Conditioning, Inference and Probability Logic

Bayesian Statistics and Machine Learning

Dominik Endres

October 29, 2015

Philipps Universität Marburg

Outline

Reminder: ingredients of probability

- Probability is built upon the concept of the *possible world* or *elementary outcome* $\in W$.
- Not all possible events need to be assigned a probability, only those that are members of a given σ-algebra $\mathcal{F}$.

Definition: A *probability space* is a tuple $(W, \mathcal{F}, P)$, where $\mathcal{F}$ is a σ-algebra over W and $P : \mathcal{F} \to [0, 1]$, with the properties:

P1 $P(W) = 1$

P2 If $U, V \in \mathcal{F}$ and $U \cap V = \emptyset$, then $P(U \cup V) = P(U) + P(V)$.

Important consequence: if $\mathcal{F} = 2^W$, then it is sufficent to specify $P(w)$ for all $w \in W$. P2 then allows you to compute $P(U)$ for any $U \subseteq W$.

From here on, we will assume that $\mathcal{F} = 2^W$!

Reminder: ingredients of probability

- Probability is built upon the concept of the *possible world* or *elementary outcome* $\in W$.
- Not all possible events need to be assigned a probability, only those that are members of a given σ-algebra $\mathcal{F}$.

Definition: A *probability space* is a tuple $(W, \mathcal{F}, P)$, where $\mathcal{F}$ is a σ-algebra over W and $P : \mathcal{F} \to [0,1]$, with the properties:

P1 $P(W) = 1$

P2 If $U, V \in \mathcal{F}$ and $U \cap V = \emptyset$, then $P(U \cup V) = P(U) + P(V)$.

Important consequence: if $\mathcal{F} = 2^W$, then it is sufficent to specify $P(w)$ for all $w \in W$. P2 then allows you to compute $P(U)$ for any $U \subseteq W$.

From here on, we will assume that $\mathcal{F} = 2^W$!

Reminder: conditioning

Updating of probability when information in the form 'the real world is in U' or 'U just happended' becomes available.

Conditioning: If $U, V \in \mathcal{F}$ and $P(U) > 0$, then $P(V|U) = \frac{P(V \cap U)}{P(U)}$.

Bayes' Rule: $P(U|V) = P(V|U)\frac{P(U)}{P(V)}$

Independence

Reminder:

Definition: A *probability space* is a tuple $(W, \mathcal{F}, P)$, where $\mathcal{F}$ is a σ-algebra over W and $P : \mathcal{F} \to [0,1]$, with the properties:

P1 $P(W) = 1$

P2 If $U, V \in \mathcal{F}$ and $U \cap V = \emptyset$, then $P(U \cup V) = P(U) + P(V)$

Two events $U, V \in \mathcal{F}$ are independent if the probability of V is not updated on learning U and vice versa:

$$P(V|U) = P(V) = \frac{P(V \cap U)}{P(U)} \quad \Rightarrow \quad P(V \cap U) = P(V)P(U)$$

$$P(U|V) = P(U) = \frac{P(U \cap V)}{P(V)} \quad \Rightarrow \quad P(U \cap V) = P(U)P(V)$$

Definition: two events $U, V \in \mathcal{F}$ are independent if and only if

$$P(U \cap V) = P(U)P(V).$$

Independence

Reminder:

Definition: A *probability space* is a tuple $(W, \mathcal{F}, P)$, where $\mathcal{F}$ is a σ-algebra over W and $P : \mathcal{F} \to [0,1]$, with the properties:

P1 $P(W) = 1$

P2 If $U, V \in \mathcal{F}$ and $U \cap V = \emptyset$, then $P(U \cup V) = P(U) + P(V)$

Two events $U, V \in \mathcal{F}$ are independent if the probability of V is not updated on learning U and vice versa:

$$P(V|U) = P(V) = \frac{P(V \cap U)}{P(U)} \quad \Rightarrow \quad P(V \cap U) = P(V)P(U)$$

$$P(U|V) = P(U) = \frac{P(U \cap V)}{P(V)} \quad \Rightarrow \quad P(U \cap V) = P(U)P(V)$$

Definition: two events $U, V \in \mathcal{F}$ are independent if and only if

$$P(U \cap V) = P(U)P(V).$$

Independence

Reminder:
Definition: A *probability space* is a tuple $(W, \mathcal{F}, P)$, where $\mathcal{F}$ is a σ-algebra over W and $P : \mathcal{F} \to [0, 1]$, with the properties:

P1 $P(W) = 1$
P2 If $U, V \in \mathcal{F}$ and $U \cap V = \emptyset$, then $P(U \cup V) = P(U) + P(V)$

Two events $U, V \in \mathcal{F}$ are independent if the probability of V is not updated on learning U and vice versa:

$$P(V|U) = P(V) = \frac{P(V \cap U)}{P(U)} \quad \Rightarrow \quad P(V \cap U) = P(V)P(U)$$

$$P(U|V) = P(U) = \frac{P(U \cap V)}{P(V)} \quad \Rightarrow \quad P(U \cap V) = P(U)P(V)$$

Definition: two events $U, V \in \mathcal{F}$ are independent if and only if

$$P(U \cap V) = P(U)P(V).$$

Products of elementary outcomes and possible worlds

Assume: the set W of possible worlds can be factorized into two sets, $W = D \times H$, i.e. the elements of W are tuples $j = (d, h)$ where $d \in D$ and $h \in H$.

For example let

- D be the possible elementary outcomes of rolling a die, $D = \{d_1, \ldots, d_6\}$.
- $H = \{h_1, h_2\}$ be the set of hypotheses $h_1 =$'the die is fair', and $h_2 =$'the die will show only the numbers 1,2,3'.
- Then, $(d_3, h_1) =$'the die showed 3 and the die is fair'.
- Let F='The die is fair'$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
- Let S_3='The die showed 3'$= \{(d_3, h_1), (d_3, h_2)\}$.
- 'The die showed 3 and the die is fair'$= (d_3, h_1) = S_3 \cap F$.

Products of elementary outcomes and possible worlds

Assume: the set W of possible worlds can be factorized into two sets, $W = D \times H$, i.e. the elements of W are tuples $j = (d, h)$ where $d \in D$ and $h \in H$.

For example let

- D be the possible elementary outcomes of rolling a die, $D = \{d_1, \ldots, d_6\}$.
- $H = \{h_1, h_2\}$ be the set of hypotheses $h_1 =$'the die is fair', and $h_2 =$'the die will show only the numbers 1,2,3'.
- Then, $(d_3, h_1) =$'the die showed 3 and the die is fair'.
- Let F='The die is fair'$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
- Let S_3='The die showed 3'$= \{(d_3, h_1), (d_3, h_2)\}$.
- 'The die showed 3 and the die is fair'$= (d_3, h_1) = S_3 \cap F$.

