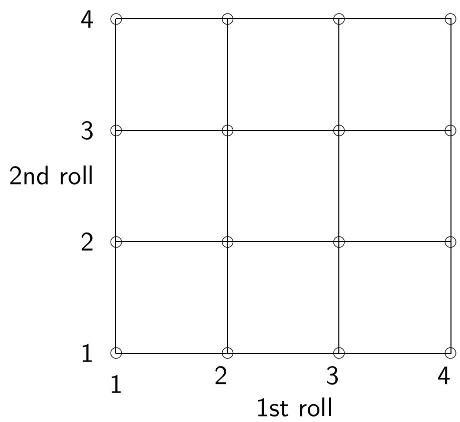
- Sequential models: For sequential experiments, the sample space can be described in terms of a tree, where each outcome corresponds to a terminal node (or a *leaf*)

  Example: Three flips of a coin



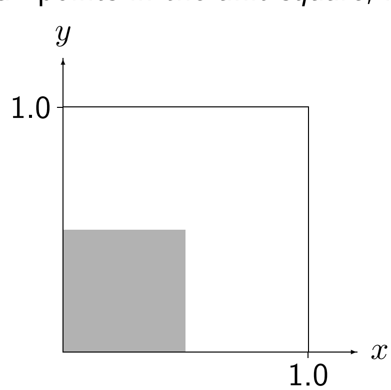
- Example: Roll a fair four-sided die twice.

  Sample space can be represented by a tree as above, or graphically



Continuous Sample Spaces

- A *continuous* sample space consists of a continuum of points and thus contains an uncountable number of points
- Examples:
  - Random number between 0 and 1: $\Omega = [0, 1]$
  - Suppose we pick two numbers at random between 0 and 1, then the sample space consists of all points in the *unit square*, i.e., $\Omega = [0, 1]^2$



- Packet arrival time: $t \in (0, \infty)$, thus $\Omega = (0, \infty)$
- Arrival times for n packets: $t_i \in (0, \infty)$, for $i = 1, 2, \ldots, n$, thus $\Omega = (0, \infty)^n$

- A sample space is said to be *mixed* if it is neither discrete nor continuous, e.g., $\Omega = [0, 1] \cup \{3\}$

Events

- Events are *subsets* of the sample space. An event occurs if the outcome of the experiment belongs to the event
- Examples:
  - Any outcome (sample point) is an event (also called an elementary event), e.g., {HTH} in three coin flips experiment or $\{0.35\}$ in the picking of a random number between 0 and 1 experiment
  - Flip coin 3 times and get exactly one H. This is a more complicated event, consisting of three sample points {TTH, THT, HTT}
  - Flip coin 3 times and get an odd number of H's. The event is {TTH, THT, HTT, HHH}
  - Pick a random number between 0 and 1 and get a number between 0.0 and 0.5. The event is $[0, 0.5]$
- An event might have *no points* in it, i.e., be the empty set $\emptyset$

Axioms of Probability

- Probability law (measure or function) is an assignment of probabilities to events (subsets of sample space Ω) such that the following three axioms are satisfied:
  1. $\mathrm{P}(A) \geq 0$, for all A (nonnegativity)
  2. $\mathrm{P}(\Omega) = 1$ (normalization)
  3. If A and B are disjoint ($A \cap B = \emptyset$), then
  $$\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) \text{ (additivity)}$$
  More generally,
  3'. If the sample space has an infinite number of points and $A_1, A_2, \ldots$ are disjoint events, then
  $$\mathrm{P}\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathrm{P}(A_i)$$

- Mimics relative frequency, i.e., perform the experiment n times (e.g., roll a die n times). If the number of occurances of A is n_A, define the relative frequency of an event A as $f_A = n_A/n$
  - Probabilities are nonnegative (like relative frequencies)
  - Probability something happens is 1 (again like relative frequencies)
  - Probabilities of *disjoint* events add (again like relative frequencies)
- Analogy: Except for normalization, probability is a *measure* much like
  - mass
  - length
  - area
  - volume

  They all satisfy axioms 1 and 3

  This analogy provides some intuition but is not sufficient to fully understand probability theory — other aspects such as conditioning, independence, etc.., are unique to probability

Probability for Discrete Sample Spaces

- Recall that sample space Ω is said to be *discrete* if it is countable
- The probability measure P can be simply defined by first assigning probabilities to outcomes, i.e., elementary events $\{\omega\}$, such that:

$$\begin{aligned} \mathrm{P}(\{\omega\}) \geq 0, \text{ for all } \omega \in \Omega, \text{ and} \\ \sum_{\omega \in \Omega} \mathrm{P}(\{\omega\}) = 1 \end{aligned}$$

