Applied Probability Theory
Why (Kolmogorov) Axioms?

- If $P$ models an \textit{objectively} observable probability, these axioms are obviously reasonable.

- However, why should an agent obey formal axioms when modeling degrees of (subjective) belief?

- Objective vs. subjective probabilities

- Axioms constrain the set of beliefs an agent can abide.

- Finetti (1931) gave one of the most plausible arguments why subjective beliefs should respect axioms:

  “When using contradictory beliefs, the agent will eventually fail.”
Unconditional Probabilities

- $P(A)$ designates the *unconditioned* or *a priori* probability that $A \subseteq \Omega$ occurs if *no* other additional information is present. For example:

\[ P(\text{cavity}) = 0.1 \]

Note: Here, *cavity* is a proposition.

- A formally different way to state the same would be via a binary random variable *Cavity*:

\[ P(\text{Cavity} = \text{true}) = 0.1 \]

- A priori probabilities are derived from statistical surveys or general rules.
Unconditional Probabilities

• In general a random variable can assume more than two values:

\[
\begin{align*}
P(\text{Weather} = \text{sunny}) &= 0.7 \\
P(\text{Weather} = \text{rainy}) &= 0.2 \\
P(\text{Weather} = \text{cloudy}) &= 0.02 \\
P(\text{Weather} = \text{snowy}) &= 0.08 \\
P(\text{Headache} = \text{true}) &= 0.1
\end{align*}
\]

• \(P(X)\) designates the vector of probabilities for the (ordered) domain of the random variable \(X\):

\[
\begin{align*}
P(\text{Weather}) &= \langle 0.7, 0.2, 0.02, 0.08 \rangle \\
P(\text{Headache}) &= \langle 0.1, 0.9 \rangle
\end{align*}
\]

• Both vectors define the respective probability distributions of the two random variables.
Conditional Probabilities

• New evidence can alter the probability of an event.

• Example: The probability for cavity increases if information about a toothache arises.

• With additional information present, the a priori knowledge must not be used!

• $P(A \mid B)$ designates the conditional or a posteriori probability of $A$ given the sole observation (evidence) $B$.

$$P(\text{cavity} \mid \text{toothache}) = 0.8$$

• For random variables $X$ and $Y$ $P(X \mid Y)$ represents the set of conditional distributions for each possible value of $Y$. 
Conditional Probabilities

- \( P(\text{Weather} \mid \text{Headache}) \) consists of the following table:

<table>
<thead>
<tr>
<th>Weather = sunny</th>
<th>( P(\text{Weather} = \text{sunny} \mid \text{Headache}) )</th>
<th>( P(\text{Weather} = \text{sunny} \mid \neg \text{Headache}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weather = rainy</td>
<td>( P(\text{Weather} = \text{rainy} \mid \text{Headache}) )</td>
<td>( P(\text{Weather} = \text{rainy} \mid \neg \text{Headache}) )</td>
</tr>
<tr>
<td>Weather = cloudy</td>
<td>( P(\text{Weather} = \text{cloudy} \mid \text{Headache}) )</td>
<td>( P(\text{Weather} = \text{cloudy} \mid \neg \text{Headache}) )</td>
</tr>
<tr>
<td>Weather = snowy</td>
<td>( P(\text{Weather} = \text{snowy} \mid \text{Headache}) )</td>
<td>( P(\text{Weather} = \text{snowy} \mid \neg \text{Headache}) )</td>
</tr>
</tbody>
</table>

- Note that we are dealing with two distributions now! Therefore each column sums up to unity!

- Formal definition:

\[
P(A \mid B) = \frac{P(A \land B)}{P(B)} \quad \text{if} \quad P(B) > 0
\]
Conditional Probabilities

\[ P(A \mid B) = \frac{P(A \land B)}{P(B)} \]

- Product Rule: \( P(A \land B) = P(A \mid B) \cdot P(B) \)
- Also: \( P(A \land B) = P(B \mid A) \cdot P(A) \)
- \( A \) and \( B \) are independent iff
  \[ P(A \mid B) = P(A) \quad \text{and} \quad P(B \mid A) = P(B) \]
- Equivalently, iff the following equation holds true:
  \[ P(A \land B) = P(A) \cdot P(B) \]
Interpretation of Conditional Probabilities

Caution! Common misinterpretation:

“\(P(A \mid B) = 0.8\) means, that \(P(A) = 0.8\), given \(B\) holds.”

