Reasoning under uncertainty

- In incomplete or imperfect settings. Not enough information.
- Why reasoning under uncertainty:
  - Modeling every detail of a complex system is difficult.
  - Not enough information to understand the system.

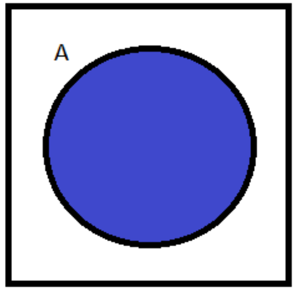
What is Probability?

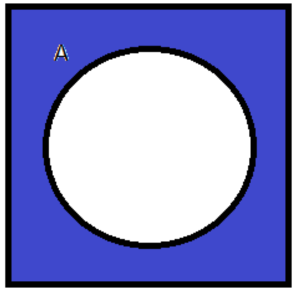
Different conditions:

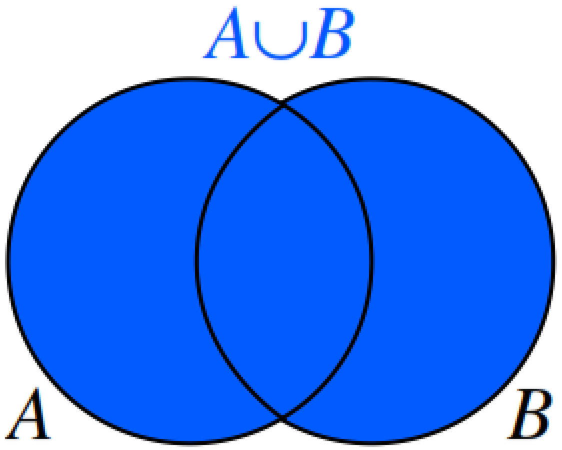
- All outcomes have equal probability.
- Randomness (such as weather)
- Frequency approach, that is probability is based on history.

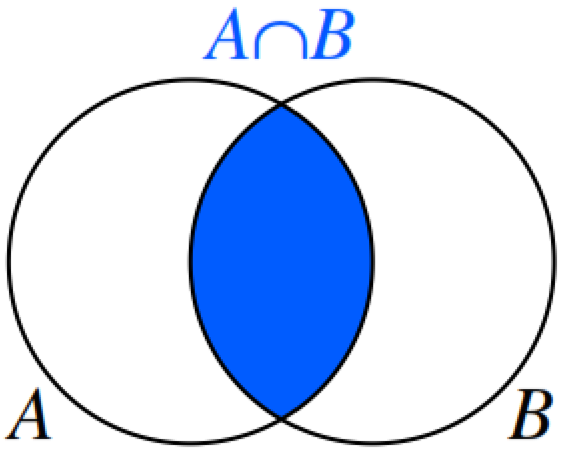
Probabilities quantify uncertainty regarding the occurrence of events.

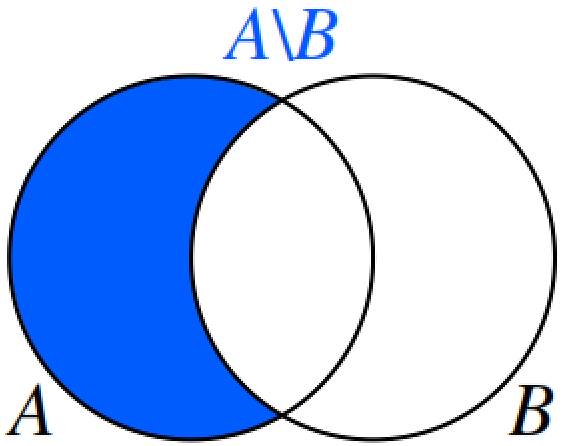
Set Theory









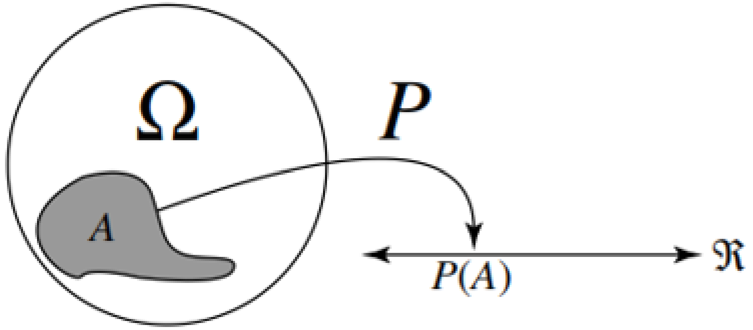


Showing all possible logical relations between a finite collection of sets.

