ARTIFICIAL INTELLIGENCE

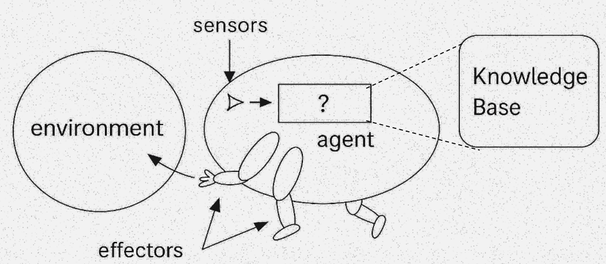
The Very Idea

Vasant G. Honavar
Dorothy Foehr Huck and J. Lloyd Huck Chair in Biomedical Data Sciences and Artificial Intelligence
Professor of Informatics and Intelligent Systems
Professor of Data Sciences and Artificial Intelligence Methods and Applications Undergraduate Programs
Professor of Informatics, Computer Science, Bioinformatics & Genomics and Neuroscience, and Public Health Sciences Graduate Programs
Director, Artificial Intelligence Research Laboratory
Director, Center for Artificial Intelligence Foundations and Scientific Applications
Director of Strategic Initiatives, Institute for Computational and Data Sciences
Pennsylvania State University

vhonavar@psu.edu
http://faculty.ist.psu.edu/vhonavar
http://ailab.ist.psu.edu

On Representing and Reasoning Under Uncertainty

- Intelligent behavior requires knowledge about the world
- In logic, we can express facts that are either true or false in the world
- Given facts that are true in the world, we can infer other facts that logically follow from what we know to be true in the world
- What if we are uncertain about what is true in the world?



On Representing and Reasoning Under Uncertainty

- What if we are uncertain about what is true in the world?
- We no longer have facts that are true or false with certainty
- We need a language to express and reason with uncertainty

Representing and Reasoning under Uncertainty

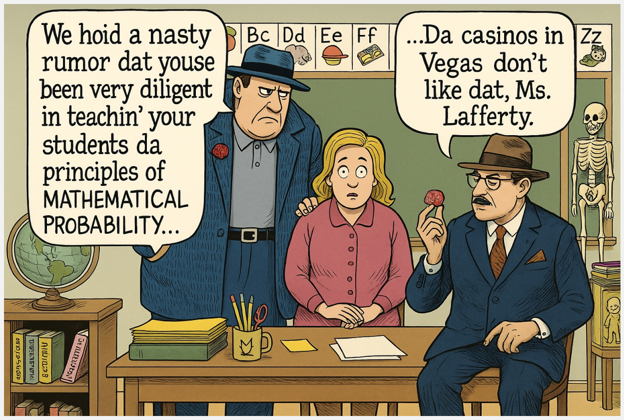
- Probability Theory provides a framework for representing and reasoning under uncertainty
  - Represent beliefs about the world as sentences (much like in propositional logic)
  - Associate probabilities with sentences
  - Reason by manipulating sentences according to sound rules of probabilistic inference
  - Results of inference are probabilities associated with conclusions justified by beliefs and data (observations)
- Allows agents to substitute thinking for acting in the world

Representing and Reasoning under Uncertainty

- Beliefs:
  - If Oksana studies,
    - there is an 60% probability that she will pass the test and
    - a 40 percent probability that she will not.
  - If Oksana does not study,
    - there is 20% percent probability that she will pass the test and
    - 80% probability that she will not.
- Observation: Oksana did not study.
- Example probabilistic reasoning task:
  - What is the chance that Oksana will pass the test?
  - What is the chance that she will fail?

Representing and Reasoning under Uncertainty

- Acting in an uncertain world is akin to gambling



- Agents that can't reason with probabilities will lose at gambling
- Casinos would be better off if no one understood probability

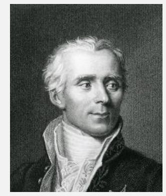
Origins of probability


Girolamo Cardano


Blaise Pascal

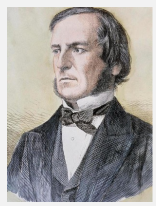

Pierre Fermat

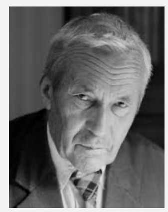

Christiaan Huygens


Pierre Laplace


Thomas Bayes


Charles S. Peirce


George Boole


Andrei Kolmogorov

On Reasoning under uncertainty

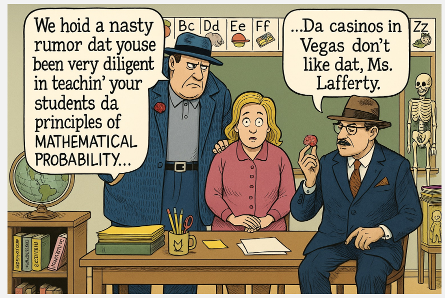
- The excitement of a gambler in making a bet is equal to the amount he might win times the probability of winning it – Blaise Pascal
- All human affairs rest upon probabilities, and the same thing is true everywhere. – CS Peirce
- The theory of probabilities is nothing but common sense reduced to calculus – Pierre Laplace
- Probability is expectation founded upon partial knowledge – George Boole
- The epistemological value of probability theory is based on the fact that chance phenomena, considered collectively and on a grand scale, create non-random regularity – Andrei Kolmogorov

Probability Theory as a Knowledge Representation

- Ontological commitments (what can we talk about?)
  - Propositions that represent the agent's beliefs about the world
- Epistemological Commitments (what can we believe?)
  - What is the probability that a given proposition true (given the beliefs and observations)?
- Syntax
  - Much like propositional logic
- Semantics
  - Relative frequency interpretation
  - Bayesian interpretation
- Inference
  - Based on laws of probability

Representing and Reasoning under Uncertainty

- Acting in an uncertain world is gambling
- Agents that don't use probabilities will lose to those who do
- Probabilities can be updated from observations
- Bayes' rule specifies how to combine data and prior knowledge.



