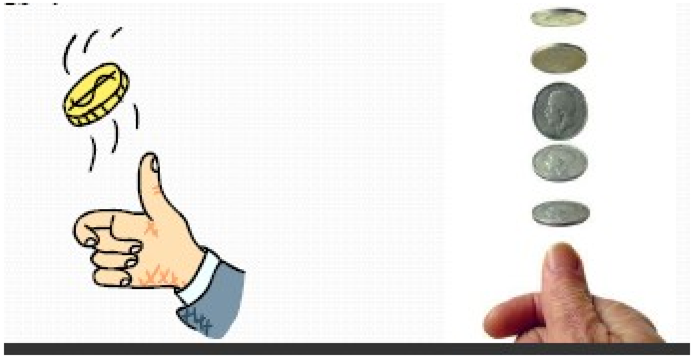
Probability

Probability is the measure of the likelihood that an event will occur. Probability is quantified as a number between 0 and 1 (where 0 indicates impossibility and 1 indicates certainty).

Example

A simple example is the toss of a fair (unbiased) coin. Since the two outcomes are equally probable, the probability of "heads" equals the probability of "tails", so the probability is 1/2 (or 50%) chance of either "heads" or "tails".



Probability Theory

Try to write rules for dental diagnosis using first-order logic, so that we can see how the logical approach breaks down

Consider the following rule:

$$\forall p \;\; Symptom(p, Toothache) \Rightarrow Disease(p, Cavity)\,.$$

The problem is that this rule is wrong. Not all patients with toothaches have cavities; some of them have gum disease, an abscess, or one of several other problems:

$$\forall p \;\; Symptom(p, Toothache) \Rightarrow$$
$$Disease(p, Cavity) \vee Disease(p, GumDisease) \vee Disease(p, Abscess) \ldots$$

Unfortunately, in order to make the rule true, we have to add an almost unlimited list of possible causes. We could try turning the rule into a causal rule

$$\forall p \;\; Disease(p, Cavity) \Rightarrow Symptom(p, Toothache) \, .$$

But this rule is not right either; not all cavities cause pain The only way to fix the rule is to make it logically exhaustive: to augment the left-hand side with all the qualifications required for a cavity to cause a toothache. Even then, for the purposes of diagnosis, one must also take into account the possibility that the patient might have a toothache and a cavity that are unconnected

Probability Theory: Variables and Events

- A random variable can be an observation, outcome or event the value of which is uncertain.
- e.g a coin. Let's use Throw as the random variable denoting the outcome when we toss the coin.
- The set of possible outcomes for a random variable is called its domain.
- The domain of Throw is {head, tail}
- A Boolean random variable has two outcomes.
- Cavity has the domain {true, false}
- Toothache has the domain {true, false}

Two (toy) examples

- I have toothache. What is the cause?

  There are many possible causes of an observed event.

- If I go to the dentist and he examines me, when the probe catches this indicates there may be a cavity, rather than another cause.

  The likelihood of a hypothesised cause will change as additional pieces of evidence arrive.

- Bob lives in San Francisco. He has a burglar alarm on his house, which can be triggered by burglars and earthquakes. He has two neighbours, John and Mary, who will call him if the alarm goes off while he is at work, but each is unreliable in their own way. All these sources of uncertainty can be quantified. Mary calls, how likely is it that there has been a burglary?

  Using probabilistic reasoning we can calculate how likely a hypothesised cause is.

- The connection between toothaches and cavities is just not a logical consequence in either direction.

- This is typical of, law, business, design, automobile repair, gardening, dating, and so on.

- The agent's knowledge can at best provide only a **degree of belief** in the relevant sentences.
- Our main tool for dealing PROBABILITY THEORY with degrees of belief will be **probability theory,** which assigns to each sentence a numerical degree of belief between 0 and 1.

- We might not know for sure what afflicts a particular patient,
- Say, an 80% chance-that is, a probability of 0.8-that the patient has a cavity if he or she has a toothache
- The missing 20% summarizes all the other possible causes of toothache that we are too lazy or ignorant to confirm or deny.

The basic element of the language is the random variable, which can be thought of as referring to a "part" of the world whose "status" is initially unknown.

Each random variable has a domain of values that it can take on. For example, the domain of *Cavzty* might be *(t rue,f a l ~ e*

Measures of Central Tendency

- Measures of central tendency yield information about "particular places or locations in a group of numbers."
- Common Measures of Location
  - Mode
  - Median
  - Mean
  - Percentiles
  - Quartiles

Measures of Central Tendency

- Mean … the average score
- Median … the value that lies in the middle after ranking all the scores
- Mode … the most frequently occurring score

Measures of Central Tendency

The measure you choose should give you a good indication of the typical score in the sample or population.

Measures of Central Tendency

Mean … the most frequently used but is sensitive to extreme scores

e.g. 1 2 3 4 5 6 7 8 9 10

Mean = 5.5 (median = 5.5)

e.g. 1 2 3 4 5 6 7 8 9 20

Mean = 6.5 (median = 5.5)

e.g. 1 2 3 4 5 6 7 8 9 100

Mean = 14.5 (median = 5.5)

Mean (ΜΕΣΟΣ)

- Is the average of a group of numbers
- Applicable for interval and ratio data, not applicable for nominal or ordinal data
- Affected by each value in the data set, including extreme values
- Computed by summing all values in the data set and dividing the sum by the number of values in the data set

Population Mean

$$\mu = \frac{\sum X}{N} = \frac{X_1 + X_2 + X_3 + \ldots + X_N}{N}$$

$$= \frac{24 + 13 + 19 + 26 + 11}{5}$$

$$= \frac{93}{5}$$

$$= 18.6$$

Measures of Central Tendency

Mode

… does not involve any calculation or ordering of data

… use it when you have categories (e.g. occupation)

Variation or Spread of Distributions

Standard Deviation

- It tells us what is happening between the minimum and maximum scores
- It tells us how much the scores in the data set vary around the mean
- It is useful when we need to compare groups using the same scale

Calculating a Mean and a Standard Deviation

	Data x	Deviation x - Mean	Absolute Deviation |x - Mean|	Squared Deviation (x-Mean)²
	10	-20	20	400
	20	-10	10	100
	30	0	0	0
	40	10	10	100
	50	20	20	400
Sums	150	0	60	1000
Means	30	0	12	200
				Variance
	Standard deviation = $\sqrt{\text{Variance}}$			14.1421356

Prior probability

The **unconditional** or **prior probability** associated with a proposition *a* is the degree of belief accorded to it in the absence of any other information.

it is written as $P(a)$.

For example,

If the prior probability that I have a cavity is 0.1, then we would write

$\boldsymbol{P(Cavity = true) = 0.1}$ **or** $\boldsymbol{P(cavity) = 0.1}$.

It is important to remember that $P(a)$ 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 will want to talk about the probabilities of all the possible values of a random variable. In that case, we will use an expression such as **P(** *Weather)*, which denotes a ***vector*** of values for the probabilities of each individual state of the weather.