This statement is wrong due to (at least) two facts:

- \(P(A)\) is *always* the a-priori probability, never the probability of \(A\) given that \(B\) holds!

- \(P(A \mid B) = 0.8\) is only applicable as long as no other evidence except \(B\) is present. If \(C\) becomes known, \(P(A \mid B \land C)\) has to be determined.

In general we have:

\[
P(A \mid B \land C) \neq P(A \mid B)
\]

E. g. \(C \rightarrow A\) might apply.
• Let $X_1, \ldots, X_n$ be random variables over the same frame of discernment $\Omega$ and event algebra $\mathcal{E}$. Then $\vec{X} = (X_1, \ldots, X_n)$ is called a \textit{random vector} with

$$\vec{X}(\omega) = (X_1(\omega), \ldots, X_n(\omega))$$

• Shorthand notation:

$$P(\vec{X} = (x_1, \ldots, x_n)) = P(X_1 = x_1, \ldots, X_n = x_n) = P(x_1, \ldots, x_n)$$

• Definition:

$$P(X_1 = x_1, \ldots, X_n = x_n) = P\left( \big\{ \omega \in \Omega \mid \bigwedge_{i=1}^{n} X_i(\omega) = x_i \big\} \right)$$

$$= P\left( \bigcap_{i=1}^{n} \{X_i = x_i\} \right)$$
Joint Probabilities

- Example: $P(\text{Headache}, \text{Weather})$ is the joint probability distribution of both random variables and consists of the following table:

<table>
<thead>
<tr>
<th>Weather</th>
<th>h $\equiv$ Headache $= \text{true}$</th>
<th>$\neg h \equiv$ Headache $= \text{false}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>sunny</td>
<td>$P(W = \text{sunny} \land h)$</td>
<td>$P(W = \text{sunny} \land \neg h)$</td>
</tr>
<tr>
<td>rainy</td>
<td>$P(W = \text{rainy} \land h)$</td>
<td>$P(W = \text{rainy} \land \neg h)$</td>
</tr>
<tr>
<td>cloudy</td>
<td>$P(W = \text{cloudy} \land h)$</td>
<td>$P(W = \text{cloudy} \land \neg h)$</td>
</tr>
<tr>
<td>snowy</td>
<td>$P(W = \text{snowy} \land h)$</td>
<td>$P(W = \text{snowy} \land \neg h)$</td>
</tr>
</tbody>
</table>

- All table cells sum up to unity.
Calculating with Joint Probabilities

All desired probabilities can be computed from a joint probability distribution.

<table>
<thead>
<tr>
<th></th>
<th>toothache</th>
<th>¬toothache</th>
</tr>
</thead>
<tbody>
<tr>
<td>cavity</td>
<td>0.04</td>
<td>0.06</td>
</tr>
<tr>
<td>¬cavity</td>
<td>0.01</td>
<td>0.89</td>
</tr>
</tbody>
</table>

• Example: \( P(\text{cavity} \lor \text{toothache}) = P(\text{cavity} \land \text{toothache}) + P(\neg\text{cavity} \land \text{toothache}) + P(\text{cavity} \land \neg\text{toothache}) = 0.11 \)

• Marginalizations: \( P(\text{cavity}) = P(\text{cavity} \land \text{toothache}) + P(\text{cavity} \land \neg\text{toothache}) = 0.10 \)

• Conditioning:
\[
P(\text{cavity} \mid \text{toothache}) = \frac{P(\text{cavity} \land \text{toothache})}{P(\text{toothache})} = \frac{0.04}{0.04 + 0.01} = 0.80
\]
Problems

- Easiness of computing all desired probabilities comes at an unaffordable price:
  Given \( n \) random variables with \( k \) possible values each, the joint probability distribution contains \( k^n \) entries which is infeasible in practical applications.

- Hard to handle.

- Hard to estimate.

