Bayesian Networks and Decision Graphs

Chapter 1

Two perspectives on probability theory

In many domains, the probability of an outcome is interpreted as a relative frequency:

- The probability of getting a three by throwing a six-sided die is $1/6$.

However, we often talk about the probability of an event without being able to specify a frequency for it:

- What is the probability that Denmark wins the world cup in 2010?

Such probabilities are called subjective probabilities

Two perspectives on probability theory

In many domains, the probability of an outcome is interpreted as a relative frequency:

- The probability of getting a three by throwing a six-sided die is $1/6$.

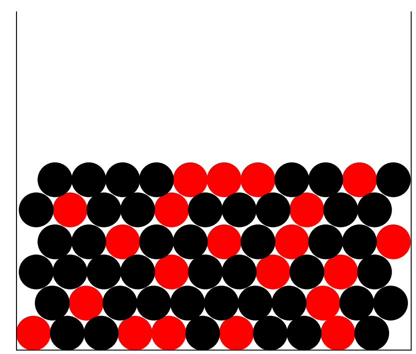
However, we often talk about the probability of an event without being able to specify a frequency for it:

- What is the probability that Denmark wins the world cup in 2010?

Such probabilities are called subjective probabilities

Possible interpretation:

- I receive Dkr 1000 if Denmark wins.
- If I draw a red ball I receive Dkr 1000.



Basic probability axioms

The set of possible outcomes of an “experiment” is called the sample space $\mathcal{S}$:

- Throwing a six sided die: $\{1, 2, 3, 4, 5, 6\}$.
- Will Denmark win the world cup: {yes,no}.
- The values in a deck of cards: $\{2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A\}$.

Basic probability axioms

The set of possible outcomes of an “experiment” is called the sample space $\mathcal{S}$:

- Throwing a six sided die: $\{1, 2, 3, 4, 5, 6\}$.
- Will Denmark win the world cup: $\{$yes,no$\}$.
- The values in a deck of cards: $\{2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A\}$.

An event $\mathcal{E}$ is a subset of the sample space:

- The event that we will get an even number when throwing a die: $\{2, 4, 6\}$.
- The event that Denmark wins: $\{$yes$\}$.
- The event that we will get a 6 or below when drawing a card: $\{2, 3, 4, 5, 6\}$.

Basic probability axioms

The set of possible outcomes of an "experiment" is called the sample space $\mathcal{S}$:

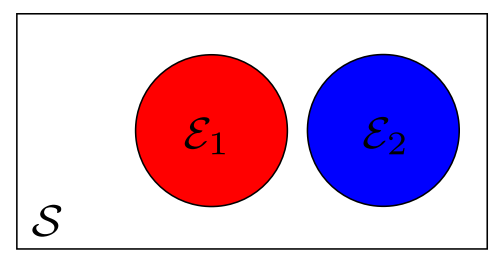
- Throwing a six sided die: $\{1, 2, 3, 4, 5, 6\}$.
- Will Denmark win the world cup: $\{$yes,no$\}$.
- The values in a deck of cards: $\{2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A\}$.

An event $\mathcal{E}$ is a subset of the sample space:

- The event that we will get an even number when throwing a die: $\{2, 4, 6\}$.
- The event that Denmark wins: $\{$yes$\}$.
- The event that we will get a 6 or below when drawing a card: $\{2, 3, 4, 5, 6\}$.

We measure our uncertainty about an experiment by assigning probabilities to each event. The probabilities must obey the following axioms:

- $P(\mathcal{S}) = 1$.
- For all events $\mathcal{E}$ it holds that $P(\mathcal{E}) \geq 0$.
- If $\mathcal{E}_1 \cap \mathcal{E}_2 = \emptyset$, then $P(\mathcal{E}_1 \cup \mathcal{E}_2) = P(\mathcal{E}_1) + P(\mathcal{E}_2)$.



Conditional probabilities

Every probability is conditioned on a context. For example, if we throw a dice:

"$P(\{\text{six}\}) = \frac{1}{6}$" = "$P(\text{six}|\text{symmetric dice}) = \frac{1}{6}$".

In general, if $\mathcal{A}$ and $\mathcal{B}$ are events and $P(\mathcal{A}|\mathcal{B}) = x$, then:

"In the context of $\mathcal{B}$ we have that $P(\mathcal{A}) = x$"

Note: It is not "whenever $\mathcal{B}$ we have $P(\mathcal{A}) = x$", but rather: if $\mathcal{B}$ and everything else known is irrelevant to $\mathcal{A}$, then $P(\mathcal{A}) = x$.