Products of elementary outcomes and possible worlds

Assume: the set W of possible worlds can be factorized into two sets, $W = D \times H$, i.e. the elements of W are tuples $j = (d, h)$ where $d \in D$ and $h \in H$.

For example let

- D be the possible elementary outcomes of rolling a die, $D = \{d_1, \ldots, d_6\}$.
- $H = \{h_1, h_2\}$ be the set of hypotheses h_1 ='the die is fair', and h_2 ='the die will show only the numbers 1,2,3'.
- Then, (d_3, h_1) ='the die showed 3 and the die is fair'.
- Let F='The die is fair'$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
- Let S_3='The die showed 3'$= \{(d_3, h_1), (d_3, h_2)\}$.
- 'The die showed 3 and the die is fair'$= (d_3, h_1) = S_3 \cap F$.

Products of elementary outcomes and possible worlds

Assume: the set W of possible worlds can be factorized into two sets, $W = D \times H$, i.e. the elements of W are tuples $j = (d, h)$ where $d \in D$ and $h \in H$.

For example let

- D be the possible elementary outcomes of rolling a die, $D = \{d_1, \ldots, d_6\}$.
- $H = \{h_1, h_2\}$ be the set of hypotheses $h_1 =$'the die is fair', and $h_2 =$'the die will show only the numbers 1,2,3'.
- Then, $(d_3, h_1) =$'the die showed 3 and the die is fair'.
- Let F='The die is fair'$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
- Let S_3='The die showed 3'$= \{(d_3, h_1), (d_3, h_2)\}$.
- 'The die showed 3 and the die is fair'$= (d_3, h_1) = S_3 \cap F$.

Products of elementary outcomes and possible worlds

Assume: the set W of possible worlds can be factorized into two sets, $W = D \times H$, i.e. the elements of W are tuples $j = (d, h)$ where $d \in D$ and $h \in H$.

For example let

- D be the possible elementary outcomes of rolling a die, $D = \{d_1, \dots, d_6\}$.
- $H = \{h_1, h_2\}$ be the set of hypotheses h_1 ='the die is fair', and h_2 ='the die will show only the numbers 1,2,3'.
- Then, (d_3, h_1) ='the die showed 3 and the die is fair'.
- Let F='The die is fair'$= \{(d_1, h_1), (d_2, h_1), \dots, (d_6, h_1)\}$
- Let S_3='The die showed 3'$= \{(d_3, h_1), (d_3, h_2)\}$.
- 'The die showed 3 and the die is fair'$= (d_3, h_1) = S_3 \cap F$.

Joint probabilities

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D$ and $h \in H$.
$F =$'The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.
'The die showed 3 and the die is fair'$= (d_3, h_1) = S_3 \cap F$

Definition: the probabilities

$$P(j) = P((d, h))$$

are called *joint probabilities* of (d, h), because they represent the probability of d and h happening together.

Marginal probabilities, sum rule

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D$ and $h \in H$.
F ='The die is fair' (i.e. h_1 is true)= $\{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
S_i='The die showed i'= $\{(d_i, h_1), (d_i, h_2)\}$.

The probability of F can be computed via P2, because $\{(d_i, h_1)\} \cap \{(d_j, h_1)\} = \emptyset$ for $i \neq j$.

Definition: the probability

$$P(F) = P(\bigcup_i \{(d_i, h_1)\}) = \sum_{i=1}^{6} P(\{(d_i, h_1)\})$$

$P(F)$ is called *marginal probability* of h_1, because the d_i have been marginalized in order to compute $P(F)$. This is the *sum rule* of probability.

Assigning joint probabilities

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.
If h_2 is true, die will show 1,2,3 only.

To apply the sum rule, we need the probabilities $P((d_i, h_1))$.
Assigning them directly might be difficult:

- Principle of indifference: $P((d, h)) = \frac{1}{|D||H|} = \frac{1}{12}$??
- But then, $P((d_5, h_2)) > 0$, but it should be impossible.
- $\Rightarrow$ this makes no sense.

Assigning joint probabilities

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.
If h_2 is true, die will show 1,2,3 only.

To apply the sum rule, we need the probabilities $P((d_i, h_1))$. Assigning them directly might be difficult:

- Principle of indifference: $P((d, h)) = \frac{1}{|D||H|} = \frac{1}{12}$??
- But then, $P((d_5, h_2)) > 0$, but it should be impossible.
- $\Rightarrow$ this makes no sense.

Product rule

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)= $\{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
S_i='The die showed i'= $\{(d_i, h_1), (d_i, h_2)\}$.

We can write $(d_3, h_1) = S_3 \cap F$

Conditioning: $P(U|V) = \frac{P(U \cap V)}{P(U)}$, therefore

$$P(U \cap V) = P(U|V)P(V)$$

This is known as the *product rule*.
Question: can we assign joint probabilities via the product rule, i.e. first determine $P(U|V)$ and then $P(V)$?

Product rule

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)= $\{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
S_i='The die showed i'= $\{(d_i, h_1), (d_i, h_2)\}$.

We can write $(d_3, h_1) = S_3 \cap F$

Conditioning: $P(U|V) = \frac{P(U \cap V)}{P(U)}$, therefore

$$P(U \cap V) = P(U|V)P(V)$$

This is known as the *product rule*.
Question: can we assign joint probabilities via the product rule, i.e. first determine $P(U|V)$ and then $P(V)$?

Assigning joint probabilities through conditionals and the product rule

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.

Question: can we assign probabilities via conditioning and the product rule?

- Given that h_1 is true, we assign $P(d_i|h_1 = \text{true}) = P(S_i|F) = \frac{1}{6}$.
- Given that h_2 is true, we assign
  - $P(d_i|h_2 = \text{true}) = P(S_i|\overline{F}) = \frac{1}{3}$ for $i = 1, 2, 3$.
  - $P(d_i|h_2 = \text{true}) = P(S_i|\overline{F}) = 0$ for $i = 4, 5, 6$

Assigning joint probabilities through conditionals and the product rule

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.

Question: can we assign probabilities via conditioning and the product rule?

- Given that h_1 is true, we assign $P(d_i|h_1 = \text{true}) = P(S_i|F) = \frac{1}{6}$.
- Given that h_2 is true, we assign
  - $P(d_i|h_2 = \text{true}) = P(S_i|\overline{F}) = \frac{1}{3}$ for $i = 1, 2, 3$.
  - $P(d_i|h_2 = \text{true}) = P(S_i|\overline{F}) = 0$ for $i = 4, 5, 6$

Assigning joint probabilities through conditionals and the product rule

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.

