Chapter 14:
Probabilistic Reasoning

CS-4365 Artificial Intelligence

Instructor
Chris Irwin Davis

Acting Under Uncertainty

- Logical agents assume propositions are
  - True
  - False
  - Unknown acting under uncertainty

Acting Under Uncertainty

- **Example**: diagnosis (for medicine) dental diagnosis using first order logic:
  - $\forall p \; Symptom(p, Toothache) \Rightarrow Disease(p, Cavity)$
  - Wrong: not all patients *p* with toothaches have cavities! Some have gum disease, an abscess or other problems
- **Conclusion**: To make the rule true, we have to add an almost unlimited list of possible causes
  - $\forall p \; Symptom(p, Toothache) \Rightarrow$
    $Disease(p, Cavity) \lor$
    $Disease(p, GumDisease) \lor$
    $Disease(p, Abscess)$

Medical Diagnosis

- Trying to use first-order logic to cope with a domain like medical diagnosis fails because:
  - **Laziness** – too much work to list the complete list of rules + too hard to use such rules. Example of causal rule:
    - $\forall p\ Disease(p, Cavity) \Rightarrow Symptom(p, Toothache)$
  - Wrong – not all cavities cause pain $\longrightarrow$ need to augment the antecedent with all conditions that cause toothaches
  - **Theoretical ignorance** – Medical science has no complete theory for the domain
  - **Practical ignorance** – Even if we know all the rules, we might be uncertain about a particular patient, because not all necessary tests have been, or can be, run

Degree of Belief

- When propositions are not known to be true or false, the agent can at best provide a degree of belief in relevant sentences.
- The main tool for dealing with degrees of belief is Probability Theory
  - **`function DT-AGENT(percept) returns an action`**

Basic Probability Notation

- **Random Variables** – thought as referring to a "part" of the world whose "status" is unknown
  - Random variables have domains → values they may take.
- Depending on the domain, random variables may be classified as
  - **Boolean random variables** have the domain ⟨ **true, false** ⟩
    - Example: Cavity = true **OR** Cavity = false (i.e. ¬cavity)
  - **Discrete random variables** take values from a countable domain
  - **Continuous random variables** values from the real numbers
    - Example: the proposition $\boldsymbol{x} = \mathbf{4.02}$ asserts that the random variable x has the exact value 4.02.
    - We can also have propositions that use inequalities like $\boldsymbol{x} \leq \mathbf{4.20}$

Atomic Events

- An atomic event is a complete specification of the state of the world about which the agent is uncertain.
- If the world is described by a set of random variables, an **atomic event** is a particular assignment of values to the random variables

Atomic Events

- Example: 2 random variables: *Cavity* and *Toothache*
- How many atomic events?
  - 4
- Hypothesis space
  - $E_1 = (cavity = \textbf{false}) \wedge (toothache = \textbf{false})$
  - $E_2 = (cavity = \textbf{false}) \wedge (toothache = \textbf{true})$
  - $E_3 = (cavity = \textbf{true}) \wedge (toothache = \textbf{false})$
  - $E_4 = (cavity = \textbf{true}) \wedge (toothache = \textbf{true})$
- Hypothesis Space Properties:
  - a) Events are mutually exclusive: at most one can be true
  - b) The set of all possible atomic events is exhaustive

Prior Probability

- Prior probability (or unconditional probability) associated with proposition a is the *degree of belief* accorded to it in the absence of any other information:
  - Example: $P(cavity = \text{true}) = 0.1$
- Important: $P(a)$ can be used only when there is no other information. As soon as some new information is known, we must reason with the *conditional probability* of a, given that new information

Prior Probability

- Sometimes we are interested in all possible values of a random variable
  - $\therefore$ use expressions such as $P(weather)$ which denotes a *vector* of values for the probabilities of each individual state of the weather
  - Example:
  - $P(weather = \textbf{sunny}) = 0.7$
  - $P(weather = \textbf{rain}) = 0.2$
  - $P(weather = \textbf{cloudy}) = 0.08$
  - $P(weather = \textbf{snow}) = 0.02$
- Also written as:
  - $P(weather) = \langle\, 0.7, 0.2, 0.08, 0.02 \,\rangle$

Probability Distributions

- **Example 2**:
- UTD_student is a binary random variable
  - Domain = ⟨true, false⟩
- Out of a sample of 1000 young adults between 19 and 21 in the Richardson area, 321 were students at UTD
  - $P(\text{UTD\_student}) = \langle 0.321, 0.679 \rangle$

Probability Distributions

- **Example 1**:
- Hair_color is a discrete random variable.
  - Domain = ⟨blond, brown, red, black, white, none⟩
- Out of a sample of 10,000 people, we find that 1872 had blond hair, 4325 had brown hair, 652 had red hair, 2135 had black hair, 321 had white hair and the rest were bald
- The probability distribution is:
  - P(hair_color) = ⟨0.1872, 0.4325, 0.0652, 0.2135, 0.0321, 0.0721⟩

Probability Distributions

- **Binary Random Variable**
  - $P(\mathit{event}{=}\mathbf{true}) = 0.23$
  - $\therefore\ \neg P(\mathit{event}{=}\mathbf{true}) = 0.77$
  - $\neg P(\mathit{event}{=}\mathbf{true}) = 0.77 \Leftrightarrow P(\mathit{event}{=}\mathbf{false}) = 0.77$
- **Discrete Random Variable**
  - $P(\mathit{weather}{=}\mathbf{rain}) = 0.2$
  - $\neg P(\mathit{weather}{=}\mathbf{rain}) = 0.8$
  - $\neg P(\mathit{weather}{=}\mathbf{rain}) = 0.8 \Leftrightarrow$
    $P(\mathit{event}{=}\mathbf{sunny}) + P(\mathit{event}{=}\mathbf{cloudy}) + P(\mathit{event}{=}\mathbf{snow}) = 0.8$

Full Joint Probability Distribution

- $P(Cavity) = \langle 0.23, 0.77 \rangle$
- $P(Toothache) = \langle 0.35, 0.65 \rangle$
- $P(Weather) = \langle 0.7, 0.2, 0.08, 0.02 \rangle$

Suppose the world consists only of the variables

Cavity	Toothache	Weather		2x2x4=16
$\downarrow$	$\downarrow$	$\downarrow$	$\Rightarrow$	
2 values (binary random variable)	2 values (binary random variable)	4 values (sunny, rain cloudy, snow)		entries for the joint distribution

Full Joint Probability Distribution

- Represent the joint probability distribution
- Compute from individual probabilities

	Cavity		¬Cavity	
Weather	**Toothache**	**¬Toothache**	**Toothache**	**¬Toothache**
Sunny	?	?	?	?
Rain	?	?	?	?
Cloudy	?	?	?	?
Snow	?	?	?	?