Event A: A student has selected the statistics course.
Event B: A student has selected the language course.
Explain set of events using the van diagrams.

Probability Space

- A *probability space* represents our uncertainty regarding an experiment
- Three parts:
  - The *sample space* Ω, which is a set of *outcomes*
  - A set l of subsets of Ω, called *events*
  - The *probability measure* $P: l \rightarrow \mathbb{R}^+$



  - A $\in l$ is a set of outcomes, an event. P(a) represents the probability that the experiment's actual outcome will be a member of A.

$\sigma - algebra$

- Let Ω be an arbitrary set. A non-empty collection l subsets of Ω satisfying the following conditions called $\sigma - algebra$:
    - $\emptyset \in l$
    - If $D \in l$ then its complement is also in l, that is $\Omega \backslash \mathrm{D} \in l$
    - If $D_1, D_2, \cdots$, is a countable collection of sets in l then their union implies $\bigcup_{i=1}^{n} D_i \in l$

Example of a probability space

- Example 1: Throwing a dice.
  - Sample Space: {1,2,3,4,5,6}
  - Event (Dice number $\leq$ 3) : {1,2,3}

Repeating the experiment for large number of times, the fraction of times that an event A occurs is called the Probability of A.

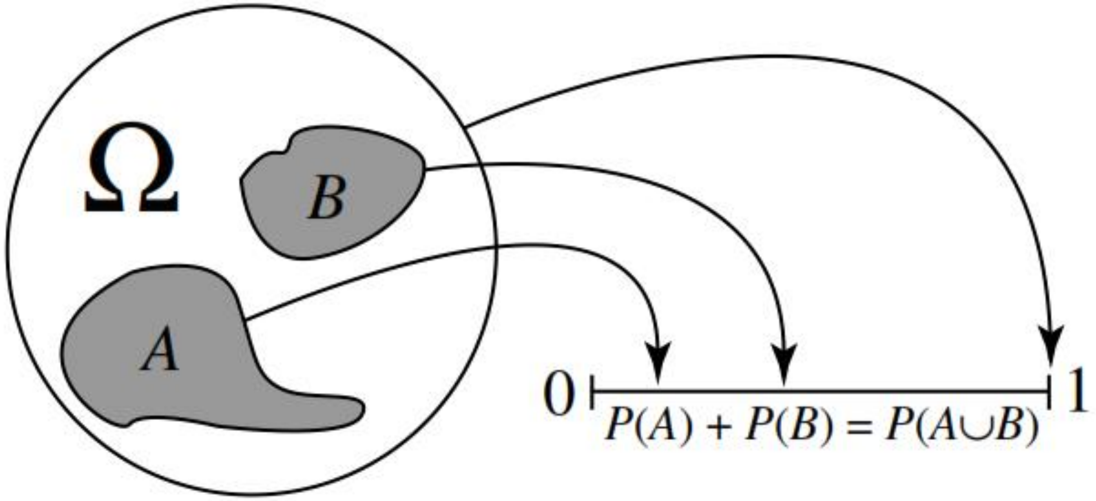
That is $P(A) = \frac{n}{N}$

Discrete sample space

- Example 2: t is the lifespan of an electrical device and shown as $S = \{t|t \geq 0\}$.
- Event: A device lasts less than 5 years shown as $A = \{t|0 \leq t < 5\}$.

Three axioms of probability space

1. $P(A) \geq 0$ for all events A
2. $P(\Omega) = 1$
3. $P(A \cup B) = P(A) + P(B)$ for disjoint events A and B



Events – [Mutually Exclusive]

- Two events are mutually exclusive if they cannot occur at the same time. Another word that means mutually exclusive is disjoint.
  - Let l be a $\sigma - algebra$ on the universe Ω.

  for sets $D_1, D_2, \cdots, D_n \in l$, that are mutually disjoint, we have

  $D_i \cap D_j = \emptyset, if\ i \neq j.$

$$P(\cup_{i=1}^{n} D_i) = \sum_{i=1} P(D_i)$$

- P(A) = 0.2, P(B) = 0.7 , A and B are disjoints.