Question: can we assign probabilities via conditioning and the product rule?

- Given that h_1 is true, we assign $P(d_i | h_1 = \text{true}) = P(S_i | F) = \frac{1}{6}$.
- Given that h_2 is true, we assign
  - $P(d_i | h_2 = \text{true}) = P(S_i | \overline{F}) = \frac{1}{3}$ for $i = 1, 2, 3$.
  - $P(d_i | h_2 = \text{true}) = P(S_i | \overline{F}) = 0$ for $i = 4, 5, 6$

Assigning joint probabilities through conditionals and the product rule

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)= $\{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
$\overline{F}$='The die shows only 1,2,3' (i.e. 'h_2 is true) = $\{(d_1, h_2), \ldots, (d_6, h_2)\} = (D \times H) \setminus F$.
S_i='The die showed i'= $\{(d_i, h_1), (d_i, h_2)\}$.

Question: what are the probabilities of h_1 and h_2, i.e. of F and $\overline{F}$?

- Not knowing anything else, we assign $P(F) = P(\overline{F}) = \frac{1}{2}$.
- Applying the principle of indifference seems justified here.

Now we can compute the joint probabilities:

$$
\begin{aligned}
P((d_2, h_1)) &= P(S_2|F)P(F) = \frac{1}{6} \cdot \frac{1}{2} \\
P((d_4, h_2)) &= P(S_4|\overline{F})P(\overline{F}) = 0 \cdot \frac{1}{2}
\end{aligned}
$$

Assigning joint probabilities through conditionals and the product rule

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)= $\{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
$\overline{F}$='The die shows only 1,2,3' (i.e. 'h_2 is true) = $\{(d_1, h_2), \ldots, (d_6, h_2)\} = (D \times H) \setminus F$.
S_i='The die showed i'= $\{(d_i, h_1), (d_i, h_2)\}$.

Question: what are the probabilities of h_1 and h_2, i.e. of F and $\overline{F}$?

- Not knowing anything else, we assign $P(F) = P(\overline{F}) = \frac{1}{2}$.
- Applying the principle of indifference seems justified here.

Now we can compute the joint probabilities:

$$P((d_2, h_1)) = P(S_2|F)P(F) = \frac{1}{6} \cdot \frac{1}{2}$$

$$P((d_4, h_2)) = P(S_4|\overline{F})P(\overline{F}) = 0 \cdot \frac{1}{2}$$

Assigning joint probabilities through conditionals and the product rule

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
$\overline{F}$='The die shows only 1,2,3' (i.e. 'h_2 is true) $= \{(d_1, h_2), \ldots, (d_6, h_2)\} = (D \times H) \setminus F$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.

Question: what are the probabilities of h_1 and h_2, i.e. of F and $\overline{F}$?

- Not knowing anything else, we assign $P(F) = P(\overline{F}) = \frac{1}{2}$.
- Applying the principle of indifference seems justified here.

Now we can compute the joint probabilities:

$$
\begin{aligned}
P((d_2, h_1)) &= P(S_2|F)P(F) = \frac{1}{6} \cdot \frac{1}{2} \\
P((d_4, h_2)) &= P(S_4|\overline{F})P(\overline{F}) = 0 \cdot \frac{1}{2}
\end{aligned}
$$

Extension of the product rule: the chain rule

Product rule so far: for $U, V \in W$, $P(U) > 0$:
$P(V \cap U) = P(V|U)P(U)$

Question: can this be generalized to more than 2 events?

Answer: assume: $U = R \cap S$. Then

$$\begin{aligned} P(V \cap R \cap S) &= P(V \cap U) = P(V|U)P(U) \\ &= P(V|R \cap S)P(R \cap S) \\ &= P(V|R \cap S)P(R|S)P(S) \end{aligned}$$

Generally, for a sequence of events or propositions S_i, $i = 1, \ldots, N$ we obtain the **chain rule of probability**

$$P(\cap_{i=1}^{N} S_i) = \prod_{i=1}^{N} P(S_i | \cap_{j=i+1}^{N} S_j)$$

with $P(S_N | \cap_{j=N+1}^{N} S_j) = P(S_n)$, which holds for any ordering of the S_i.

Extension of the product rule: the chain rule

Product rule so far: for $U, V \in W$, $P(U) > 0$:
$P(V \cap U) = P(V|U)P(U)$
Question: can this be generalized to more than 2 events?
Answer: assume: $U = R \cap S$. Then

$$\begin{aligned} P(V \cap R \cap S) &= P(V \cap U) = P(V|U)P(U) \\ &= P(V|R \cap S)P(R \cap S) \\ &= P(V|R \cap S)P(R|S)P(S) \end{aligned}$$

Generally, for a sequence of events or propositions S_i, $i = 1, \ldots, N$ we obtain the **chain rule of probability**

$$P(\cap_{i=1}^{N} S_i) = \prod_{i=1}^{N} P(S_i | \cap_{j=i+1}^{N} S_j)$$

with $P(S_N | \cap_{j=N+1}^{N} S_j) = P(S_n)$, which holds for any ordering of the S_i.

Extension of the product rule: the chain rule

Product rule so far: for $U, V \in W$, $P(U) > 0$:
$P(V \cap U) = P(V|U)P(U)$
Question: can this be generalized to more than 2 events?
Answer: assume: $U = R \cap S$. Then

$$
\begin{aligned}
P(V \cap R \cap S) &= P(V \cap U) = P(V|U)P(U) \\
&= P(V|R \cap S)P(R \cap S) \\
&= P(V|R \cap S)P(R|S)P(S)
\end{aligned}
$$

Generally, for a sequence of events or propositions S_i, $i = 1, \ldots, N$ we obtain the **chain rule of probability**

$$P(\cap_{i=1}^{N} S_i) = \prod_{i=1}^{N} P(S_i| \cap_{j=i+1}^{N} S_j)$$

with $P(S_N| \cap_{j=N+1}^{N} S_j) = P(S_n)$, which holds for any ordering of the S_i.

Extension of the product rule: the chain rule

For a sequence of events or propositions $S_i,\ i = 1, \ldots, N$ we obtain the **chain rule of probability**

$$P(\cap_{i=1}^{N} S_i) = \prod_{i=1}^{N} P(S_i | \cap_{j=i+1}^{N} S_j)$$

with $P(S_N | \cap_{j=N+1}^{N} S_j) = P(S_n)$.