Full Joint Probability Distribution

- Represent the joint probability distribution
- Compute from individual probabilities

	Cavity		¬Cavity	
Weather	**Toothache**	**¬Toothache**	**Toothache**	**¬Toothache**
Sunny	?	?	?	?
Rain	?	?	?	?
Cloudy	?	?	?	?
Snow	?	?	?	?

Probability Density Functions

- For **Continuous Random Variables**, it is not possible to write out the entire distribution as a table, because there are infinitely many values.
- Instead, we define the probability that a random variable takes on *some* value x as a **parameterized function** of x.
- Example: Let the random variable X denote tomorrow's maximum rain fall in Dallas.
  - The sentence $P(X=x) = U[1,3](x)$ expresses the belief that X is distributed uniformly between 1 and 3, inclusive.

Conditional Probability

- When new evidence concerning a previously unknown random variable is found, prior probabilities no longer apply.
- We use conditional probabilities:
  - If a and b are random variables
  - $P(a|b)$ denotes "the probability of a, given that all we know is b"
- e.g. $P(cavity|toothache) = 0.8$
- How do we compute $P(a|b)$?
  - The Product Rule

$$P(a \mid b) = \frac{P(a \wedge b)}{P(b)}$$

Conditional Distributions

- If two random variables X and Y define the world, $P(X|Y)$ gives the values for $P(X{=}x_i|Y{=}y_j)$ for each possible i and j.
- Expressed with the product rule, this becomes:
  - $P(X{=}x_1 \wedge Y{=}\mathbf{y_1}) \quad = \quad P(X{=}x_1|Y{=}\mathbf{y_1})P(Y{=}\mathbf{y_1})$
  - $P(X{=}x_1 \wedge Y{=}\mathbf{y_2}) \quad = \quad P(X{=}x_1|Y{=}\mathbf{y_2})P(Y{=}\mathbf{y_2})$
- This can be combined in a single equation:
  - $P(X,Y) = P(X|Y)P(Y)$
- This denotes a set of equations relating the corresponding individual entries in the tables, ***not*** a matrix multiplication of the tables!

$$P(T|E) = \frac{P(E|T)P(T)}{P(E|T)P(T) + P(E|\neg T)P(\neg T)}$$

$$P(T|E) = \frac{P(E|T)P(T)}{P(E|T)P(T) + P(E|\neg T)P(\neg T)}$$

$$P(E) = \sum_{i=1}^{n} P(E|T_i)P(T_i)$$

Axioms of Probability

- $0 \leq P(a) \leq 1$
- $P(\text{true}) = 1$, $P(\text{false}) = 0$
- $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$
- We can deduce $P(\neg a) = 1 - P(a)$ because
  - $P(\text{true}) = P(a) + P(\neg a) - P(\text{false})$
- $a \vee \neg a = 1$
- $a \wedge \neg a = 0$
- $P(a) + P(\neg a) = 1$

Inference Using Full Joint Distribution

- **Example**: The domains consist of three boolean variables: Toothache, Cavity and Catch (i.e. the dentist's nasty steel probe catches in your tooth)

	Toothache		**¬Toothache**		
	catch	**¬catch**	**catch**	**¬catch**	
Cavity	0.108	0.012	0.072	0.008	**0.200**
¬Cavity	0.016	0.064	0.144	0.576	**0.800**
	0.124	**0.076**	**0.216**	**0.584**	**1.000**

Inference Using Full Joint Distribution

- How many atomic events?
- $2^3 = 8$ (as many entries as in the table!)

e_1	(cavity=false)∧(toothache=false)∧(catch=false)	$P(e_1)=0.576$
e_2	(cavity=false)∧(toothache=false)∧(catch=true)	$P(e_2)=0.144$
e_3	(cavity=false)∧(toothache=true)∧(catch=false)	$P(e_3)=0.064$
e_4	(cavity=false)∧(toothache=true)∧(catch=true)	$P(e_4)=0.016$
e_5	(cavity=true)∧(toothache=false)∧(catch=false)	$P(e_5)=0.008$
e_6	(cavity=true)∧(toothache=false)∧(catch=true)	$P(e_6)=0.072$
e_7	(cavity=true)∧(toothache=true)∧(catch=false)	$P(e_7)=0.012$
e_8	(cavity=true)∧(toothache=true)∧(catch=true)	$P(e_8)=0.108$

Inferring Probabilities

- Given any proposition *a*, we can derive its probability as the **sum** of the probabilities of the atomic events in which it holds

$$P(a) = \sum_{e_i \in e(a)} P(e_i)$$

- $P(cavity) = 0.10$ … $= \mathbf{0.2}$
- $P(\neg cavity) = 0.0$ … $6 = \mathbf{0.8}$
- $P(cavity \lor toothache) = ?$
- $P(cavity \land catch) = ?$

	Toothache		¬Toothache	
	catch	**¬catch**	**catch**	**¬catch**
Cavity	0.108	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

Marginalization and Conditioning Rules

- Given any two random variables Y and Z,
- **Marginalization rule**
  - $P(Y) = \sum_{z} P(Y, Z)$

Marginalization and Conditioning Rules

- Given any two random variables *Y* and *Z*,
- **Conditioning rule**
  - $P(Y) = \sum_{z} P(Y \mid Z)P(Z)$

Marginalization and Conditioning Rules

- Probabilities
  - $P(cavity) = \mathbf{0.2}$
  - $P(\neg cavity) = \mathbf{0.8}$
  - $P(toothache) = 0.108 + 0.012 + 0.016 + 0.06 = \mathbf{0.2}$
  - $P(\neg toothache) = \mathbf{0.8}$
  - $P(catch) = 0.108 + 0.016 + 0.072 + 0.14 = \mathbf{0.34}$
  - $P(\neg catch) = \mathbf{0.66}$

Marginalization and Conditioning Rules

- If we have three random variables X, Y and Z

$$\Rightarrow P(X) = \sum_y \sum_z P(X, y, z) \quad \longleftarrow \text{ marginalization rule}$$

$$P(X) = \sum_y \sum_z \underbrace{P(X \mid y, z) P(y \mid z) P(z)}_{\text{Why?}}$$