We can write the four equations

$$P(\ Weather = sunny) = 0.7$$

$$P(\ Weather = rain) = 0.2$$

$$P(\ Weather = cloudy) = 0.08$$

$$P(\ Weather = snow) = 0.02\ .$$

we may simply write

$$P(\ Weather) = (0.7, 0.2, 0.08, 0.02)$$

This statement defines a prior **probability distribution** for the random variable *Weather*

Probability Theory: Conditional Probability

- A conditional probability expresses the likelihood that one event a will occur if b occurs. We denote this as follows

$$P(a \mid b)$$

- e.g.

$$P(Toothache = true) = 0.2$$

$$P(Toothache = true \mid Cavity = true) = 0.6$$

- So conditional probabilities reflect the fact that some events make other events more (or less) likely
- If one event doesn't affect the likelihood of another event they are said to be independent and therefore

$$P(a \mid b) = P(a)$$

- E.g. if you roll a 6 on a die, it doesn't make it more or less likely that you will roll a 6 on the next throw. The rolls are independent.

Conditional Probability

A **conditional probability** measures the probability of an event given that (by assumption, presumption, assertion or evidence) another event has occurred.

If the event of interest is *A* and the event *B* is known or assumed to have occurred, "the conditional probability of *A* given *B*", or "the probability of *A* under the condition *B*", For example,

- **Probability that any given person has a cough on any given day may be only 5%.**
- But if we know or assume that the person has a cold, then they are much more likely to be coughing.

- The conditional probability of coughing given that you have a cold might be a much higher 75%

Conditional Probability

- $P(A|B)$ may or may not be equal to $P(A)$ (the unconditional probability of A). If $P(A|B) = P(A)$, then events A and B are said to be independent:.
- $P(A|B)$ (the conditional probability of A given B) is not equal to $P(B|A)$.
- For example, **if a person has dengue they might have a 90%** chance of testing positive for dengue. In this case what is being measured is that if event B ("having dengue") has occurred,
- the probability of A (*test is positive*) given that B (*having dengue*) occurred is 90%: that is, $P(A|B) = 90\%$

Conditional Probability

Conditional probabilities can be defined in terms of unconditional probabilities defining equation is

$$P(a|b) = \frac{P(a,b)}{P(b)}$$

which holds whenever $P(b) > 0$. This equation can also be written as

$$P(a,b) = P(a|b)\,P(b)$$

which is called the **product rule.** The product rule is perhaps easier to remember:

it comes from the fact that, for a and b to be true, we need b to be true, and we also need a to be true

given b. We can also have it the other way around:

$$P(a,b) = P(b|a)\,P(a)$$

Combining Probabilities: the product rule

- How we can work out the likelihood of two events occurring together given their base and conditional probabilities?

$$P(a \wedge b) = P(a \mid b)P(b) = P(b \mid a)P(a)$$

- So in our toy example 1:

$$P(toothache \wedge cavity) = P(toothache \mid cavity)P(cavity)$$

$$= P(cavity \mid toothache)P(toothache)$$

- But this doesn't help us answer our question:

  "I have toothache. Do I have a cavity?"

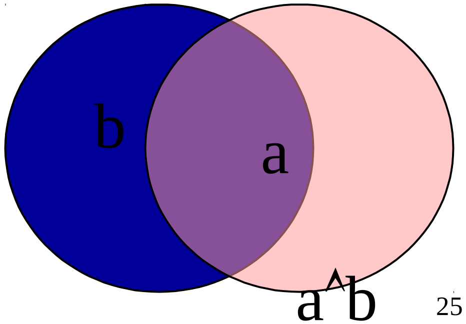
KOLMOGOROV'S AXIOMS?

- Kolmogorov showed that three simple axioms lead to the rules of probability theory

1. All probabilities are between 0 and 1:
   - $0 \leq P(a) \leq 1$
2. Necessarily true (i.e., valid) propositions have probability 1, and necessarily false 0 $P(true) = 1$; $P(false) = 0$

- Next, we need an axiom that connects the probabilities of logically related propositions. The simplest way disjunction as follows

3. The probability of a disjunction is given by:
   - $P(a \vee b) = P(a) + P(b) - P(a \wedge b)$



Using the axioms of probability

The familiar rule for negation follows by substituting $\neg a$ for b in axiom 3, giving us:

$$
\begin{aligned}
P(a \lor \neg a) &= P(a) + P(\neg a) - P(a \land \neg a) && \text{(by axiom 3 with } b = \neg a\text{)} \\
P(true) &= P(a) + P(\neg a) - P(false) && \text{(by logical equivalence)} \\
1 &= P(a) + P(\neg a) && \text{(by axiom 2)} \\
P(\neg a) &= 1 - P(a) && \text{(by algebra).}
\end{aligned}
$$

The third line of this derivation is itself a useful fact and can be extended from the Boolean case to the general discrete case.
Let the discrete variable D have the domain $(d_1, \ldots, d_n)$.
Then it is easy to show

$$\sum_{i=1}^{n} P(D = d_i) = 1$$

That is, any probability distribution on a single variable must sum to 1

That is, any probability distribution on a single variable must sum to 1

Basic Formulas for Probabilities

- **Product Rule :** probability P(AB) of a conjunction of two events A and B:

$$P(A,B) = P(A \mid B)P(B) = P(B \mid A)P(A)$$

- **Sum Rule:** probability of a disjunction of two events A and B:

$$P(A+B) = P(A) + P(B) - P(AB)$$

- **Theorem of Total Probability** : if events A1, …., An are mutually exclusive with

$$P(B) = \sum_{i=1}^{n} P(B \mid A_i)P(A_i)$$

BAYES RULE

We define the product rule and pointed out that it can be written in two forms because of the commutativity of conjunction:

$$P(b|a) = P(a|b)P(b)$$
$$P(a|b) = P(b|a)P(a)$$

Equating the two right-hand sides and dividing by *P(a)*, we get

Equating the two right-hand sides and dividing by *P(a)*, we get

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

This equation is known as **Bayes' rule** (also Bayes' law or Bayes' Theorem.

This is simple equation underlies all modern A1 systems for probabilistic inference. The more general case of multivalued variables can be written in the P notation as

$$P(Y|X) = \frac{P(X|Y)}{P(X)}$$

BAYES RULE

where again this is to be taken as representing a set of equations, each dealing with specific values of the variables.

We will also have occasion to use a more general version conditionalized on some background evidence **e:**

$$P(Y|X,e) = \frac{P(X|Y)P(Y|e)}{P(X|e)}$$

- Bayes' rule does not seem very useful. It requires three terms-a conditional probability and two unconditional probabilities-just
- to compute one conditional probability.

where again this is to be taken as representing a set of equations, each dealing with specific values of the variables.

We will also have occasion to use a more general version conditionalized on some background evidence **e:**

- Bayes' rule does not seem very useful. It requires three terms-a conditional probability and two unconditional probabilities-just
- to compute one conditional probability.

Bayes' rule

$$P(a \wedge b) = P(a|b)P(b) = P(b|a)P(a)$$

- We can rearrange the two parts of the product rule:

$$P(a|b)P(b) = P(b|a)P(a)$$

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

- This is known as Bayes' rule.
- It is the cornerstone of modern probabilistic AI.
- But why is it useful?

Bayes' rule

- We can think about some events as being "hidden" causes: not necessarily directly observed (e.g. a cavity).
- If we model how likely observable effects are given hidden causes (how likely toothache is given a cavity)
- Then Bayes' rule allows us to use that model to infer the likelihood of the hidden cause (and thus answer our question)

$$P(cause \mid effect) = \frac{P(effect \mid cause) P(cause)}{P(effect)}$$

- In fact good models of $P(effect \mid cause)$ are often available to us in real domains (e.g. medical diagnosis)

BAYES RULE

The Bayes Theorem was developed and named for Thomas Bayes(1702-1761)

Show the Relation between one conditional probability and its inverse.