A consequence is (assume $k < N$):

$$\begin{aligned} P(\cap_{i=1}^{N} S_i) &= \prod_{i=1}^{k} P(S_i | \cap_{j=i+1}^{N} S_j) \prod_{i=k+1}^{N} P(S_i | \cap_{j=i+1}^{N} S_j) \\ &= P(\cap_{i=1}^{k} S_i | \cap_{i=k+1}^{N} S_i) P(\cap_{i=k+1}^{N} S_i) \end{aligned}$$

E.g.: $P(R \cap S \cap T) = P(R | S \cap T) P(S \cap T) = P(R \cap S | T) P(T)$

Extension of the product rule: the chain rule

For a sequence of events or propositions $S_i, \; i = 1, \ldots, N$ we obtain the **chain rule of probability**

$$P(\cap_{i=1}^{N} S_i) = \prod_{i=1}^{N} P(S_i | \cap_{j=i+1}^{N} S_j)$$

with $P(S_N | \cap_{j=N+1}^{N} S_j) = P(S_n)$.

A consequence is (assume $k < N$):

$$\begin{aligned} P(\cap_{i=1}^{N} S_i) &= \prod_{i=1}^{k} P(S_i | \cap_{j=i+1}^{N} S_j) \prod_{i=k+1}^{N} P(S_i | \cap_{j=i+1}^{N} S_j) \\ &= P(\cap_{i=1}^{k} S_i | \cap_{i=k+1}^{N} S_i) P(\cap_{i=k+1}^{N} S_i) \end{aligned}$$

E.g.: $P(R \cap S \cap T) = P(R | S \cap T) P(S \cap T) = P(R \cap S | T) P(T)$

Extension of the product rule: the chain rule

For a sequence of events or propositions $S_i,\ i = 1, \ldots, N$ we obtain the **chain rule of probability**

$$P(\cap_{i=1}^{N} S_i) = \prod_{i=1}^{N} P(S_i | \cap_{j=i+1}^{N} S_j)$$

with $P(S_N | \cap_{j=N+1}^{N} S_j) = P(S_n)$.

A consequence is (assume $k < N$):

$$\begin{aligned} P(\cap_{i=1}^{N} S_i) &= \prod_{i=1}^{k} P(S_i | \cap_{j=i+1}^{N} S_j) \prod_{i=k+1}^{N} P(S_i | \cap_{j=i+1}^{N} S_j) \\ &= P(\cap_{i=1}^{k} S_i | \cap_{i=k+1}^{N} S_i) P(\cap_{i=k+1}^{N} S_i) \end{aligned}$$

E.g.: $P(R \cap S \cap T) = P(R | S \cap T) P(S \cap T) = P(R \cap S | T) P(T)$

Inverting conditioning, inference via Bayes' rule

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
$\overline{F}$='The die shows only 1,2,3' (i.e. 'h_2 is true) $= \{(d_1, h_2), \ldots, (d_6, h_2)\} = (D \times H) \setminus F$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.

Assume we observe that the die showed a 3, i.e. S_3 happened.

Question: what does this tell us about whether the die is fair (F) or not $(\overline{F})$?

Answer: in probabilistic terms: what is $P(F|S_3)$?

This answer could be provided by Bayes' rule

$$P(F|S_3) = P(S_3|F)\frac{P(F)}{P(S_3)} = \frac{P(S_3 \cap F)}{P(S_3)}$$

if we knew $P(S_3)$.

Inverting conditioning, inference via Bayes' rule

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
$\overline{F}$='The die shows only 1,2,3' (i.e. 'h_2 is true) $= \{(d_1, h_2), \ldots, (d_6, h_2)\} = (D \times H) \setminus F$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.

Assume we observe that the die showed a 3, i.e. S_3 happened.
Question: what does this tell us about whether the die is fair (F) or not ($\overline{F}$) ?
Answer: in probabilistic terms: what is $P(F|S_3)$?
This answer could be provided by Bayes' rule

$$P(F|S_3) = P(S_3|F)\frac{P(F)}{P(S_3)} = \frac{P(S_3 \cap F)}{P(S_3)}$$

if we knew $P(S_3)$.

Inverting conditioning, inference via Bayes' rule

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)= $\{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
$\overline{F}$='The die shows only 1,2,3' (i.e. 'h_2 is true) = $\{(d_1, h_2), \ldots, (d_6, h_2)\} = (D \times H) \setminus F$.
S_i='The die showed i'= $\{(d_i, h_1), (d_i, h_2)\}$.

Assume we observe that the die showed a 3, i.e. S_3 happened.
Question: what does this tell us about whether the die is fair (F) or not ($\overline{F}$) ?
Answer: in probabilistic terms: what is $P(F|S_3)$?
This answer could be provided by Bayes' rule

$$P(F|S_3) = P(S_3|F)\frac{P(F)}{P(S_3)} = \frac{P(S_3 \cap F)}{P(S_3)}$$

if we knew $P(S_3)$.

Determining the marginal probability $P(S_i)$

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
$\overline{F}$='The die shows only 1,2,3' (i.e. 'h_2 is true) $= \{(d_1, h_2), \ldots, (d_6, h_2)\} = (D \times H) \setminus F$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.

Question: what is the marginal probability $P(S_3)$?

Answer: apply the sum rule

$$\begin{aligned} P(S_3) &= P(\{(d_3, h_1), (d_3, h_2)\}) \\ &= \sum_{i=1}^{2} P(d_3, h_i) \\ &= P(S_3|F)P(F) + P(S_3|\overline{F})P(\overline{F}) \\ &= \frac{1}{6}\frac{1}{2} + \frac{1}{3}\frac{1}{2} = \frac{1}{4} \end{aligned}$$

Determining the marginal probability $P(S_i)$

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
$\overline{F}$='The die shows only 1,2,3' (i.e. 'h_2 is true) $= \{(d_1, h_2), \ldots, (d_6, h_2)\} = (D \times H) \setminus F$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.

Question: what is the marginal probability $P(S_3)$?
Answer: apply the sum rule

$$
\begin{aligned}
P(S_3) &= P(\{(d_3, h_1), (d_3, h_2)\}) \\
&= \sum_{i=1}^{2} P(d_3, h_i) \\
&= P(S_3|F)P(F) + P(S_3|\overline{F})P(\overline{F}) \\
&= \frac{1}{6}\frac{1}{2} + \frac{1}{3}\frac{1}{2} = \frac{1}{4}
\end{aligned}
$$

Inference: computing $P(F|S_i)$

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)= $\{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
$\overline{F}$='The die shows only 1,2,3' (i.e. 'h_2 is true) $= \{(d_1, h_2), \ldots, (d_6, h_2)\} = (D \times H) \setminus F$.
S_i='The die showed i'= $\{(d_i, h_1), (d_i, h_2)\}$.