- From the product rule:

$$P(a \mid bc) = \frac{P(a,b,c)}{P(b,c)} = \frac{P(a,b,c)}{P(b \mid c)P(c)}$$

$$\Rightarrow P(a, b, c) = P(a \mid bc) P(b \mid c) P(c)$$

Conditional Probabilities

$$P(cavity \mid toothache) = \frac{P(cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	0.108	0.012	0.072	0.008
¬ cavity	0.016	0.064	0.144	0.576

$$\text{Also}: P(\neg cavity \mid toothache) = \frac{P(\neg cavity \wedge toothache)}{P(toothache)}$$

$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

- $P(toothache)$ remains the same in both calculations
- It acts like a normalization constant for the distribution
  - $P(cavity|toothache)$

General Inference Procedure

- A notation:
  - X is a query variable (*cavity* in the example)
  - E is the set of evidence variables (*toothache* in the example)
  - e are the observed values for the evidence
- The query: $P(X|e)$
- Evaluated as
  - $P(X|e) = \alpha P(X, e) = \alpha \sum_y P(X, e, y)$

Independence

	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	0.108	0.012	0.072	0.008
¬ cavity	0.016	0.064	0.144	0.576

- Let us add a fourth variable → *weather*
  - The full distribution becomes $P(toothache, catch, cavity, weather)$
  - which has 32 entries (8 before × 4 values for *weather*)
  - This table contains four "editions" of the full table, one for each kind of weather
- What relations do these editions have to each other and to the original 3-variable table?

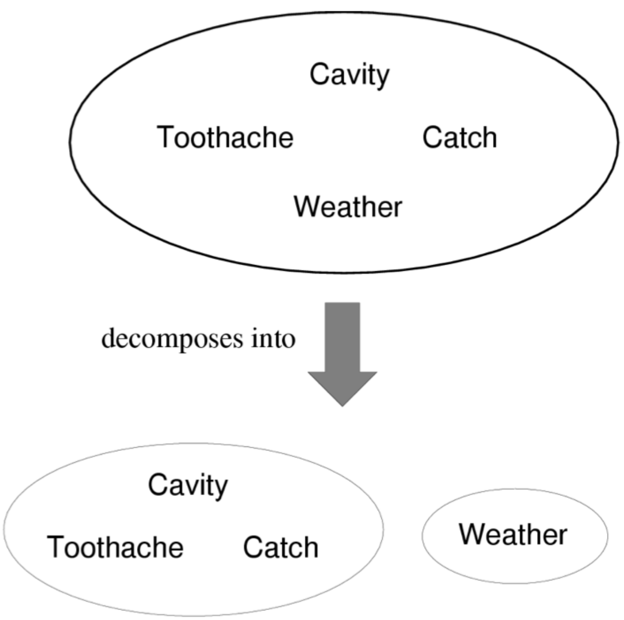
Independence

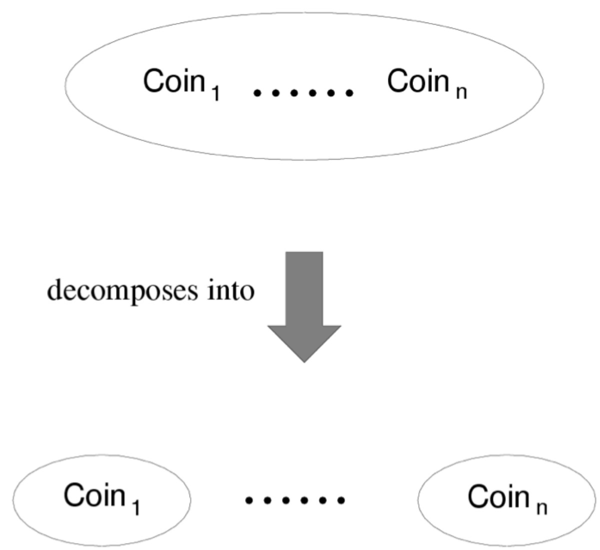
	toothache		*¬ toothache*	
	catch	*¬ catch*	*catch*	*¬ catch*
cavity	0.108	0.012	0.072	0.008
¬ cavity	0.016	0.064	0.144	0.576

- How are *P*(*toothache*, *catch*, *cavity*, *weather*=cloudy) **and** *P*(*toothache*, *catch*, *cavity*) related?
- Use the product rule:
  - *P*(*toothache*, *catch*, *cavity*, *weather*=cloudy) == *P*(*weather*=cloudy | *toothache*,*catch*,*cavity*) * *P*(*toothache*,*catch*,*cavity*)
- But it turns out that…
  - *P*(*weather*=cloudy | *toothache*,*catch*,*cavity*) == *P*(*weather*=cloudy)

Absolute Independence

- Weather is independent of one's dental problems.
  - $P(toothache, catch, cavity, weather) = P(toothache, catch, cavity) * P(weather)$
- The 32-element table can be constructed from one 8-element table and one 4-element table.





Independence In Equations

- If propositions a and b are independent
  - $P(a \wedge b) = P(a)\, P(b)$
  - $P(a \mid b) = P(a)$

$$\left(P(a \mid b) = \frac{P(a)P(b)}{P(b)} \right)$$

- Independence between variables X and Y is written:
  - $P(X, Y) = P(X)\, P(Y)$
  - $P(X \mid Y) = P(X)$
  - $P(Y \mid X) = P(Y)$

Bayes' Rule

- From the product rule:
  - $P(a \wedge b)$
    $= P(a \mid b)\ P(b)$
    $= P(b \mid a)\ P(a)$
- Bayes' Rule

$$P(b \mid a) = \frac{P(a \mid b) P(b)}{P(a)}$$

Examples

$$P(Cavity \mid Toothache) = \frac{P(Toothache \mid Cavity)P(Cavity)}{P(Toothache)}$$

$$P(cavity \mid toothache) = \frac{P(toothache \mid cavity)P(cavity)}{P(toothache)}$$

$$P(\neg cavity \mid toothache) = \frac{P(toothache \mid \neg cavity)P(\neg cavity)}{P(toothache)}$$

$$P(cavity \mid \neg toothache) = \frac{P(\neg toothache \mid cavity)P(cavity)}{P(\neg toothache)}$$

$$P(\neg cavity \mid \neg toothache) = \frac{P(\neg toothache \mid \neg cavity)P(\neg cavity)}{P(\neg toothache)}$$