	B	Ω\B
A	0.00	0.20
Ω\A	0.70	0.10

Events – [Independent Events]

- Two events are independent if the occurrence of one does not change the probability of the other occurring.
- If events are independent, then the probability of them both occurring is the product of the probabilities of each occurring.
- $P(A \cap B) = P(A) * P(B)$

- A and B are independent events
  - P(A and B) = P(A) * P(B)
  - P(A|B) = P(A)
  - P(B|A) = P(B)

P(A) = 0.20, P(B) = 0.70. => P(Ω\A) = 0.80, P(Ω\B) = 0.30

	B	Ω\B
A	0.14	0.06
Ω\A	0.56	0.24

Independent and identically distributed data

- Pick multiple random samples $s_1, s_2, \ldots, s_n \in \Omega$
- Probability that $s \in \Omega$ is picked: $P(s)$
- The sampling is independent and identically distributed iff
  - All samples are picked from the same probability space (Ω, l, P) (= identically distributed)
  - Samples are mutually independently picked, i.e.
    $$P(s_1, s_2) = P(s_1)P(s_2)$$
- The overall probability then can be written as
  $$\mathrm{P}(s_1, s_2, \ldots, s_n) = \prod_{i=1}^{n} P(s_i)$$

Dependent Events

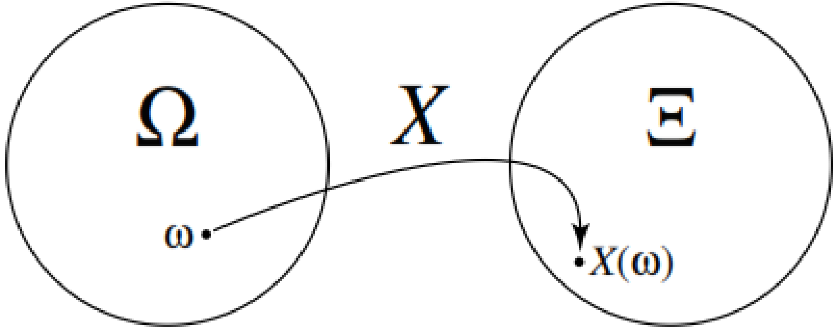
If the occurrence of one event does affect the probability of the other occurring, then the events are dependent.

- Conditional probability allows us to reason with *partial information.*
- When $P(B) > 0$, the *conditional probability of A given B* is defined as

$$P(A \mid B) \triangleq \frac{P(A \cap B)}{P(B)}$$

Random variables

- It is often useful to "pick out" aspects of the experiment's outcomes.
- A *random variable* X is a function from the sample space Ω.

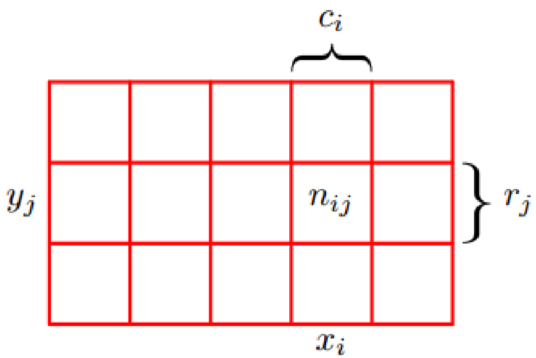


- Random variables can define events, e.g., $\{\omega \in \Omega : X(\omega) = \mathit{true}\}$.
- One will often see expressions like $P\{X = 1, Y = 2\}$ or $P(X = 1, Y = 2)$. These both mean $P(\{\omega \in \Omega : X(\omega) = 1, Y(\omega) = 2\})$.

