Chapter 3: Probability Calculus

Adnan Darwiche[1]

September 25, 2009

[1] Lecture slides for *Modeling and Reasoning with Bayesian Networks*, Adnan Darwiche, Cambridge University Press, 2009.

Degrees of Belief

- Assign a degree of belief or probability in $[0, 1]$ to each world ω and denote it by $\Pr(\omega)$.
- The belief in, or probability of, a sentence α:

$$\Pr(\alpha) \stackrel{def}{=} \sum_{\omega \models \alpha} \Pr(\omega).$$

world	Earthquake	Burglary	Alarm	$\Pr(.)$
ω_1	true	true	true	.0190
ω_2	true	true	false	.0010
ω_3	true	false	true	.0560
ω_4	true	false	false	.0240
ω_5	false	true	true	.1620
ω_6	false	true	false	.0180
ω_7	false	false	true	.0072
ω_8	false	false	false	.7128

Degrees of Belief

world	Earthquake	Burglary	Alarm	$\Pr(.)$
ω_1	true	true	true	.0190
ω_2	true	true	false	.0010
ω_3	true	false	true	.0560
ω_4	true	false	false	.0240
ω_5	false	true	true	.1620
ω_6	false	true	false	.0180
ω_7	false	false	true	.0072
ω_8	false	false	false	.7128

$$\begin{aligned}
\Pr(\text{Earthquake}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_3) + \Pr(\omega_4) = .1 \\
\Pr(\text{Burglary}) &= .2 \\
\Pr(\neg\text{Burglary}) &= .8 \\
\Pr(\text{Alarm}) &= .2442
\end{aligned}$$

Properties of Beliefs

- A bound on the belief in any sentence:

$$0 \leq \Pr(\alpha) \leq 1 \qquad \text{for any sentence } \alpha.$$

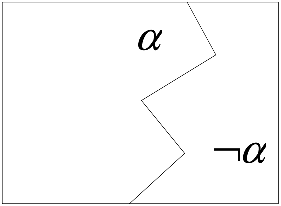
- A baseline for inconsistent sentences:

$$\Pr(\alpha) = 0 \qquad \text{when } \alpha \text{ is inconsistent.}$$

- A baseline for valid sentences:

$$\Pr(\alpha) = 1 \qquad \text{when } \alpha \text{ is valid.}$$

Properties of Beliefs



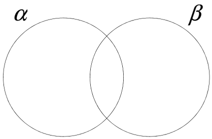
- The belief in a sentence given the belief in its negation:

$$\Pr(\alpha) + \Pr(\neg\alpha) = 1.$$

Example

$$\begin{aligned}
\Pr(\text{Burglary}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_5) + \Pr(\omega_6) = .2 \\
\Pr(\neg\text{Burglary}) &= \Pr(\omega_3) + \Pr(\omega_4) + \Pr(\omega_7) + \Pr(\omega_8) = .8
\end{aligned}$$

Properties of Beliefs



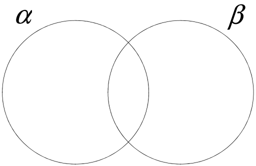
- The belief in a disjunction:

$$\Pr(\alpha \vee \beta) = \Pr(\alpha) + \Pr(\beta) - \Pr(\alpha \wedge \beta).$$

- Example:

$$\begin{aligned}
\Pr(\text{Earthquake}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_3) + \Pr(\omega_4) = .1 \\
\Pr(\text{Burglary}) &= \Pr(\omega_1) + \Pr(\omega_2) + \Pr(\omega_5) + \Pr(\omega_6) = .2 \\
\Pr(\text{Earthquake} \wedge \text{Burglary}) &= \Pr(\omega_1) + \Pr(\omega_2) = .02 \\
\Pr(\text{Earthquake} \vee \text{Burglary}) &= .1 + .2 - .02 = .28
\end{aligned}$$

Properties of Beliefs



- The belief in a disjunction:

$$\Pr(\alpha \vee \beta) = \Pr(\alpha) + \Pr(\beta) \quad \text{when } \alpha \text{ and } \beta \text{ are mutually exclusive.}$$

Quantifying Uncertainty

	Earthquake	Burglary	Alarm
true	.1	.2	.2442
false	.9	.8	.7558

Most certain about whether an earthquake has occurred and least certain about whether an alarm has triggered.