Flavors of uncertainty

- In a deterministic world, uncertainty may be due to
  - Laziness: failure to enumerate exceptions, qualifications, etc. that may be too numerous to state explicitly
  - Sensory limitations
  - Ignorance: lack of relevant facts etc.
- In a stochastic world, uncertainty may be due to
  - Inherent uncertainty (as in quantum physics)

Probability theory is agnostic about the source of uncertainty

Representing and Reasoning under Uncertainty

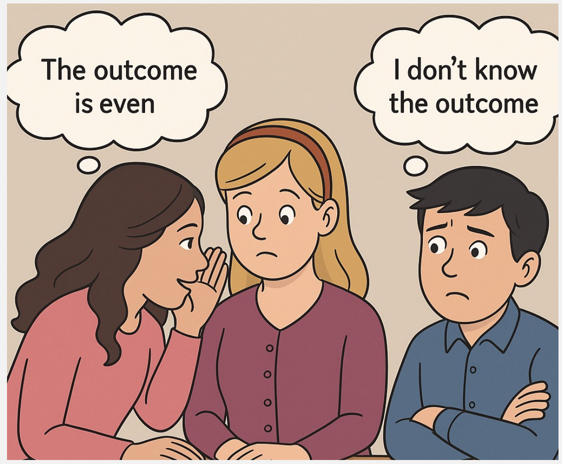
- Probability theory generalizes propositional logic
  - Probabilities lie in the interval [0,1] as opposed to being 0 or 1 (exclusively) in propositional logic
  - Belief in proposition f can be quantified by the probability of f, a number between 0 and 1
    - When we say the probability of f is 0, we mean that f is believed to be definitely False.
    - When we say the probability of f is 1, we mean that f is believed to be definitely True.
    - In general, the probability of any proposition is between 0 and 1 (inclusive).
- Probability is a measure of an agent's ignorance.

States of the world

- From the point of view of an agent Bob who can sense
  - only 3 colors – red, green, blue and
  - only 2 shapes – square and circle
- The world with a single object in it can be in only one of 6 states
  - (red, □), (red, ○), (green, □), (green, ○)(blue, □), (blue, ○)
- A world state is a complete specification of the state of the agent' s world (modulo the agent's sensors)
- Relative to the agent's sensing abilities, the world states are
  - mutually exclusive and
  - exhaustive

Probability semantics – Subjective measure of belief

- Suppose there are 3 agents – Oksana, Cornelia, Jun, in a world where a fair dice has been tossed.
- Oksana observes that the outcome is a "6" and whispers to Cornelia that the outcome is "even"
- Neither Oksana nor Cornelia tell Jun nothing about the outcome.

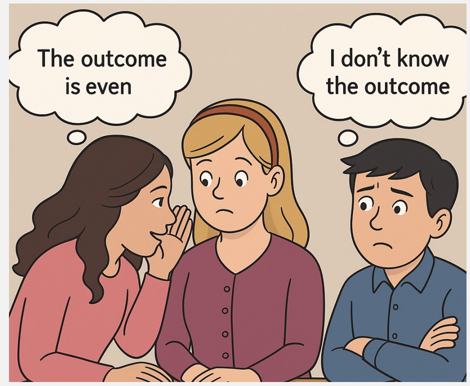


Probability semantics – Subjective measure of belief

- Set of possible mutually exclusive and exhaustive world states at the outset = {1, 2, 3, 4, 5, 6}
- Set of possible world states from Oksana's perspective based on what she knows = {6}
- Set of possible states of the world from Cornelia's perspective based on what she was told by Oksana = {2, 4, 6}
- Probability is a measure of belief over possible worlds given what an agent knows.

Probability as a subjective measure of belief

- Probability is a measure of belief over possible worlds given what an agent knows.
- Because Oksana, Cornelia, and Jun have differing amounts of information about the actual state of the world
    - Oksana knows the actual state of the world.
    - Cornelia has partial information about the actual state of the world.
    - Jun has no information about the actual state of the world



They must assign different probabilities to the possible states of the world based on what they know

Probability as a subjective measure of belief

- Probability is a measure over possible worlds, not ruled out by the available evidence
- When the evidence available differs across agents, so does the probability measure over possible worlds from the agent's perspective

$$Possibleworlds_{Oksana} = \{6\}, Possibleworlds_{Cornelia} = \{2,4,6\}$$
$$Possibleworlds_{Jun} = \{1,2,3,4,5,6\}$$

$$\Pr_{Oksana}(worldstate = 6) = 1$$
$$\Pr_{Cornelia}(worldstate = 6) = \frac{1}{3}$$
$$\Pr_{Jun}(worldstate = 6) = \frac{1}{6}$$

Oksana, Cornelia, and Jun assign different probabilities to the same world state because of differences in what they know

Probability as a subjective measure of belief

- Probability is a measure over possible worlds, not ruled out by the available evidence
- When the evidence available differs across agents, so do the subjective probabilities assigned by each of them to different world states
- In our example, given the evidence Oksana has through direct observation, she must assign probability 1 to the outcome 6 and 0 to all other outcomes
- Given the evidence communicated to Cornelia by Oksana, Cornelia must assign probability of 0 for the three odd outcomes and probability 1/3 to the three even outcomes
- Given what Jun knows, all he can do is to maintain his current belief that each of the 6 possible outcomes is equally likely.

Syntax of probability theory

Basic element: random variable

- Similar to propositional logic
- Possible worlds defined by assignment of values to random variables.
- Domain of values of a random variable must be exhaustive and mutually exclusive
- For example, the domain of $Weather$ may be $\{Sunny, Cloudy\}$
- Atomic propositions correspond to assignment of values to a random variable
  - $(Weather\ =\ Sunny)$ abbreviated as $Sunny$ takes values $True$ or $False.$
  - When $Sunny$ is True, $\neg Sunny$ is $False$ and vice versa.
- Complex sentences are formed from atomic propositions and standard logical connectives

Syntax and Semantics

- Possible world: A complete specification of the state of the world about which the agent is uncertain
- Possible worlds correspond to possible states of affairs described by the propositions (much like in the case of propositional logic)

  E.g., if the world consists of only two Boolean propositions Red and $Rose$, then there are 4 distinct possible worlds:

$$Red = False \wedge Rose = False$$
$$Red = False \wedge Rose = True$$
$$Red = True \wedge Rose = False$$
$$Red = True \wedge Rose = True$$