Thus, we obtain

$$P(F|S_3) = P(S_3|F)\frac{P(F)}{P(S_3)} = \frac{1}{6}\frac{\frac{1}{2}}{\frac{1}{4}} = \frac{1}{3}$$

$\Rightarrow P(F|S_3) < P(F)$, i.e. we are now more certain that the die is *not* fair.

Inference: computing $P(F|S_i)$

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \dots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \dots, (d_6, h_1)\}$.
$\overline{F}$='The die shows only 1,2,3' (i.e. 'h_2 is true) $= \{(d_1, h_2), \dots, (d_6, h_2)\} = (D \times H) \setminus F$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.

Thus, we obtain

$$P(F|S_3) = P(S_3|F)\frac{P(F)}{P(S_3)} = \frac{1}{6}\frac{\frac{1}{2}}{\frac{1}{4}} = \frac{1}{3}$$

$\Rightarrow P(F|S_3) < P(F)$, i.e. we are now more certain that the die is *not* fair.

Inference, observing S_4

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
$\overline{F}$='The die shows only 1,2,3' (i.e. 'h_2 is true) $= \{(d_1, h_2), \ldots, (d_6, h_2)\} = (D \times H) \setminus F$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.

Assume we observed S_4, i.e. the die shows a 4.
Now we find

$$
\begin{aligned}
P(S_4) &= P(\{(d_4, h_1), (d_4, h_2)\}) \\
&= P(S_4|F)P(F) + P(S_4|\overline{F})P(\overline{F}) \\
&= \frac{1}{6} \cdot \frac{1}{2} + 0 \cdot \frac{1}{2} \\
&= \frac{1}{12}
\end{aligned}
\tag{1}
$$

Inference, observing S_4

Reminder:
The set W of possible worlds can be factorized into two sets, $W = D \times H$.
The elements of W are tuples $j = (d, h)$ where $d \in D = \{d_1, \ldots, d_6\}$ and $h \in H = \{h_1, h_2\}$.
F ='The die is fair' (i.e. h_1 is true)$= \{(d_1, h_1), (d_2, h_1), \ldots, (d_6, h_1)\}$.
$\overline{F}$='The die shows only 1,2,3' (i.e. 'h_2 is true) $= \{(d_1, h_2), \ldots, (d_6, h_2)\} = (D \times H) \setminus F$.
S_i='The die showed i'$= \{(d_i, h_1), (d_i, h_2)\}$.

Assume we observed S_4, i.e. the die shows a 4.
The probabilites $P(F|S_4)$ and $P(\overline{F}|S_4)$ now are

$$\begin{aligned} P(F|S_4) &= P(S_4|F)\frac{P(F)}{P(S_4)} = \frac{1}{6} \cdot \frac{\frac{1}{2}}{\frac{1}{12}} = 1 \\ P(\overline{F}|S_4) &= P(S_4|\overline{F})\frac{P(\overline{F})}{P(S_4)} = 0 \cdot \frac{\frac{1}{2}}{\frac{1}{12}} = 0 \end{aligned} \tag{2}$$

$\Rightarrow$ we are now *certain* that the die is fair.

Summary: conditioning, joint and marginal probabilities

- Events V, U are independent iff $P(V \cap U) = P(V)P(U)$.
- If U, V independent: $P(V|U) = P(V)$ and $P(U|V) = P(U)$.
- Assume the set W of possible worlds can be factorized into two sets, $W = D \times H$.
  - Elementary outcomes are tuples $w = (d, h)$, $d \in D$, $h \in H$.
  - Joint probability $P((d, h))$ that d and h happen together.
  - Marginal probability $P(h) = \sum_d P((d, h))$.
  - Product rule: $P(U \cap V) = P(U|V)P(V)$
  - Useful for assigning joint probabilities!
  - Chain rule $P(\cap_{i=1}^N S_i) = \prod_{i=1}^N P(S_i| \cap_{j=i+1}^N S_j)$
- Conditioning can be inverted via Bayes' rule: $P(U|V) = P(V|U)\frac{P(U)}{P(V)}$
- Inversion can be used to condition on "data".

Summary: conditioning, joint and marginal probabilities

- Events V, U are independent iff $P(V \cap U) = P(V)P(U)$.
- If U, V independent: $P(V|U) = P(V)$ and $P(U|V) = P(U)$.
- Assume the set W of possible worlds can be factorized into two sets, $W = D \times H$.
    - Elementary outcomes are tuples $w = (d, h)$, $d \in D$, $h \in H$.
    - Joint probability $P((d, h))$ that d and h happen together.
    - Marginal probability $P(h) = \sum_d P((d, h))$.
    - Product rule: $P(U \cap V) = P(U|V)P(V)$
    - Useful for assigning joint probabilities!
    - Chain rule $P(\cap_{i=1}^N S_i) = \prod_{i=1}^N P(S_i | \cap_{j=i+1}^N S_j)$
- Conditioning can be inverted via Bayes' rule: $P(U|V) = P(V|U)\frac{P(U)}{P(V)}$
- Inversion can be used to condition on "data".

Summary: conditioning, joint and marginal probabilities

- Events V, U are independent iff $P(V \cap U) = P(V)P(U)$.
- If U, V independent: $P(V|U) = P(V)$ and $P(U|V) = P(U)$.
- Assume the set W of possible worlds can be factorized into two sets, $W = D \times H$.
  - Elementary outcomes are tuples $w = (d, h)$, $d \in D$, $h \in H$.
  - Joint probability $P((d, h))$ that d and h happen together.
  - Marginal probability $P(h) = \sum_d P((d, h))$.
  - Product rule: $P(U \cap V) = P(U|V)P(V)$
  - Useful for assigning joint probabilities!
  - Chain rule $P(\cap_{i=1}^N S_i) = \prod_{i=1}^N P(S_i| \cap_{j=i+1}^N S_j)$
- Conditioning can be inverted via Bayes' rule: $P(U|V) = P(V|U)\frac{P(U)}{P(V)}$
- Inversion can be used to condition on "data".

Probabilistic Machine Learning terminology

- D: observed variables, the *data*.
- H: hypotheses or hidden variables or latent variables.
  - often assigned via Principle of Indifference.
- $P(F)$: **prior** probability of F='the coin is fair', i.e. before observing data.
- $P(S_3|F)$: **likelihood** of F. Conditional prob. of data (S_3 ='die shows three') given hidden/latent $F \subseteq D \times H$.
- $P(S_3)$ (if S_3 was observed): marginal probability of data. Also: evidence.
- $P(F|S_3)$: **posterior** probability of F, i.e. after observing data.
- Hierarchical model: $P(S_3, F) = P(S_3|F)P(F)$.
  - Decompose difficult-to-assign $P(S_i, F)$ into more accessible $P(S_i|F)$ and $P(F)$.