Therefore:

1. Is there a more dense representation of joint probability distributions?

2. Is there a more efficient way of processing this representation?

- The answer is no for the general case, however, certain dependencies and independencies can be exploited to reduce the number of parameters to a practical size.
Stochastic Independence

- Two events $A$ and $B$ are *stochastically independent* iff
  \[ P(A \land B) = P(A) \cdot P(B) \]
  \[ \iff \]
  \[ P(A \mid B) = P(A) = P(A \mid \overline{B}) \]

- Two random variables $X$ and $Y$ are *stochastically independent* iff
  \[ \forall x \in \text{dom}(X) : \forall y \in \text{dom}(Y) : \quad P(X = x, Y = y) = P(X = x) \cdot P(Y = y) \]
  \[ \iff \]
  \[ \forall x \in \text{dom}(X) : \forall y \in \text{dom}(Y) : \quad P(X = x \mid Y = y) = P(X = x) \]

- Shorthand notation: $P(X, Y) = P(X) \cdot P(Y)$.
  Note the formal difference between $P(A) \in [0, 1]$ and $P(X) \in [0, 1]^{|\text{dom}(X)|}$. 

Rudolf Kruse, Matthias Steinbrecher, Pascal Held
Bayesian Networks
• Let $X$, $Y$ and $Z$ be three random variables. We call $X$ and $Y$ conditionally independent given $Z$, iff the following condition holds:

\[
\forall x \in \text{dom}(X) : \forall y \in \text{dom}(Y) : \forall z \in \text{dom}(Z) : \\
P(X = x, Y = y \mid Z = z) = P(X = x \mid Z = z) \cdot P(Y = y \mid Z = z)
\]

• Shorthand notation: $X \perp\!\!\!\!\!\!\perp P \ Y \mid Z$

• Let $X = \{A_1, \ldots, A_k\}$, $Y = \{B_1, \ldots, B_l\}$ and $Z = \{C_1, \ldots, C_m\}$ be three disjoint sets of random variables. We call $X$ and $Y$ conditionally independent given $Z$, iff

\[
P(X, Y \mid Z) = P(X \mid Z) \cdot P(Y \mid Z) \iff P(X \mid Y, Z) = P(X \mid Z)
\]

• Shorthand notation: $X \perp\!\!\!\!\!\!\perp P \ Y \mid Z$
• The complete condition for $\mathbf{X} \perp_{P} \mathbf{Y} \mid \mathbf{Z}$ would read as follows:

\[
\forall a_1 \in \text{dom}(A_1): \cdots \forall a_k \in \text{dom}(A_k): \\
\forall b_1 \in \text{dom}(B_1): \cdots \forall b_l \in \text{dom}(B_l): \\
\forall c_1 \in \text{dom}(C_1): \cdots \forall c_m \in \text{dom}(C_m): \\
\begin{align*}
P(A_1 = a_1, \ldots, A_k = a_k, B_1 = b_1, \ldots, B_l = b_l \mid C_1 = c_1, \ldots, C_m = c_m) \\
= P(A_1 = a_1, \ldots, A_k = a_k \mid C_1 = c_1, \ldots, C_m = c_m) \\
\cdot P(B_1 = b_1, \ldots, B_l = b_l \mid C_1 = c_1, \ldots, C_m = c_m)
\end{align*}
\]

• Remarks:

1. If $\mathbf{Z} = \emptyset$ we get (unconditional) independence.

2. We do not use curly braces ($\{\}$) for the sets if the context is clear. Likewise, we use $\mathbf{X}$ instead of $\mathbf{X}$ to denote sets.
(Weak) Dependence in the entire dataset: $X$ and $Y$ dependent.
Conditional Independence — Example 1

No Dependence in Group 1: $X$ and $Y$ conditionally independent given Group 1.
Conditional Independence — Example 1

No Dependence in Group 2: $X$ and $Y$ conditionally independent given Group 2.
Conditional Independence — Example 2

- \( \text{dom}(G) = \{ \text{mal, fem} \} \)  
  Geschlecht (gender)

- \( \text{dom}(S) = \{ \text{sm, \overline{sm}} \} \)  
  Raucher (smoker)

- \( \text{dom}(M) = \{ \text{mar, \overline{mar}} \} \)  
  Verheiratet (married)

- \( \text{dom}(P) = \{ \text{preg, \overline{preg}} \} \)  
  Schwanger (pregnant)

<table>
<thead>
<tr>
<th>( p_{\text{GSMP}} )</th>
<th>( G = \text{mal} )</th>
<th>( G = \text{fem} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( S = \text{sm} )</td>
<td>( S = \overline{\text{sm}} )</td>
</tr>
<tr>
<td>( M = \text{mar} )</td>
<td>( P = \text{preg} )</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>( P = \overline{\text{preg}} )</td>
<td>0.04</td>
</tr>
<tr>
<td>( M = \overline{\text{mar}} )</td>
<td>( P = \text{preg} )</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>( P = \overline{\text{preg}} )</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Conditional Independence — Example 2

\[ P(G = \text{fem}) = P(G = \text{mal}) = 0.5 \quad P(P = \text{preg}) = 0.08 \]
\[ P(S = \text{sm}) = 0.25 \quad P(M = \text{mar}) = 0.4 \]