- The probability of any other event A (by the additivity axiom) is simply

$$\mathrm{P}(A) = \sum_{\omega \in A} \mathrm{P}(\{\omega\})$$

- Examples:
  - For the coin flipping experiment, assign

    $$\mathrm{P}(\{H\}) = p \text{ and } \mathrm{P}(\{T\}) = 1 - p, \text{ for } 0 \leq p \leq 1$$

    Note: p is the *bias* of the coin, and a coin is *fair* if $p = \frac{1}{2}$
  - For the die rolling experiment, assign

    $$\mathrm{P}(\{i\}) = \frac{1}{6}, \text{ for } i = 1, 2, \ldots, 6$$

    The probability of the event "the outcome is even", $A = \{2, 4, 6\}$, is

    $$\mathrm{P}(A) = \mathrm{P}(\{2\}) + \mathrm{P}(\{4\}) + \mathrm{P}(\{6\}) = \frac{1}{2}$$

- If Ω is countably infinite, we can again assign probabilities to elementary events
  Example: Assume $\Omega = \{1, 2, \ldots\}$, assign probability 2^{-k} to event $\{k\}$
  The probability of the event "the outcome is even"

$$\begin{aligned}
\mathrm{P}(\text{ outcome is even }) &= \mathrm{P}(\{2,4,6,8,\ldots\}) \\
&= \mathrm{P}(\{2\}) + \mathrm{P}(\{4\}) + \mathrm{P}(\{6\}) + \ldots \\
&= \sum_{k=1}^{\infty} \mathrm{P}(\{2k\}) \\
&= \sum_{k=1}^{\infty} 2^{-2k} = \frac{1}{3}
\end{aligned}$$

Probability for Continuous Sample Space

- Recall that if a sample space is *continuous*, Ω is uncountably infinite
- For continuous Ω, we cannot in general define the probability measure P by first assigning probabilities to outcomes
- To see why, consider assigning a uniform probability measure to $\Omega = (0, 1]$
  - In this case the probability of each single outcome event is zero
  - How do we find the probability of an event such as $A = \left[\frac{1}{2}, \frac{3}{4}\right]$?
- For this example we can define uniform probability measure over $[0, 1]$ by assigning to an event A, the probability

$$\mathrm{P}(A) = \text{length of } A,$$

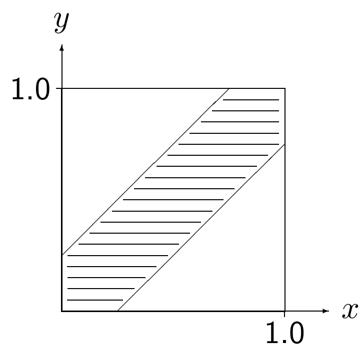
  e.g., $\mathrm{P}([0, 1/3] \cup [2/3, 1]) = 2/3$

  Check that this is a legitimate assignment

- Another example: Romeo and Juliet have a date. Each arrives late with a random delay of up to 1 hour. Each will wait only 1/4 of an hour before leaving. What is the probability that Romeo and Juliet will meet?

  Solution: The pair of delays is equivalent to that achievable by picking two random numbers between 0 and 1. Define probability of an event as its *area*

  The event of interest is represented by the cross hatched region



  Probability of the event is:

$$\text{area of crosshatched region} = 1 - 2 \times \frac{1}{2}(0.75)^2 = 0.4375$$

Basic Properties of Probability

- There are several useful properties that can be derived from the axioms of probability:

  1. $\mathrm{P}(A^c) = 1 - \mathrm{P}(A)$
     - $\mathrm{P}(\emptyset) = 0$
     - $\mathrm{P}(A) \leq 1$
  2. If $A \subseteq B$, then $\mathrm{P}(A) \leq \mathrm{P}(B)$
  3. $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$
  4. $\mathrm{P}(A \cup B) \leq \mathrm{P}(A) + \mathrm{P}(B)$, or in general

$$\mathrm{P}(\cup_{i=1}^n A_i) \leq \sum_{i=1}^n \mathrm{P}(A_i)$$

     This is called the *Union of Events Bound*

- These properties can be proved using the axioms of probability and visualized using Venn diagrams