- Possible worlds are mutually exclusive and exhaustive

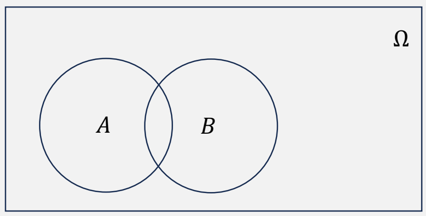
Possible world semantics

- Ω is the set of possible worlds.
- A random variable is a function over subsets of possible worlds
- A random variable with a countable (or finite) range is a discrete random variable.
- A variable with range $\{True, False\}$ is a Boolean random variable.
- $\omega \vDash (X = v)$ means the random variable X has value v in world ω.
- $\omega \vDash (X > v)$ means the random variable X is greater than v in world ω.
- Logical connectives have their standard meaning:
  - $\omega \vDash \alpha \wedge \beta$ if $\omega \vDash$ AND $\omega \vDash \beta$
  - $\omega \vDash \alpha \vee \beta$ if $\omega \vDash \alpha$ OR $\omega \vDash \beta$
  - $\omega \nvDash \neg\alpha$ (ω does not entail α) if α is not $True$ in ω

ω |= ε ↑θ if ω |= ε or ω |= θ ω |= ¬ε if ω ↓/= ε

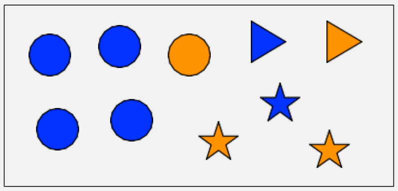
Possible worlds semantics

- Probability is a measure μ over sets of possible worlds
- $\mu(\Omega) = 1$ (
- $\mu(A \vee B)$= $\mu(A) + \mu(B) - \mu(A \wedge B)$
- $\mu(\emptyset)$= 0



- $P(\alpha) = \mu(\{\omega | \omega \vDash \alpha\})$

Possible Worlds Semantics



- There are 10 samples as shown, each with measure 0.1

- What is the probability of circle?
- What is the probability of a star?
- What is the probability of a triangle?
- What is the probability of orange?
- What is the probability of blue?
- What is the probability of orange circle?

Probability as a Measure over Possible worlds

- Associated with each possible world is a measure.
- When there are only a finite number of possible worlds, the measure of the world ω, denoted by $\mu(\omega)$ has the following properties:

$$\forall \omega \in \Omega,\ 0 \le \mu(\omega)$$
$$\sum_{\omega \in \Omega} \mu(\omega) = 1$$

The probability of the state of affairs described by a sentence s, written as $P(s)$ is the sum of the measures of the possible words (models) in which s is True.

$$P(s) = \sum_{\omega \models s} \mu(\omega)$$

Probability as a measure over possible worlds

- Suppose I have two coins – one a normal fair coin, and the other with 2 heads.
- I pick a coin at *random* and toss it. What is the probability that the outcome is a head?

$$\Omega = \{(Fair, H), (Fair, T), (Rigged, H), (Rigged, T)\}$$

$$\mu = \left\{\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0\right\}$$

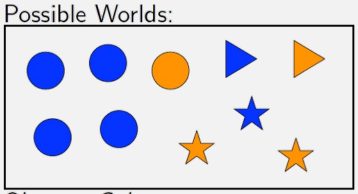
$$P(H) = \sum_{\omega \models H} \mu(\omega) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$$

What is $P\big((H \vee T) \wedge Fair\big)$?

Evidence rules out some possible worlds

- There are 10 possible worlds as shown, each with measure 0.1
- A given piece of evidence *e* selects subset of possible worlds that are compatible with *e*

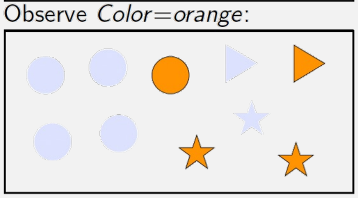
Possible Worlds:



$P(Shape{=}star) = 0.3$

$P(Shape{=}circle) = 0.5$

Observe $Color{=}orange$:

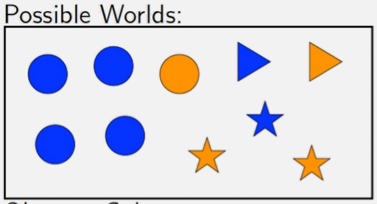


$P(Shape{=}star \mid Color{=}orange) = 0.5$

Evidence rules out some possible worlds

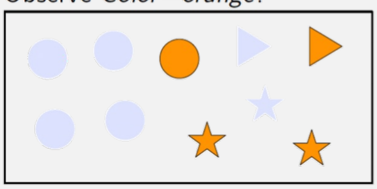
- A given piece of evidence e selects subset of possible worlds that are compatible with e

Possible Worlds:



$$P(Shape{=}star) = 0.3$$
$$P(Shape{=}circle) = 0.5$$

Observe $Color{=}orange$:



$$P(Shape{=}star \mid Color{=}orange) = 0.5$$
$$P(Shape{=}circle \mid Color{=}orange) = 0.25$$

Revising beliefs in response to evidence

- Conditional probability governs the revision of an agent's beliefs in response to new information
- An agent builds a picture of the world based on what it knows (analogous to axioms of a propositional knowledge base)
- The agent's model assigns prior probabilities to possible worlds (and hence to all sentences that describe states of affairs of the world)
- As it receives further information (analogous to facts or assumptions in a propositional knowledge base), it needs to revise those probabilities.
- If $P(h)$ is the prior probability of sentence h, $P(h|e)$ denotes the conditional (posterior) probability of h given evidence e

Evidence rules out some possible worlds

- A given piece of evidence e rules out all possible worlds that are incompatible with e or selects the possible worlds in which e is $True$. Evidence e *induces* a new measure μ_e.

$$\mu_e(\omega) = \begin{cases} \frac{1}{P(e)} \mu(\omega) \text{ if } \omega \models e \\ 0 \text{ if } \omega \not\models e \end{cases}$$

$$P(h|e) = \sum_{\omega \models h} \mu_e(\omega) = \frac{1}{P(e)} \sum_{\omega \models h \wedge e} \mu(\omega) = \frac{P(h \wedge e)}{P(e)}$$