- **Gender** and **Smoker** are not independent:
  \[ P(G = \text{fem} \mid S = \text{sm}) = 0.44 \neq 0.5 = P(G = \text{fem}) \]

- **Gender** and **Marriage** are marginally independent but conditionally dependent given **Pregnancy**:
  \[ P(\text{fem, mar} \mid \overline{\text{preg}}) \approx 0.152 \neq 0.169 \approx P(\text{fem} \mid \overline{\text{preg}}) \cdot P(\text{mar} \mid \overline{\text{preg}}) \]
Bayes Theorem

- Product Rule (for events $A$ and $B$):
  \[
P(A \cap B) = P(A \mid B)P(B) \quad \text{and} \quad P(A \cap B) = P(B \mid A)P(A)
  \]

- Equating the right-hand sides:
  \[
P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}
  \]

- For random variables $X$ and $Y$:
  \[
  \forall x \forall y : \quad P(Y = y \mid X = x) = \frac{P(X = x \mid Y = y)P(Y = y)}{P(X = x)}
  \]

- Generalization concerning background knowledge/evidence $E$:
  \[
P(Y \mid X, E) = \frac{P(X \mid Y, E)P(Y \mid E)}{P(X \mid E)}
  \]
\[ P(\text{toothache} \mid \text{cavity}) = 0.4 \]
\[ P(\text{cavity}) = 0.1 \]
\[ P(\text{cavity} \mid \text{toothache}) = \frac{0.4 \cdot 0.1}{0.05} = 0.8 \]
\[ P(\text{toothache}) = 0.05 \]

Why not estimate \( P(\text{cavity} \mid \text{toothache}) \) right from the start?

- Causal knowledge like \( P(\text{toothache} \mid \text{cavity}) \) is more robust than diagnostic knowledge \( P(\text{cavity} \mid \text{toothache}) \).
- The causality \( P(\text{toothache} \mid \text{cavity}) \) is independent of the a priori probabilities \( P(\text{toothache}) \) and \( P(\text{cavity}) \).
- If \( P(\text{cavity}) \) rose in a caries epidemic, the causality \( P(\text{toothache} \mid \text{cavity}) \) would remain constant whereas both \( P(\text{cavity} \mid \text{toothache}) \) and \( P(\text{toothache}) \) would increase according to \( P(\text{cavity}) \).
- A physician, after having estimated \( P(\text{cavity} \mid \text{toothache}) \), would not know a rule for updating.
Assumption:
We would like to consider the probability of the diagnosis GumDisease as well.

\[
P(\text{toothache} \mid \text{gumdisease}) = 0.7 \\
P(\text{gumdisease}) = 0.02
\]

Which diagnosis is more probable?

If we are interested in relative probabilities only (which may be sufficient for some decisions), \( P(\text{toothache}) \) needs not to be estimated:

\[
\frac{P(C \mid T)}{P(G \mid T)} = \frac{P(T \mid C)P(C)}{P(T)} \cdot \frac{P(T)}{P(T \mid G)P(G)}
\]

\[
= \frac{P(T \mid C)P(C)}{P(T \mid G)P(G)} = \frac{0.4 \cdot 0.1}{0.7 \cdot 0.02} = 28.57
\]
Normalization

If we are interested in the absolute probability of $P(C \mid T)$ but do not know $P(T)$, we may conduct a complete case analysis (according $C'$) and exploit the fact that $P(C \mid T) + P(\neg C \mid T) = 1$.

$$P(C \mid T) = \frac{P(T \mid C)P(C)}{P(T)}$$

$$P(\neg C \mid T) = \frac{P(T \mid \neg C)P(\neg C)}{P(T)}$$

$$1 = P(C \mid T) + P(\neg C \mid T) = \frac{P(T \mid C)P(C)}{P(T)} + \frac{P(T \mid \neg C)P(\neg C)}{P(T)}$$

$$P(T) = P(T \mid C)P(C) + P(T \mid \neg C)P(\neg C)$$
Normalization

- Plugging into the equation for $P(C \mid T)$ yields:

$$
P(C \mid T) = \frac{P(T \mid C)P(C)}{P(T \mid C)P(C) + P(T \mid \neg C)P(\neg C)}
$$

- For general random variables, the equation reads:

$$
P(Y = y \mid X = x) = \frac{P(X = x \mid Y = y)P(Y = y)}{\sum_{\forall y' \in \text{dom}(Y)} P(X = x \mid Y = y')P(Y = y')}
$$

- Note the “loop variable” $y'$. Do not confuse with $y$.  