Applying Bayes' Rule: The Simple Case

UTD

- Medical diagnosis
  - A doctor knows that meningitis causes a stiff neck 50% of the time in meningitis patients: $P(s|m) = 0.5$
- The doctor also knows some unconditional facts:
  - The prior probability that the patient has meningitis is 1/50,000: $\Rightarrow P(m) = 0.00002$
  - The prior probability that any patient has a stiff neck is 1/20 $\Rightarrow P(s) = 0.05$

$$P(m \mid s) = \frac{P(s \mid m) P(m)}{P(s)} = \frac{0.5 \times 1/50{,}000}{1/20} = 0.0002$$

Another Way

- We can still compute $P(m|s)$ without knowing $P(s)$
- Instead, compute the **posterior probability** for each value of the query variable (here m and $\neg m$) and normalizing the results:
  - $P(s) = P(s|m)\,P(m) + P(s|\neg m)\,P(\neg m)$

Then: $$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s \mid m)P(m) + P(s \mid \neg m)P(\neg m)} = \frac{1}{1 + \dfrac{P(s \mid \neg m)P(\neg m)}{P(s \mid m)P(m)}}$$

Similarly: $$P(\neg m \mid s) = \frac{1}{1 + \dfrac{P(s \mid m)P(m)}{P(s \mid \neg m)P(\neg m)}}$$

$\mathbf{P}(M|s) = \alpha < P(s|m)P(m),\ P(s|\neg m)P(\neg m)>$

this can be obtained also from applying Bayes' Rule with normalization

- $P(m|s) = \alpha \langle P(s|m)P(m), \; P(s|\neg m)P(\neg m) \rangle$
  - This can be obtained also from applying Bayes' Rule with normalization
- $P(Y|X) = \alpha \, P(X|Y) \, P(Y)$
  - α is a normalization constant needed to make the entries in $P(Y|X)$ sum to 1

Example of Bayes' Rule with Normalization

- We have two discrete random variables:
  - X describing weather conditions, with the domain:
    $X = \{sunny, rain, cloudy, snow\}$
  - Y describing clothes:
    $Y = \{t\text{-}shirt, long\text{-}sleeves, coat\}$
  - The distributions of X and Y are:
    $X = \langle\ 0.5, 0.2, 0.29, 0.01\ \rangle$
    $Y = \langle\ 0.5, 0.3, 0.2\ \rangle$
- We also have values of conditional probabilities:

$P(\text{t-shirt} \mid \text{sunny}) = 0.32$
$P(\text{t-shirt} \mid \text{rain}) = 0.08$
$P(\text{t-shirt} \mid \text{cloudy}) = 0.09$
$P(\text{t-shirt} \mid \text{snow}) = 0.01$

$P(\text{long-sleeves} \mid \text{sunny}) = 0.01$
$P(\text{long-sleeves} \mid \text{rain}) = 0.15$
$P(\text{long-sleeves} \mid \text{cloudy}) = 0.05$
$P(\text{long-sleeves} \mid \text{snow}) = 0.09$

$P(\text{coat} \mid \text{sunny}) = 0.001$
$P(\text{coat} \mid \text{rain}) = 0.03$
$P(\text{coat} \mid \text{cloudy}) = 0.019$
$P(\text{coat} \mid \text{snow}) = 0.15$

Computing the Normalization Constant

- Trying to guess the weather from the clothes people wear:
  - $P(X \mid Y) = \alpha\, P(Y \mid X)\ P(X)$
- $P(\text{sunny} \mid \text{t-shirt}) = \alpha\, P(\text{t-shirt} \mid \text{sunny})\, P(\text{sunny})$
  - $= \alpha \times 0.32 \times 0.5$
  - $= \alpha \times 0.16$
- $P(\text{sunny} \mid \text{long-sleeves}) = \alpha\, P(\text{long-sleeves} \mid \text{sunny})\, P(\text{sunny})$
  - $= \alpha \times 0.01 \times 0.5$
  - $= \alpha \times 0.005$
- etc.

Computing the Normalization Constant

- P(sunny|t-shirt) = α P(t-shirt|sunny) P(sunny) = α × 0.32×0.5 = **α × 0.16**
- P(sunny|long-sleeves) = α P(long-sleeves|sunny) P(sunny) = α × 0.01×0.5 = **α × 0.005**
- P(sunny|coat) = α P(coat|sunny) P(sunny) = α × 0.001×0.5 = **α × 0.0005**
- P(rain|t-shirt) = α P(t-shirt|rain) P(rain) = α × 0.08×0.2 = **α × 0.016**
- P(rain|long-sleeves) = α P(long-sleeves|rain) P(rain) = α × 0.15×0.2 = **α × 0.03**
- P(rain|coat) = α P(coat|rain) P(rain) = α × 0.03×0.2 = **α × 0.006**
- P(cloudy|t-shirt) = α P(t-shirt|cloudy) P(cloudy) = α × 0.09×0.29 = **α × 0.0261**
- P(cloudy|long-sleeves) = α P(long-sleeves|cloudy) P(cloudy) = α × 0.05×0.29 = **α × 0.014**
- P(cloudy|coat) = α P(coat|cloudy) P(cloudy) = α × 0.019×0.29 = **α × 0.00551**
- P(snow|t-shirt) = α P(t-shirt|snow) P(snow) = α × 0.01×0.01 = **α × 0.0001**
- P(snow|long-sleeves) = α P(long-sleeves|snow) P(snow) = α × 0.09×0.01 = **α × 0.0009**
- P(snow|coat) = α P(coat|snow) P(snow) = α × 0.15×0.01 = **α × 0.0015**
  - ⇒ α (0.16 + 0.005 + 0.0005 + 0.016 + 0.03 + 0.006 + 0.0261 + 0.014 + 0.00551 + 0.0001 + 0.0009 + 0.0015) = **1**

Solving for the Normalization Constant

- $\alpha\,(0.16 + 0.005 + 0.0005 + 0.016 + 0.03 + 0.006 + 0.0261 + 0.014 + 0.00551 + 0.0001 + 0.0009 + 0.0015) = 1$
- $\alpha\, 0.2656 = 1$
- $\alpha = 1/0.2656$
- $\alpha = 3.765$

Solving for the Normalization Constant

- $P(\text{sunny}|\text{t-shirt}) = \alpha \times 0.16$
  - $3.765 \times 0.16 = 0.6024$
- $P(\text{sunny}|\text{long-sleeves}) = \alpha \times 0.005$
  - $3.765 \times 0.005 = 0.018825$
- $P(\text{sunny}|\text{coat}) = \alpha \times 0.0005$
  - $3.765 \times 0.0005 = 0.0018825$

More on Bayes' rule

- What happens if one has the conditional probability in one direction but not the other?
- **Example**: meningitis domain
  - The doctor knows that a stiff neck implies meningitis in 1 of 5000 cases. The doctor has quantitative information in the diagnostic direction from symptoms to causes.
  - <u>Note</u>: Unfortunately, diagnostic knowledge is often more fragile than causal knowledge.