Conditional Probability

- Conditional probability allows us to compute probabilities of events based on partial knowledge of the outcome of a random experiment
- Examples:
  - We are told that the sum of the outcomes from rolling a die twice is 9. What is the probability the outcome of the first die was a 6?
  - A spot shows up on a radar screen. What is the probability that there is an aircraft?
  - You receive a 0 at the output of a digital communication system. What is the probability that a 0 was sent?
- As we shall see, conditional probability provides us with two methods for computing probabilities of events: the *sequential* method and the *divide-and-conquer* method
- It is also the basis of *inference* in statistics: make an observation and reason about the cause

- In general, given an event B has occurred, we wish to find the probability of another event A, $\mathrm{P}(A|B)$
- If all elementary outcomes are equally likely, then

$$\mathrm{P}(A|B) = \frac{\#\text{ of outcomes in both } A \text{ and } B}{\#\text{ of outcomes in } B}$$

- In general, if B is an event such that $\mathrm{P}(B) \neq 0$, the *conditional probability* of any event A given B is defined as

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}, \text{ or } \frac{\mathrm{P}(A, B)}{\mathrm{P}(B)}$$

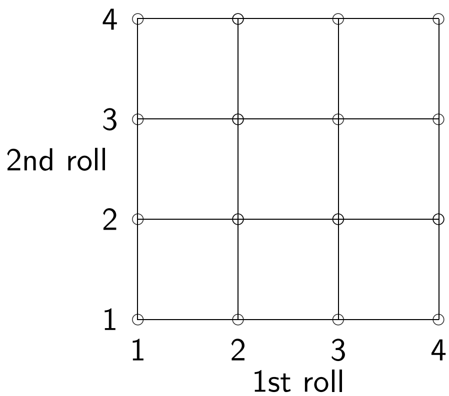
- The function $\mathrm{P}(\cdot|B)$ for fixed B specifies a probability law, i.e., it satisfies the axioms of probability

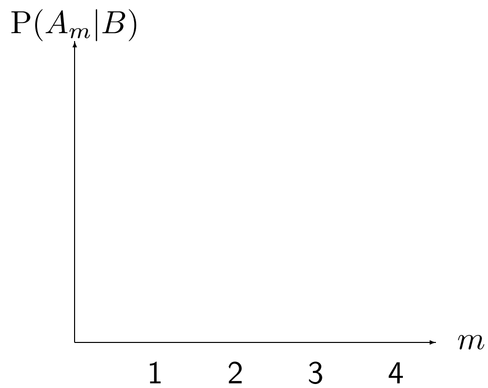
Example

- Roll a fair four-sided die twice. So, the sample space is $\{1, 2, 3, 4\}^2$. All sample points have probability 1/16

  Let B be the event that the minimum of the two die rolls is 2 and A_m, for $m = 1, 2, 3, 4$, be the event that the maximum of the two die rolls is m. Find $\mathrm{P}(A_m|B)$

- Solution:





Conditional Probability Models

- Before: Probability law $\Rightarrow$ conditional probabilities
- Reverse is often more natural: Conditional probabilities $\Rightarrow$ probability law
- We use the *chain rule* (also called *multiplication rule*):

  By the definition of conditional probability, $\mathrm{P}(A \cap B) = \mathrm{P}(A|B)\mathrm{P}(B)$. Suppose that $A_1, A_2, \ldots, A_n$ are events, then

$$
\begin{aligned}
&\mathrm{P}(A_1 \cap A_2 \cap A_3 \cdots \cap A_n) \\
&= \mathrm{P}(A_1 \cap A_2 \cap A_3 \cdots \cap A_{n-1}) \times \mathrm{P}(A_n | A_1 \cap A_2 \cap A_3 \cdots \cap A_{n-1}) \\
&= \mathrm{P}(A_1 \cap A_2 \cap A_3 \cdots \cap A_{n-2}) \times \mathrm{P}(A_{n-1} | A_1 \cap A_2 \cap A_3 \cdots \cap A_{n-2}) \\
&\quad \times \mathrm{P}(A_n | A_1 \cap A_2 \cap A_3 \cdots \cap A_{n-1}) \\
&\vdots \\
&= \mathrm{P}(A_1) \times \mathrm{P}(A_2|A_1) \times \mathrm{P}(A_3|A_1 \cap A_2) \cdots \mathrm{P}(A_n|A_1 \cap A_2 \cap A_3 \cdots \cap A_{n-1}) \\
&= \prod_{i=1}^{n} \mathrm{P}(A_i | A_1, A_2, \ldots, A_{i-1}),
\end{aligned}
$$