• The patient complains about a toothache. From this first evidence the dentist infers:

\[ P(\text{cavity} | \text{toothache}) = 0.8 \]

• The dentist palpates the tooth with a metal probe which catches into a fracture:

\[ P(\text{cavity} | \text{fracture}) = 0.95 \]

• Both conclusions might be inferred via Bayes rule. But what does the combined evidence yield? Using Bayes rule further, the dentist might want to determine:

\[
P(\text{cavity} | \text{toothache} \land \text{fracture}) = \frac{P(\text{toothache} \land \text{fracture} | \text{cavity}) \cdot P(\text{cavity})}{P(\text{toothache} \land \text{fracture})}
\]
Problem:
He needs $P(\text{toothache} \land \text{catch} \mid \text{cavity})$, i.e. diagnostics knowledge for all combinations of symptoms in general. Better incorporate evidences step-by-step:

$$P(Y \mid X, E) = \frac{P(X \mid Y, E)P(Y \mid E)}{P(X \mid E)}$$

Abbreviations:
- $C$ — cavity
- $T$ — toothache
- $F$ — fracture

Objective:
Computing $P(C \mid T, F)$ with just causal statements of the form $P(\cdot \mid C)$ and under exploitation of independence relations among the variables.
Multiple Evidences

- A priori: \( P(C) \)

- Evidence toothache: \( P(C \mid T) = P(C) \frac{P(T \mid C)}{P(T)} \)

- Evidence fracture: \( P(C \mid T, F) = P(C \mid T) \frac{P(F \mid C, T)}{P(F \mid T)} \)

\[
T \perp\!\!\!\!\perp F \mid C \iff P(F \mid C, T) = P(F \mid C)
\]

\[
P(C \mid T, F) = P(C) \frac{P(T \mid C)}{P(T)} \frac{P(F \mid C)}{P(F \mid T)}
\]

Seems that we still have to cope with symptom inter-dependencies?!
Compound equation from last slide:

\[
P(C \mid T, F) = P(C) \frac{P(T \mid C) P(F \mid C)}{P(T) P(F \mid T)}
\]

\[
= P(C) \frac{P(T \mid C) P(F \mid C)}{P(F, T)}
\]

- \(P(F, T)\) is a normalizing constant and can be computed if \(P(F \mid \neg C)\) and \(P(T \mid \neg C)\) are known:

\[
P(F, T) = \frac{P(F, T \mid C)}{P(F \mid C) P(T \mid C)} P(C) + \frac{P(F, T \mid \neg C)}{P(F \mid \neg C) P(T \mid \neg C)} P(\neg C)
\]

- Therefore, we finally arrive at the following solution...
### Multiple Evidences

Given a set of evidences $F$ and $T$, the probability of a cause $C$ is given by:

$$P(C | F, T) = \frac{P(C) \cdot P(T | C) \cdot P(F | C)}{P(F | C) \cdot P(T | C) \cdot P(C) + P(F | \neg C) \cdot P(T | \neg C) \cdot P(\neg C)}$$

Note that we only use causal probabilities $P(\cdot | C)$ together with the a priori (marginal) probabilities $P(C)$ and $P(\neg C)$. 
Multiple Evidences — Summary

Multiple evidences can be treated by reduction on

- a priori probabilities
- (causal) conditional probabilities for the evidence
- under assumption of conditional independence

General rule:

\[
P(Z \mid X,Y) = \alpha P(Z) P(X \mid Z) P(Y \mid Z)
\]

for \( X \) and \( Y \) conditionally independent given \( Z \) and with normalizing constant \( \alpha \).
Monty Hall Puzzle

Marylin Vos Savant in her riddle column in the New York Times:

You are a candidate in a game show and have to choose between three doors. Behind one of them is a Porsche, whereas behind the other two there are goats. After you chose a door, the host Monty Hall (who knows what is behind each door) opens another (not your chosen one) door with a goat. Now you have the choice between keeping your chosen door or choose the remaining one.