Causal Knowledge

- Unfortunately, diagnostic knowledge is often more fragile than causal knowledge.
- Why?
  - If there is a sudden epidemic of meningitis, the prior probability of meningitis $P(m)$ will go up.
  - Because of this, the doctor who designed the diagnostic probability $P(m|s)$ directly from statistical information, will not know how to update $P(m|s)$.
- But

$$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)}$$

Causal Knowledge

$$P(m|s) = \frac{P(s|m)P(m)}{P(s)}$$

- $P(m|s)$ should go up proportionally with $P(m)$
- $P(s|m)$ is unaffected by epidemic because it reflects how meningitis works.
- **Conclusions**:
  - Using Causal or model-based knowledge provides robustness
  - Feasible probabilistic reasoning

Combining Evidence

- Until now, we considered probabilistic information available in the form $P(\text{effect}|\text{cause})$
- What happens when there are multiple pieces of evidence?
- Example: dentist domain
  - $P(cavity \mid toothache \wedge catch) = \alpha \langle 0.108, 0.016 \rangle \cong \langle 0.871, 0.129 \rangle$
  - This will not scale up to larger numbers of variables
  - $P(cavity \mid toothache \wedge catch) =$
    $$\alpha\, P(toothache \wedge catch \mid cavity)\, P(cavity)$$
  - We need to know the values of the conditional probabilities of the conjunction *toothache* $\wedge$ *catch* for all values of *cavity*
  - If we have n possible evidence variables (X rays, diet, oral hygiene,...) these are 2^n possible combinations of observed values, and for each we need to know the conditional probabilities.

Solution

- Consider the notion of independence
- Three variables: *cavity*, *toothache*, *catch*
  — which are independent?
  - (*cavity*, *catch*)?
    - If the probe catches in the tooth, it probably has a cavity and that probably causes a toothache
  - (*toothache*, *catch*)?
    - If there is a cavity, there will be a toothache, regardless of the probable catching the tooth
  - (*toothache*, *cavity*)?
    - If there is a cavity, it might cause a toothache, but toothaches are not only caused by cavities

Bayesian Networks

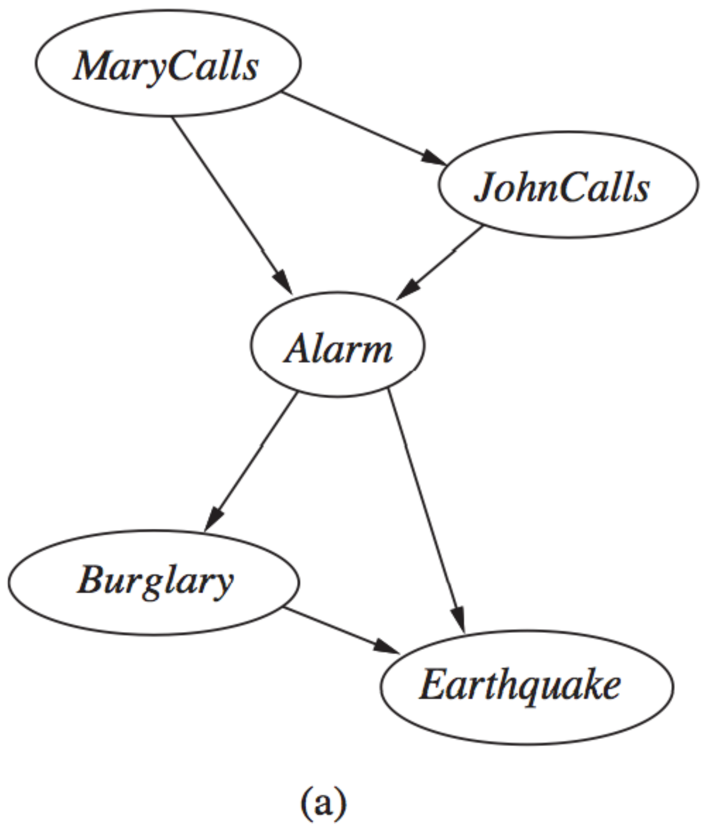
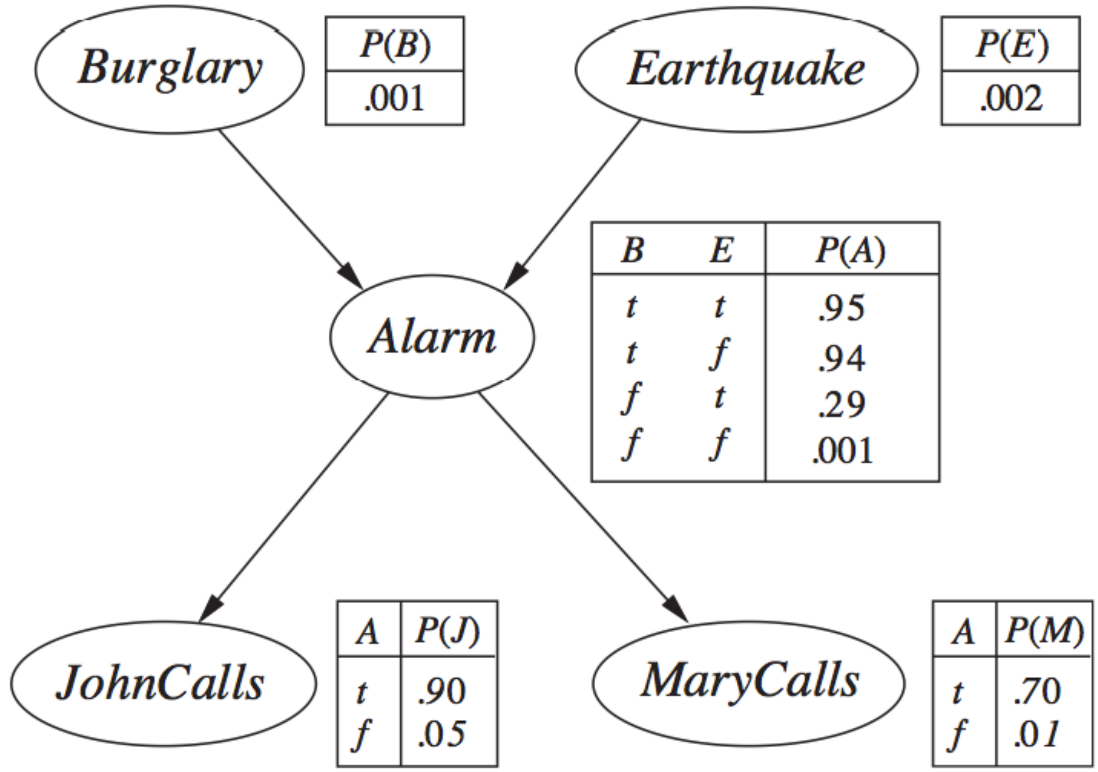


Figure 14.3 Network structure depends on order of introduction. In each network, we have introduced nodes in top-to-bottom order.