where $A_0 = \emptyset$

Sequential Calculation of Probabilities

- Procedure:
  1. Construct a tree description of the sample space for a sequential experiment
  2. Assign the conditional probabilities on the corresponding branches of the tree
  3. By the chain rule, the probability of an outcome can be obtained by multiplying the conditional probabilities along the path from the root to the leaf node corresponding to the outcome
- Example (Radar Detection): Let A be the event that an aircraft is flying above and B be the event that the radar detects it. Assume $\mathrm{P}(A) = 0.05$, $\mathrm{P}(B|A) = 0.99$, and $\mathrm{P}(B|A^c) = 0.1$

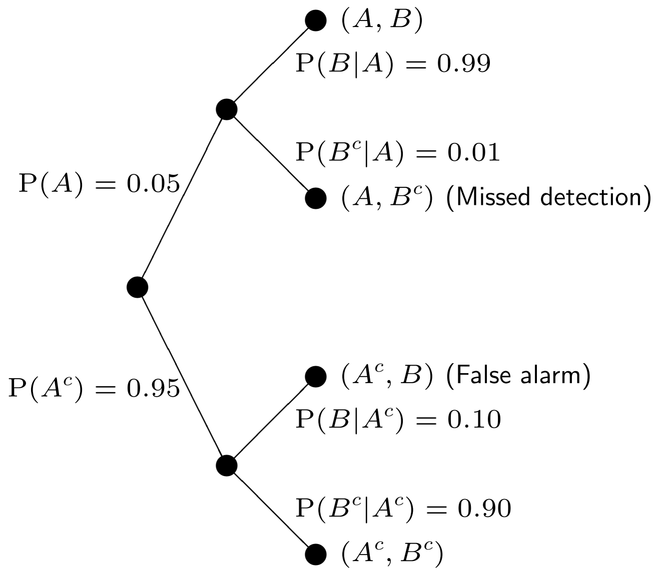
  What is the probability of

  ◦ Missed detection?, i.e., $\mathrm{P}(A \cap B^c)$

  ◦ False alarm?, i.e., $\mathrm{P}(B \cap A^c)$

  The sample space is: $\Omega = \{(A, B), (A^c, B), (A, B^c), (A^c, B^c)\}$

Solution: Represent the sample space by a tree with conditional probabilities on its edges

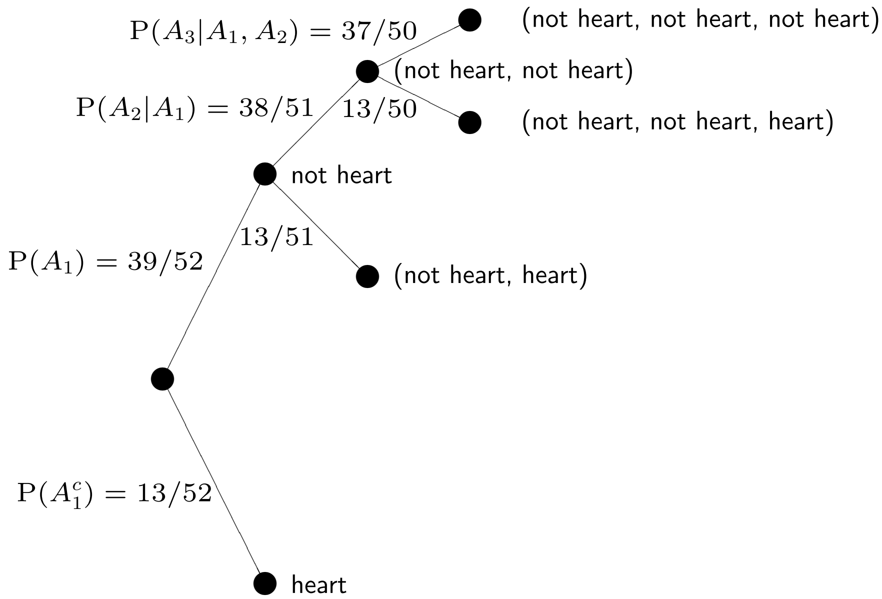


Example: Three cards are drawn at random (without replacement) from a deck of cards. What is the probability of not drawing a heart?