Entropy

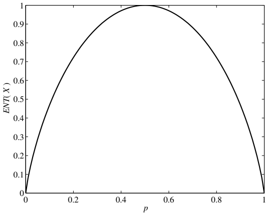
Quantify uncertainty about a variable X using the notion of entropy:

$$\mathrm{ENT}(X) \stackrel{def}{=} -\sum_x \Pr(x) \log_2 \Pr(x),$$

where $0 \log 0 = 0$ by convention.

	Earthquake	Burglary	Alarm
true	.1	.2	.2442
false	.9	.8	.7558
ENT(.)	.469	.722	.802

Entropy



- The entropy for a binary variable X and varying $p = \Pr(X)$.
- Entropy is non-negative.
- When $p = 0$ or $p = 1$, the entropy of X is zero and at a minimum, indicating no uncertainty about the value of X.
- When $p = \frac{1}{2}$, we have $\Pr(X) = \Pr(\neg X)$ and the entropy is at a maximum (indicating complete uncertainty).

Updating Beliefs

- Evidence will be represented by an arbitrary event, say β.
- Our goal is to update the state of belief $\Pr(.)$ into a new state of belief, which we will denote by $\Pr(.|\beta)$.

Updating Beliefs

- We must have $\Pr(\beta|\beta) = 1$ and $\Pr(\neg\beta|\beta) = 0$.

Updating Beliefs

- We must have $\Pr(\beta|\beta) = 1$ and $\Pr(\neg\beta|\beta) = 0$.
- Hence,

$$\sum_{\omega \models \beta} \Pr(\omega|\beta) = 1.$$

and

$$\Pr(\omega|\beta) = 0 \qquad \text{for all } \omega \models \neg\beta.$$

Updating Beliefs

- We must have $\Pr(\beta|\beta) = 1$ and $\Pr(\neg\beta|\beta) = 0$.
- Hence,

$$\sum_{\omega \models \beta} \Pr(\omega|\beta) = 1.$$

and

$$\Pr(\omega|\beta) = 0 \qquad \text{for all } \omega \models \neg\beta.$$

- Zero probability worlds continue to have zero probability:

$$\Pr(\omega|\beta) = 0 \qquad \text{for all } \omega \text{ where } \Pr(\omega) = 0.$$

Updating Beliefs

- We must have $\Pr(\beta|\beta) = 1$ and $\Pr(\neg\beta|\beta) = 0$.
- Hence,

$$\sum_{\omega \models \beta} \Pr(\omega|\beta) = 1.$$

  and

$$\Pr(\omega|\beta) = 0 \qquad \text{for all } \omega \models \neg\beta.$$

- Zero probability worlds continue to have zero probability:

$$\Pr(\omega|\beta) = 0 \qquad \text{for all } \omega \text{ where } \Pr(\omega) = 0.$$

- Relative beliefs in positive probability worlds stay the same:

$$\frac{\Pr(\omega)}{\Pr(\omega')} = \frac{\Pr(\omega|\beta)}{\Pr(\omega'|\beta)} \qquad \text{for all } \omega, \omega' \models \beta, \Pr(\omega) > 0, \Pr(\omega') > 0.$$

Updating Beliefs

- Previous constraints lead to:

$$\Pr(\omega|\beta) = \frac{\Pr(\omega)}{\Pr(\beta)} \qquad \text{for all } \omega \models \beta.$$

- Our new state of belief is now completely defined:

$$\Pr(\omega|\beta) \stackrel{\mathit{def}}{=} \begin{cases} 0, & \text{if } \omega \models \neg\beta; \\ \dfrac{\Pr(\omega)}{\Pr(\beta)}, & \text{if } \omega \models \beta. \end{cases}$$

- $\Pr(.|\beta)$ is the result of conditioning $\Pr$ on evidence β.