P(B)
.001

P(E)
.002

B	E	P(A)
t	t	.95
t	f	.94
f	t	.29
f	f	.001

A	P(J)
t	.90
f	.05

A	P(M)
t	.70
f	.01

Figure 14.2 A typical Bayesian network, showing both the topology and the conditional probability tables (CPTs). In the CPTs, the letters B, E, A, J, and M stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MaryCalls*, respectively.

Bayesian Networks

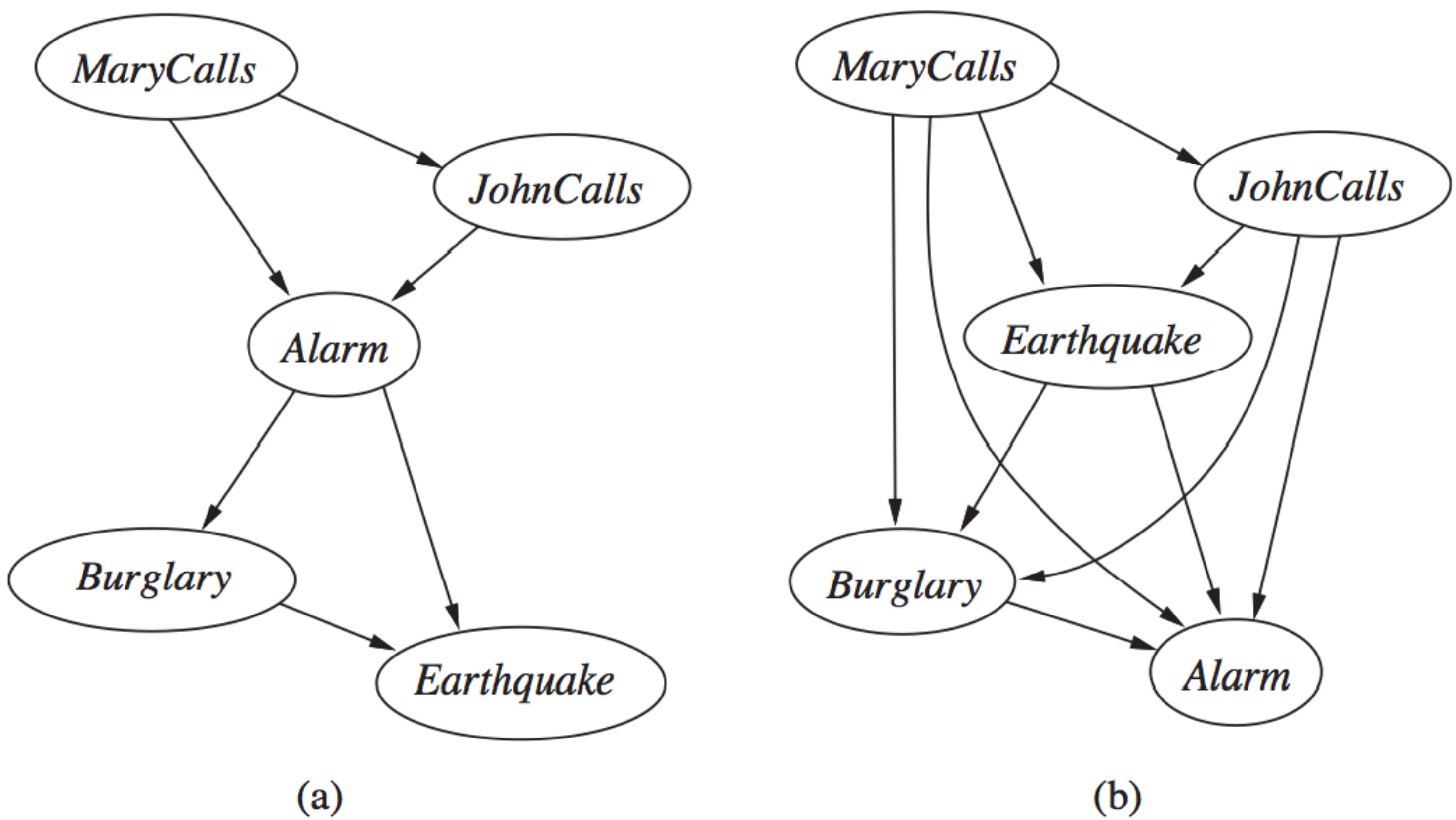


Figure 14.3 Network structure depends on order of introduction. In each network, we have introduced nodes in top-to-bottom order.

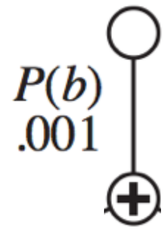


Figure 14.8 The structure of the expression shown in Equation (14.4). The evaluation proceeds top down, multiplying values along each path and summing at the "+" nodes. Notice the repetition of the paths for j and m.

Bayesian Networks

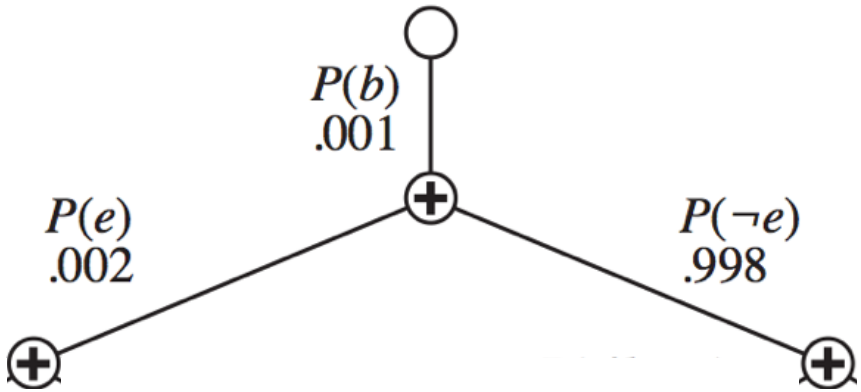


Figure 14.8 The structure of the expression shown in Equation (14.4). The evaluation proceeds top down, multiplying values along each path and summing at the "+" nodes. Notice the repetition of the paths for j and m.

