Bayesian Epistemology

Stephan Hartmann
London School of Economics

Motivation

- *Bayesian Epistemology* is a quite active research program.
- It has many attractions, and I hope to convince you of its superiority over reliablism…
- Needless to say, there are different approaches to *Bayesian Epistemology*.
- In this lecture and the next, I will present my take on the subject.

Overview

I. Why Bayesianism?
II. Probability and Its Interpretations
III. The Elements of Bayesianism
IV. Success Stories
V. Further Topics and Limitations
VI. Modeling in Science
VII. Bayesian Networks

I. Why Bayesianism?

Two traditions in philosophy of science

1. Normativism
   - Examples: Falsificationism, Bayesianism
   - Typically motivated on *a priori* grounds
2. Descriptivism
   - Examples: Kuhn, naturalized philosophies of science
   - Examine specific case studies that contribute to a better understanding of science

… and their problems

. Normativism

- challenged by insights from the history of science (e.g. Popper and the stability of normal science)
- often "too far away" from real science

2. Descriptivism

- not clear how generalizable insights or normative standards can be provided

Desiderata for a methodology of science

1. It should be *normative* and provide a defensible *general* account of scientific rationality.
2. It should provide a framework to *illuminate* “intuitively correct judgments in the history of science and explains the incorrectness of those judgments that seem clearly intuitively incorrect (and shed light on ‘grey cases’)” (John Worrall)

Goal

- Develop a version of Bayeisanism that does the job.
- This version if Bayesianism will help us to explain features of the methodology of science by fitting it in a normative framework theory.
- Which features are explained:

  - specific episodes, or

  - more general features?

II. Probability and Its Interpretations

- Bayesianism uses probabilities to represent the degree of belief of an agent.
- To understand this better, we have to get clearer about

  (i) what probabilities are, and

  (ii) what probabilities mean.
- Note: Probabilities are *mathematical objects,* which are defined in an *axiomatic* way. How can these objects relate to something in the world, i.e. outside the mathematical realm?

1. Probability Theory

- Let S be a collection of sentences, and P is a probability function. It satisfies the *Kolmogorov axioms*:

  1. $P(A) \geq 0$
  2. $P(A) = 1$ if A true in all models
  3. $P(A \vee B) = P(A) + P(B)$ if A, B mutually exclusive

- Some consequences:

  1. $P(\neg A) = 1 - P(A)$
  2. $P(A) = P(B)$ if $A \Leftrightarrow B$
  3. $P(A \vee B) = P(A) + P(B) - P(A, B)$; note: $P(A, B) := P(A \wedge B)$

Conditional Probabilities

- Definition:

  $P(A|B) = P(A, B)/P(B)$ if $P(B) \neq 0$

- Bayes' Theorem:

  $P(B|A) = P(A|B)\, P(B)/P(A)$

  $= P(A|B)\, P(B)/[P(A|B)\, P(B) + P(A|\neg B)\, P(\neg B)]$

  $= P(B)/[P(B) + P(\neg B)\, x]$

with the *likelihood ratio*

$x := P(A|\neg B)/P(A|B)$

Conditional Independence

- A and B are (unconditionally) independent iff
  $P(A, B) = P(A)\, P(B) \Leftrightarrow P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B)$
- Example: throw a dice; A = observe even number, B = observe number ≤ 4.
- A is conditionally independent of B given C iff
  $P(A|B, C) = P(A|C)$
- Example: A=yellow fingers, B=lung cancer, C=smoking
- In symbols: $A \perp\!\!\!\perp B | C$
- Note: The relation $A \perp\!\!\!\perp B | C$ is symmetrical:
  $$A \perp\!\!\!\perp B | C \Leftrightarrow B \perp\!\!\!\perp A | C$$

Examples

1. What is the probability of throwing either 5 or 6? (One cannot have both, that is why the events are mutually exclusive.) Answer: the probability of throwing 6 plus the probability of throwing 5.
2. The general multiplication rule: $P(B \wedge C) = P(B) P(C|B)$

   What is your chance of meeting someone who is both good looking (*B*) and intelligent (*C*)? Answer: the probability of meeting someone who is good looking, $P(B)$, multiplied by the probability that this person is also intelligent, $P(C|B)$.

More example

3. The probability of having rain tomorrow (*C*) does not depend on today's seminar taking place; the two events are independent. Your knowledge of Bayesianism tomorrow (hopefully!) does depend on that. These two events are not independent.

4. What is the probability of the GNP going up next year and that the sun shines tomorrow? Answer: just multiply the respective probabilities, since the two events are independent.

5. What is the chance of picking a philosophical or entertaining book when picking a book randomly from a box on the market? These two events are not mutually exclusive because some philosophical books are entertaining. The answer is: The probability of picking a philosophical book plus the probability of picking a entertaining book minus the probability of picking one that is both.

2. What are Probabilities?

- So far, probabilities are only defined mathematically (through the axioms, i.e implicitly).
- But we want more: We have to specify what probabilities are and how the relate to events in the world.

(a) Classical interpretation (Laplace)

Definition: The probability of an event is the ratio of favourable cases to the number of equally possible cases.

Examples:

- What is the probability of getting 6 when throwing a dice? Answer 1/6.
- What is the probability of getting an even number when throwing a dice? Answer: 3/6=1/2.

Problems:

The definition rests not only on possible outcomes but on *equally possible* outcomes. Example: biased coin (one for which the probability of getting heads is 1/3, for instance). How can we specify what equally possible outcomes are?