Definition: For two events $\mathcal{A}$ and $\mathcal{B}$ we have:

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

Example:

$$P(\mathcal{A} = \{4\}|\mathcal{B} = \{2,4,6\}) = \frac{P(\mathcal{A} \cap \mathcal{B} = \{4\})}{P(\mathcal{B} = \{2,4,6\})} = \frac{1/6}{3/6} = \frac{1}{3}.$$

Basic probability calculus: the fundamental rule

Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be events.

The fundamental rule: $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A}|\mathcal{B})P(\mathcal{B})$.

The fundamental rule, conditioned: $P(\mathcal{A} \cap \mathcal{B}|\mathcal{C}) = P(\mathcal{A}|\mathcal{B} \cap \mathcal{C})P(\mathcal{B}|\mathcal{C})$.

Proof: Derived directly from the definition of conditional probability.

Basic probability calculus: Bayes' rule

Bayes rule:

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

Proof:

$$P(\mathcal{B}|\mathcal{A})P(\mathcal{A}) = P(\mathcal{B} \cap \mathcal{A}) = P(\mathcal{A}|\mathcal{B})P(\mathcal{B})$$

Bayes rule, conditioned:

$$P(\mathcal{B}|\mathcal{A} \cap \mathcal{C}) = \frac{P(\mathcal{A}|\mathcal{B} \cap \mathcal{C})P(\mathcal{B}|\mathcal{C})}{P(\mathcal{A}|\mathcal{C})}$$

Example: We have two diseases $\mathcal{A}_1$ and $\mathcal{A}_2$ that are the only diseases that can cause the symptoms $\mathcal{B}$. If

- $\mathcal{A}_1$ and $\mathcal{A}_2$ are equally likely ($P(\mathcal{A}_1) = P(\mathcal{A}_2)$)
- $P(\mathcal{B}|\mathcal{A}_1) = 0.9$
- $P(\mathcal{B}|\mathcal{A}_2) = 0.3$

what are then the probabilities $P(\mathcal{A}_1|\mathcal{B})$ and $P(\mathcal{A}_2|\mathcal{B})$?

Basic probability calculus

Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be events.

Conditional probability: $P(\mathcal{A}|\mathcal{B}) = \frac{P(\mathcal{A} \cap \mathcal{B})}{P(\mathcal{B})}$

The fundamental rule: $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A}|\mathcal{B})P(\mathcal{B})$.

The conditional fundamental rule: $P(\mathcal{A} \cap \mathcal{B}|\mathcal{C}) = P(\mathcal{A}|\mathcal{B} \cap \mathcal{C})P(\mathcal{B}|\mathcal{C})$.

Basic probability calculus

Let $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ be events.

Conditional probability: $P(\mathcal{A}|\mathcal{B}) = \frac{P(\mathcal{A}\cap\mathcal{B})}{P(\mathcal{B})}$

The fundamental rule: $P(\mathcal{A} \cap \mathcal{B}) = P(\mathcal{A}|\mathcal{B})P(\mathcal{B})$.

The conditional fundamental rule: $P(\mathcal{A} \cap \mathcal{B}|\mathcal{C}) = P(\mathcal{A}|\mathcal{B} \cap \mathcal{C})P(\mathcal{B}|\mathcal{C})$.

Bayes rule: $P(\mathcal{B}|\mathcal{A}) = \frac{P(\mathcal{A}|\mathcal{B})P(\mathcal{B})}{P(\mathcal{A})}$.

Proof: $P(\mathcal{B}|\mathcal{A})P(\mathcal{A}) = P(\mathcal{B} \cap \mathcal{A}) = P(\mathcal{A}|\mathcal{B})P(\mathcal{B})$.

Bayes rule, conditioned: $P(\mathcal{B}|\mathcal{A} \cap \mathcal{C}) = \frac{P(\mathcal{A}|\mathcal{B}\cap\mathcal{C})P(\mathcal{B}|\mathcal{C})}{P(\mathcal{A}|\mathcal{C})}$.

Conditional independence: If $P(\mathcal{A}|\mathcal{B} \cap \mathcal{C}) = P(\mathcal{A}|\mathcal{C})$ then $P(\mathcal{A} \cap \mathcal{B}|\mathcal{C}) = P(\mathcal{A}|\mathcal{C}) \cdot P(\mathcal{B}|\mathcal{C})$.

Probability calculus for variables

A is a variable with states $a_1, \ldots, a_n$; B is a variable with states $b_1, \ldots, b_m$.

$P(A) = (x_1, \ldots, x_n)$ is a probability distribution ; $x_i \geq 0$;$\sum_{i=1}^n x_i = 1$ ($\sum_A P(A) = 1$).