Solution: Let A_i, $i = 1, 2, 3$, represent the event of no heart in the ith draw. We can represent the sample space as:

$$\Omega = \{(A_1, A_2, A_3),\ (A_1^c, A_2, A_3), \ldots, (A_1^c, A_2^c, A_3^c)\}$$

To find the probability law, we represent the sample space by a tree, write conditional probabilities on branches, and use the chain rule



Total probability – Divide and Conquer Method

- Let $A_1, A_2, \ldots, A_n$ be events that partition Ω, i.e., that are disjoint ($A_i \cap A_j = \emptyset$ for $i \neq j$) and $\cup_{i=1}^{n} A_i = \Omega$. Then for any event B

$$\mathrm{P}(B) = \sum_{i=1}^{n} \mathrm{P}(A_i \cap B) = \sum_{i=1}^{n} \mathrm{P}(A_i)\mathrm{P}(B|A_i)$$

  This is called the Law of Total Probability. It also holds for $n = \infty$. It allows us to compute the probability of a complicated event from knowledge of probabilities of simpler events

- Example: Chess tournament, 3 types of opponents for a certain player.
  - $\mathrm{P}(\text{Type } 1) = 0.5$, $\mathrm{P}(\text{Win } |\text{Type1}) = 0.3$
  - $\mathrm{P}(\text{Type } 2) = 0.25$, $\mathrm{P}(\text{Win } |\text{Type2}) = 0.4$
  - $\mathrm{P}(\text{Type } 3) = 0.25$, $\mathrm{P}(\text{Win } |\text{Type3}) = 0.5$

  What is probability of player winning?

Solution: Let B be the event of winning and A_i be the event of playing Type i, $i = 1, 2, 3$:

$$\begin{aligned} \mathrm{P}(B) &= \sum_{i=1}^{3} \mathrm{P}(A_i)\mathrm{P}(B|A_i) \\ &= 0.5 \times 0.3 + 0.25 \times 0.4 + 0.25 \times 0.5 \\ &= 0.375 \end{aligned}$$

Bayes Rule

- Let $A_1, A_2, \ldots, A_n$ be nonzero probability events (the causes) that partition Ω, and let B be a nonzero probability event (the effect)
- We often know the *a priori* probabilities $\mathrm{P}(A_i)$, $i = 1, 2, \ldots, n$ and the conditional probabilities $\mathrm{P}(B|A_i)$s and wish to find the *a posteriori probabilities* $\mathrm{P}(A_j|B)$ for $j = 1, 2, \ldots, n$
- From the definition of conditional probability, we know that

$$\mathrm{P}(A_j|B) = \frac{\mathrm{P}(B, A_j)}{\mathrm{P}(B)} = \frac{\mathrm{P}(B|A_j)}{\mathrm{P}(B)}\mathrm{P}(A_j)$$

By the law of total probability

$$\mathrm{P}(B) = \sum_{i=1}^{n} \mathrm{P}(A_i)\mathrm{P}(B|A_i)$$

Substituting we obtain *Bayes rule*

$$\mathrm{P}(A_j|B) = \frac{\mathrm{P}(B|A_j)}{\sum_{i=1}^{n} \mathrm{P}(A_i)\mathrm{P}(B|A_i)} \mathrm{P}(A_j) \text{ for } j = 1, 2, \ldots, n$$

- Bayes rule also applies when the number of events $n = \infty$
- Radar Example: Recall that A is event that the aircraft is flying above and B is the event that the aircraft is detected by the radar. What is the probability that an aircraft is actually there given that the radar indicates a detection?

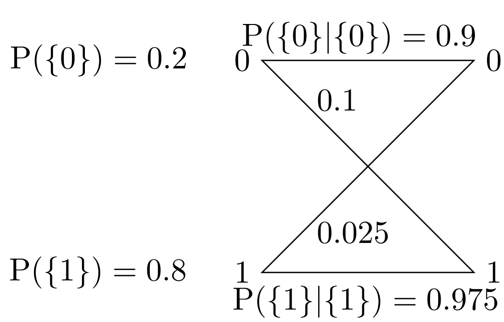
  Recall $\mathrm{P}(A) = 0.05$, $\mathrm{P}(B|A) = 0.99$, $\mathrm{P}(B|A^c) = 0.1$. Using Bayes rule:

$$\begin{aligned}
\mathrm{P}(\text{there is an aircraft}|\text{radar detects it}) &= \mathrm{P}(A|B) \\
&= \frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(B|A)\mathrm{P}(A) + \mathrm{P}(B|A^c)\mathrm{P}(A^c)} \\
&= \frac{0.05 \times 0.99}{0.05 \times 0.99 + 0.95 \times 0.1} \\
&= 0.3426
\end{aligned}$$