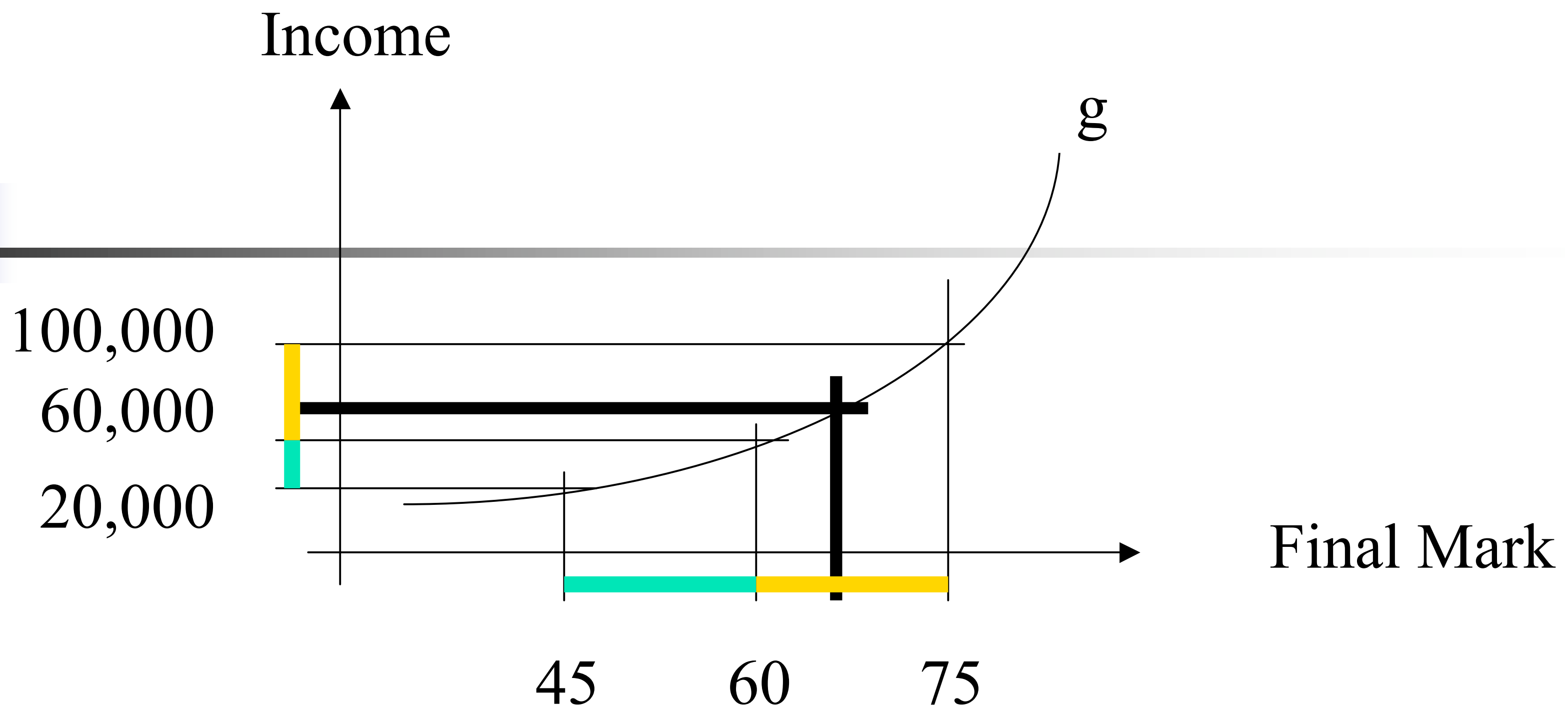
Answer: Laplace's (in)famous *principle of indifference*, which says that two events are equally possible if we have no reason to prefer one to the other.

Example: in the case of a perfectly symmetrical coin we have no reason to prefer one side to the other.

Problem 1: Can we make this principle precise?

Problem 2: It can lead to inconsistencies because it can be applied in different ways to the same situation yielding incompatible values for the same probability.

Example:



Now pick a student at random.

- What is her final mark? Applying the principle of indifference to the final mark gives p=0.5 for a mark between 45 and 60 and also p=0.5 for one between 60 and 75.
- What is her income? Applying the principle of indifference to the 'Income' gives p=0.5 for an income between 20,000 and 60,000 and also p=0.5 for one between 60,000 and 100,000.
- Given that income and final mark are related by function *g* these two sets of probabilities are incompatible!

(b) Frequency Interpretation

Definition: The probability of an event is the relative frequency with which it occurs.

Example: Toss a coin: H T H H T T H H H T H H T T T T H …

The probability for ‘heads’ in this process is the relative frequency with which ‘heads’ occurs in this sequence; that is, the number of heads divided by the total number of tosses.

Problems

1. How long does the sequence have to be? Answer: Frequencies are defined in the *long run*. But how long is the long run? What are we to make of infinity?
2. How do we get the actual value of a probability if we need an infinite limit? Does the limit exist at all?
3. Single cases? (E.g. the probability that there is no rain tomorrow)

Subjective Probabilities

Unlike classical and frequency interpretations, which take probabilities to be features of the real world, the subjective interpretation takes them to be *degrees of belief* of an agent.

Example: My degree of belief that it will rain tomorrow.

Problem: Individual degrees of belief may be such that the axioms of probability are violated.

Example: Many people believe that the possibility of at least getting one six when throwing a dice three times is .5. This is false (Exercise: show why!).

Answer: Yes, but assume there are people who have beliefs that are consistent with the axioms.

Then degrees of belief are indeed an interpretation of the calculus!

Subjective Probabilities (cont'd)

- Whether or not such humans exist does not matter – it is a normative account! (Compare deductive logic!) A set of beliefs that violates the axioms of probability is said to be ***incoherent***.
- Penalties for incoherence: ***Dutch books!*** A Dutch book is an set of bets such that, no matter what the outcome, the person loses.
- **Example**: James believes that outcome of the next toss of a coin will be heads with probability 2/3 and also believes that it will come up tails with probability 2/3. So James should be willing to bet at odds of 2 to 1 that the coin will come up heads and similarly for tails. That is, Problem: whatever the outcome, James loses 1 EURO.
- **Theorem**: A person is subject to Dutch books if, and only if, he holds an incoherent set of beliefs.

Problems

- Where to get the values of the probabilities from? The calculus does not provide us with values (with the exception of trivial cases: logically necessary and contradictory beliefs).
- There need be little connection between personal probabilities and what goes on in the world. What then are these probabilities good for? For instance, James can believe that the next coin toss will be heads with probability 9/10, but that does not make a him a good gambler.

III. Textbook Bayesianism

What is Bayesianism?

- Quantitative confirmation theory
- When (and how much) does a piece of evidence *E* confirm a hypothesis *H*?
- Typically formulated in terms of probabilities = subjective degrees of belief
- Normative theory (see Dutch books)
- Textbooks: Howson & Urbach: *Scientific Reasoning*, and Earman: *Bayes or Bust*.

The mathematical machinery

- Confirmation = positive relevance between *H* and *E*
- Start with a prior probability $P(H)$
- Updating-rule: $P_{new}(H) := P(H|E)$
- By mathematics ("Bayes' Theorem") we get:

$$P_{new}(H) = P(E|H)\,P(H)/P(E)$$

- H confirms E iff $P_{new}(H) > P(H)$
- H disconfirms E iff $P_{new}(H) < P(H)$

Some terminology

Remember: $P_{new}(H) = P(H)\,P(E|H)\,/\,P(E)$

- $P(H|E)$: *Posterior probability*: the probability of the hypothesis H judged in the light of evidence E.
- $P(H)$: *Prior probability*: the probability of the hypothesis *without* taking evidence E into account.
- $P(E|H)$: *Likelihood of evidence E*: the probability of E under the assumption that H is correct.
- $P(E)$: *Expectedness of E*: the probability that E would obtain regardless of whether H is correct or not.

Examples of Textbook Bayesianism

- Let's assume that H entails E: $P(H|E) = 1$.
- Then $P_{new}(H) = P(H)/P(E)$
- (i) It is easy to see that Bayesianism can account for the insight that surprising evidence (i.e. if $P(E)$ is small) confirms better.
- (ii) Let's assume we have different pieces of evidence: $P(E) = P(E_1, E_2, \ldots, E_n) = P(E_1) \cdot P(E_2|E_1) \cdot P(E_3|E_2, E_1) \cdot \ldots \cdot P(E_n|E_{n-1}, \ldots, E_1)$. Hence, less correlated pieces of evidence confirm better: the variety-of-evidence thesis.