Bayes Conditioning

world	Earthquake	Burglary	Alarm	$\Pr(.)$	$\Pr(.\|\text{Alarm})$
ω_1	true	true	true	.0190	.0190/.2442
ω_2	true	true	false	.0010	0
ω_3	true	false	true	.0560	.0560/.2442
ω_4	true	false	false	.0240	0
ω_5	false	true	true	.1620	.1620/.2442
ω_6	false	true	false	.0180	0
ω_7	false	false	true	.0072	.0072/.2442
ω_8	false	false	false	.7128	0

Example

Our belief in Burglary increases:

$$\begin{aligned}\Pr(\text{Burglary}) &= .2\\ \Pr(\text{Burglary}|\text{Alarm}) &\approx .741 \uparrow\end{aligned}$$

And so does our belief in Earthquake:

$$\begin{aligned}\Pr(\text{Earthquake}) &= .1\\ \Pr(\text{Earthquake}|\text{Alarm}) &\approx .307 \uparrow\end{aligned}$$

Bayes Conditioning

Closed form for Bayes conditioning:

$$\Pr(\alpha|\beta) = \frac{\Pr(\alpha \wedge \beta)}{\Pr(\beta)}.$$

Defined only when $\Pr(\beta) \neq 0$.

Belief Change

Conditioning on evidence Earthquake:

$$
\begin{aligned}
\Pr(\text{Burglary}) &= .2 \\
\Pr(\text{Burglary}|\text{Earthquake}) &= .2 \\
\\
\Pr(\text{Alarm}) &= .2442 \\
\Pr(\text{Alarm}|\text{Earthquake}) &\approx .75 \uparrow
\end{aligned}
$$

The belief in Burglary is not changed, but the belief in Alarm increases.

Belief Change

Conditioning on evidence Burglary:

$$
\begin{aligned}
\Pr(\text{Alarm}) &= .2442 \\
\Pr(\text{Alarm}|\text{Burglary}) &\approx .905 \uparrow \\
\\
\Pr(\text{Earthquake}) &= .1 \\
\Pr(\text{Earthquake}|\text{Burglary}) &= .1
\end{aligned}
$$

The belief in Alarm increases in this case, but the belief in Earthquake stays the same.

Belief Change

- The above belief dynamics are a property of the given state of belief in and may not hold for other states of beliefs.
- It is possible to conceive of a reasonable state of belief in which information about Earthquake would change the belief about Burglary and vice versa.
- One of the central questions in building automated reasoning systems is that of synthesizing states of beliefs that are faithful, i.e., those that correspond to the beliefs held by some human expert.
- The Bayesian network, which we shall introduce in the following chapter, can be viewed as a modeling tool for synthesizing faithful states of beliefs.

Belief Change

The belief in Burglary increases when accepting the evidence Alarm. How would such a belief change further upon obtaining more evidence?

- Confirming that an Earthquake took place:

$$\begin{aligned} \Pr(\text{Burglary}|\text{Alarm}) &\approx .741 \\ \Pr(\text{Burglary}|\text{Alarm} \wedge \text{Earthquake}) &\approx .253 \downarrow \end{aligned}$$

  We now have an explanation of Alarm.

- Confirming that there was no Earthquake:

$$\begin{aligned} \Pr(\text{Burglary}|\text{Alarm}) &\approx .741 \\ \Pr(\text{Burglary}|\text{Alarm} \wedge \neg\text{Earthquake}) &\approx .957 \uparrow \end{aligned}$$

  New evidence will further establish burglary as an explanation.

Conditional Entropy

To quantify the average uncertainty about the value of X after observing the value of Y.

Conditional entropy of a variable X given another variable Y

$$\mathrm{ENT}(X|Y) \stackrel{def}{=} \sum_{y} \Pr(y)\mathrm{ENT}(X|y),$$

where

$$\mathrm{ENT}(X|y) \stackrel{def}{=} -\sum_{x} \Pr(x|y) \log_2 \Pr(x|y).$$

- Entropy never increases after conditioning:

$$\mathrm{ENT}(X|Y) \leq \mathrm{ENT}(X).$$

- Observing the value of Y reduces our uncertainty about X.
- For a particular value y, we may have $\mathrm{ENT}(X|y) > \mathrm{ENT}(X)$.