Bayesian Networks

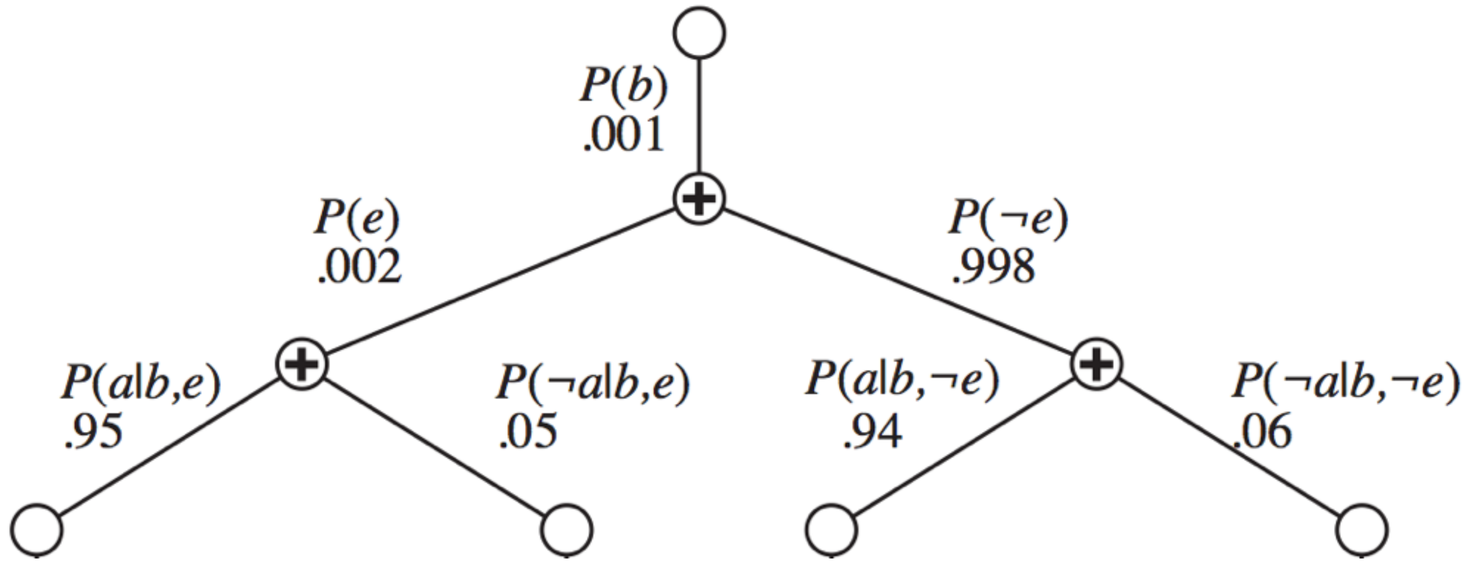


Figure 14.8 The structure of the expression shown in Equation (14.4). The evaluation proceeds top down, multiplying values along each path and summing at the "+" nodes. Notice the repetition of the paths for j and m.