We conclude

- Textbook Bayesianism *explains* very general features of science.
- However, it turns out to be too general as it does not take into account *de facto constraints* of the scientific practice (such as dependencies between measurement instruments). Taking them into account might lead to different conclusions.
- Moreover, Textbook Bayesianism does not have an account of what a scientific theory is.

Jon Dorling’s version

- Jon Dorling aims at reconstructing specific episodes in the history of science and fitting them into the Bayesian apparatus.
- To do so, one has to assign specific probability values to the hypotheses and likelihoods in question.
- This variant of Textbook Bayesianism is too specific.

Upshot

- We need a version of Bayesianism that is not too general (to connect to the practice of science), and not too specific (to gain some philosophical insight).
- It would also be nice to have an account that has a somewhat wider scope, i.e. an account that reaches beyond confirmation theory.

IV. (More) Success Stories

- The larger picture seems appealing: evidence confirms hypotheses as it raises their probabilities in a certain way. Moreover, the account provides a quantitative account of confirmation, that is, it affords us with a means to determine the *degree* of to which a hypothesis is confirmed.

1. Bayesianism & hypothetico-deductivism

- Assume $1 > P(H) > 0$ and $1 > P(E) > 0$.
- Since E is a logical consequence of H, $P(E|H) = 1$
- Bayes' theorem reduces to $p(H|E) = p(H) / p(E)$.
- Since $P(E) < 1$, $P(H|E)$ is greater than $p(H)$.
- Hence every non-trivial deductive consequence of H confirms H.
- So the Bayesian account captures the intuition of the H-D account that a hypothesis is confirmed (or corroborated, as Popper preferred to say) by its consequences.

2. Popper’s risky predictions

- Of two deductive consequences of a hypothesis, the more improbable one confirms it more strongly because

  $P(H|E_1) = P(H) / P(E_1) > P(H) / P(E_2) = P(H|E_2)$ where $P(E_1) < P(E_2)$

- Hence the Bayesian account captures the intuition that an unexpected (or unlikely) piece of evidence confirms the hypothesis stronger than evidence that was to be expected. And this was Popper’s point that scientists must make risky predictions to test their theory.

An example from 19th century optics

Poisson deduced that the wave theory of light predicts that the shadow of a small circular disk by a narrow beam has a white spot in the middle. This was considered to be extremely unlikely. So when Arago experimentally confirmed this prediction, the wave theory quickly received widespread acceptability.

3. Evading the raven paradox

- Since there are much more non-ravens in the world than ravens, the prior probability of ‘x is not a raven & x is not black’ is much greater than probability of ‘x is a raven & x is black’. (This follows from our background knowledge about the world.)
- But we have just seen that the more unlikely a piece of evidence is, the more it confirms the hypothesis.
- Hence black ravens confirm the hypothesis that all ravens are black much more than white shoes.
- Question: Is this a satisfying solution?

V. Further Topics and Limitations

1. Induction and convergence

- The Bayesian approach to confirmation seems to be able to circumvent Hume's problem.
- **Next case induction**:

  (H) 'all observed *a* are *B*' → 'the next observed *a* is *B*'.
- Weak projectability (WP):

  $\lim_{n\to\infty} p(Ba_{n+1} \mid Ba_1 \,\&\, Ba_2 \,\&\, \ldots \,\&\, Ba_n) = 1.$
- Claim: granted that the prior probability of H is not zero, WP follows from Bayesian confirmation theory.

Proof

- Set $E = Ba_1 \;\&\; Ba_2 \;\&\; \ldots \;\&\; Ba_{n+1}$, and notice that $p(E|H) = 1$, since E is a logical consequence of H. Then Bayes' theorem reads
- $P(H|E) = p(H) \,/\, p(E)$.
- Iterative expansion of the denominator yields:
- $P(H|\, Ba_1 \;\&\; Ba_2 \;\&\; \ldots \;\&\; Ba_{n+1}) =$

  $p(H) \,/\, p(Ba_1)\, p(Ba_2|\, Ba_1) \ldots p(Ba_{n+1}|\, Ba_1 \;\&\; Ba_2 \;\&\; \ldots \;\&\; Ba_n)$

Proof (cont'd)

- Now: if $p(Ba_{n+1}| Ba_1 \& Ba_2 \& \ldots \& Ba_n)$ does not approach 1 as *n* tends towards infinity, then the right hand side of the above expression tends towards infinity. But this contradicts the axioms of probability, which stipulate that $0 \leq P(H| Ba_1 \& Ba_2 \& \ldots \& Ba_{n+1}) \leq 1$. Therefore $p(Ba_{n+1}| Ba_1 \& Ba_2 \& \ldots \& Ba_n)$ must approach 1 as *n* tends towards infinity.
- Hence $\lim_{n\to\infty} p(Ba_{n+1} | Ba_1 \& Ba_2 \& \ldots \& Ba_n) = 1$, which is what we need as a reply to Hume.

Question: Does this really solve the problem of induction?

The problem of the priors

- The Bayesian account is quantitative. That is, it is a procedure to calculate new probabilities from old ones. But in order to apply Bayes' formula we need concrete values for the *expectedness of E*, the *prior probability of H* and the *likelihood of E*.
- **Problem**: Where do we get these values from? *('problem of the priors')*
- An answer to this question depends on how one understands the probabilities involved:
  1.) Objective-empirical (relative frequency)
  2.) Subjective empirical (degrees of belief)

1. Frequentism

- At the beginning we don’t know whether a hypothesis is true or not. So how do we get a frequency? The basic idea is that we place a hypothesis in a *reference class* of similar hypotheses. Then we look at how many of these were successful in the past. The ratio of the successful ones to the total number of hypotheses is the prior probability of the hypothesis.
- **Problem**: What is the relevant reference class? That is, what does it mean to say that a hypothesis *H* is similar to another hypothesis? Is it a matter of mathematical form? If so, what about hypotheses that are not formulated in terms of mathematics?

2. Subjectivism

- **Idea**: Probabilities are degrees of belief. Hence, the priors are a measure of the *initial degree of belief* in the truth of the hypothesis.
- **Problems**:

  1. Again, how do we determine the initial degree of belief? Just ask scientists what they believe? Who, then, is qualified to have a view on the probability?

  2. This is to admit a thoroughly subjective element in a theory of confirmation. Is this a sound procedure?

Replies on behalf of the subjectivist

- **Reply 1**: Salmon’s ‘tempered personalism’ (Shimony, Salmon) or objective Bayesianism (Jaynes, Williamson)
- **Reply 2**: ‘*Washing out of the priors*’: One can show that even if different prior probabilities are assumed, as evidence accumulates the posterior probabilities get closer and closer to each other. That is, the influence of the prior probabilities on the posterior probabilities decreases as more evidence becomes available.
- **Question**: Is this enough to solve the problem of the priors?

3. Zero priors

- **Problem**: It follows immediately from Bayes' formula that the posterior probability *p*(*H*|*E*) will always equal zero if either the prior probability *p*(*H*) or the likelihood *p*(*E*|*H*) equal zero. How can we know which hypotheses (if any) should be assigned zero probability?
- **Example**: Swinburne, Grünbaum and the existence of God.