Exercise

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$

$$P(flu \wedge sneeze) = \sum_{Flu=True,\, Sneeze=True} P(flu, sneeze, snore)$$

$$= 0.064 + 0.096 = 0.16$$

Exercise

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$

Exercise

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(\mathit{flu} \wedge \mathit{sneeze})$ 0.16

(b) $P(\mathit{flu} \wedge \neg \mathit{sneeze})$

$$P(flu \wedge \neg sneeze) = \sum_{Flu=True,\, Sneeze=True} P(flu, \neg sneeze, snore)$$

$$= 0.016 + 0.024 = 0.04$$

Exercise

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$

Exercise

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ 0.2

(d) $P(sneeze \mid flu)$

$$P(sneeze|flu) = \frac{P(flu \wedge sneeze)}{P(flu)}$$

$$= \frac{0.16}{0.2}$$

$$= 0.8$$

Exercise

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ 0.2

(d) $P(sneeze \mid flu)$ 0.8

Exercise

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$

Exercise

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$

Exercise

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$

Exercise

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$ 0.4

(h) $P(sneeze \mid flu \wedge snore)$

Exercise

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \land sneeze)$ 0.16

(b) $P(flu \land \neg sneeze)$ 0.04

(c) $P(flu)$ 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \land sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$ 0.4

(h) $P(sneeze \mid flu \land snore)$ 0.8

(i) $P(flu \mid sneeze \land snore)$

Exercise

Flu	*Sneeze*	*Snore*	*P*
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

What is:

(a) $P(flu \wedge sneeze)$ 0.16

(b) $P(flu \wedge \neg sneeze)$ 0.04

(c) $P(flu)$ 0.2

(d) $P(sneeze \mid flu)$ 0.8

(e) $P(\neg flu \wedge sneeze)$ 0.24

(f) $P(sneeze)$ 0.4

(g) $P(flu \mid sneeze)$ 0.4

(h) $P(sneeze \mid flu \wedge snore)$ 0.8

(i) $P(flu \mid sneeze \wedge snore)$ 0.4

Chain rule

$$P(h \mid e) = \frac{P(h \wedge e)}{P(e)}$$

Hence

$$P(h \wedge e) = P(e)\, P(h \mid e)$$

$$P(h_1 \wedge \cdots h_{n-1} \wedge h_n) = P(h_n \mid h_1 \wedge h_2 \cdots \wedge h_{n-1}) P(h_{n-1} \wedge \cdots \wedge h_1)$$
$$= P(h_n \mid h_1 \wedge h_2 \cdots \wedge h_{n-1})\, P(h_{n-1} \mid h_1 \wedge h_2 \cdots \wedge h_{n-2})\, P(h_{n-2} \wedge \cdots \wedge h_1)$$
$$\vdots$$
$$= P(h_1) \prod_{i=2}^{n} P(h_i \mid h_1 \wedge \cdots \wedge h_{i-1})$$

Expected Value

- The expected value of numerical random variable X with respect to probability P is
  $\mathbb{E}_P(X) = \sum_{v \in Domain(X)} vP(v)$

  when the domain is X is finite or countable.
- When the domain is continuous, the sum becomes an integral.
- If α is a proposition, by representing *True* as 1 and *False* as 0, we have
  $\mathbb{E}_P(\alpha) = P(\alpha)$, the probability that α is $True$.

Bayes Theorem

$$P(h \mid e) = \frac{P(h \land e)}{P(e)}$$

$$P(e \mid h) = \frac{P(e \land h)}{P(h)}$$

$P(h \land e) = P(e \land h)$ because $\land$ is commutative

So $P(h \mid e)P(e) = P(e \mid h)\, P(h)$

$$P(h \mid e) = \frac{P(e \mid h)\, P(h)}{P(e)}$$

Why is Bayes Theorem useful?

- Suppose we have causal knowledge
  - $P(symptom|\ \ disease)$
  - $P(alarm\,|\,fire\ )$
  - $P(pass\,|\,study)$
- We want to do evidential reasoning, that is, calculate
  - $P(disease\ |symptom)$
  - $P(fire\,|\,alarm)$
  - $P(study\,|\,pass)$

Example: Cancer diagnosis

Does patient have cancer or not given a positive test result?

- A patient takes a lab test and the result comes back positive.
- The test returns
  - a correct positive result in only 98% of the cases in which the disease is actually present, and
  - a correct negative result in only 97% of the cases in which the disease is not present.
- Moreover, only .8% of the entire population have this cancer.

Cancer diagnosis

Does patient have cancer or not given a positive test?

$$P(cancer) = 0.008 \qquad P(\neg cancer) = 0.992$$

$$P(+ \mid cancer) = 0.98 \qquad P(- \mid cancer) = 0.02$$

$$P(+ \mid \neg cancer) = 0.03 \qquad P(- \mid \neg cancer) = 0.97$$

$$P(+) = P(+|cancer)P(cancer) + P(+|\neg cancer)P(\neg cancer)$$

$$P(+) = (0.98)(0.008) + (0.03)(0.992) = 0.0078 + 0.0298$$

$$P(cancer|+) = \frac{P(+|cancer)P(cancer)}{P(+)}$$

$$P(cancer|+) = \frac{0.0078}{0.0078 + 0.0298} = 0.21$$

$$P(\neg cancer|+) = 1 - P(cancer|+) = 1 - 0.21 = 0.79$$

Accident investigation

A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

- 85% of the cabs in the city are Green and 15% are Blue.
- A witness identified the cab as Blue. The court tested the reliability of the witness in the circumstances that existed on the night of the accident and concluded that the witness correctly identifies each one of the two colors 80% of the time and failed 20% of the time.

What is the probability that the cab involved in the accident was Blue?

[From D. Kahneman, Thinking Fast and Slow, 2011, p. 166.]