Provide a mathematical rule for revising an estimate or forecast in light of experience and observation.

In the 18th Century , Thomas Bayes,

Ponder this question:

"Does God really exist?"

- Being interested in the mathematics, he attempt to develop a formula to arrive at the probability that God does exist based on the evidence that was available to him on earth.

Later, **Laplace** refined **Bayes' work** and gave it the name "Bayes' Theorem".

Example-1 BAYES RULE

A doctor knows that the disease meningitis causes the patient to have a stiff neck, say, 50% of the time. The doctor also knows some unconditional facts: the prior probability that a patient has meningitis is 50,000, and the prior probability that any patient has a stiff neck is 20. Letting *s* be the proposition that the patient has a stiff neck and *m* be the proposition that the patient has meningitis, we have

$$
\begin{aligned}
P(s|m) &= 0.5 \\
P(m) &= 1/50000 \\
P(s) &= 1/20 \\
P(m|s) &= \frac{P(s|m)P(m)}{P(s)} = \frac{0.5 \times 1/50000}{1/20} = 0.0002\,.
\end{aligned}
$$

That is, we expect only 1 in 5000 patients with a stiff neck to have meningitis

Bayes' rule can capture causal models

- Suppose a doctor knows that a meningitis causes a stiff neck in 50% of cases

$$P(s \mid m) = 0.5$$

- She also knows that the probability in the general population of someone having a stiff neck at any time is 1/20

$$P(s) = 0.05$$

- She also has to know the incidence of meningitis in the population (1/50,000)

$$P(m) = 0.00002$$

- Using Bayes' rule she can calculate the probability the patient has meningitis:

$$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.5 \times 0.00002}{0.05} = 0.0002 = 1/5000$$

$$P(cause \mid effect) = \frac{P(effect \mid cause)P(cause)}{P(effect)}$$

Example-2 of Bayes Rule

Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding?

Solution…

The sample space is defined by two mutually-exclusive events - it rains or it does not rain. Additionally, a third event occurs when the weatherman predicts rain. Notation for these events appears below.

Event A1. It rains on Marie's wedding.
Event A2. It does not rain on Marie's wedding.
Event B. The weatherman predicts rain

In terms of probabilities, we know the following:

$P(A1) = 5/365 = 0.0136985$ [It rains 5 days out of the year.]

$P(A2) = 360/365 = 0.9863014$ [It does not rain 360 days out of the year.]

$P(B \mid A1) = 0.9$ [When it rains, the weatherman predicts rain 90% of the time.]

$P(B \mid A2) = 0.1$ [When it does not rain, the weatherman predicts rain 10% of the time.]

We want to know $P(A1 \mid B)$, the probability it will rain on the day of Marie's wedding, given a forecast for rain by the weatherman. The answer can be determined from Bayes' theorem, as shown below.

$$P(A_1 \mid B) = \frac{P(A_1)\,P(B \mid A_1)}{P(A_1)\,P(B \mid A_1) + P(A_2)\,P(B \mid A_2)}$$

$$P(A_1 \mid B) = (0.014)(0.9) \,/\, [\,(0.014)(0.9) + (0.986)(0.1)\,]$$

$$P(A_1 \mid B) = 0.111$$

Joint Probability Mass Function

- Completely specifies all beliefs in a problem domain.
- Joint prob Distribution is an n-dimensional table with a probability in each cell of that state occurring.
- Written as $P(X_1, X_2, X_3 \ldots, X_n)$
- When instantiated as $P(x_1, x_2 \ldots, x_n)$

Random variables: sunshine S ∈ {0, 1}, rain R ∈ {0, 1}

Joint distribution:

$\mathbb{P}(S, R) =$

s	r	$\mathbb{P}(S = s, R = r)$
0	0	0.20
0	1	0.08
1	0	0.70
1	1	0.02

Marginal distribution:

$\mathbb{P}(S) =$

s	$\mathbb{P}(S = s)$
0	0.28
1	0.72

(aggregate rows)

Conditional distribution:

$\mathbb{P}(S \mid R = 1) =$

s	$\mathbb{P}(S = s \mid R = 1)$
0	0.8
1	0.2

(select rows, normalize)

Suppose we have two boolean random variables, S and R representing sunshine and rain. Think of an assignment to (S, R) as representing a possible state of the world.

- The joint distribution specifies a probability for each assignment to (S, R) (state of the world).
- We use lowercase letters (e.g., s and r) to denote values and uppercase letters (e.g., S and R) to denote random variables.
- Note that P(S = s, R = r) is a probability (a number) while P(S, R) is a distribution (a table of probabilities).
- We don’t know what state of the world we’re in, but we know what the probabilities are (there are no unknown unknowns).

- **The joint distribution contains all the information and acts as the central source of truth**

- The conditional distribution selects rows of the table matching the condition (right of the bar), and then normalizes the probabilities so that they sum to 1. The interpretation is that we observe the condition ($R = 1$) and are interested in S. This is the conditioning that we saw for factor graphs, but where we normalize the selected rows to get probabilities

What is a Bayesian Network ?

A graphical model that efficiently encodes the joint probability distribution for a large set of variables

Bayesian network Or Belief Networks

A Bayesian network is a directed graph in which each node is annotated with quantitative probability information. The full specification is as follows:

1. **A set of random variables makes up the nodes** of the network. Variables may be discrete or continuous.
2. **A set of directed links or arrows connects pairs of nodes.** If there is an arrow from node *X* to node *Y*, X is said to be a parent of *Y*.
3. **Each node X_i has a conditional probability distribution** $P(X_i \mid Parents(X_i))$ that quantifies the effect of the parents on the node.
4. **The graph has no directed cycles** (and hence is a directed, acyclic graph, or DAG).

Bayesian network

Conditional independence of *Toothache* and *Catch* given *Cavity* is indicated by the absence of a link between *Toothache* and *Catch*.

Intuitively, the network represents the fact that *Cavity* is a direct cause of *Toothache* and *Catch*, whereas no direct causal relationship exists between *Toothache* and *Catch*.

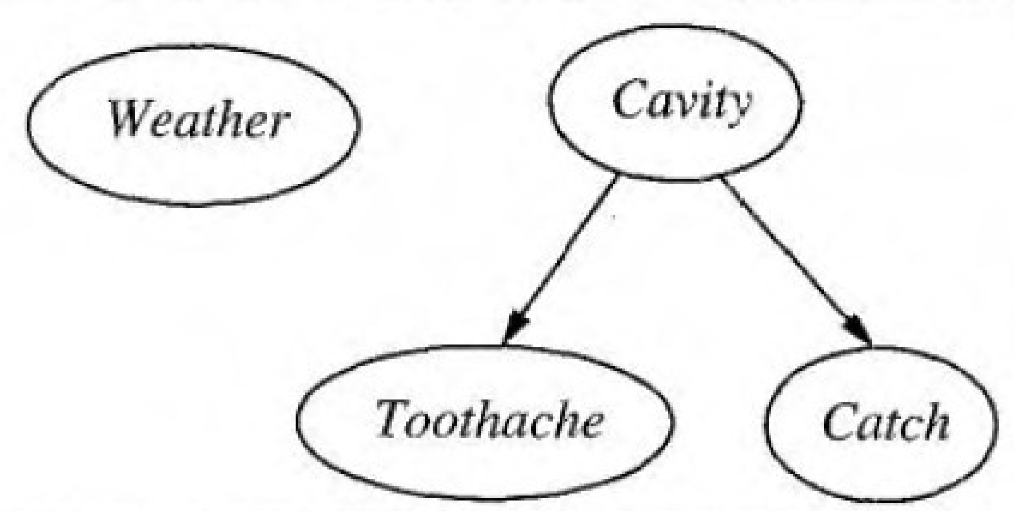
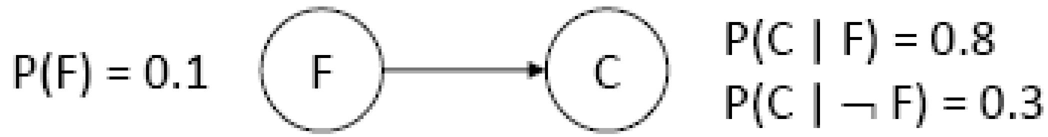