4. The problem of the Old Evidence

- **Problem**: we have always assumed so far that $p(E) < 1$, i.e. that there must be some surprise to *E*. However, if *E* is already known when the theory is formulated, there is no surprise to *E* and we have: $p(E) = p(E|H) = 1$. Bayes' formula then reads $p(H|E) = p(H)$ and no confirmation takes place.
- This seems to contradict our intuitions about confirmation as well as historical facts. The perihelion of Mercury was old news when Einstein came up with GTR but it was taken to be confirming the theory.

Possible solutions

(1) Bite the bullet and admit that old evidence doesn't confirm.

(2) *p*(*E*) does not equal one because we have to conditionalise on background knowledge:

- Evaluate *p*(*E*) not with respect to someone's actual background knowledge, but with respect to someone's *background knowledge would* E *not yet be known.*
- Consider what *p*(*E*) was for someone in the past, at a time when *E* was indeed not yet known.

Another solution of the OE problem

(3) Admit that $p(E) = 1$ and propose a new criterion of confirmation: what we are really after is the knowledge that *H* implies *E*, and this may not be known even though *E* itself is. This is a new insight and hence confirms the theory. Then confirmation takes place!

5. Conflict with actual scientific practice

- A quote from Glymour: '... probabilistic analyses [of confirmation] remain at too great a distance from the history of scientific practice to be really informative about that practice, and in part they do so exactly because they are probabilistic.'
- In short: no scientist ever gives probabilistic arguments for a theory.
- **Question**: Can we dismiss Bayesian confirmation theory on the basis that is that 'real' scientists do not reason in this way?

6. The updating rule is unrealistic

- The updating rule is unrealistic as it supposes that the evidence is certain.
- Way out: Jeffrey conditionalisation
- Other problem: There are no convincing Dutch book arguments for the updating rule.

Outline

I. Why Bayesianism?
II. Probability and Its Interpretations
III. The Elements of Bayesianism
IV. Success Stories
V. Further Topics and Limitations
VI. Modeling in Science
VII. Bayesian Networks

VI. Modeling in Science

The ubiquity of models in science

- Highschool: Bohr model of the atom, model of the pendulum,…
- University: Standard Big Bang Model, Standard Model of particle physics,…
- Even later: Physicists like Lisa Randall (Harvard) use models to learn about the most fundamental question of the universe.
- We observe a shift from theories to models in science, and this shift is reflected in the work of philosophers of science in the last 25 years.

Theories and models

- Scientists often use the words *theory* and *model* interchangeably.
- Example: the Standard Model of particle physics.
- So how can one distinguish between theories and models?

Features of theories

- Examples: Newtonian Mechanics, Quantum Mechanics
- General and unversal in scope
- Abstract
- Often difficult to solve (example: QCD)
- No idealizations involved (ideally...)

Features of models

- Examples: the model of a pendulum, the Bohr model of the atom, gas models, the MIT Bag model,...
- Specific and limited in scope
- Concrete
- Intuitive and visualizable
- Can be solved
- Involve idealizing assumptions

Two kinds of models

- Models of a theory

  - example: the treatment of the pendulum is a model of Newtonian Mechanics

  - the theory acts as a modeling framework (and not much follows from the theory w/o specifying a model)

- Phenomenological models

  - example: the MIT Bag model

  - there is no theory into which a model can be embedded.

Modeling in philosophy

- So why not constructing philosophical models?
- Some authors do this already, see, e.g., the work of Brian Skyrms on the social contract and Clark Glymour et al.'s work on causal discovery.

- Goal: Show that models are also valuable tools for *understanding the methodology of science*.
- More specifically: Construct Bayesian Network models.

Glymour‘s distintction

- **Platonic tradition** in philosophy: formulate necessary and sufficient conditions for conepts such as knowledge, virtue and the good. Comes up with general and universal theories, which aim at explaining or accounting for everything... and end up explaining nothing!
- **Euclidian tradition** in philosophy: make idealized assumptions and explore the consequences of these assumptions... Just like scientists do it!
- Should we go for the Euclidian program in philosophy?

  - perhaps not always, but sometimes

  - pluralism of methods

VII. Bayesian Networks

1. Probability Theory

- Let S be a collection of sentences, and P is a probability function. It satisfies the *Kolmogorov axioms*:

  1. $P(A) \geq 0$
  2. $P(A) = 1$ if A true in all models
  3. $P(A \vee B) = P(A) + P(B)$ if A, B mutually exclusive

- Some consequences:

  1. $P(\neg A) = 1 - P(A)$
  2. $P(A) = P(B)$ if (in all models) $A \Leftrightarrow B$
  3. $P(A \vee B) = P(A) + P(B) - P(A, B)$; note: $P(A, B) := P(A \wedge B)$

Conditional Probabilities

- Definition:

  $P(A|B) = P(A, B)/P(B)$ if $P(B) \neq 0$

- Bayes' Theorem:

  $P(B|A) = P(A|B)\, P(B)/P(A)$

  $= P(A|B)\, P(B)/[P(A|B)\, P(B) + P(A|\neg B)\, P(\neg B)]$

  $= P(B)/[P(B) + P(\neg B)\, x]$

with the *likelihood ratio*

  $x := P(A|\neg B)/P(A|B)$

Conditional Independence

- A and B are (unconditionally) independent iff
  $P(A, B) = P(A)\, P(B) \Leftrightarrow P(A|B) = P(A) \Leftrightarrow P(B|A) = P(B)$
- Example: throw a dice; A = observe even number, B = observe number ≤ 4.
- A is conditionally independent of B given C iff
  $P(A|B, C) = P(A|C)$
- Example: A=yellow fingers, B=lung cancer, C=smoking
- In symbols: $A \perp\!\!\!\perp B | C$
- Note: The relation $A \perp\!\!\!\perp B | C$ is symmetrical:

$$A \perp\!\!\!\perp B | C \quad \Leftrightarrow \quad B \perp\!\!\!\perp A | C$$

The Semi-Graphoid Axioms

- The conditional independence relation satisfies the semi-graphoid axioms:

  1. Symmetry: $X \perp\!\!\!\perp Y | Z \leftrightarrow Y \perp\!\!\!\perp X | Z$
  2. Decomposition: $X \perp\!\!\!\perp Y, W | Z \rightarrow X \perp\!\!\!\perp Y | Z$
  3. Weak Union: $X \perp\!\!\!\perp Y, W | Z \rightarrow X \perp\!\!\!\perp Y, Z | W$
  4. Contraction: $X \perp\!\!\!\perp W | Y, Z \ \& \ X \perp\!\!\!\perp Y | Z \rightarrow X \perp\!\!\!\perp Y, W | Z$