Random variables - Example

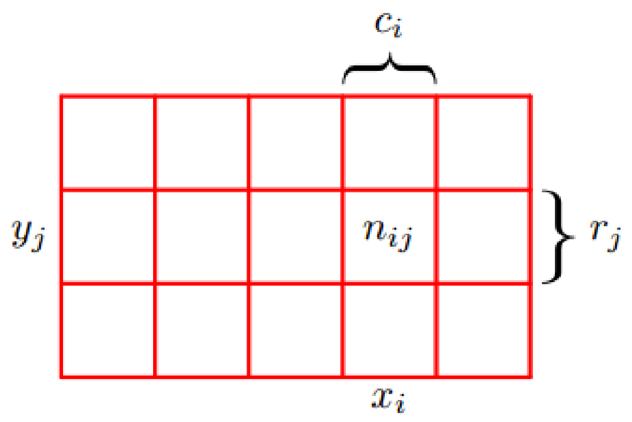
- Toss a coin 10 times.
- Sample Space (S)= {HTHT …}, all different configurations of H & T.
- Random variable X = number of heads. $X: S \to R$
- $X: S \to \{0,1,\cdots,10\}$.

Product and Sum rule



- Random variable X can take values $\{x_i\}, i = 1, 2, \ldots, M$ $(M = 5)$, Y can take values $\{y_j\}, j = 1, 2, \ldots, L$ $(L = 3)$.
- Consider a total number N of instances of these variables, and denote the number of instances where $X = x_i$ and $Y = y_i$ by n_{ij} (number of points in the corresponding cell of the array).

Product and Sum rule



- Number of points in column i, corresponding to $X = x_i$, is denoted by c_i.
- Number of points in row j, corresponding to $Y = y_i$, is denoted by r_j.

Sum Rule

Number of instances in column i is just the sum of the number of instances in each cell of that column, we have $c_i = \sum_j n_{ij}$ and hence

$$\begin{aligned} p(X = x_i) &= \frac{c_i}{N} \\ &= \frac{\sum_{j=1}^{L} n_{ij}}{N} \\ &= \sum_{j=1}^{L} p(X = x_i, Y = y_j). \end{aligned}$$

Sometimes also called *marginal* probability, because it is obtained by marginalizing, or summing out, the other variables (here Y).

Product Rule

Now we are interested in finding the fraction of those points in column i that fall in cell i, j, that is

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}.$$

By means of this *conditional* probability, we can derive the following relationship

$$\begin{aligned} p(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\ &= p(Y = y_j | X = x_i) p(X = x_i) \end{aligned}$$

which is the product rule of probabilities.

Two Important Rule in Probability Theory

- The product Rule
  - The probability that A and B both happens is given as the probability that A happens times the probability that B happens, given A has occurred.
    - $P(A \cap B) = P(A)P(B|A)$

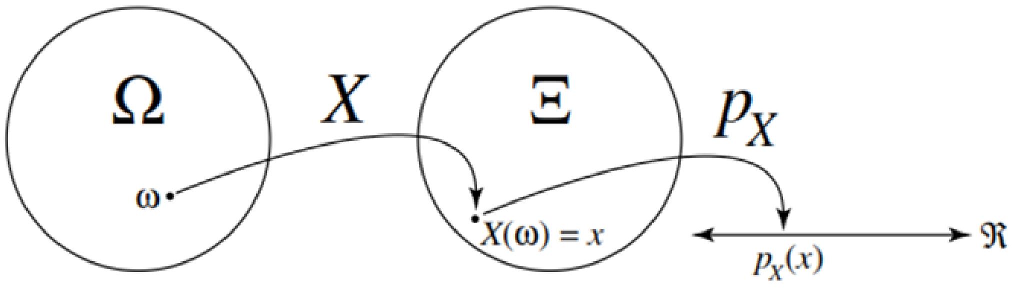
Densities

- We considered probabilities defined over discrete sets of events.
- Consider now probabilities with respect to *continuous* variables.

If the probability of a real-valued variable x falling in the interval $(x, x + \delta x)$ is given by $p(x)\delta x$ for $\delta x \to 0$, then $p(x)$ is called the *probability density* over x.

- The *probability* that x will lie in an interval (a, b) is given by

$$p(x \in (a,b)) = \int_a^b p(x)dx.$$



Expected Value

Let X be a numerically-values discrete random variable with sample space of Ω and the distribution function m(x) . The expected value is given as:

$$E(X) = \sum_{x \in \Omega} x m(x)$$

Let an experiment consist of tossing a fair coin three times. Let X denote the number of heads which appear. Then the possible values of X are $0, 1, 2$ and 3. The corresponding probabilities are $1/8, 3/8, 3/8$, and $1/8$. Thus, the expected value of X equals