Probability logic

Reminder:
Probability space $(W, \mathcal{F}, P)$, where $\mathcal{F} = 2^W$.

Propositional logic

- Propositions A, B
- $A \in \{\text{true}, \text{false}\}$
- Logical operators
  - AND: $A \wedge B$
  - OR: $A \vee B$
  - NOT: $\sim A$
- Deductive reasoning. Validity, soundness.

Probability

- Propositions $A, B \in \mathcal{F}$
- $P(A) \in [0, 1]$
- Set operations
  - Intersection: $A \cap B$
  - Union: $A \cup B$
  - Complement: $\bar{A}$
- Probabilistic reasoning? How does it work?

Probability logic

Reminder:
Probability space $(W, \mathcal{F}, P)$, where $\mathcal{F} = 2^W$.

Propositional logic

- Propositions A,B
- $A \in \{\text{true}, \text{false}\}$
- Logical operators
  - AND: $A \wedge B$
  - OR: $A \vee B$
  - NOT: $\sim A$
- Deductive reasoning. Validity, soundness.

Probability

- Propositions $A, B \in \mathcal{F}$
- $P(A) \in [0, 1]$
- Set operations
  - Intersection: $A \cap B$
  - Union: $A \cup B$
  - Complement: $\bar{A}$
- Probabilistic reasoning? How does it work?

Probability logic

Reminder:
Probability space $(W, \mathcal{F}, P)$, where $\mathcal{F} = 2^W$.

Propositional logic

- Propositions A,B
- $A \in \{\text{true}, \text{false}\}$
- Logical operators
  - AND: $A \wedge B$
  - OR: $A \vee B$
  - NOT: $\sim A$
- Deductive reasoning. Validity, soundness.

Probability

- Propositions $A, B \in \mathcal{F}$
- $P(A) \in [0, 1]$
- Set operations
  - Intersection: $A \cap B$
  - Union: $A \cup B$
  - Complement: $\bar{A}$
- Probabilistic reasoning? How does it work?

Probability logic

Reminder:
Probability space $(W, \mathcal{F}, P)$, where $\mathcal{F} = 2^W$.

Propositional logic

- Propositions A, B
- $A \in \{\text{true}, \text{false}\}$
- Logical operators
  - AND: $A \wedge B$
  - OR: $A \vee B$
  - NOT: $\sim A$
- Deductive reasoning. Validity, soundness.

Probability

- Propositions $A, B \in \mathcal{F}$
- $P(A) \in [0, 1]$
- Set operations
  - Intersection: $A \cap B$
  - Union: $A \cup B$
  - Complement: $\bar{A}$
- Probabilistic reasoning? How does it work?

Deductive vs. probabilistic reasoning: example

Let the propositions A, B. Assume you are doing a Neuroscience experiment.

- $A =$"stimulus X is presented"
- $B =$"neuron Y fires"

- If stimulus X is presented, then neuron Y fires.
- Stimulus X is presented.
- Therefore, neuron Y fires.

$$\frac{\begin{array}{c} A \Rightarrow B \\ A \end{array}}{B}$$

This is an example of *modus ponens*.
The $\Rightarrow$ is a (material) implication.
The conclusion B is true *given* that the premises $A \Rightarrow B$ and A are both true.
Question: how do we translate this into probability?

Deductive vs. probabilistic reasoning: example

Let the propositions A, B. Assume you are doing a Neuroscience experiment.

- $A =$"stimulus X is presented"
- $B =$"neuron Y fires"

- If stimulus X is presented, then neuron Y fires.
- Stimulus X is presented.
- Therefore, neuron Y fires.

$$\begin{array}{c} A \Rightarrow B \\ A \\ \hline B \end{array}$$

This is an example of *modus ponens*.
The $\Rightarrow$ is a (material) implication.
The conclusion B is true *given* that the premises $A \Rightarrow B$ and A are both true.
Question: how do we translate this into probability?

Translating deductive into probabilistic reasoning

Reminder:

Propositions A ="stimulus X is presented", B ="neuron Y fires". Modus ponens: $\dfrac{\begin{array}{c}A \Rightarrow B\\ A\end{array}}{B}$

Modus ponens: The conclusion B is true *given* that the premises $A \Rightarrow B$ and A are both true.
Probability: $P(B|A \cap (A \Rightarrow B)) = 1$.
Thus: need to express the implication $A \Rightarrow B$ with set operators.

Truth table of $A \Rightarrow B$

A	B	$A \Rightarrow B$	$\sim A \vee B$
true	true	true	true
true	false	false	false
false	true	true	true
false	false	true	true

Thus,

$$A \Rightarrow B \quad = \quad \sim A \vee B.$$
$$A \Rightarrow B \quad = \quad \bar{A} \cup B$$

Translating deductive into probabilistic reasoning

Reminder:

Propositions A ="stimulus X is presented", B ="neuron Y fires". Modus ponens: $\dfrac{\begin{array}{c}A \Rightarrow B\\ A\end{array}}{B}$

Modus ponens:The conclusion B is true *given* that the premises $A \Rightarrow B$ and A are both true.

Probability: $P(B|A \cap (A \Rightarrow B)) = 1$.

Thus: need to express the implication $A \Rightarrow B$ with set operators.

Truth table of $A \Rightarrow B$

A	B	$A \Rightarrow B$	$\sim A \vee B$
true	true	true	true
true	false	false	false
false	true	true	true
false	false	true	true

Thus,

$$A \Rightarrow B = \sim A \vee B.$$
$$A \Rightarrow B = \bar{A} \cup B$$

Translating deductive into probabilistic reasoning

Reminder:

Propositions A ="stimulus X is presented", B ="neuron Y fires". Modus ponens: $\dfrac{\begin{array}{c} A \Rightarrow B \\ A \end{array}}{B}$

Modus ponens: The conclusion B is true *given* that the premises $A \Rightarrow B$ and A are both true.

Probability: $P(B|A \cap (A \Rightarrow B)) = 1$.

Thus: need to express the implication $A \Rightarrow B$ with set operators.