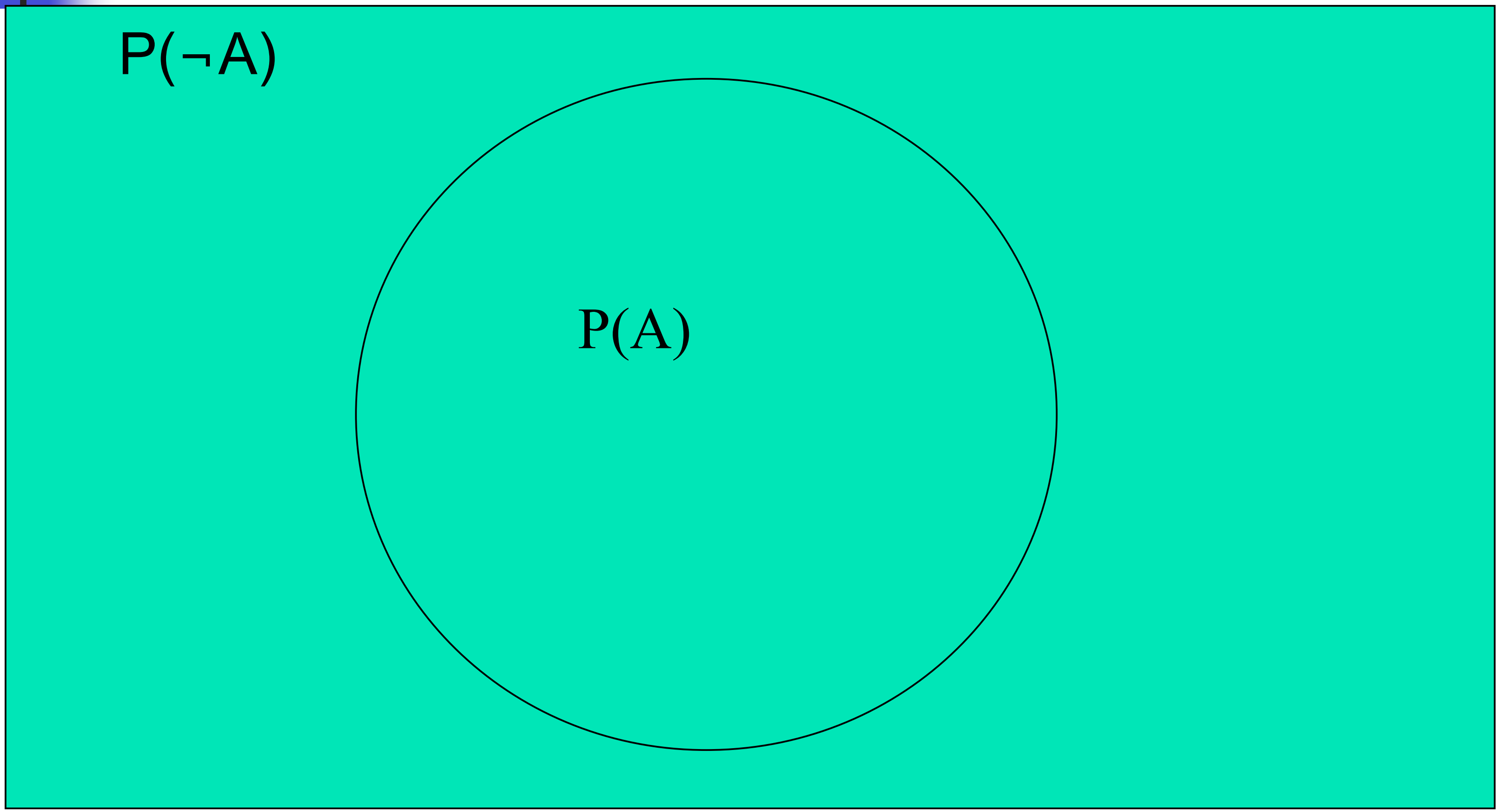
- With these axioms, new conditional independencies can be obtained from known independencies.

Joint and Marginal Probability

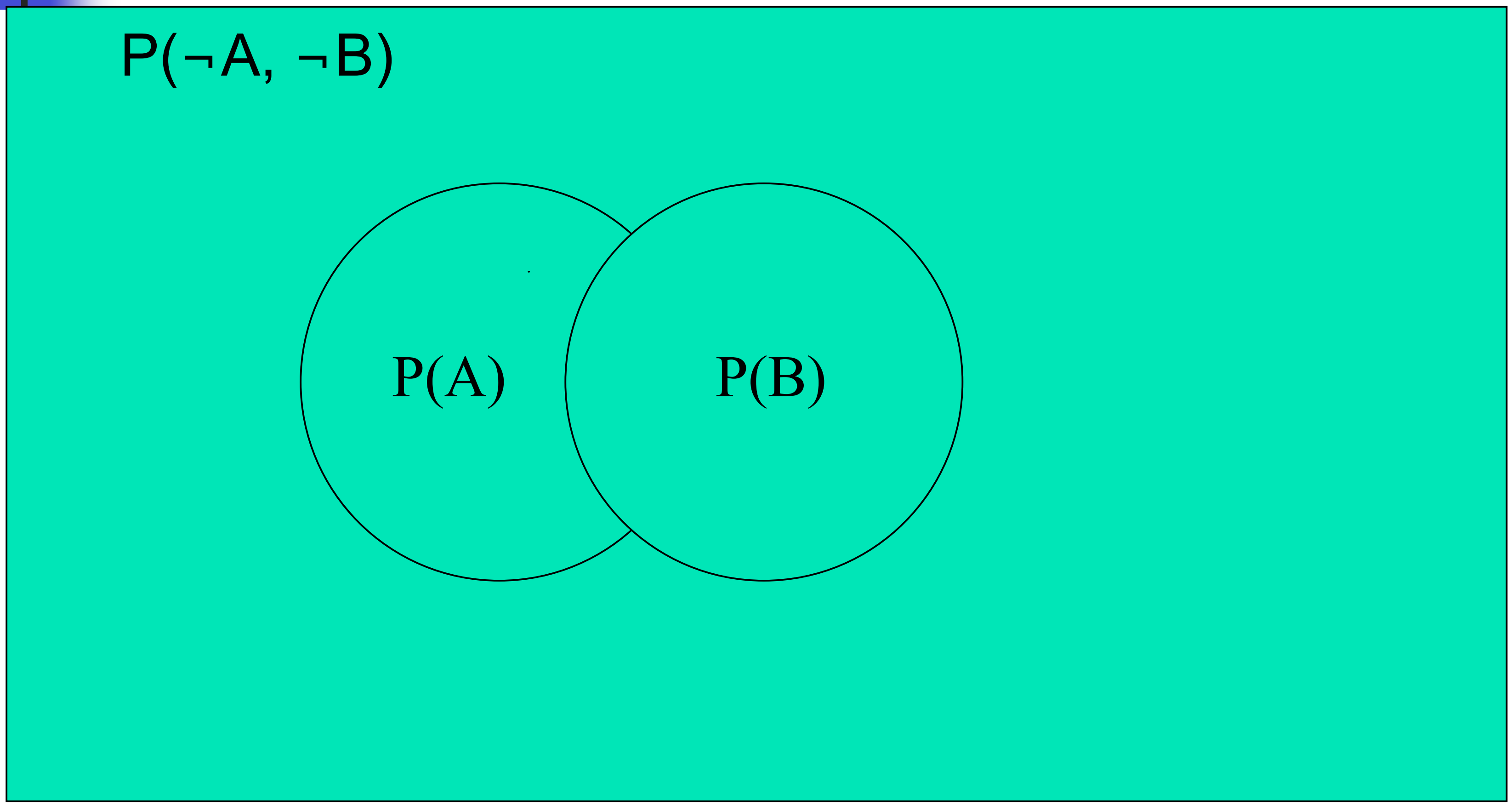
- To specify the *joint probability distribution* over two *binary propositional variables A, B*, three probabilities have to be specified (if we do not have any further knowledge):