$P(A|B)$ is a $n \times m$ table containing the numbers $P(a_i|b_j)$.

Note: $\sum_A P(A|b_j) = 1$ for all b_j.

		B		
		b_1	b_2	b_3
A	a_1	0.4	0.3	0.6
	a_2	0.6	0.7	0.4

$P(A, B)$ is a $n \times m$ table too; $\sum_{A,B} P(A, B) = 1$.

		B		
		b_1	b_2	b_3
A	a_1	0.16	0.12	0.12
	a_2	0.24	0.28	0.08

The fundamental rule for variables

$P(A|B)P(B)$: $n \times m$ multiplications $P(a_i|b_j)P(b_j) = P(a_i, b_j)$

	b_1	b_2	b_3
a_1	0.4	0.3	0.6
a_2	0.6	0.7	0.4

$P(A|B)$

b_1	b_2	b_3
0.4	0.4	0.2

$P(B)$

=

	b_1	b_2	b_3
a_1	0.16	0.12	0.12
a_2	0.24	0.28	0.08

$P(A, B)$

The fundamental rule for variables

$P(A|B)P(B)$: $n \times m$ multiplications $P(a_i|b_j)P(b_j) = P(a_i, b_j)$

	b_1	b_2	b_3
a_1	0.4	0.3	0.6
a_2	0.6	0.7	0.4

$P(A|B)$

b_1	b_2	b_3
0.4	0.4	0.2

$P(B)$

=

	b_1	b_2	b_3
a_1	0.16	0.12	0.12
a_2	0.24	0.28	0.08

$P(A, B)$

A is independent of B given C if $P(A|B, C) = P(A|C)$.

	b_1	b_2	b_3
c_1	$(0.4, 0.6)$	$(0.4, 0.6)$	$(0.4, 0.6)$
c_2	$(0.7, 0.3)$	$(0.7, 0.3)$	$(0.7, 0.3)$

$P(A|B, C)$

=

	a_1	a_2
c_1	0.4	0.6
c_2	0.7	0.3

$P(A|C)$

Marginalization

We have $P(A, B)$ and we need $P(A)$.

	b_1	b_2	b_3		
a_1	0.16	0.12	0.12	$\rightarrow$	0.4
a_2	0.24	0.28	0.08	$\rightarrow$	0.6

B is marginalized out of $P(A, B)$:

$$
\begin{aligned}
"A = a_1" &= ("A = a_1" \wedge "B = b_1") \vee ("A = a_1" \wedge "B = b_2") \vee ("A = a_1" \wedge "B = b_3") \\
&= 0.16 + 0.12 + 0.12 = 0.4 \\
"A = a_2" &= ("A = a_2" \wedge "B = b_1") \vee ("A = a_2" \wedge "B = b_2") \vee ("A = a_2" \wedge "B = b_3") \\
&= 0.24 + 0.28 + 0.08 = 0.6
\end{aligned}
$$

Notation: $P(A) = \sum_B P(A, B)$

Notation

A potential ϕ is a table of real numbers over a set of variables, dom(ϕ).

A table of probabilities is a probability potential.

Notation

A potential ϕ is a table of real numbers over a set of variables, dom(ϕ).

A table of probabilities is a probability potential.

Multiplication

	b_1	b_2
a_1	2	1
a_2	3	4

	b_1	b_2
a_1	1	2
a_2	5	6

=

	b_1	b_2
a_1		
a_2		

Notation

A potential ϕ is a table of real numbers over a set of variables, dom(ϕ).

A table of probabilities is a probability potential.

Multiplication

	b_1	b_2
a_1	2	1
a_2	3	4

	b_1	b_2
a_1	1	2
a_2	5	6

=

	b_1	b_2
a_1	2	
a_2		

Notation

A potential ϕ is a table of real numbers over a set of variables, dom(ϕ).

A table of probabilities is a probability potential.

Multiplication

	b_1	b_2
a_1	2	1
a_2	3	4

	b_1	b_2
a_1	1	2
a_2	5	6

=

	b_1	b_2
a_1	2	2
a_2		

Notation

A potential ϕ is a table of real numbers over a set of variables, dom(ϕ).

A table of probabilities is a probability potential.

Multiplication

	b_1	b_2
a_1	2	1
a_2	3	4

	b_1	b_2
a_1	1	2
a_2	5	6

=

	b_1	b_2
a_1	2	2
a_2	15	

Notation

A potential ϕ is a table of real numbers over a set of variables, dom(ϕ).

A table of probabilities is a probability potential.