Accident investigation

$$P(\neg Blue) = 0.85;\ P(Blue) = 0.15$$

$$P(BlueID \mid Blue) = 0.8;\ P(\neg BlueID | Blue) = 0.2$$
$$P(\neg BlueID | \neg Blue) = 0.8;\ P(BlueID \mid \neg Blue) = 0.2$$
$$BlueID = True$$

$$P(Blue|BlueID) = \frac{P(BlueID \mid Blue)P(Blue)}{P(BlueID)}$$

$$= \frac{P(BlueID \mid Blue)P(Blue)}{P(BlueID \mid Blue)P(Blue) + P(BlueID \mid \neg Blue)P(\neg Blue)}$$

$$= \frac{0.8 \times 0.15}{0.8 \times 0.15 + 0.20 \times 0.85} = 0.4138$$

Marginalization

- We know $P(x, y)$, what is $P(x)$?
- We can use the law of total probability, why?

$$P(x) = \sum_{y} P(x, y) = \sum_{y} P(y)P(x|y)$$

Note here that $P(X)$ denotes a probability distribution. If X takes 3 values, $P(X)$ has 3 elements, one for each of the 3 values of X. We use $P(x)$ to denote the probability $P(X = x)$

Marginalization

- We know $P(X, Y, Z)$, what is $P(x)$?

$$P(x) = \sum_{y,z} P(x, y, z) = \sum_{y} P(y, z) P(x|y, z)$$

Marginalization

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

Given $P(Flu, Sneeze, Snore)$
What is $P(Flu)$?

$$P(flu = true)$$
$$= \sum_{sneeze,snore} P(flu = true, sneeze, snore)$$
$$= 0.064 + 0.096 + 0.016 + 0.024$$
$$= 0.2$$

$$P(flu = false)$$
$$= \sum_{sneeze,snore} P(flu = false, sneeze, snore)$$
$$= 0.096 + 0.144 + 0.224 + 0.336$$
$$= 0.8$$

Flu	$P(flu)$
true	0.2
false	0.8

Bayes Rule

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

Given $P(Flu, Sneeze, Snore)$
What is $P(Flu, Sneeze)$?

Flu	$Sneeze$	$P(flu, sneeze)$
true	true	0.16
true	false	0.04
false	true	0.24
false	false	0.56

Bayes Rule

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

Given $P(Flu, Sneeze, Snore)$
What is $P(Flu, Sneeze)$?

Flu	$Sneeze$	$P(flu, sneeze)$
true	true	
true	false	
false	true	
false	false	

$$P(flu, sneeze) = \sum_{snore} P(flu, sneeze, snore)$$

$$P(flu = true, sneeze = true) = \sum_{snore} P(flu = true, sneeze = true, snore)$$

$$= 0.064 + 0.096 = 0.16$$

Conditioning

Given $P(Flu, Sneeze, Snore)$
What is $P(Sneeze|Flu)$?

Flu	*Sneeze*	*Snore*	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

Flu	$P(flu)$
true	0.2
false	0.8

Flu	$Sneeze$	$P(flu, sneeze)$
true	true	0.16
true	false	0.04
false	true	0.24
false	false	0.56

Conditioning

Given $P(Flu, Sneeze, Snore)$
What is $P(Sneeze|Flu)$?

Flu	$Sneeze$	$Snore$	μ
true	true	true	0.064
true	true	false	0.096
true	false	true	0.016
true	false	false	0.024
false	true	true	0.096
false	true	false	0.144
false	false	true	0.224
false	false	false	0.336

Flu	$P(flu)$
true	0.2
false	0.8

$$P(sneeze|flu) = \frac{P(flu, sneeze)}{P(flu)}$$

Flu	$Sneeze$	$P(sneeze\|flu)$
true	true	0.8
true	false	0.2
false	true	0.3
false	false	0.7

Flu	$Sneeze$	$P(flu, sneeze)$
true	true	0.16
true	false	0.04
false	true	0.24
false	false	0.56

Bayes Rule

Given $P(Flu)$, P(Sneeze), and
Given $P(Sneeze|Flu)$, what is $P(flu|sneeze)$?

flu	$P(flu)$
true	0.2
false	0.8

$sneeze$	$P(sneeze)$
true	0.4
false	0.6

flu	$sneeze$	$P(sneeze\|flu)$
true	true	0.8
true	false	0.2
false	true	0.3
false	false	0.7

flu	$sneeze$	$P(flu\|sneeze)$
true	true	
true	false	
false	true	
false	false	

Bayes Rule

Given $P(Flu)$, P(Sneeze), and
Given $P(Sneeze|Flu)$, what is $P(flu|sneeze)$?

flu	$P(flu)$
true	0.2
false	0.8

$sneeze$	$P(sneeze)$
true	0.4
false	0.6

flu	$sneeze$	$P(flu\|sneeze)$
true	true	0.4
true	false	0.067
false	true	0.6
false	false	0.933

Flu	$Sneeze$	$P(sneeze\|flu)$
true	true	0.8
true	false	0.2
false	true	0.3
false	false	0.7

$$P(flu = true|sneeze = true)$$
$$= \frac{P(sneeze = true|flu = true)P(flu = true)}{P(sneeze = true)}$$
$$= \frac{(0.8)(0.2)}{0.4} = 0.4$$

$\perp$

Independence

- X is independent of Y means that knowing Y does not change our belief about X.
  - $P(X|Y = y) \; = \; P(X)$ for all y
  - $P(X = x, Y = y) \; = \; P(X = x)\, P(Y = y)$
  - The above should hold for all x, y
  - X is independent of Y is written as $X \perp Y$
  - If $X \perp Y$ then $Y \perp X$

Independence

Let X and Y denote the random variable that denote the outcomes two distinct coin tosses

Is X independent of Y?

Independence

Let X and Y denote the random variable that denote the outcomes two distinct tosses of the same coin.

Suppose the coin has memory of the outcome of its previous toss; If the the previous outcome was heads, the new coin comes up tails and vice versa.

Is X independent of Y?