  E.g. $P(A, B) = .2$, $P(A, \neg B) = .1$, $P(\neg A, B) = .6$
- Note: $\Sigma_{A, B} P(A, B) = 1 \Rightarrow P(\neg A, \neg B) = .1$
- *Marginal probability*: $P(A) = \Sigma_B P(A, B)$
- This allows us to calculate whatever marginal or conditional probability we are interested in from the joint probability:
- $P(A_1,\dots A_m | A_{m+1},\dots A_n) = P(A_1,\dots A_n)/P(A_{m+1},\dots A_n)$
  $= P(A_1,\dots A_n)/\Sigma_{A_{m+1}, A_n} P(A_1,\dots A_n)$

Representing a Joint Probability Distribution



Representing… (cont'd)



2. Bayesian Networks

- Venn diagrams and the specification of all entries in P(A_1,…,A_n) are not the most efficient ways to represent a joint probability distribution.
- There is also a problem of computational complexity: Specifying the joint probability distribution over n variables requires the specification of 2^n-1 numbers.
- The trick: Use information about conditional independencies that hold between (sets of) variables. This will reduce the number of numbers that have to be specified (and stored in a computer).

An Example from Medicine

- T: Patient has tuberculosis
- X: Positive X-ray
- Given information:

  $t := P(T) = .01$

  $p := P(X|T) = .95 = 1 - P(\neg X|T) = 1 -$ rate of *false negatives*

  $q := P(X|\neg T) = .02$ = rate of *false positives*

- What is $P(T|X)$? ⇒Apply Bayes' Theorem

  $$P(T|X) = P(X|T)\, P(T) / [P(X|T)\, P(T) + P(X|\neg T)\, P(\neg T)] =$$

  $$= p\,t / [p\,t + q\,(1-t)] = t/[t + (1-t)\,x] \quad \text{with } x := q/p$$

  $$= .32$$

A Bayesian Network Representation

P(T) = .1

P(X|T) = .95

P(X|¬T) = .2

Parlance:

"*T* causes *X*"

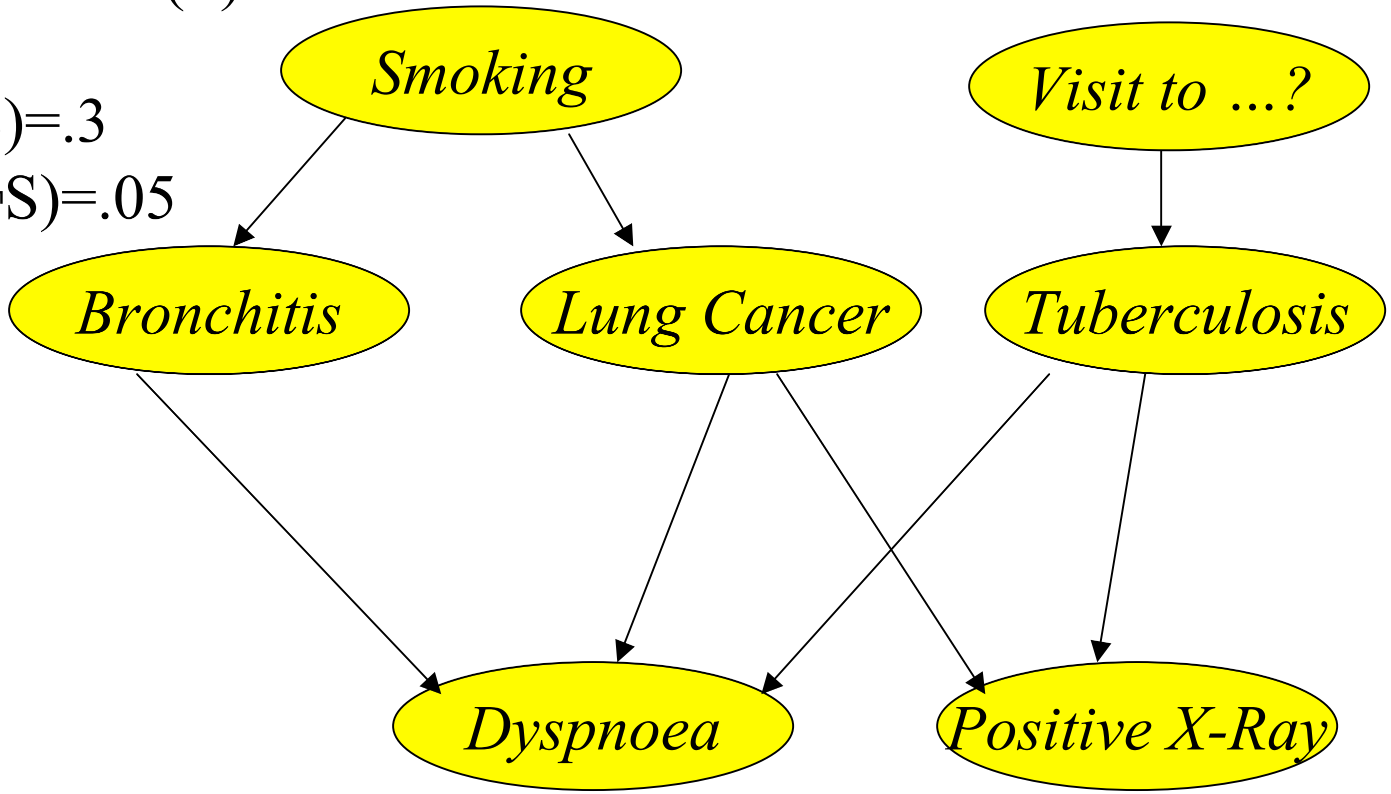
"*T* directly influences *X*"

A More Complicated (=Realistic) Scenario

$P(\text{S}) = .23$

$P(\text{B}|\text{S})=.3$

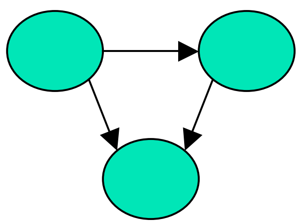
$P(\text{B}|\neg\text{S})=.05$

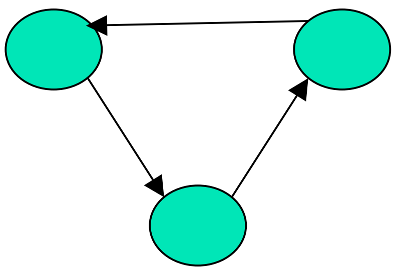


Directed Acyclic Graphs

- A directed graph *G* (*V*, *E*) consists of a finite set of nodes *V* and an irreflexive binary relation *E* on *V*.