Figure 14.1 A simple Bayesian network in which *Weather* is independent of the other three variables and *Toothache* and *Catch* are conditionally independent, given ***Cavity***

Bayesian Networks

1) Consider the following Bayesian network, where F = having the flu and C = coughing:



a) Write down the joint probability table specified by the Bayesian network.

Answer:

F	C	
t	t	$0.1 \times 0.8 = 0.08$
t	f	$0.1 \times 0.2 = 0.02$
f	t	$0.9 \times 0.3 = 0.27$
f	f	$0.9 \times 0.7 = 0.63$

b) Determine the probabilities for the following Bayesian network

so that it specifies the same joint probabilities as the given one.

Answer:

$$P(C) = 0.08 + 0.27 = 0.35$$

$$P(F \mid C) = P(F, C) \;/\; P(C) = 0.08/0.35 \sim 0.23$$

$$P(F \mid \neg C) = P(F, \neg C) \;/\; P(\neg C) = 0.02/0.65 \sim 0.03$$

F	C	
t	t	$0.1 \times 0.8 = 0.08$
t	f	$0.1 \times 0.2 = 0.02$
f	t	$0.9 \times 0.3 = 0.27$
f	f	$0.9 \times 0.7 = 0.63$

C) Which Bayesian network would you have specified using the rules learned in class?

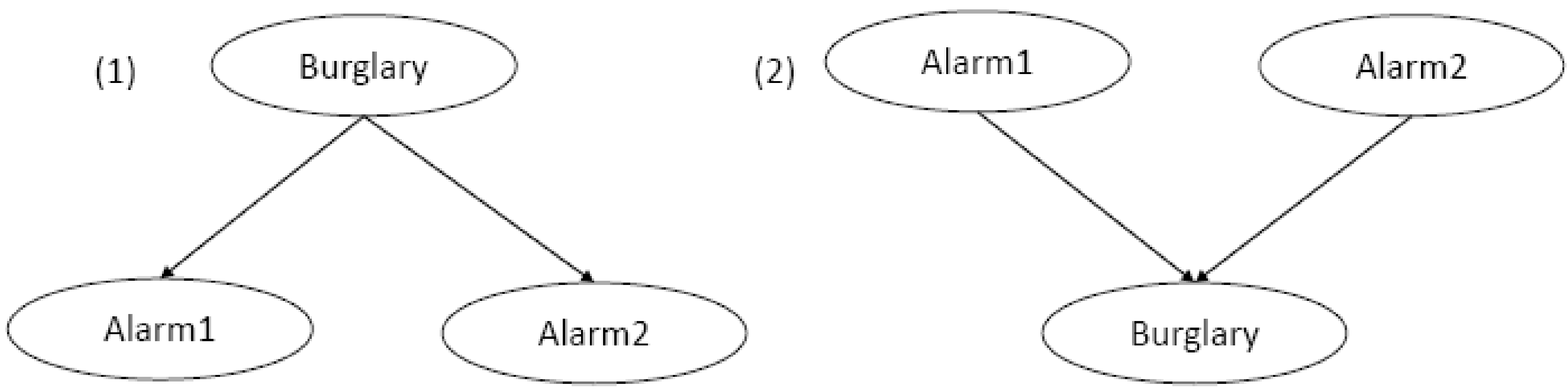
Answer:
The first one. It is good practice to add nodes that correspond to causes before nodes that correspond to their effects.

d) Are C and F independent in the given Bayesian network?

Answer:
No, since (for example) P(F) = 0.1 but P(F / C) 0.23
e) Are C and F independent in the Bayesian network from Question b?
Answer:
No, for the same reason.

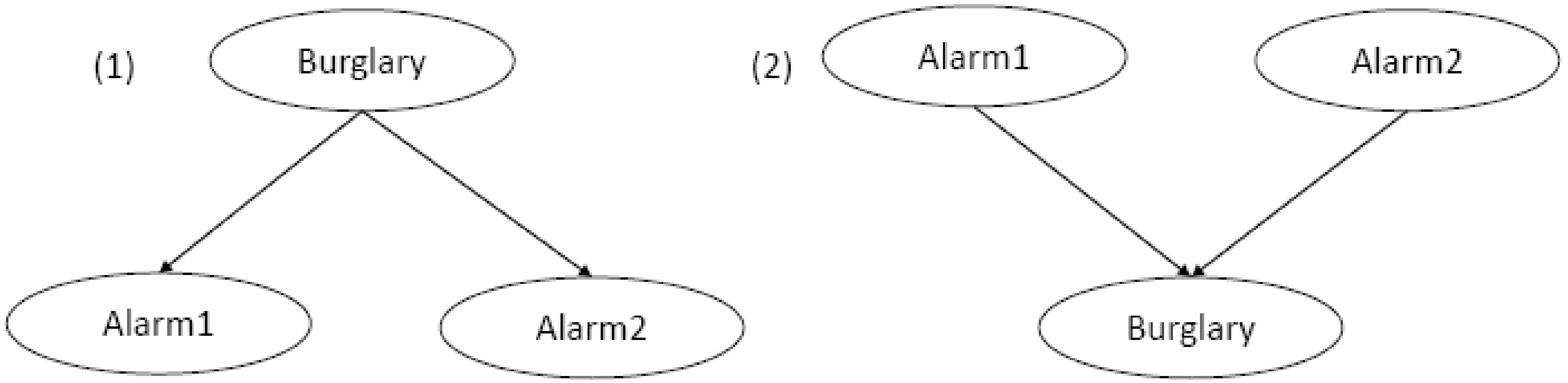
2) **To safeguard your house, you recently installed two deferent alarm systems by two deferent reputable manufacturers that use completely deferent sensors for their alarm systems**.

a) Which one of the two Bayesian networks given below **makes independence assumptions that are not true?** Explain all of your reasoning. **Alarm1 means that the first alarm system rings, Alarm2 means that the second alarm system rings**, and Burglary means that a burglary is in progress.



Answer: The second one falsely assumes that **Alarm1 and Alarm2 are independent if the value of Burglary is unknown.** However, if the alarms are working as intended, it should be more likely that Alarm1 rings if Alarm2 rings (that is, they should not be independent).

2) **To safeguard your house, you recently installed two deferent alarm systems by two deferent reputable manufacturers that use completely deferent sensors for their alarm systems.**



b) **Consider the first Bayesian network. How many probabilities need to be specified for its conditional probability tables?** How many probabilities would need to be given if the same joint probability distribution were specified in a joint probability table?