Truth table of $A \Rightarrow B$

A	B	$A \Rightarrow B$	$\sim A \vee B$
true	true	true	true
true	false	false	false
false	true	true	true
false	false	true	true

Thus,

$$A \Rightarrow B \quad = \quad \sim A \vee B.$$

$$A \Rightarrow B \quad = \quad \bar{A} \cup B$$

Translating deductive into probabilistic reasoning

Reminder:

Propositions A ="stimulus X is presented", B ="neuron Y fires". Modus ponens: $\dfrac{\begin{array}{c} A \Rightarrow B \\ A \end{array}}{B}$

$A \Rightarrow B = \bar{A} \cup B$
Need to show: $P(B|A \cap (A \Rightarrow B)) = 1$.

We find:

$$\begin{aligned} A \cap (A \Rightarrow B) &= A \cap (\bar{A} \cup B) \\ &= (A \cap \bar{A}) \cup (A \cap B) \\ &= \emptyset \cup (A \cap B) \\ &= A \cap B \end{aligned}$$

Therefore,
$P(B|A \cap (A \Rightarrow B)) = P(B|A \cap B) = \frac{P(B \cap A \cap B)}{P(A \cap B)} = \frac{P(A \cap B)}{P(A \cap B)} = 1 \quad \square$

Translating deductive into probabilistic reasoning

Reminder:

Propositions A ="stimulus X is presented", B ="neuron Y fires". Modus ponens: $\dfrac{\begin{array}{c} A \Rightarrow B \\ A \end{array}}{B}$

$A \Rightarrow B = \bar{A} \cup B$

Need to show: $P(B|A \cap (A \Rightarrow B)) = 1$.

We find:

$$\begin{aligned} A \cap (A \Rightarrow B) &= A \cap (\bar{A} \cup B) \\ &= (A \cap \bar{A}) \cup (A \cap B) \\ &= \emptyset \cup (A \cap B) \\ &= A \cap B \end{aligned}$$

Therefore,

$$P(B|A \cap (A \Rightarrow B)) = P(B|A \cap B) = \frac{P(B \cap A \cap B)}{P(A \cap B)} = \frac{P(A \cap B)}{P(A \cap B)} = 1 \quad \square$$

Uncertain premises

Reminder:

Propositions A ="stimulus X is presented", B ="neuron Y fires". Modus ponens: $\dfrac{\begin{array}{c}A \Rightarrow B\\ A\end{array}}{B}$

$A \Rightarrow B = \bar{A} \cup B$.
$P(B|A \cap (A \Rightarrow B) = P(B|A \cap B) = 1$.

Question: what if we are not sure of the premises, i.e. they are not *given* ?

Propositional logic: we're stuck. *No answer.*

Probability: what can we say about the *marginal* probability of the conclusion, $P(B)$?

Assume

- $P(A \Rightarrow B) = p_I$
- $P(A) = p_A$

Question: What is $P(B) = P(A \cap B) + P(\bar{A} \cap B)$

Uncertain premises

Reminder:

Propositions A ="stimulus X is presented", B ="neuron Y fires". Modus ponens: $\dfrac{\begin{array}{c}A \Rightarrow B\\ A\end{array}}{B}$

$A \Rightarrow B = \bar{A} \cup B$.
$P(B|A \cap (A \Rightarrow B) = P(B|A \cap B) = 1$.

Question: what if we are not sure of the premises, i.e. they are not *given* ?
Propositional logic: we're stuck. *No answer*.
Probability: what can we say about the *marginal* probability of the conclusion, $P(B)$?

Assume

- $P(A \Rightarrow B) = p_I$
- $P(A) = p_A$

Question: What is $P(B) = P(A \cap B) + P(\bar{A} \cap B)$

Uncertain premises

Reminder:

Propositions $A =$"stimulus X is presented", $B =$"neuron Y fires". Modus ponens: $\dfrac{\begin{array}{c}A \Rightarrow B\\ A\end{array}}{B}$

$A \Rightarrow B = \bar{A} \cup B$.
$P(B|A \cap (A \Rightarrow B) = P(B|A \cap B) = 1$.

Question: what if we are not sure of the premises, i.e. they are not *given* ?
Propositional logic: we're stuck. *No answer.*
Probability: what can we say about the *marginal* probability of the conclusion, $P(B)$?

Assume

- $P(A \Rightarrow B) = p_I$
- $P(A) = p_A$

Question: What is $P(B) = P(A \cap B) + P(\bar{A} \cap B)$

Uncertain premises: evaluating $P(B)$

Reminder:
$A \Rightarrow B = \bar{A} \cup B$.
$P(B) = P(A \cap B) + P(\bar{A} \cap B)$
$P(A \Rightarrow B) = p_I$ and $P(A) = p_A$.

We compute $P(A \cap B) = P(B|A)P(A)$ via the product rule.

Note that

$$\begin{aligned} P(A \Rightarrow B) &= P(\bar{A} \cup B) \\ &= P\left(\overline{\overline{\bar{A} \cup B}}\right) \\ &= P\left(\overline{A \cap \bar{B}}\right) \\ &= 1 - P(A \cap \bar{B}) \end{aligned}$$

Thus $P(\bar{B} \cap A) = P(\bar{B}|A)P(A) = 1 - p_I$ and $P(B|A) = 1 - \frac{1-p_I}{p_A}$.

Whence $\boxed{P(B \cap A) = P(B|A)P(A) = p_A + p_I - 1}$.

Uncertain premises: evaluating $P(B)$

Reminder:
$A \Rightarrow B = \bar{A} \cup B$.
$P(B) = P(A \cap B) + P(\bar{A} \cap B)$
$P(A \Rightarrow B) = p_I$ and $P(A) = p_A$.

We compute $P(A \cap B) = P(B|A)P(A)$ via the product rule.
Note that

$$
\begin{aligned}
P(A \Rightarrow B) &= P(\bar{A} \cup B) \\
&= P\left(\overline{\overline{\bar{A} \cup B}}\right) \\
&= P\left(\overline{A \cap \bar{B}}\right) \\
&= 1 - P(A \cap \bar{B})
\end{aligned}
$$

Thus $P(\bar{B} \cap A) = P(\bar{B}|A)P(A) = 1 - p_I$ and $P(B|A) = 1 - \frac{1-p_I}{p_A}$.

Whence $\boxed{P(B \cap A) = P(B|A)P(A) = p_A + p_I - 1}$.

Uncertain premises: evaluating $P(B)$

Reminder:
$A \Rightarrow B = \bar{A} \cup B$.
$P(B) = P(A \cap B) + P(\bar{A} \cap B)$
$P(A \Rightarrow B) = p_I$ and $P(A) = p_A$.

We compute $P(A \cap B) = P(B|A)P(A)$ via the product rule.
Note that

$$\begin{aligned} P(A \Rightarrow B) &= P(\bar{A} \cup B) \\ &= P\left(\overline{\overline{\bar{A} \cup B}}\right) \\ &= P\left(\overline{A \cap \bar{B}}\right) \\ &= 1 - P(A \cap \bar{B}) \end{aligned}$$

Thus $P(\bar{B} \cap A) = P(\bar{B}|A)P(A) = 1 - p_I$ and $P(B|A) = 1 - \frac{1-p_I}{p_A}$.