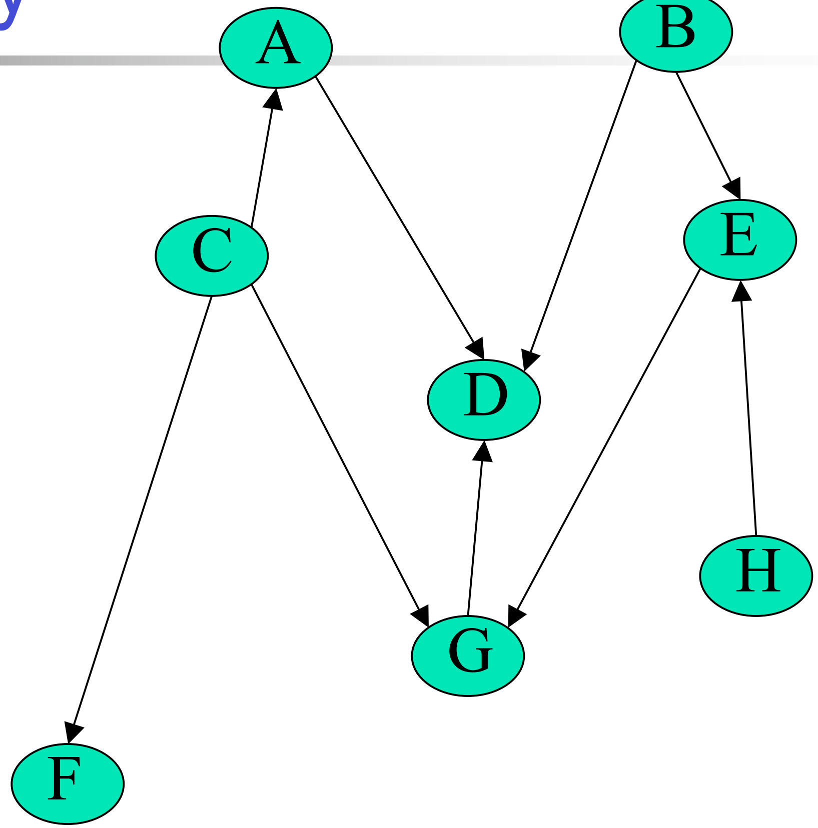
- A directed acyclic graph (DAG) is a directed graph which does not contain cycles.





Some Vocabulary

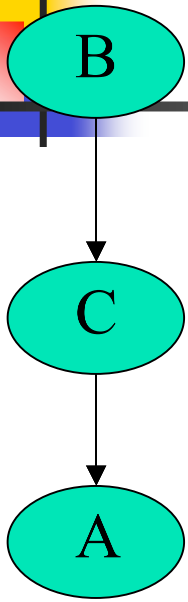
- Parents of node *A*: par(*A*)
- Ancestor
- Child node
- Descendents
- Non-Descendents
- Root node



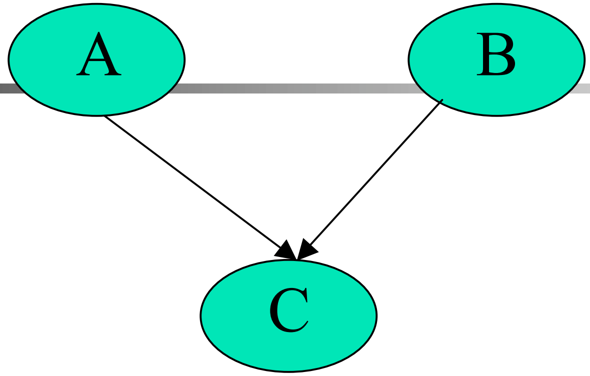
The Parental Markov Condition

- The Parental Markov Condition (PMC) requires that a variable is conditionally independent of its non-descendents given its parents.
- Definition: A Bayesian Network is a DAG with a probability distribution which respects the PMC.

Three Examples

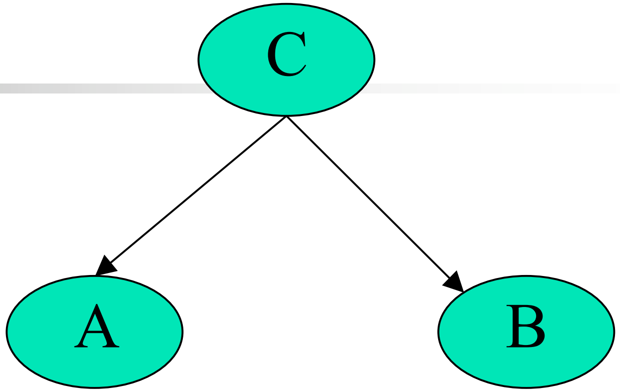


$A \perp\!\!\!\perp B | C$



$A \perp\!\!\!\perp B$

“collider”



$A \perp\!\!\!\perp B | C$

“common cause”

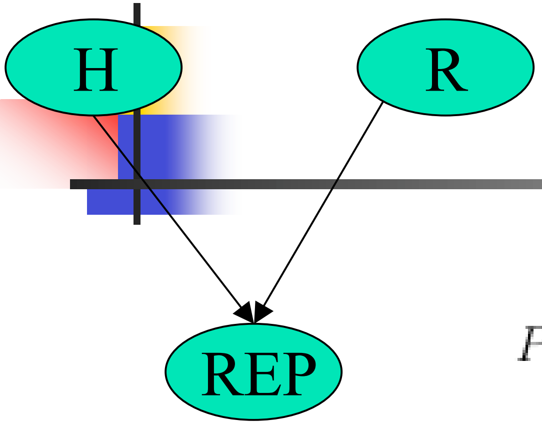
Bayesian Networks at Work

- How can one calculate probabilities with B.N.?
- Here is the answer:

$$P(A_1, \ldots, A_n) = \Pi_i \, P(A_i | \text{par}(A_i))$$

- I.e. the joint probability distribution is determined by the prior probability of the root nodes (par(A) = ∅) and the conditional probabilities of all other nodes.
- This typically requires the specification of much less numbers than 2^n-1.

An Example



P(*H, R, REP*) = P(*H*) P(*R*)P(*REP*|*H, R*)

What is P(H|REP)?