Answer:

We need to specify 5 probabilities,
namely P(Burglary),
$P(Alarm1|Burglary)$,
$P(Alarm1|\neg Burglary)$,
$P(Alarm2|Burglary)$ and
$P(Alarm2|\neg Burglary)$.

A joint probability table would need $2^3 - 1$ probabilities.

c) Consider the second Bayesian network. Assume that:

$P(\text{Alarm1}) = 0.1$

$P(\text{Alarm2}) = 0.2$

$P(\text{Burglary} \mid \text{Alarm1}, \text{Alarm2}) = 0.8$

$P(\text{Burglary} \mid \text{Alarm1}, \neg \text{Alarm2}) = 0.7$

$P(\text{Burglary} \mid \neg \text{Alarm1}, \text{Alarm2}) = 0.6$

$P(\text{Burglary} \mid \neg \text{Alarm1}, \neg \text{Alarm2}) = 0.5$

Calculate $P(\text{Alarm2} \mid \text{Burglary}, \text{Alarm1})$. Show all of your reasoning.

Answer:

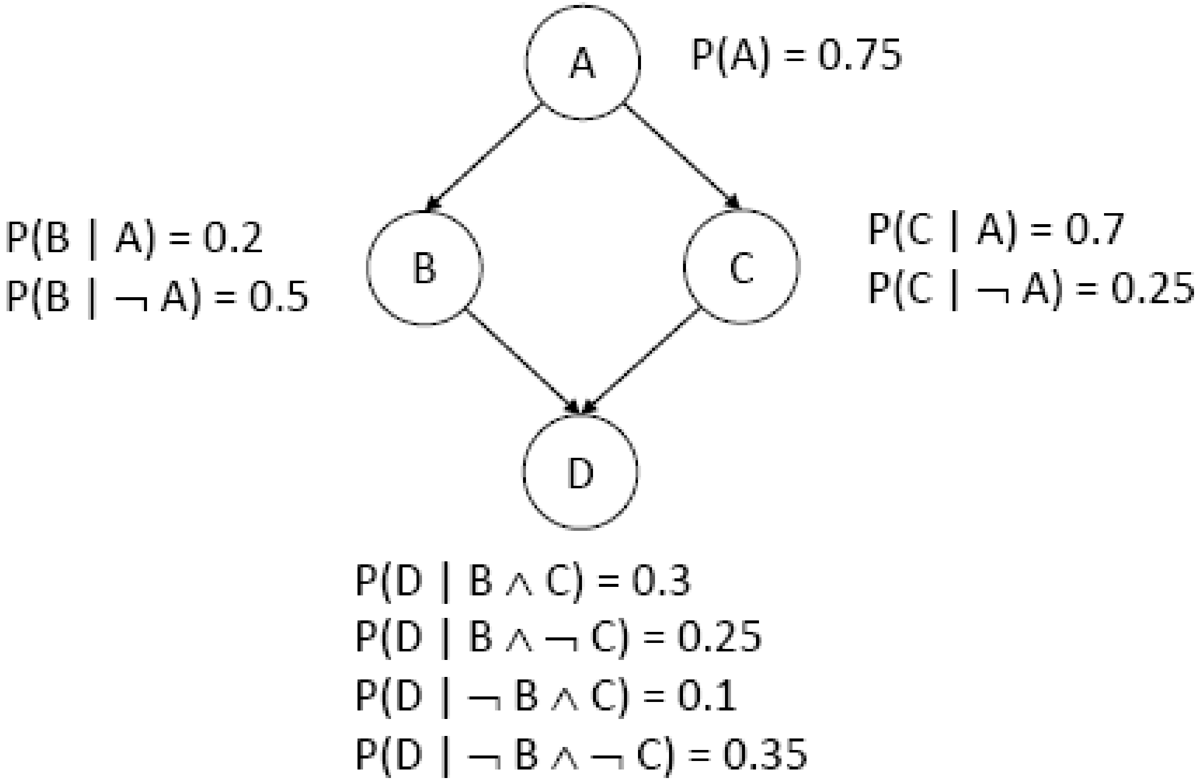
P(Alarm2 | Burglary, Alarm1) = P(Alarm1, Alarm2, Burglary) / P(Burglary, Alarm1) = $0.016/0.072 \sim 0.22$ with

P(Alarm1, Alarm2, Burglary) = P(Alarm1) P(Alarm2) P(Burglary | Alarm1, Alarm2) = $0.1 \times 0.2 \times 0.8 = 0.016$

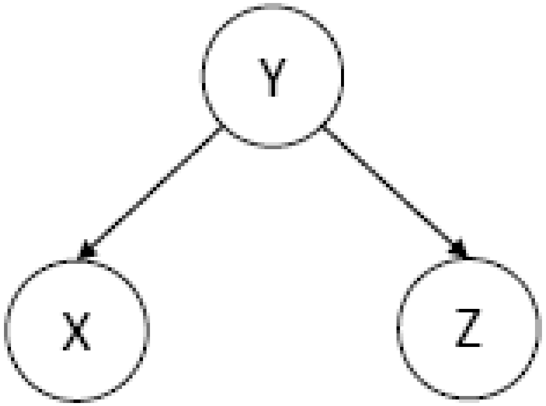
P(Alarm1, ¬ Alarm2, Burglary) = P(Alarm1) P(¬ Alarm2) P(Burglary | Alarm1, ¬ Alarm2) = $0.1 \times 0.8 \times 0.7 = 0.056$

P(Burglary, Alarm1) = P(Alarm1, Alarm2, Burglary) + P(Alarm1, ¬ Alarm2, Burglary) = $0.016 + 0.056 = 0.072$

4) Consider the following Bayesian network. A, B, C, and D are Boolean random variables. If we know that A is true, what is the probability of D being true?

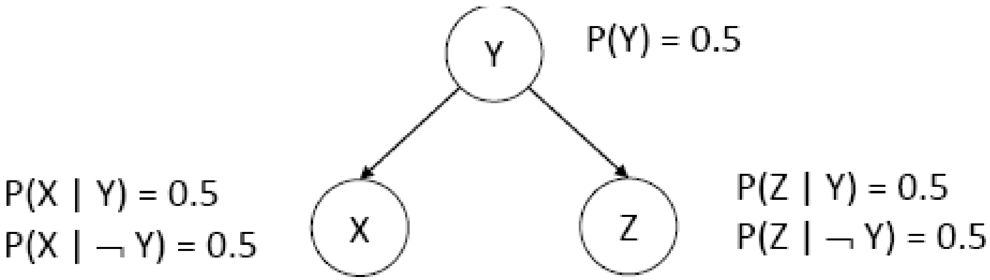


5) For the following Bayesian network



We know that X and Z are not guaranteed to be independent if the value of Y is unknown.
This means that, depending on the probabilities, X and Z can be independent or dependent if the value of Y is unknown.