Independence

If X and Y are independent, we have:

$$P(X,Y) = P(X)P(Y)$$

That is, for x in the domain of X and for all y in the domain of Y

$$P(x,y) = P(x)P(y)$$

Independence

x	$P(x)$
0	0.2
1	0.8

y	$P(y)$
1	0.3
2	0.2
3	0.5

Suppose X is independent of Y
Calculate $P(X, Y)$

Independence

x	$P(x)$
0	0.2
1	0.8

y	$P(y)$
1	0.3
2	0.2
3	0.5

If X is independent of Y, we have

x	y	$P(x,y)$
0	1	0.06
0	2	0.04
0	3	0.10
1	1	0.24
1	2	0.16
1	3	0.40

Independence

Let X, Y, and Z are independent

$$P(X,Y) = P(X)P(Y)$$
$$P(Y,Z) = P(Y)P(Z)$$
$$P(X,Z) = P(X)P(Z)$$
$$P(X,Y,Z) = P(X)P(Y)P(Z)$$

The above must hold for all x in the domain of X, for all y in the domain of Y, and for all z in the domain of z.

If *N* random variables are independent, then the joint probability of any subset of them can be written as the product of their individual probabilities.

Conditional Independence

- We say that X and Y are conditionally independent given Z if it is the case that

$$P(X|Y,Z) = P(X|Z) \text{ and } P(Y|X,Z) = P(Y|Z)$$

- In other words, when we condition on Z, knowing X tells me nothing about Y and vice versa.
- Like in the case of unconditional independence, this represents multiple equations for all possible values of the random variables X, Y, and Z

Independence and conditional independence

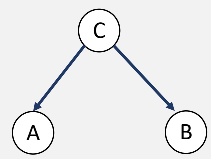
- A department store has a nice stock of umbrellas.
- John and Jane are two unrelated customers who walk into the store.
- Let
  - A be the event John buys an umbrella
  - B be the event that Jane buys an umbrella, and
  - C the event that it is raining when John and Jane enter the store.
- Is A independent of C?
- Is B independent of C?

Independence and conditional independence

- A department store has a nice stock of umbrellas.
- John and Jane are two unrelated customers who walk into the store.
- Let
  - A be the event John buys an umbrella
  - B be the event that Jane buys an umbrella, and
  - C the event that it is raining when John and Jane enter the store.
- Is A independent of C? – No!
- Is B independent of C? – No!
- Are A and B independent?

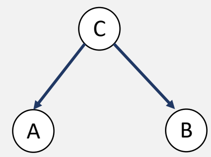
Independence and conditional independence

- A department store has a nice stock of umbrellas.
- John and Jane are two unrelated customers who walk into the store.
- Let
  - A be the event John buys an umbrella
  - B be the event that Jane buys an umbrella, and
  - C the event that it is raining when John and Jane enter the store.
- Are A and B independent?
- No!
- Why?
- Because both A and B aare under the influence of C



Independence and conditional independence

- A department store has a nice stock of umbrellas.
- John and Jane are two unrelated customers who walk into the store.
- Let
  - A be the event John buys an umbrella
  - B be the event that Jane buys an umbrella, and
  - C the event that it is raining when John and Jane enter the store.
- A and B are not (unconditionally) independent.
- What if we condition on C?
- For example, if we know that it is raining when John and Jane enter the store, are A and B independent?
- Knowing C, makes A and B independent
- That is, A is conditionally independent of B given C



Independence and conditional independence

$$P(A|C) = 0.60, P(A|\neg C) = 0.20$$
$$P(B|C) = 0.50, P(B|\neg C) = 0.15$$
$$P(C) = 0.30$$

Suppose A and B are conditionally independent given C

$$P(A) = ?$$
$$P(B) = ?$$
$$P(AB) = ?$$

Are A, B independent?

Independence and conditional independence

$$P(A|C) = 0.60, P(A|\neg C) = 0.20$$
$$P(B|C) = 0.50, P(B|\neg C) = 0.15$$
$$P(C) = 0.30$$

$$P(A) = P(A|C)P(C) + P(A|\neg C)P(\neg C) = (0.6)(0.3) + (0.2)(.7) = 0.32$$
$$P(B) = P(B|C)P(C) + P(B|\neg C)P(\neg C) = (0.5)(0.3) + (0.15)(.7) = 0.255$$
$$P(AB) = P(A, B|C)P(C) + P(A, B|\neg C)P(\neg C)$$

Because A and B are conditionally independent given C

$$P(AB) = P(A|C)P(B|C)P(C) + P(A|\neg C)P(B|\neg C)P(\neg C)$$
$$= (0.60)(0.50)(0.30) + (0.2)(0.15)(0.7)$$
$$= 0.111$$
$$P(A)P(B) = (0.32)(0.255) = 0.0816$$

Are A, B independent? No, because $P(AB) \neq P(A)P(B)$

What does independence or conditional independence buy us?

- Suppose we have N binary random variables.
- In the absence of any other information, what is the size of the joint probability table?
  - We have to specify a probability value for every combination of binary values for the N random variables
  - This requires a table of 2^N entries (actually, $2^N - 1$ because the probabilities must sum up to 1

What does independence or conditional independence buy us?

- Suppose we have N binary random variables.
- In the absence of any other information, the the joint probability table must provide 2^N entries
- What if the N variables are independent?
  - The joint probabilities can be written as products of the probabilities of each of the binary variables
  - For each variable X_i we need $P(x_i)$
  - $P(\neg x_i)$ is simply $1 - P(x_i)$
  - Then, because of independence, we can calculate

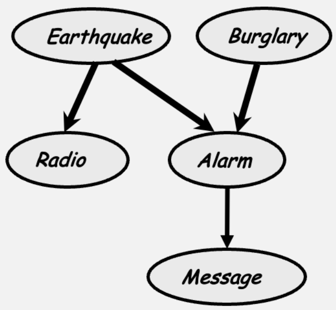
$$P(X_1, X_2, \cdots X_{N-1}, X_N) = P(X_1)\,P(X_2) \cdots P(X_{N-1})\,P(X_N)$$

This requires only N instead of 2^N entries

Bayesian networks

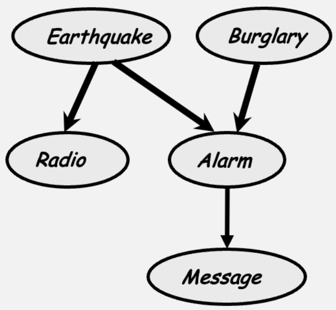
- Bayesian networks provide compact graphical representation of conditional independence assumptions about random variables of interest
  - Nodes represent random variables
  - Directed links represent direct dependencies
  - Each node is conditionally independent of all other nodes given its parents in the graph
  - We only need to specify probability distributions of each node conditioned on its parents in the graph
- The conditional independence assumptions may be based on domain knowledge or learned (from data)
- Conditional independence relations dramatically simplify reasoning under uncertainty