Bayesian Networks

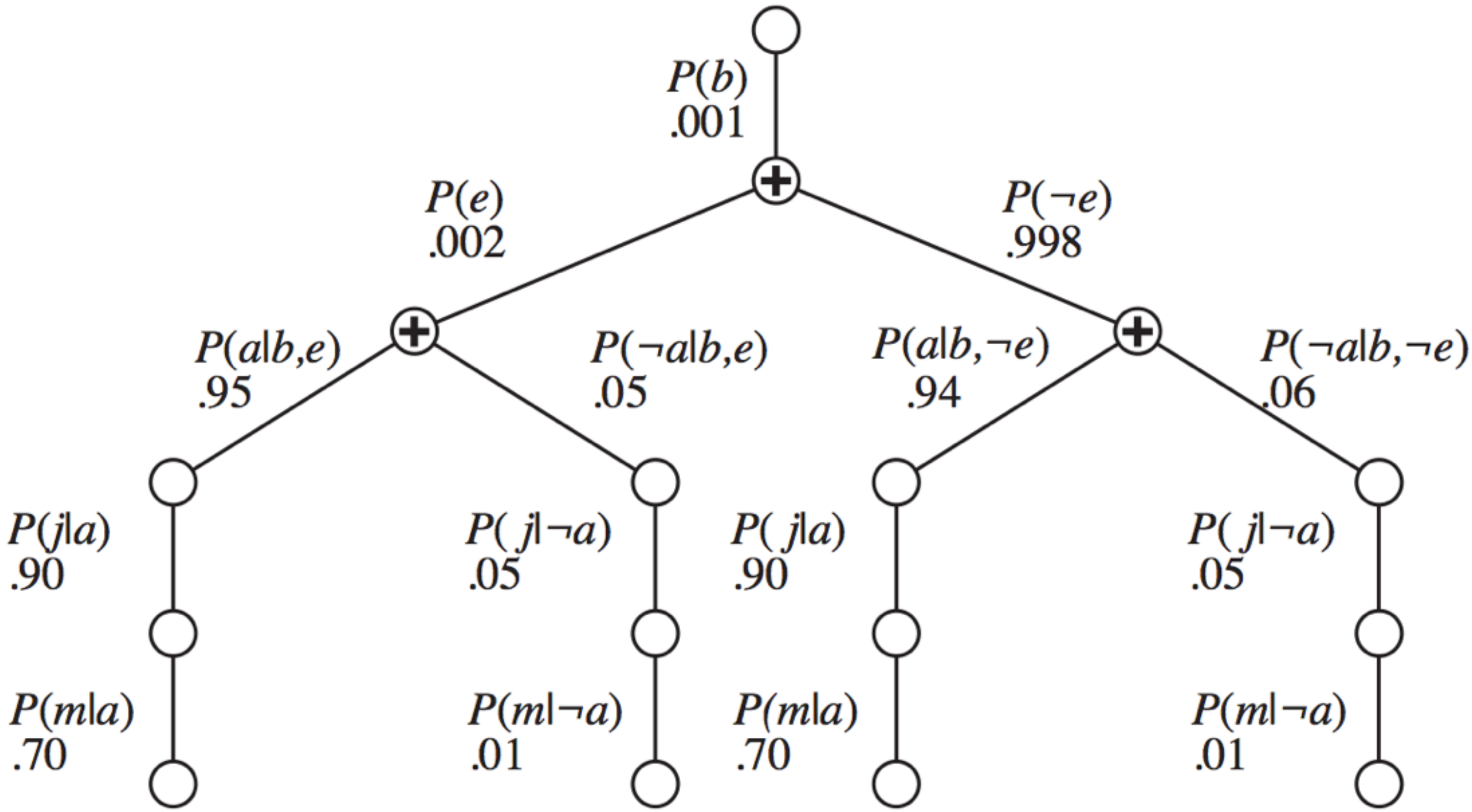


Figure 14.8 The structure of the expression shown in Equation (14.4). The evaluation proceeds top down, multiplying values along each path and summing at the "+" nodes. Notice the repetition of the paths for j and m.

Bayes Example

- Anatomy of a Test
  - 1% of women have breast cancer
  - 80% of mammograms detect breast cancer when it is there
  - 9.6% of mammograms detect breast cancer when it's not there

Bayes Example

- How do we read it?
  - 1% of people have cancer
  - If you already have cancer, you are in the first column. There's an 80% chance you will test positive. There's a 20% chance you will test negative.
  - If you don't have cancer, you are in the second column. There's a 9.6% chance you will test positive, and a 90.4% chance you will test negative.

	Cancer (1%)	No Cancer (99%)
Test Pos	80%	9.6%
Test Neg	20%	90.4%

Bayes Example

- How Accurate Is The Test?
- Now suppose you get a positive test result. What are the chances you have cancer?
  - 80%?
  - 99%?
  - 1%?

Bayes Example

- How accurate is the test?
  - Ok, we got a positive result. It means we're somewhere in the top row of our table. Let's not assume anything — it could be a true positive or a false positive.
  - The chances of a true positive = chance you have cancer * chance test caught it = 1% * 80% = .008
  - The chances of a false positive = chance you don't have cancer * chance test caught it anyway = 99% * 9.6% = 0.09504

	Cancer (1%)	No Cancer (99%)
Test Pos	True Pos: 1% * 80%	False Pos: 99% * 9.6%
Test Neg	False Neg: 1% * 20%	True Neg: 99% * 90.4%

Bayes Example

- And what was the question again?
  - Oh yes: what's the chance we really have cancer if we get a positive result. The chance of an event is the number of ways it could happen given all possible outcomes:
  - Probability = desired event / all possibilities
  - The chance of getting a real, positive result is .008. The chance of getting any type of positive result is the chance of a true positive plus the chance of a false positive (.008 + 0.09504 = .10304).
  - So, our chance of cancer is .008/.10304 = 0.0776, or about 7.8%.

Bayes Example

- Interesting — a positive mammogram only means you have a 7.8% chance of cancer, rather than 80% (the supposed accuracy of the test).
- It might seem strange at first but it makes sense: the test gives a false positive 10% of the time, so there will be a ton of false positives in any given population. There will be so many false positives, in fact, that most of the positive test results will be wrong.

Bayes Example

- Let's test our intuition by drawing a conclusion from simply eyeballing the table.
- If you take 100 people, only 1 person will have cancer (1%), and they're nearly guaranteed to test positive (80% chance).
- Of the 99 remaining people, about 10% will test positive, so we'll get roughly 10 false positives.
- Considering all the positive tests, just 1 in 11 is correct, so there's a 1/11 chance of having cancer given a positive test. The real number is 7.8% (closer to 1/13, computed above), but we found a reasonable estimate without a calculator.

	Cancer (1%)	No Cancer (99%)
Test Pos	True Pos: 1% * 80%	False Pos: 99% * 9.6%
Test Neg	False Neg: 1% * 20%	True Neg: 99% * 90.4%

Bayesian Net Example

- Consider a person who might suffer from a back injury, an event represented by the variable Back (denoted by B).
- Such an injury can cause a backache, an event represented by the variable Ache (denoted by A).
- The back injury might result from a wrong sport activity, represented by the variable Sport (denoted by S) or from new uncomfortable chairs installed at the person's office, represented by the variable Chair (denoted by C).
- In the latter case, it is reasonable to assume that a coworker will suffer and report a similar backache syndrome, an event represented by the variable Worker (denoted by W).

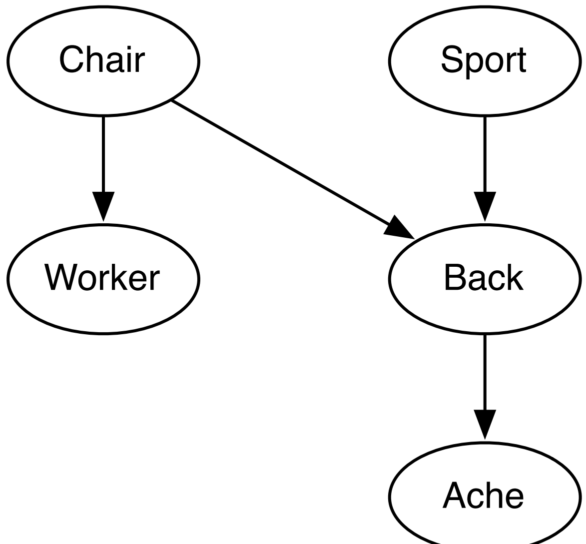
Bayesian Net Example

- All variables (for example X) are binary; thus, they are either **true** or **false**, denoted by X and $\neg X$, respectively.
- The conditional probability table (CPT) of each node is listed besides the node.
- In this example the parents of the variable *Back* are the nodes *Chair* and *Sport*.
- The child of *Back* is *Ache*, and the parent of *Worker* is *Chair*.
- Following the BN independence assumption, several independence statements can be observed in this case.

Bayesian Net Example

P(C)	P(¬C)
0.8	0.2

P(S)	P(¬S)
0.02	0.98



	$P(W \mid C)$	$P(\neg W \mid C)$
C	0.9	0.1
¬C	0.01	0.99

		$P(B \mid C,S)$	$P(\neg B \mid C,S)$
C	**S**	0.9	0.1
C	**¬S**	0.2	0.8
¬C	**S**	0.9	0.1
¬C	**¬S**	0.01	0.99

	$P(A \mid B)$	$P(\neg A \mid B)$
B	0.9	0.1
¬B	0.01	0.99