Construct probabilities where X and Z are independent if the value of Y is unknown, and show that they are indeed independent.



$$\mathrm{P(X) = P(Y)\ P(X \mid Y) + P(\neg\ Y)\ P(X \mid \neg\ Y)} = 0.5 \times 0.5 + 0.5 \times 0.5 = 0.5$$

$$\mathrm{P(Z) = P(Y)\ P(Z \mid Y) + P(\neg\ Y)\ P(Z \mid \neg\ Y)} = 0.5 \times 0.5 + 0.5 \times 0.5 = 0.5$$

$$\mathrm{P(X,\ Z) = P(X,\ Y,\ Z) + P(X,\ \neg\ Y,\ Z)}$$

$$= \mathrm{P(Y)\ P(X \mid Y)\ P(Z \mid Y) + P(\neg\ Y)\ P(X \mid \neg\ Y)\ P(Z} \mid \neg Z)$$

$$= 0.5 \times 0.5 \times 0.5 + 0.5 \times 0.5 \times 0.5 = 0.25$$

Therefore, P(X) P(Z) = P(X,Z). We can similarly show that P(X) P(¬ Z) = P(X, ¬ Z), P(¬ X) P(Z) = P(¬ X, Z) and P(¬ X) P(¬ Z) = P(¬ X, ¬ Z) to prove that X and Z are independent if the value of Y is unknown.

Bayesian network earthquake example

- You have a new burglar alarm installed
- It is reliable about detecting burglary, but responds to minor earthquakes
- Two neighbors (John, Mary) promise to call you at work when they hear the alarm
  - John always calls when hears alarm, but confuses alarm with phone ringing (and calls then also)
  - Mary likes loud music and sometimes misses alarm!
- Given evidence about who has and hasn’t called, estimate the probability of a burglary

Bayesian network Another Example

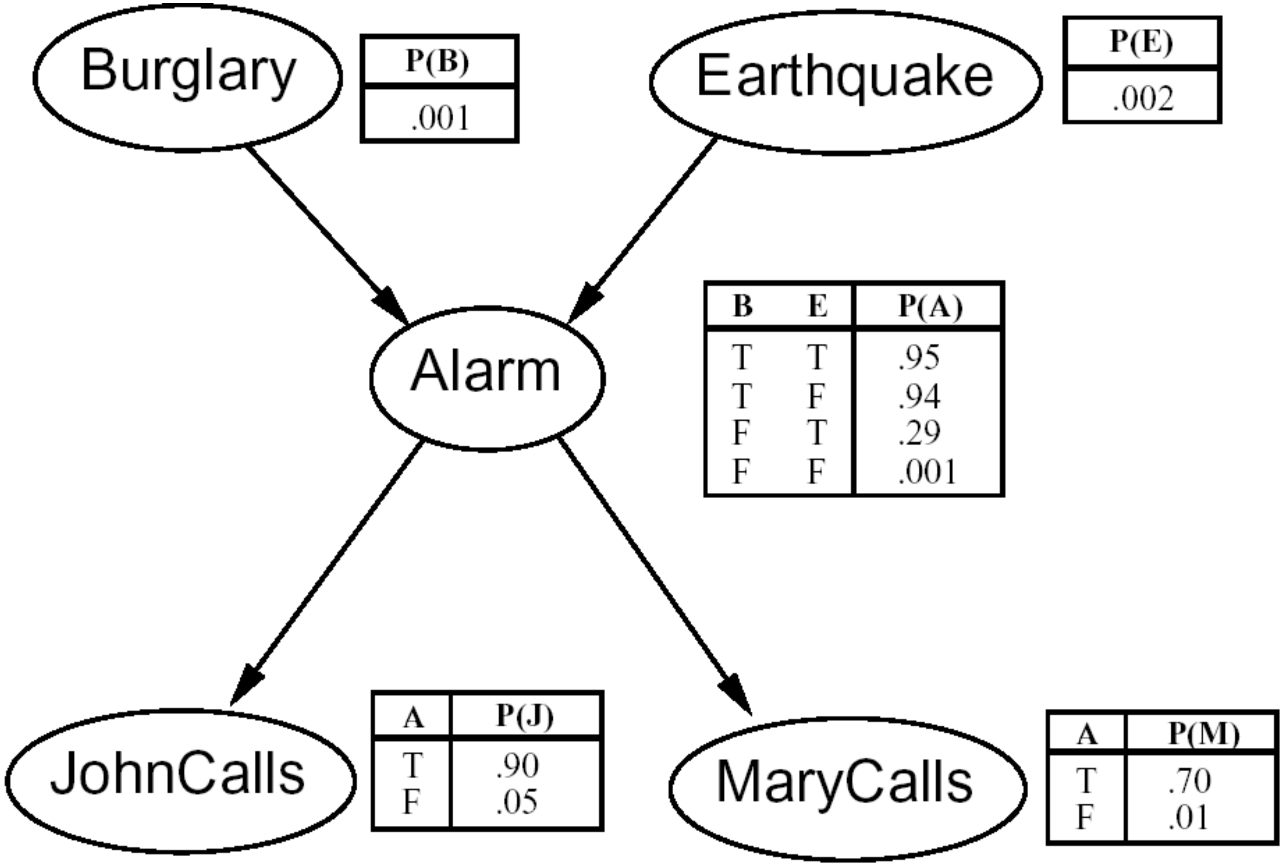
- or has not called, we would like to estimate the probability of a burglary. The Bayesian network for this domain appears in Figure 14.2.

The network

I´m at work, John calls to say my alarm is ringing, Mary doesn´t call. Is there a burglary?

5 Variables

network topology reflects *causal* knowledge



In the CPTs, the letters B, E, A, J, and *M* stand for *Burglary*, *Earthquake*, *Alarm*, *JohnCalls*, and *MarCalls* *re*spectively.

Calculations on the belief network

Using the network in the example, suppose you want to calculate:

$P(A = true, B = true, C = true, D = true)$

$= P(A = true) * P(B = true \mid A = true) *$

$P(C = true \mid B = true)\ P(D = true \mid B = true)$

$= (0.4)*(0.3)*(0.1)*(0.95)$

This is from the graph structure

These numbers are from the conditional probability tables

So let's see how you can calculate P(John called) if there was a burglary?

- Inference from effect to cause; Given a burglary, what is P(J|B)?

$$P(J \mid B) = ?$$

$$P(A \mid B) = P(B)P(\neg E)(0.94) + P(B)P(E)(0.95)$$

$$P(A \mid B) = 1(0.998)(0.94) + 1(0.002)(0.95)$$

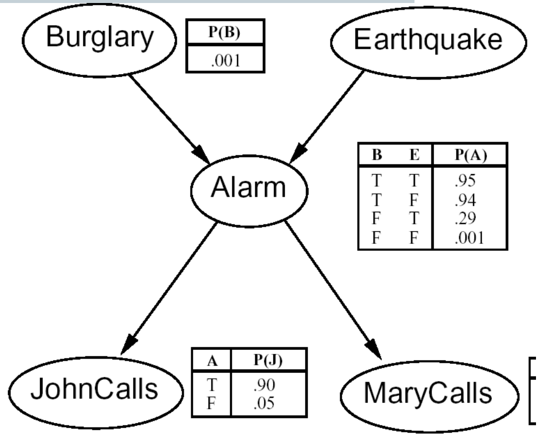
$$P(A \mid B) = 0.94$$

$$P(J \mid B) = P(A)(0.9) + P(\neg A)(0.05)$$

$$P(J \mid B) = (0.94)(0.9) + (0.06)(0.05)$$

$$= 0.85$$

- Can also calculate P (M|B) = 0.67



Why Bayesian Networks?

Bayesian Probability represents the degree of belief in that event while Classical Probability (or frequents approach) deals with true or physical probability of an event

•**Bayesian Network**

•Handling of Incomplete or missing Data Sets
•Over-fitting of data can is avoidable when using Bayesian networks and Bayesian statistical methods.