Binary Communication Channel

- Consider a noisy binary communication channel, where 0 or 1 is sent and 0 or 1 is received. Assume that 0 is sent with probability 0.2 (and 1 is sent with probability 0.8)

  The channel is noisy. If a 0 is sent, a 0 is received with probability 0.9, and if a 1 is sent, a 1 is received with probability 0.975
- We can represent this channel model by a *probability transition diagram*



Given that 0 is received, find the probability that 0 was sent

- This is a random experiment with sample space $\Omega = \{(0,0), (0,1), (1,0), (1,1)\}$, where the first entry is the bit sent and the second is the bit received

- Define the two events

$$A = \{0 \text{ is sent}\} = \{(0,1), (0,0)\}, \text{ and}$$
$$B = \{0 \text{ is received}\} = \{(0,0), (1,0)\}$$

- The probability measure is defined via the $\mathrm{P}(A)$, $\mathrm{P}(B|A)$, and $\mathrm{P}(B^c|A^c)$ provided on the probability transition diagram of the channel

- To find $\mathrm{P}(A|B)$, the *a posteriori* probability that a 0 was sent. We use Bayes rule
$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\mathrm{P}(A)}{\mathrm{P}(A)\mathrm{P}(B|A) + \mathrm{P}(A^c)\mathrm{P}(B|A^c)},$$
to obtain
$$\mathrm{P}(A|B) = \frac{0.9}{0.2} \times 0.2 = 0.9,$$
which is much higher than the $a\ priori$ probability of A $(= 0.2)$

Independence

- It often happens that the knowledge that a certain event B has occurred has no effect on the probability that another event A has occurred, i.e.,

$$\mathrm{P}(A|B) = \mathrm{P}(A)$$

In this case we say that the two events are statistically independent

- Equivalently, two events are said to be *statistically independent* if

$$\mathrm{P}(A, B) = \mathrm{P}(A)\mathrm{P}(B)$$

So, in this case, $\mathrm{P}(A|B) = \mathrm{P}(A)$ and $\mathrm{P}(B|A) = \mathrm{P}(B)$

- Example: Assuming that the binary channel of the previous example is used to send two bits independently, what is the probability that both bits are in error?

Solution:

- Define the two events

$$E_1 = \{\text{First bit is in error}\}$$
$$E_2 = \{\text{Second bit is in error}\}$$

- Since the bits are sent independently, the probability that both are in error is

$$\mathrm{P}(E_1, E_2) = \mathrm{P}(E_1)\mathrm{P}(E_2)$$

Also by symmetry, $\mathrm{P}(E_1) = \mathrm{P}(E_2)$
To find $\mathrm{P}(E_1)$, we express E_1 in terms of the events A and B as

$$E_1 = (A_1 \cap B_1^c) \cup (A_1^c \cap B_1),$$

- Thus,

$$\begin{aligned}\mathrm{P}(E_1) &= \mathrm{P}(A_1, B_1^c) + \mathrm{P}(A_1^c, B_1)\\ &= \mathrm{P}(A_1)\mathrm{P}(B_1^c|A_1) + \mathrm{P}(A_1^c)\mathrm{P}(B_1|A_1^c)\\ &= 0.2 \times 0.1 + 0.8 \times 0.025 = 0.04\end{aligned}$$

- The probability that the two bits are in error

$$\mathrm{P}(E_1, E_2) = (0.04)^2 = 16 \times 10^{-4}$$

- In general $A_1, A_2, \ldots, A_n$ are mutually independent if for *each* subset of the events $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$

$$\mathrm{P}(A_{i_1}, A_{i_2}, \ldots, A_{i_k}) = \prod_{j=1}^{k} \mathrm{P}(A_{i_j})$$

- Note: $\mathrm{P}(A_1, A_2, \ldots, A_n) = \prod_{j=1}^{n} \mathrm{P}(A_i)$ alone is not sufficient for independence

Example: Roll two fair dice independently. Define the events

$$\begin{aligned} A &= \{\text{First die } = 1, 2, \text{ or } 3\} \\ B &= \{\text{First die } = 2, 3, \text{ or } 6\} \\ C &= \{\text{Sum of outcomes } = 9\} \end{aligned}$$

Are A, B, and C independent?