Conditional Entropy

	Burglary	Burglary$\mid$Alarm$=$true	Burglary$\mid$Alarm$=$false
true	.2	.741	.025
false	.8	.259	.975
$\text{ENT}(.)$	.722	.825	.169

The conditional entropy of Burglary given Alarm is then:

$$
\begin{aligned}
&\text{ENT}(\mathsf{Burglary}|\mathsf{Alarm}) \\
&\quad = \text{ENT}(\mathsf{Burglary}|\mathsf{Alarm}=\mathsf{true})\Pr(\mathsf{Alarm}=\mathsf{true}) + \\
&\quad \text{ENT}(\mathsf{Burglary}|\mathsf{Alarm}=\mathsf{false})\Pr(\mathsf{Alarm}=\mathsf{false}) \\
&\quad = .329,
\end{aligned}
$$

indicating a decrease in the uncertainty about variable Burglary.

Independence

Example

$$\begin{aligned} \Pr(\text{Earthquake}) &= .1 \\ \Pr(\text{Earthquake}|\text{Burglary}) &= .1 \end{aligned}$$

The state of belief $\Pr$ finds the Earthquake event independent of the Burglary event.

$\Pr$ finds event α independent of event β iff

$$\Pr(\alpha|\beta) = \Pr(\alpha) \quad \text{or } \Pr(\beta) = 0.$$

Independence is Symmetric

Example

$$\begin{aligned}
\Pr(\text{Earthquake}) &= .1 \\
\Pr(\text{Earthquake}|\text{Burglary}) &= .1 \\
\\
\Pr(\text{Burglary}) &= .2 \\
\Pr(\text{Burglary}|\text{Earthquake}) &= .2
\end{aligned}$$

A general property: $\Pr$ must find event α independent of event β if it also finds β independent of α.

Independence is Dynamic

Events are independent

$$\begin{aligned} \Pr(\text{Burglary}) &= .2 \\ \Pr(\text{Burglary}|\text{Earthquake}) &= .2 \end{aligned}$$

Events no longer independent given Alarm

$$\begin{aligned} \Pr(\text{Burglary}|\text{Alarm}) &\approx .741 \\ \Pr(\text{Burglary}|\text{Alarm} \wedge \text{Earthquake}) &\approx .253 \end{aligned}$$

Conditional Independence

world	Temp	Sensor1	Sensor2	$\Pr(.)$
ω_1	normal	normal	normal	.576
ω_2	normal	normal	extreme	.144
ω_3	normal	extreme	normal	.064
ω_4	normal	extreme	extreme	.016
ω_5	extreme	normal	normal	.008
ω_6	extreme	normal	extreme	.032
ω_7	extreme	extreme	normal	.032
ω_8	extreme	extreme	extreme	.128

Conditional Independence

Example

$$
\begin{aligned}
\Pr(\text{Temp}=\text{normal}) &= .80 \\
\Pr(\text{Sensor1}=\text{normal}) &= .76 \\
\Pr(\text{Sensor2}=\text{normal}) &= .68 \\
\\
\Pr(\text{Sensor2}=\text{normal}|\text{Sensor1}=\text{normal}) &\approx .768 \uparrow \\
\Pr(\text{Sensor2}=\text{normal}|\text{Temp}=\text{normal}) &= .80 \\
\Pr(\text{Sensor2}=\text{normal}|\text{Temp}=\text{normal}, \text{Sensor1}=\text{normal}) &= .80
\end{aligned}
$$

Even though the sensor readings were initially dependent, they become independent once we know the state of temperature.