Which decision yields the best chance of winning the Porsche?
Monty Hall Puzzle

\( G \quad \text{You win the Porsche.} \)
\( R \quad \text{You revise your decision.} \)
\( A \quad \text{Behind your initially chosen door is (and remains) the Porsche.} \)

\[
P(G \mid R) = P(G, A \mid R) + P(G, \bar{A} \mid R)
\]
\[
= P(G \mid A, R)P(A \mid R) + P(G \mid \bar{A}, R)P(\bar{A} \mid R)
\]
\[
= 0 \cdot P(A \mid R) + 1 \cdot P(\bar{A} \mid R)
\]
\[
= P(\bar{A} \mid R) = P(\bar{A}) = \frac{2}{3}
\]

\[
P(G \mid \bar{R}) = P(G, A \mid \bar{R}) + P(G, \bar{A} \mid \bar{R})
\]
\[
= P(G \mid A, \bar{R})P(A \mid \bar{R}) + P(G \mid \bar{A}, \bar{R})P(\bar{A} \mid \bar{R})
\]
\[
= 1 \cdot P(A \mid \bar{R}) + 0 \cdot P(\bar{A} \mid \bar{R})
\]
\[
= P(A \mid \bar{R}) = P(A) = \frac{1}{3}
\]
Simpson’s Paradox

Example:  \( C = \) Patient takes medication, \( E = \) patient recovers

<table>
<thead>
<tr>
<th></th>
<th>( E )</th>
<th>( \neg E )</th>
<th>( \sum )</th>
<th>Recovery rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C )</td>
<td>20</td>
<td>20</td>
<td>40</td>
<td>50%</td>
</tr>
<tr>
<td>( \neg C )</td>
<td>16</td>
<td>24</td>
<td>40</td>
<td>40%</td>
</tr>
<tr>
<td>( \sum )</td>
<td>36</td>
<td>44</td>
<td>80</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Men</th>
<th>( E )</th>
<th>( \neg E )</th>
<th>( \sum )</th>
<th>Rec.rate</th>
<th>Women</th>
<th>( E )</th>
<th>( \neg E )</th>
<th>( \sum )</th>
<th>Rec.rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C )</td>
<td>18</td>
<td>12</td>
<td>30</td>
<td>60%</td>
<td>( C )</td>
<td>2</td>
<td>8</td>
<td>10</td>
<td>20%</td>
</tr>
<tr>
<td>( \neg C )</td>
<td>7</td>
<td>3</td>
<td>10</td>
<td>70%</td>
<td>( \neg C )</td>
<td>9</td>
<td>21</td>
<td>30</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>15</td>
<td>40</td>
<td></td>
<td></td>
<td>11</td>
<td>29</td>
<td>40</td>
<td></td>
</tr>
</tbody>
</table>

\[
P(E \mid C) > P(E \mid \neg C)
\]

but \[
P(E \mid C, M) < P(E \mid \neg C, M)
\]

\[
P(E \mid C, W) < P(E \mid \neg C, W)
\]
Probabilistic Reasoning

- Probabilistic reasoning is difficult and may be problematic:
  - $P(A \land B)$ is not determined simply by $P(A)$ and $P(B)$:
    $P(A) = P(B) = 0.5 \Rightarrow P(A \land B) \in [0, 0.5]
  - $P(C \mid A) = x, P(C \mid B) = y \Rightarrow P(C \mid A \land B) \in [0, 1]
    Probabilistic logic is *not truth functional!*

- Central problem: How does additional information affect the current knowledge? I.e., if $P(B \mid A)$ is known, what can be said about $P(B \mid A \land C)$?

- High complexity: $n$ propositions $\rightarrow 2^n$ full conjunctives

- Hard to specify these probabilities.
Summary

- Uncertainty is inevitable in complex and dynamic scenarios that force agents to cope with ignorance.

- Probabilities express the agent’s inability to vote for a definitive decision. They model the degree of belief.

- If an agent violates the axioms of probability, it may exhibit irrational behavior in certain circumstances.

- The Bayes rule is used to derive unknown probabilities from present knowledge and new evidence.

- Multiple evidences can be effectively included into computations exploiting conditional independencies.