Solution:

Since the dice are fair and the experiments are performed independently, the probability of any pair of outcomes is $1/36$, and

$$\mathrm{P}(A) = \frac{1}{2}, \quad \mathrm{P}(B) = \frac{1}{2}, \quad \mathrm{P}(C) = \frac{1}{9}$$

Since $A \cap B \cap C = \{(3, 6)\}$, $\mathrm{P}(A, B, C) = \frac{1}{36} = \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C)$

But are A, B, and C independent? Let's find

$$\mathrm{P}(A, B) = \frac{2}{6} = \frac{1}{3} \neq \frac{1}{4} = \mathrm{P}(A)\mathrm{P}(B),$$

and thus A, B, and C are not independent!

- Also, independence of subsets does not necessarily imply independence

  Example: Flip a fair coin twice independently. Define the events:

  - A: First toss is Head
  - B: Second toss is Head
  - C: First and second toss have different outcomes

  A and B are independent, A and C are independent, and B and C are independent

  Are A, B, C mutually independent?

  Clearly not, since if you know A and B, you know that C could not have occured, i.e., $\mathrm{P}(A, B, C) = 0 \neq \mathrm{P}(A)\mathrm{P}(B)\mathrm{P}(C) = 1/8$

Counting

- Discrete uniform law:
  - Finite sample space where all sample points are equally probable:

$$\mathrm{P}(A) = \frac{\text{number of sample points in } A}{\text{total number of sample points}}$$

  - Variation: all outcomes in A are equally likely, each with probability p. Then,

$$\mathrm{P}(A) = p \times (\text{number of elements of } A)$$

- In both cases, we compute probabilities by *counting*

Basic Counting Principle

- Procedure (multiplication rule):
  - r steps
  - n_i choices at step i
  - Number of choices is $n_1 \times n_2 \times \cdots \times n_r$
- Example:
  - Number of license plates with 3 letters followed by 4 digits:
  - With no repetition (replacement), the number is:
- Example: Consider a set of objects $\{s_1, s_2, \ldots, s_n\}$. How many subsets of these objects are there (including the set itself and the empty set)?
  - Each object may or may not be in the subset (2 options)
  - The number of subsets is:

De Méré's Paradox

- Counting can be *very* tricky
- Classic example: Throw three 6-sided dice. Is the probability that the sum of the outcomes is 11 equal to the the probability that the sum of the outcomes is 12?
- De Méré's argued that they are, since the number of different ways to obtain 11 and 12 are the same:
  - Sum=11: $\{6,4,1\}$, $\{6,3,2\}$, $\{5,5,1\}$, $\{5,4,2\}$, $\{5,3,3\}$, $\{4,4,3\}$
  - Sum=12: $\{6,5,1\}$, $\{6,4,2\}$, $\{6,3,3\}$, $\{5,5,2\}$, $\{5,4,3\}$, $\{4,4,4\}$
- This turned out to be false. Why?

Basic Types of Counting

- Assume we have n distinct objects $a_1, a_2, \ldots, a_n$, e.g., digits, letters, etc.
- Some basic counting questions:
  - How many different *ordered* sequences of k objects can be formed out of the n objects *with replacement*?
  - How many different *ordered* sequences of $k \leq n$ objects can be formed out of the n objects *without replacement*? (called k-*permutations*)
  - How many different *unordered* sequences (subsets) of $k \leq n$ objects can be formed out of the n objects *without replacement*? (called k-*combinations*)
  - Given r nonnegative integers $n_1, n_2, \ldots, n_r$ that sum to n (the number of objects), how many ways can the n objects be *partitioned* into r subsets (unordered sequences) with the ith subset having exactly n_i objects? (caller *partitions*, and is a generalization of combinations)

Ordered Sequences With and Without Replacement

- The number of ordered k-sequences from n objects with replacement is $n \times n \times n \ldots \times n$ k times, i.e., n^k

  Example: If $n = 2$, e.g., binary digits, the number of ordered k-sequences is 2^k
- The number of different ordered sequences of k objects that can be formed from n objects without replacement, i.e., the k-permutations, is

  $$n \times (n-1) \times (n-2) \cdots \times (n-k+1)$$

  If $k = n$, the number is

  $$n \times (n-1) \times (n-2) \cdots \times 2 \times 1 = n! \text{ (n-factorial)}$$

  Thus the number of k-permutations is: $n!/(n-k)!$
- Example: Consider the alphabet set $\{A, B, C, D\}$, so $n = 4$