Limitations of Bayesian Networks

Typically require initial knowledge of many probabilities…quality and extent of prior knowledge play an important role

- Significant computational cost(NP hard task)
- Unanticipated probability of an event is not taken care of.

Decision Network

- Decision making under uncertainty
  - Expected utility
    - Utility theory and rationality
  - Utility functions
  - Decision networks
    - Value of information

Uncertain Outcome of Actions

- Some **actions may have uncertain outcomes**
  - Action: spend $10 to buy a lottery which pays $10,000 to the winner
  - Outcome: {win, not-win}
- Each **outcome is associated with some merit (utility)**
  - Win: gain $9990
  - Not-win: lose $10
- There is a probability distribution associated with the outcomes of this action (0.0001, 0.9999).
- Should I take this action?

Expected Utility

- Random variable X with n values $x_1,\ldots,x_n$ and distribution $(p_1,\ldots,p_n)$
  - X is the outcome of performing action A (i.e., the state reached after A is taken)
- Function U of X
  - U is a mapping from states to numerical utilities (values)
- The expected utility of performing action A is

$$EU[A] = \Sigma_{i=1,\ldots,n}\, p(x_i|A)U(x_i)$$

Probability of each outcome   Utility of each outcome

MEU Principle

- Decision theory: A rational agent should choose the action that maximizes the agent's expected utility
- Maximizing expected utility (MEU) is a normative criterion for rational choices of actions
- Must have **complete** model of:
  - Actions
  - Utilities
  - States

Decision networks

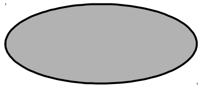
- **Extend Bayes nets** to **handle actions and utilities**
  - a.k.a. **influence diagrams**
- Make use of **Bayes net inference**
- **Useful** application: **Value of Information**

Decision Networks

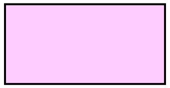
- A **decision network** is a graphical representation of a finite sequential decision problem.

- A decision network is an extension of the Bayes' Net representation that allows us to calculated expected utilities for the actions that we take.

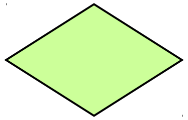
Decision network representation



- **Chance** nodes: random variables, as in Bayes nets



- **Decision** nodes: actions that decision maker can take



- **Utility/value** nodes: the utility of the outcome state.

Decision Networks

Decision networks combine Bayesian networks with additional node types for **actions and utilities.** We, **will use airport siting** as an **example**.

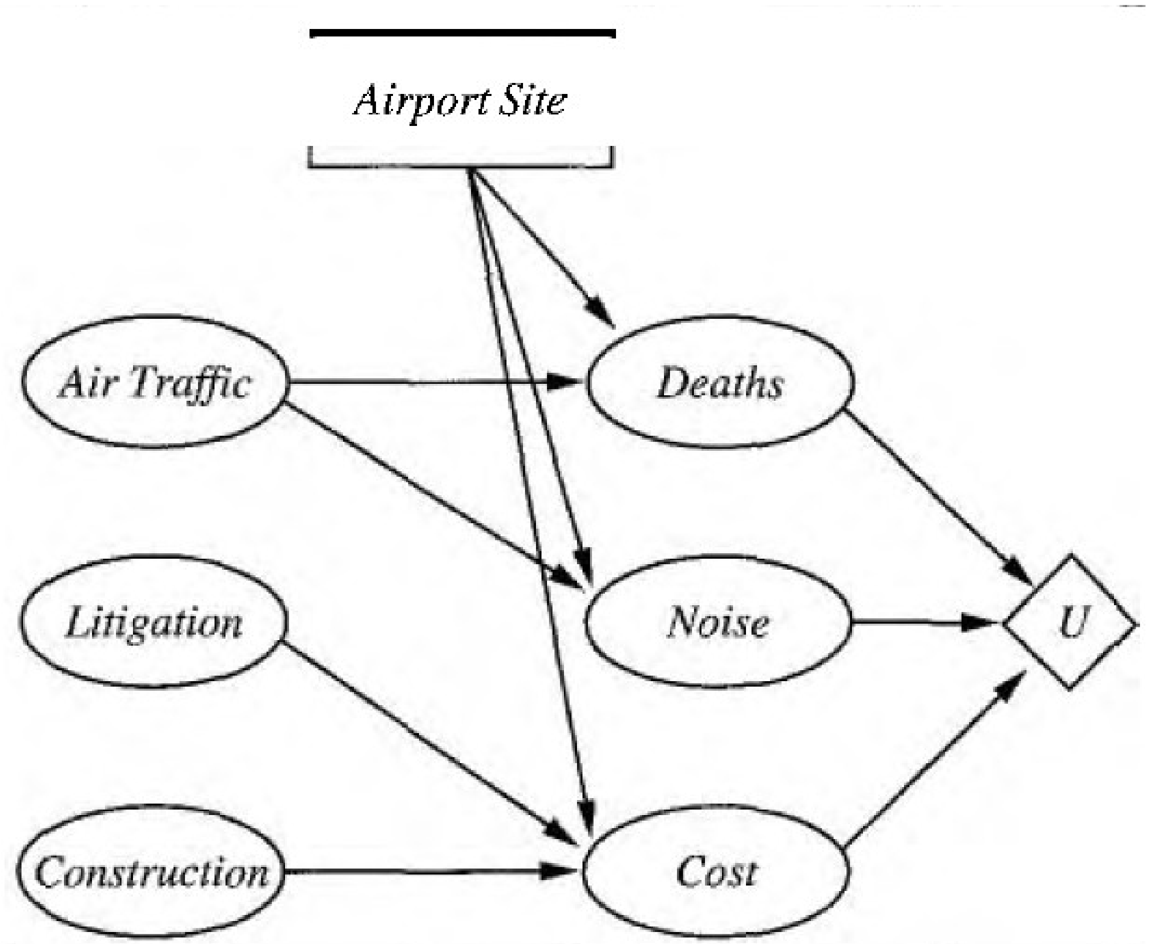


Figure 16.5 A simple decision network for the airport-siting problem

Representing a decision problem with a decision network

Decision network represents information about the **agent's current state, its possible actions, the state that will result from the agent's** action, and the **utility** of that state.

Chance nodes (ovals) represent random variables, **just as they do in Bayes nets**.

The **agent could be uncertain about the construction cost,** the **level of air traffic** and the **potential for litigation**, and the **Deaths, Noise, and total Cost variables**, each of which also depends on the site chosen.

Decision nodes (rectangles) **represent points where the decision-maker has a choice of actions.** In this case, the **Airport site action can take on a different value for each site under consideration**. The choice influences the cost, safety, and noise that will result.

Utility nodes (diamonds) **represent the agent's utility function** .

The utility node has as parents **all variables describing the outcome that directly affect utility.**

Associated with **the utility node is a description of the agent's utility** as a function of the **parent attributes**.

The description could be **just a tabulation of the function**, or **it might be a parameterized additive** or multi linear function.