Bayesian networks



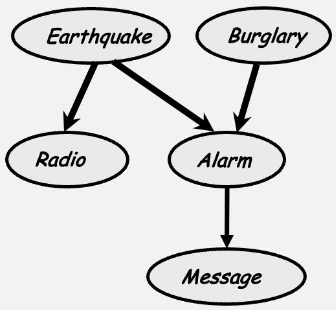
- You have a home in a region of California where burglaries are frequent and earthquakes are rare
- You install a Burglar Alarm
- Unfortunately, the alarm is buggy and may be triggered by Earthquake
- You are on vacation

Bayesian networks



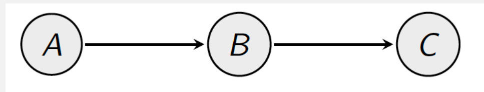
- Earthquake and Burglary are independent
- Alarm may be triggered by Burglary or Earthquake
- Triggered alarm may lead to a cell phone alert message
- Earthquake may be covered by radio news
- How should this impact your worry about burglary at your home?
- You should be less worried about burglary but more concerned about earthquake!

Bayesian networks



- You are on vacation overseas
- You get a message on your cell phone about alarm having gone off
- You are now worried about burglary at your home
- Then you hear news on the radio that there was an earthquake in California

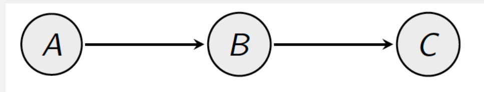
Chain of Mediation



- Rumor A on a given day ($P(A = 1) = 0.5$),
- Person B knows the rumor ($B = 1$) sometimes,
- Person B sometimes spreads rumor to person C ($C = 1$).
- B or C do not invent rumors beyond those propagated by A
- You measure A, B, C for multiple distinct days.
  - Will C be informative about A?
    - Yes, when C knows a rumor, a rumor A definitely occurred
    - $P(A = 1|C = 1) = 1 > P(A = 1) = 0.5$
    - C is not independent of A!

- What happens to this dependence when we only look at days when $B = 1$?

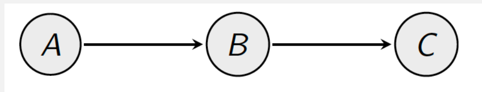
Chain of Mediation



- What happens to this dependence when we only look at days where $B = 1$?

- $P(A = 1 | B = 1) = 1$ because B is truthful and does not invent rumors.
- We are sure there was a rumor when we know $B = 1$
- So $P(A = 1 | B = 1) = 1$ regardless of C
- C is independent of A given B

Chain of Mediation



- What happens to this dependence when we only look at days where $B = 0$?

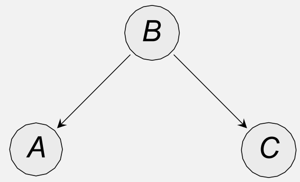
- $P(C = 1 | B = 0) = 0$ because C is truthful and does not invent rumors.
- When we know $B = 0$
- So $P(C = 1 | B = 0) = 0$ regardless of A
- C is independent of A given B

Chain of Mediation

$$A \rightarrow B \rightarrow C$$

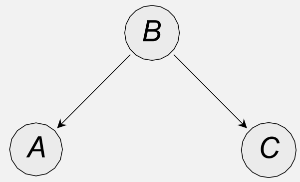
- In chain of mediation:
  - A and C are dependent
  - A and C become independent when conditioned on B
- We say this path is **open** unconditionally, but conditional on the middle node it is **blocked**

Common Cause



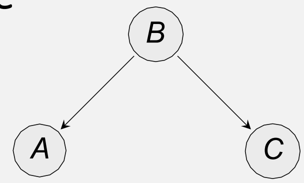
- Example: *B* denotes Season; A, daily icecream sales in the city; and C, daily number of drownings in the city
- Question: Does knowing *C* help you predict *A*?
  - *B* is a common cause of both *A* and *C*
  - *A* and *C* are correlated
  - Knowing the number of drownings is high, implies probably *B* = summer, and that means ice cream sales *C* are relatively high
  - So $P(A) \neq P(A|C)$, A is not independent of C

Common Cause



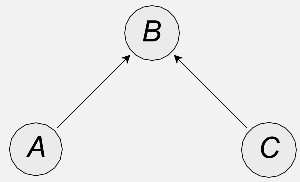
- What if we look at a subset of the data where season is held constant, e.g. *B* = summer?
  - During summer, you see variation in ice cream sales and drownings
  - But once you fix *B*, changes in B cannot influence A or C.
  - So A is independent of C when conditioned on B

Common Cause



- In the common fork graph
- Unconditionally, the path is **open,** making A and C dependent
- Conditional on the middle node, it is **blocked**
- This is exactly like in the chain of mediation, but different "causal story"

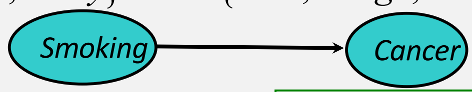
Collider



- Example: *A* and *C* denote independent coin flips (results are 0 or 1).
  - Suppose *B* is the sum of A and B, so 0, 1, or 2
- Can you predict *A* when you know *C*?
  - No! Because *A* is independent of *C*
- But suppose you know one coin *A* = 1 and *B* = 1. Can you then predict the other coin *C*?
  - Yes, it HAS to be 0
- *A* and *C* are independent, but become dependent when conditioned on *B*

Bayesian Networks

$S \in \{no, light, heavy\}$ $C \in \{none, benign, malignant\}$



Smoking	*no*	*light*	*heavy*
P(C=none)	0.96	0.88	0.60
P(C=benign)	0.03	0.08	0.25
P(C=malig)	0.01	0.04	0.15

P(S=no)	0.80
P(S=light)	0.15
P(S=heavy)	0.05

Product Rule

$$P(C,S) = P(C|S)\,P(S)$$

$S\Downarrow \quad C\Rightarrow$	$none$	$benign$	$malignant$
no	0.768	0.024	0.008
$light$	0.132	0.012	0.006
$heavy$	0.035	0.010	0.005