  The number of $k = 2$-permutations of $n = 4$ objects is 12

  They are: $AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB$, and DC

Unordered Sequences Without Replacement

- Denote by $\binom{n}{k}$ (n choose k) the number of unordered k-sequences that can be formed out of n objects without replacement, i.e., the k-combinations
- Two different ways of constructing the k-permutations:
  1. Choose k objects ($\binom{n}{k}$), then order them ($k!$ possible orders). This gives $\binom{n}{k} \times k!$
  2. Choose the k objects one at a time:

$$n \times (n-1) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!} \text{ choices}$$

- Hence $\binom{n}{k} \times k! = \frac{n!}{(n-k)!}$, or $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
- Example: The number of $k = 2$-combinations of $\{A, B, C, D\}$ is 6

  They are AB, AC, AD, BC, BD, CD
- What is the number of binary sequences of length n with exactly $k \le n$ ones?
- Question: What is $\sum_{k=0}^{n} \binom{n}{k}$?

Finding Probablity Using Counting

- Example: *Die Rolling*

  Roll a fair 6-sided die 6 times independently. Find the probability that outcomes from all six rolls are different

  Solution:
  - # outcomes yielding this event =
  - # of points in sample space =
  - Probability =

- Example: *The Birthday Problem*

  k people are selected at random. What is the probability that all k birthdays will be different (neglect leap years)

  Solution:

  - Total # ways of assigning birthdays to k people:

  - # of ways of assigning birthdays to k people with no two having the same birthday:

  - Probability:

Binomial Probabilities

- Consider n independent coin tosses, where $\mathrm{P}(H) = p$ for $0 < p < 1$
- Outcome is a sequence of Hs and Ts of length n

$$\mathrm{P}(\text{sequence}) = p^{\#\text{heads}}(1-p)^{\#\text{tails}}$$

- The probability of k heads in n tosses is thus

$$\begin{aligned}\mathrm{P}(k \text{ heads}) &= \sum_{\text{sequences with } k \text{ heads}} \mathrm{P}(\text{sequence}) \\ &= \#(k\text{−head sequences}) \times p^k(1-p)^{n-k} \\ &= \binom{n}{k} p^k (1-p)^{n-k}\end{aligned}$$

Check that it sums to 1

- Example: Toss a coin with bias p independently 10 times.

  Define the events $B = \{3 \text{ out of 10 tosses are heads}\}$ and $A = \{\text{first two tosses are heads}\}$. Find $\mathrm{P}(A|B)$

  Solution: The conditional probability is

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A,B)}{\mathrm{P}(B)}$$

  All points in B have the same probability $p^3(1-p)^7$, so we can find the conditional probability by *counting*:

  - \# points in B beginning with two heads =
  - \# points in B =
  - Probability =

Partitions

- Let $n_1, n_2, n_3, \ldots, n_r$ be such that

$$\sum_{i=1}^{r} n_i = n$$

  How many ways can the n objects be *partitioned* into r subsets (unordered sequences) with the ith subset having exactly n_i objects?

- If $r = 2$ with $n_1, n - n_1$ the answer is the n_1-combinations $\binom{n}{n_1}$
- Answer in general:

$$
\begin{aligned}
\binom{n}{n_1\; n_2 \ldots n_r} &= \binom{n}{n_1} \times \binom{n-n_1}{n_2} \times \binom{n-(n_1+n_2)}{n_3} \times \ldots \binom{n-\sum_{i=1}^{r-1} n_i}{n_r} \\
&= \frac{n!}{n_1!(n-n_1)!} \times \frac{(n-n_1)!}{n_2!(n-(n_1+n_2))!} \times \cdots \\
&= \frac{n!}{n_1! n_2! \cdots n_r!}
\end{aligned}
$$

- Example: *Balls and bins*

  We have n balls and r bins. We throw each ball independently and at random into a bin. What is the probability that bin $i = 1, 2, \ldots, r$ will have exactly n_i balls, where $\sum_{i=1}^{r} n_i = n$?

  Solution:

  - The probability of each outcome (sequence of n_i balls in bin i) is:

  - # of ways of partitioning the n balls into r bins such that bin i has exactly n_i balls is:

  - Probability:

- Example: *Cards*

  Consider a perfectly shuffled 52-card deck dealt to 4 players. Find $\mathrm{P(each\ player\ gets\ an\ ace)}$

  Solution:

  - Size of the sample space is:

  - # of ways of distributing the four aces:

  - # of ways of dealing the remaining 48 cards:

  - Probability =