Conditional Independence

$\Pr$ finds α conditionally independent of β given γ iff

$$\Pr(\alpha|\beta \wedge \gamma) = \Pr(\alpha|\gamma) \quad \text{or } \Pr(\beta \wedge \gamma) = 0.$$

Another definition

$$\Pr(\alpha \wedge \beta|\gamma) = \Pr(\alpha|\gamma)\Pr(\beta|\gamma) \quad \text{or } \Pr(\gamma) = 0.$$

Variable Independence

$\Pr$ finds $\mathbf{X}$ independent of $\mathbf{Y}$ given $\mathbf{Z}$, denoted $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$, means that $\Pr$ finds $\mathbf{x}$ independent of $\mathbf{y}$ given $\mathbf{z}$ for all instantiations $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$.

Example

$\mathbf{X} = \{A, B\}$, $\mathbf{Y} = \{C\}$ and $\mathbf{Z} = \{D, E\}$, where A, B, C, D and E are all propositional variables. The statement $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ is then a compact notation for a number of statements about independence:

$$
\begin{array}{l}
A \wedge B \text{ is independent of } C \text{ given } D \wedge E; \\
A \wedge \neg B \text{ is independent of } C \text{ given } D \wedge E; \\
\vdots \\
\neg A \wedge \neg B \text{ is independent of } \neg C \text{ given } \neg D \wedge \neg E;
\end{array}
$$

That is, $I_{\Pr}(\mathbf{X}, \mathbf{Z}, \mathbf{Y})$ is a compact notation for $4 \times 2 \times 4 = 32$ independence statements of the above form.

Mutual Information

Mutual information quantifies the impact of observing one variable on the uncertainty in another:

$$\mathrm{MI}(X;Y) \stackrel{def}{=} \sum_{x,y} \Pr(x,y) \log_2 \frac{\Pr(x,y)}{\Pr(x)\Pr(y)}.$$

- Mutual information is non-negative.
- Equal to zero iff variables X and Y are independent.
- Measures the extent to which observing one variable will reduce the uncertainty in another:

$$\begin{aligned} \mathrm{MI}(X;Y) &= \mathrm{ENT}(X) - \mathrm{ENT}(X|Y) \\ &= \mathrm{ENT}(Y) - \mathrm{ENT}(Y|X). \end{aligned}$$

Mutual Information

Conditional mutual information can be defined as follows

$$\mathrm{MI}(X;Y|Z) \stackrel{def}{=} \sum_{x,y,z} \mathrm{Pr}(x,y,z) \log_2 \frac{\mathrm{Pr}(x,y|z)}{\mathrm{Pr}(x|z)\mathrm{Pr}(y|z)},$$

leading to

$$\begin{aligned} \mathrm{MI}(X;Y|Z) &= \mathrm{ENT}(X|Z) - \mathrm{ENT}(X|Y,Z) \\ &= \mathrm{ENT}(Y|Z) - \mathrm{ENT}(Y|X,Z). \end{aligned}$$

Entropy and Mutual Information on Sets of Variables

Entropy and mutual information can be extended to sets of variables in the obvious way.

Entropy can be generalized to a set of variables $\mathbf{X}$ as follows

$$\mathrm{ENT}(\mathbf{X}) = -\sum_{\mathbf{x}} \Pr(\mathbf{x}) \log_2 \Pr(\mathbf{x}).$$

Further Properties of Beliefs

Chain rule

$$
\begin{aligned}
& \Pr(\alpha_1 \wedge \alpha_2 \wedge \ldots \wedge \alpha_n) \\
= \quad & \Pr(\alpha_1 | \alpha_2 \wedge \ldots \wedge \alpha_n) \Pr(\alpha_2 | \alpha_3 \wedge \ldots \wedge \alpha_n) \ldots \Pr(\alpha_n).
\end{aligned}
$$

Case analysis (law of total probability)

$$\Pr(\alpha) = \sum_{i=1}^{n} \Pr(\alpha \wedge \beta_i),$$

where the events $\beta_1, \ldots, \beta_n$ are mutually exclusive and exhaustive.

Further Properties of Beliefs

Another version of case analysis

$$\Pr(\alpha) = \sum_{i=1}^{n} \Pr(\alpha|\beta_i)\Pr(\beta_i),$$

where the events $\beta_1, \ldots, \beta_n$ are mutually exclusive and exhaustive.