Whence $\boxed{P(B \cap A) = P(B|A)P(A) = p_A + p_I - 1}$.

Uncertain premises: evaluating $P(B)$

Reminder:
$A \Rightarrow B = \bar{A} \cup B$.
$P(B) = P(A \cap B) + P(\bar{A} \cap B)$
$P(A \Rightarrow B) = p_I$ and $P(A) = p_A$.

We compute $P(A \cap B) = P(B|A)P(A)$ via the product rule.
Note that

$$
\begin{aligned}
P(A \Rightarrow B) &= P(\bar{A} \cup B) \\
&= P\left(\overline{\overline{\bar{A} \cup B}}\right) \\
&= P\left(\overline{A \cap \bar{B}}\right) \\
&= 1 - P(A \cap \bar{B})
\end{aligned}
$$

Thus $P(\bar{B} \cap A) = P(\bar{B}|A)P(A) = 1 - p_I$ and $P(B|A) = 1 - \frac{1-p_I}{p_A}$.
Whence $\boxed{P(B \cap A) = P(B|A)P(A) = p_A + p_I - 1}$.

Uncertain premises: evaluating $P(B)$, part 2

Reminder:
$A \Rightarrow B = \bar{A} \cup B$.
$P(B) = P(A \cap B) + P(\bar{A} \cap B)$
$P(A \Rightarrow B) = p_I$ and $P(A) = p_A$.
$\boxed{P(B \cap A) = P(B|A)P(A) = p_A + p_I - 1.}$

We cannot compute $P(\bar{A} \cap B)$ from our premises, but we know $P(\bar{A} \cap B) \geq 0$.
Thus

$$\boxed{P(B) \geq p_I + p_A - 1}$$

i.e. we know a lower bound on $P(B)$.
In the limit of certainty, i.e. $p_I = p_A = 1$, we obtain $P(B) = 1$.
I.e. the deductive result is recovered.

Uncertain premises: evaluating $P(B)$, part 2

Reminder:
$A \Rightarrow B = \bar{A} \cup B$.
$P(B) = P(A \cap B) + P(\bar{A} \cap B)$
$P(A \Rightarrow B) = p_I$ and $P(A) = p_A$.
$\boxed{P(B \cap A) = P(B|A)P(A) = p_A + p_I - 1.}$

We cannot compute $P(\bar{A} \cap B)$ from our premises, but we know $P(\bar{A} \cap B) \geq 0$.
Thus

$$\boxed{P(B) \geq p_I + p_A - 1}$$

i.e. we know a lower bound on $P(B)$.
In the limit of certainty, i.e. $p_I = p_A = 1$, we obtain $P(B) = 1$.
I.e. the deductive result is recovered.

Justifying probability: the logical perspective

Goal: extend *propositional* logic to a *probability* logic. Basic ingredients:

- Propositions with truth values: T,F.
- Logical connectives: $\sim, \vee, \wedge$. $\vee$ is not necessary.
- Inference schemes: modus ponens, tollens etc.

We'd like a more fine-grained representation of uncertainty, which should reduce to propositional reasoning in the limit of certainty.

Question: [Cox 1948, Jaynes 2003] what kind of axioms/desiderata are required for this?

Justifying probability: the logical perspective

Goal: extend *propositional* logic to a *probability logic.* Basic ingredients:

- Propositions with truth values: T,F.
- Logical connectives: $\sim, \vee, \wedge$. $\vee$ is not necessary.
- Inference schemes: modus ponens, tollens etc.

We'd like a more fine-grained representation of uncertainty, which should reduce to propositional reasoning in the limit of certainty.

Question: [Cox 1948, Jaynes 2003] what kind of axioms/desiderata are required for this?

The desiderata of a probabilistic logic

D1: Degrees of plausibility are represented by real numbers.

- Not the only possible choice.
- Further convention: greater plausibility corresponds to a greater number.
- Continuity property: an infintesimally greater plausibility corresponds to an infinitesimally greater number.
- Conditional plausibility of A given B: $A|B$.

The desiderata of a probabilistic logic

D1: Degrees of plausibility are represented by real numbers.

- Not the only possible choice.
- Further convention: greater plausibility corresponds to a greater number.
- Continuity property: an infintesimally greater plausibility corresponds to an infinitesimally greater number.
- Conditional plausibility of A given B: $A|B$.

The desiderata of a probabilistic logic

D2: Qualitative correspondence with common sense.

- This desideratum is about reasoning with conjunctions.
- Assume plausibilities $(A|C') > (A|C)$ and $(B|A \wedge C') = (B|A \wedge C)$.
- Then $(A \wedge B|C') > (A \wedge B|C)$, never the other way around!
- Likewise, if $(A|C') = (A|C)$ and $(B|A \wedge C') > (B|A \wedge C)$.
- This does not specify the magnitude of the increase!

The desiderata of a probabilistic logic

D2: Qualitative correspondence with common sense.

- This desideratum is about reasoning with conjunctions.
- Assume plausibilities $(A|C') > (A|C)$ and $(B|A \wedge C') = (B|A \wedge C)$.
- Then $(A \wedge B|C') > (A \wedge B|C)$, never the other way around!
- Likewise, if $(A|C') = (A|C)$ and $(B|A \wedge C') > (B|A \wedge C)$.
- This does not specify the magnitude of the increase!

The desiderata of a probabilistic logic

D3: Consistency.

1. If a conclusion can be reasoned out in more than one way, then every possible way must lead to the same result.
2. Equivalent states of knowledge are assigned the same plausibilities.
3. No available information (relevant to the conclusion) is ignored.

The quantitative rules of probabilistic logic

The desiderata **D1-3** can be shown to result in the following *quantitative* rules for plausible reasoning, where the plausibility is denote by P, and is (of course) a probability:

1. $P(A \wedge B|C) = P(B|A \wedge C)P(A|C) = P(A|B \wedge C)P(B|C)$
2. $P(A|B) + P(\sim A|B) = 1.$

These rules for conjunction and negation are enough to construct all other logical connectives. *Inference* can be done by conditioning the conclusion on the premises.

Summary: probability logic

- Propositions A, B in propositional logic correspond to propositions in probability theory.
- Logical operators correspond to set operations.
- Deductive reasoning can be expressed probabilistically.
- Probability provides answers beyond propositional logic.
- In the limit of certainty, deductive results are recovered.