Marginalization

$S\Downarrow \quad C\Rightarrow$	*none*	*benign*	*malig*	total
no	0.768	0.024	0.008	.80
light	0.132	0.012	0.006	.15
heavy	0.035	0.010	0.005	.05
total	0.935	0.046	0.019	

P(Smoke)

P(Cancer)

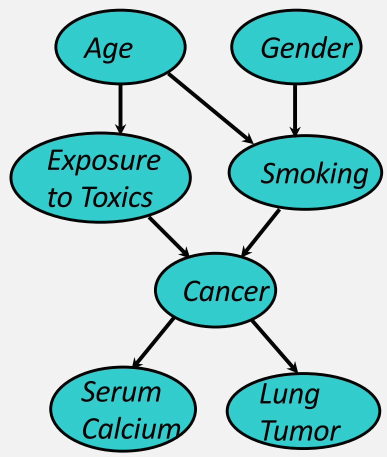
Bayes Rule Revisited

$$P(S \mid C) = \frac{P(C \mid S)P(S)}{P(C)} = \frac{P(C,S)}{P(C)}$$

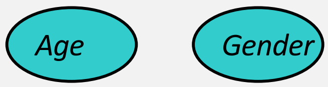
$S\Downarrow \quad C\Rightarrow$	*none*	*benign*	*malig*
no	0.768/.935	0.024/.046	0.008/.019
light	0.132/.935	0.012/.046	0.006/.019
heavy	0.030/.935	0.015/.046	0.005/.019

Cancer=	*none*	*benign*	*malignant*
P(S=no)	0.821	0.522	0.421
P(S=light)	0.141	0.261	0.316
P(S=heavy)	0.037	0.217	0.263

A Bayesian Network



Independence



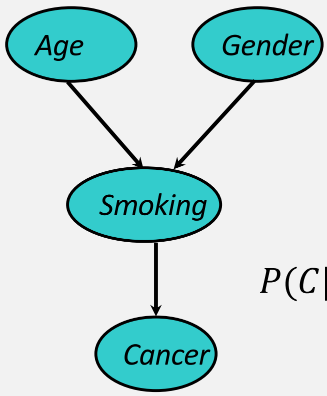
Age and *Gender* are independent.

$$P(A, G) = P(G)P(A)$$

$$P(A|G) = P(A) \quad A \perp G$$
$$P(G|A) = P(G) \quad G \perp A$$

$$P(A, G) = P(G|A)\,P(A) = P(G)P(A)$$
$$P(A, G) = P(A|G)\,P(G) = P(A)P(G)$$

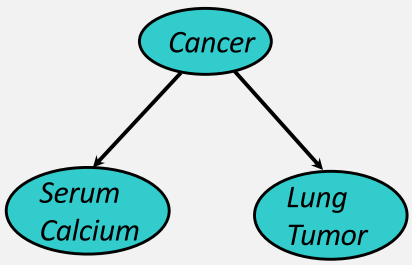
Conditional Independence



Cancer is independent of *Age* and *Gender* given *Smoking*.

$P(C|A,G,S) = P(C|S)$ $C \perp A,G \mid S$

More Conditional Independence

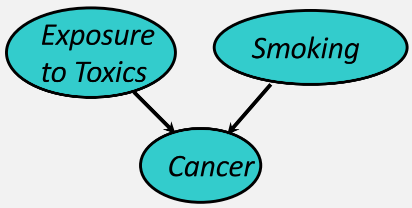


Serum Calcium and *Lung Tumor* are dependent

Serum Calcium is independent of *Lung Tumor*, given *Cancer*

$$P(L|SC,C) = P(L|C)$$

More Conditional Independence: Explaining Away



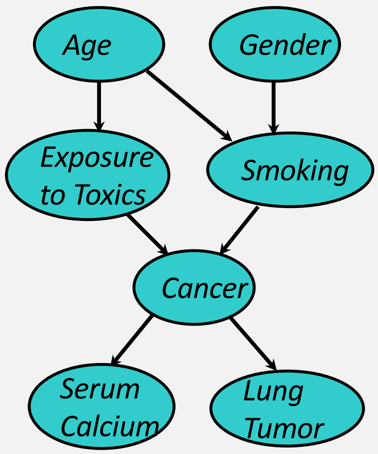
Exposure to Toxics and *Smoking* are independent

$$E \perp S$$

Exposure to Toxics is **dependent** on *Smoking*, given *Cancer*

$$P(E = heavy \mid C = malignant) > P(E = heavy \mid C = malignant, S=heavy)$$

Stitching it all together…

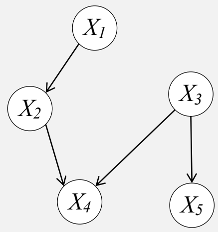


$$P(A, G, E, S, C, L, SC) =$$

$$P(A) \cdot P(G) \cdot$$

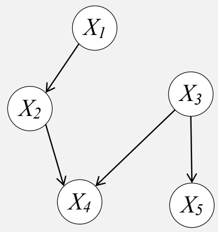
$$P(E \mid A) \cdot P(S \mid A, G) \cdot$$
$$P(C \mid E, S) \cdot$$
$$P(SC \mid C) \cdot P(L \mid C)$$

Bayesian Network



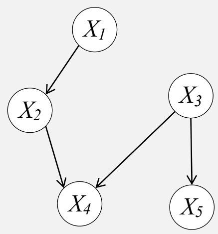
$$P(X_1 \ldots X_n) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$$

Bayesian Network



$$p(X_1, X_2, X_3, X_4, X_5) = p(X_5|X_3)p(X_4|X_2, X_3) p(X_3)p(X_2|X_1)p(X_1)$$

Bayesian Network



$$p(X_1, X_2, X_3, X_4, X_5) = p(X_5|X_3)p(X_4|X_2, X_3) p(X_3)p(X_2|X_1)p(X_1)$$

- A Bayesian Network is a **directed graphical model**
- It consists of a graph *G* and the conditional probabilities **P**
- These two parts full specify the distribution:
  - Qualitative Specification: *G*
  - Quantitative Specification: **P**