Two simple and useful forms of case analysis are these:

$$\begin{aligned} \Pr(\alpha) &= \Pr(\alpha \wedge \beta) + \Pr(\alpha \wedge \neg\beta) \\ \Pr(\alpha) &= \Pr(\alpha|\beta)\Pr(\beta) + \Pr(\alpha|\neg\beta)\Pr(\neg\beta). \end{aligned}$$

The main value of case analysis is that, in many situations, computing our beliefs in the cases is easier than computing our beliefs in α. We shall see many examples of this phenomena in later chapters.

Further Properties of Beliefs

Bayes rule

$$\Pr(\alpha|\beta) = \frac{\Pr(\beta|\alpha)\Pr(\alpha)}{\Pr(\beta)}.$$

- Classical usage: α is perceived to be a cause of β.
- Example: α is a disease and β is a symptom–
- Assess our belief in the cause given the effect.
- Belief in an effect given its cause, $\Pr(\beta|\alpha)$, is usually more readily available than the belief in a cause given one of its effects, $\Pr(\alpha|\beta)$.

Bayes Rule

Example

A patient who was just tested for a particular disease and the test came out positive. We know that one in every thousand people has this disease. We also know that the test is not reliable: it has a false positive rate of 2% and a false negative rate of 5%. Our goal is then to assess our belief in the patient having the disease given that the test came out positive.

If we let the propositional variable D stand for "the patient has the disease," and the propositional variable T stand for "the test came out positive," our goal is then to compute $\Pr(D|T)$.

Bayes Rule

- From the given information, we know that

$$\Pr(D) = \frac{1}{1000}$$

since one in every thousand has the disease.

- Since the false positive rate of the test is 2%, we know that

$$\Pr(T|\neg D) = \frac{2}{100} \text{ and } \Pr(\neg T|\neg D) = \frac{98}{100}.$$

- Similarly, since the false negative rate of the test is 5%, we know that

$$\Pr(\neg T|D) = \frac{5}{100} \text{ and } \Pr(T|D) = \frac{95}{100}.$$

- Using Bayes rule, we now have

$$\Pr(D|T) = \frac{\frac{95}{100} \times \frac{1}{1000}}{\Pr(T)}.$$

- Using case analysis:

$$\begin{aligned} \Pr(T) &= \Pr(T|D)\Pr(D) + \Pr(T|\neg D)\Pr(\neg D) \\ &= \frac{95}{100} \times \frac{1}{1000} + \frac{2}{100} \times \frac{999}{1000} = \frac{2093}{100000}, \end{aligned}$$

which leads to

$$\Pr(D|T) = \frac{95}{2093} \approx 4.5\%.$$

Bayes Rule

Another way to solve the above problem is to construct the state of belief completely and then use it to answer queries. This is feasible because we have only two events of interest T and D, leading to only four worlds.

world	D	T	
ω_1	true	true	has disease, test positive
ω_2	true	false	has disease, test negative
ω_3	false	true	has no disease, test positive
ω_4	false	false	has no disease, test negative

Bayes Rule

To compute the beliefs in the above worlds, use the chain rule.

$$\begin{aligned}
\Pr(\omega_1) &= \Pr(T \wedge D) = \Pr(T|D)\Pr(D)\\
\Pr(\omega_2) &= \Pr(\neg T \wedge D) = \Pr(\neg T|D)\Pr(D)\\
\Pr(\omega_3) &= \Pr(T \wedge \neg D) = \Pr(T|\neg D)\Pr(\neg D)\\
\Pr(\omega_4) &= \Pr(\neg T \wedge \neg D) = \Pr(\neg T|\neg D)\Pr(\neg D).
\end{aligned}$$

All quantities are available directly from the problem statement.

world	D	T	$\Pr(.)$			
ω_1	true	true	$95/100$	$\times$	$1/1000$	$= .00095$
ω_2	true	false	$5/100$	$\times$	$1/1000$	$= .00005$
ω_3	false	true	$2/100$	$\times$	$999/1000$	$= .01998$
ω_4	false	false	$98/100$	$\times$	$999/1000$	$= .97902$