Multiplication

	b_1	b_2
a_1	2	1
a_2	3	4

	b_1	b_2
a_1	1	2
a_2	5	6

=

	b_1	b_2
a_1	2	2
a_2	15	24

Multiplication of potentials

	b_1	b_2
a_1	1	3
a_2	4	5

	b_1	b_2
c_1	6	7
c_2	8	9

=

	b_1	b_2
a_1	$(\_, \_)$	$(\_, \_)$
a_2	$(\_, \_)$	$(\_, \_)$

$\phi_1(A, B)$

$\phi_2(C, B)$

$\phi_3(A, B, C) = \phi_1(A, B) \cdot \phi_2(C, B)$

Multiplication of potentials

	b_1	b_2
a_1	1	3
a_2	4	5

	b_1	b_2
c_1	6	7
c_2	8	9

=

	b_1	b_2
a_1	$(6_{c_1}, 8_{c_2})$	$(\_, \_)$
a_2	$(\_, \_)$	$(\_, \_)$

$\phi_1(A, B)$ $\phi_2(C, B)$ $\phi_3(A, B, C) = \phi_1(A, B) \cdot \phi_2(C, B)$

Multiplication of potentials

	b_1	b_2
a_1	1	3
a_2	4	5

	b_1	b_2
c_1	6	7
c_2	8	9

=

	b_1	b_2
a_1	$(6_{c_1}, 8_{c_2})$	$(21_{c_1}, 27_{c_2})$
a_2	$(\_, \_)$	$(\_, \_)$

$\phi_1(A, B)$ $\phi_2(C, B)$ $\phi_3(A, B, C) = \phi_1(A, B) \cdot \phi_2(C, B)$

Multiplication of potentials

	b_1	b_2
a_1	1	3
a_2	4	5

$\phi_1(A, B)$

	b_1	b_2
c_1	6	7
c_2	8	9

$\phi_2(C, B)$

=

	b_1	b_2
a_1	$(6_{c_1}, 8_{c_2})$	$(21_{c_1}, 27_{c_2})$
a_2	$(24_{c_1}, 32_{c_2})$	$(\_, \_)$

$\phi_3(A, B, C) = \phi_1(A, B) \cdot \phi_2(C, B)$

Multiplication of potentials

	b_1	b_2
a_1	1	3
a_2	4	5

	b_1	b_2
c_1	6	7
c_2	8	9

=

	b_1	b_2
a_1	$(6_{c_1}, 8_{c_2})$	$(21_{c_1}, 27_{c_2})$
a_2	$(24_{c_1}, 32_{c_2})$	$(35_{c_1}, 45_{c_2})$

$\phi_1(A, B)$ $\qquad$ $\phi_2(C, B)$ $\qquad$ $\phi_3(A, B, C) = \phi_1(A, B) \cdot \phi_2(C, B)$

Marginalization of potentials

$$\sum_{B} \left(\begin{array}{c|cc} & b_1 & b_2 \\ \hline a_1 & 2 & 3 \\ a_2 & 1 & 4 \end{array} \right) = \begin{array}{c|c} a_1 & - \\ a_2 & - \end{array}$$

Marginalization of potentials

$$\sum_{B} \left(\begin{array}{c|cc} & b_1 & b_2 \\ \hline a_1 & 2 & 3 \\ a_2 & 1 & 4 \end{array} \right) = \begin{array}{c|c} a_1 & 5 \\ a_2 & 5 \end{array}$$

Marginalization of potentials

$$\sum_{B} \left(\begin{array}{c|cc} & b_1 & b_2 \\ \hline a_1 & 2 & 3 \\ a_2 & 1 & 4 \end{array} \right) = \begin{array}{c|c} a_1 & 5 \\ a_2 & 5 \end{array}$$

$$\sum_{A} \left(\begin{array}{c|cc} & b_1 & b_2 \\ \hline a_1 & 2 & 3 \\ a_2 & 1 & 4 \end{array} \right) = \begin{array}{c|c} b_1 & - \\ b_2 & - \end{array}$$

Marginalization of potentials

$$\sum_B \left(\begin{array}{c|cc} & b_1 & b_2 \\ \hline a_1 & 2 & 3 \\ a_2 & 1 & 4 \end{array} \right) = \begin{array}{c|c} a_1 & 5 \\ a_2 & 5 \end{array}$$

$$\sum_A \left(\begin{array}{c|cc} & b_1 & b_2 \\ \hline a_1 & 2 & 3 \\ a_2 & 1 & 4 \end{array} \right) = \begin{array}{c|c} b_1 & 3 \\ b_2 & 7 \end{array}$$