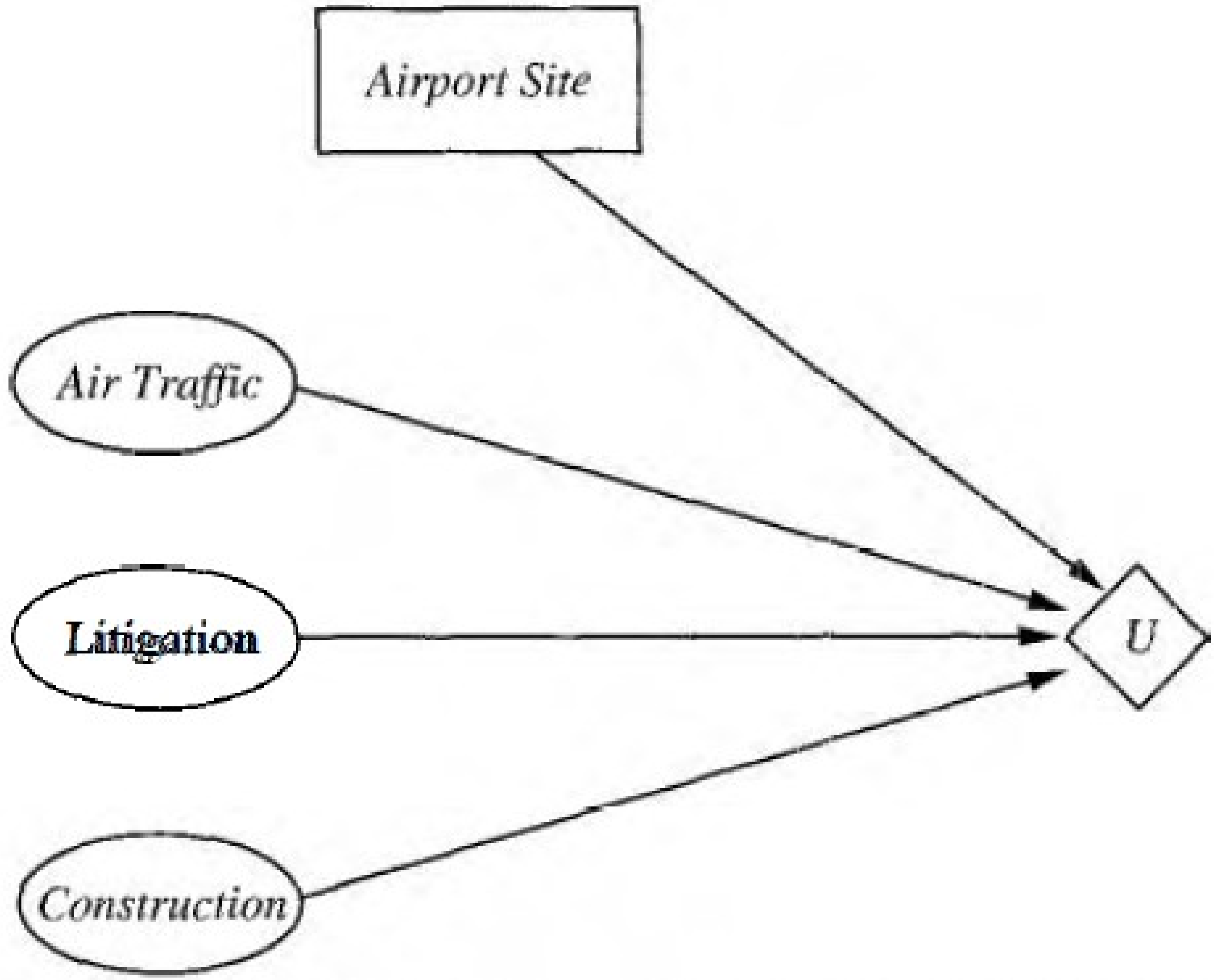


Figure 16.6 A **simplified representation of the airport-siting problem**. Chance nodes corresponding to outcome states have been factored out.

Evaluating decision networks

Actions are selected by evaluating the decision network for each possible setting of the decision node. **Once the decision node is set, it behaves exactly like a chance node that has been set as an evidence variable. The algorithm for evaluating decision networks is the following:**

I. Set the evidence variables for the current state.

2. For each possible value of the decision node;

(a) Set the decision node to that value

b) Calculate the posterior probabilities for the parent nodes of the utility node, using a standard probabilistic inference algorithm.

(c) Calculate the resulting utility for the action.

3. Return the action with the highest utility.

Evaluating decision networks

- **Set the evidence variables for the current state**.
- For each **possible value of the decision node** (assume just one):
  - **Set the decision node to that value**.
  - **Calculate the posterior probabilities for the parent nodes of the utility node**, using BN inference.
  - **Calculate the resulting utility for the action**.
- **Return the action with the highest utility**.

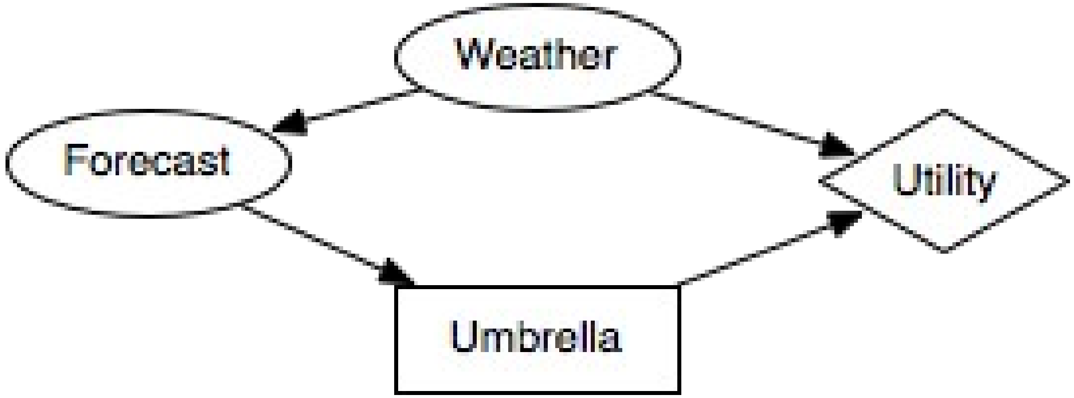


Figure 9.7: Decision network for decision of whether to take an umbrella

Example 9.11: Figure 9.7 shows a simple decision network for a decision of whether the **agent should take an umbrella when it goes out. The agent's utility depends on the weather and whether it takes an umbrella.** However**, it does not get to observe the weather. It only gets to observe the forecast**. The forecast probabilistically depends on the weather.

- Designer must specify the domain for each random variable and the domain for each decision variable.

- Suppose the random variable *Weather* has domain *{norain, rain}*,
- the **random variable *Forecast*** has domain *{sunny, rainy, cloudy}*,

- **Decision variable *Umbrella*** has domain *{takeIt, leav eIt}*.

- There is **no domain associated with the utility node**.
- The designer also must specify the probability of the random variables given their parents.

- Suppose *P(Weather)* is defined by *P(Weather=rain)=0.3*.

P(Forecast | Weather) is given by

Weather	*Forecast*	Probability
norain	*sunny*	0.7
norain	*cloudy*	0.2
norain	*rainy*	0.1
rain	*sunny*	0.15
rain	*cloudy*	0.25
rain	*rainy*	0.6

Suppose the utility function, *Utility(Weather, Umbrella)*, is

Weather	*Umbrella*	*Utility*
norain	*takeIt*	20
norain	*leaveIt*	100
rain	*takeIt*	70
rain	*leaveIt*	0

There is no table specified for the *Umbrella* decision variable. It is the task of the planner to determine which value of *Umbrella* to select, depending on the forecast.