P(H) = *h*

P(R) = *r*

P(REP|H, R) = 1

P(REP|¬H, R) = 0

P(REP|H, ¬R) = *a*

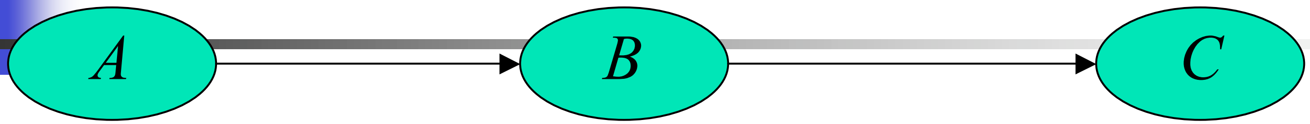
P(REP|¬H, ¬R) = *a*

$$
\begin{aligned}
P(\mathrm{H}|\mathrm{REP}) &= \frac{P(\mathrm{H}, \mathrm{REP})}{P(\mathrm{REP})} \\
&= \frac{\sum_R P(\mathrm{H}, R, \mathrm{REP})}{\sum_{H,R} P(H, R, \mathrm{REP})} \\
&= \frac{P(\mathrm{H}) \sum_R P(R) P(\mathrm{REP}|\mathrm{H}, R)}{\sum_{H,R} P(H, R, \mathrm{REP})} \\
&= \frac{h(r \cdot 1 + (1-r) \cdot a)}{h(r \cdot 1 + (1-r) \cdot a) + (1-h)(r \cdot 0 + (1-r) \cdot a)} \\
&= \frac{h(r + a(1-r))}{hr + a(1-r)}
\end{aligned}
$$

Some More Theory: *d*-Separation

- There are more independencies in a Bayesian Network than the ones accounted for by the Parental Markov Condition.
- Is there a systematic way to find all independences that hold in a given Bayesian Network?
- Yes! *d*-separation
- Let *A, B*, and *C* be sets of variables. Then:
  $A \perp\!\!\!\perp B | C$ iff *C d*-separates *A* from *B*.
- So what is *d*-separation?

Example 1



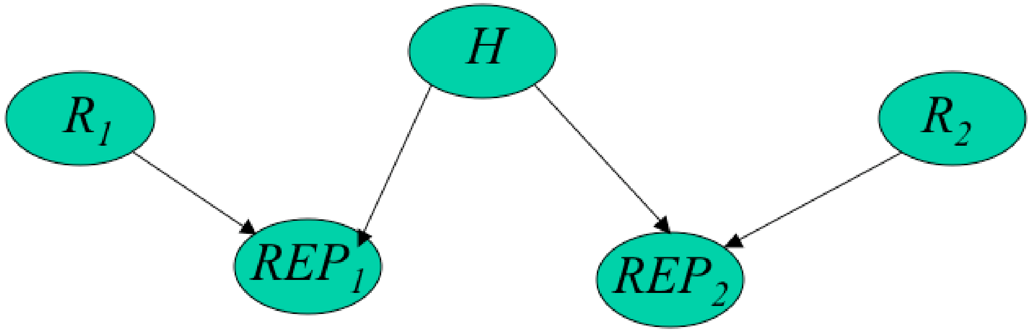
PMC $\Rightarrow C \perp\!\!\!\perp A \mid B$

But it is also the case that $A \perp\!\!\!\perp C \mid B$

This does *not* follow from PMC.

It can, however, be derived from the symmetry axiom for semi-graphoids.

Example 2



PMC $\Rightarrow$ $REP_1 \perp\!\!\!\perp REP_2 | H, R_1$ (*)

But not $REP_1 \perp\!\!\!\perp REP_2 | H$

However: PMC $\Rightarrow$ $R_1 \perp\!\!\!\perp H, REP_2$

Weak Union $\Rightarrow$ $R_1 \perp\!\!\!\perp REP_2 | H$ (**)

(*), (**) and *Contraction* $\Rightarrow R_1, REP_1 \perp\!\!\!\perp REP_2 | H$

Decomposition $\Rightarrow$ $REP_1 \perp\!\!\!\perp REP_2 | H$

d-Separation

- A path *p* is *d*-separated (or blocked) by (a set) Z iff there is a node *w* satisfying either:

  1. *w* has converging arrows ($u \rightarrow w \leftarrow v$) and none of *w* or its descendents are in Z.

  2. *w* does not have converging arrows and $w \in Z$.
- If Z blocks every path from X to Y, then Z *d*-separates X from Y and $X \perp\!\!\!\perp Y | Z$.

How to Construct a Bayesian Network

1. Specify all relevant variables.
2. Specify all conditional independences which hold between them.
3. Construct a Bayesian Network which exhibits these independencies.
4. Check other (perhaps unwanted) independencies with the *d*-separation criterion. Modify the networks.
5. Specify the prior probabilities of all root nodes and the conditional probabilities of all other (child) nodes given their parents.
6. Calculate any probability you are interested in.

3. Modeling Partially Reliable Information Sources

When we receive information from *independent* and *partially reliable* sources, what is our degree of confidence that this information is true?

- Independence?
- Partial reliability?

A. Independence

REP_i

is independent of

F_1, REP_1, F_{i-1}, REP_{i-1}, F_{i+1}, REP_{i+1}, F_n, REP_n

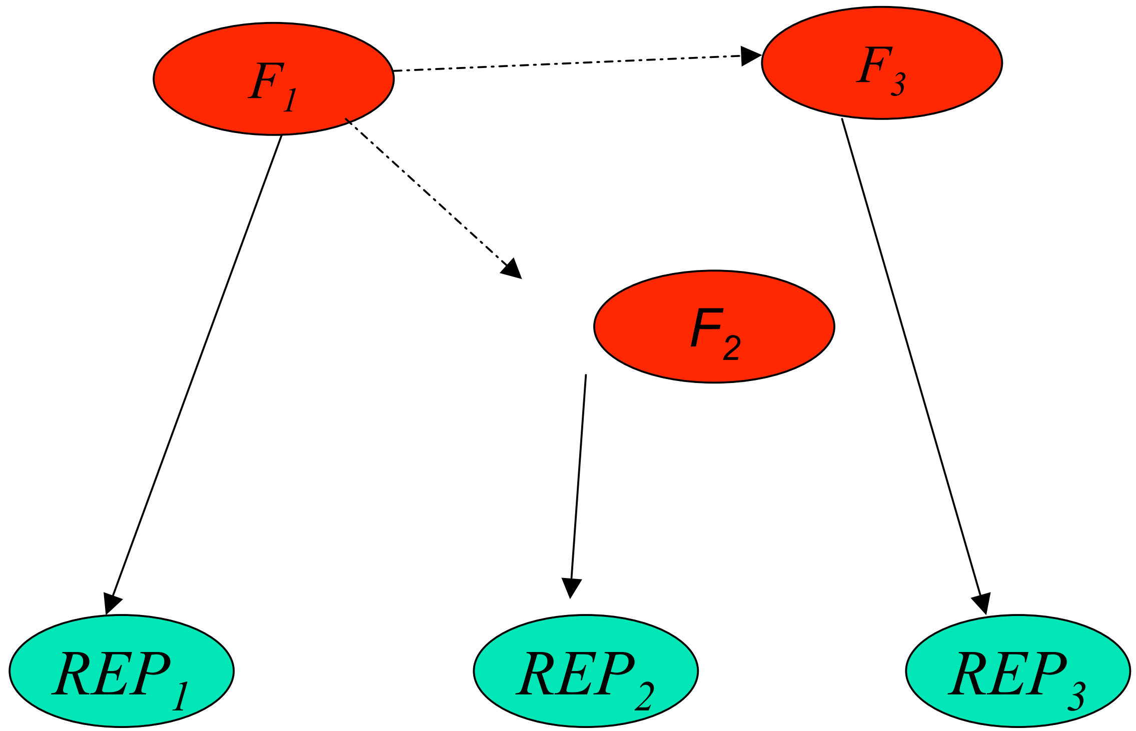
given F_i

B. Partial Reliability

To model partially reliable information sources, additional model assumptions have to be made.

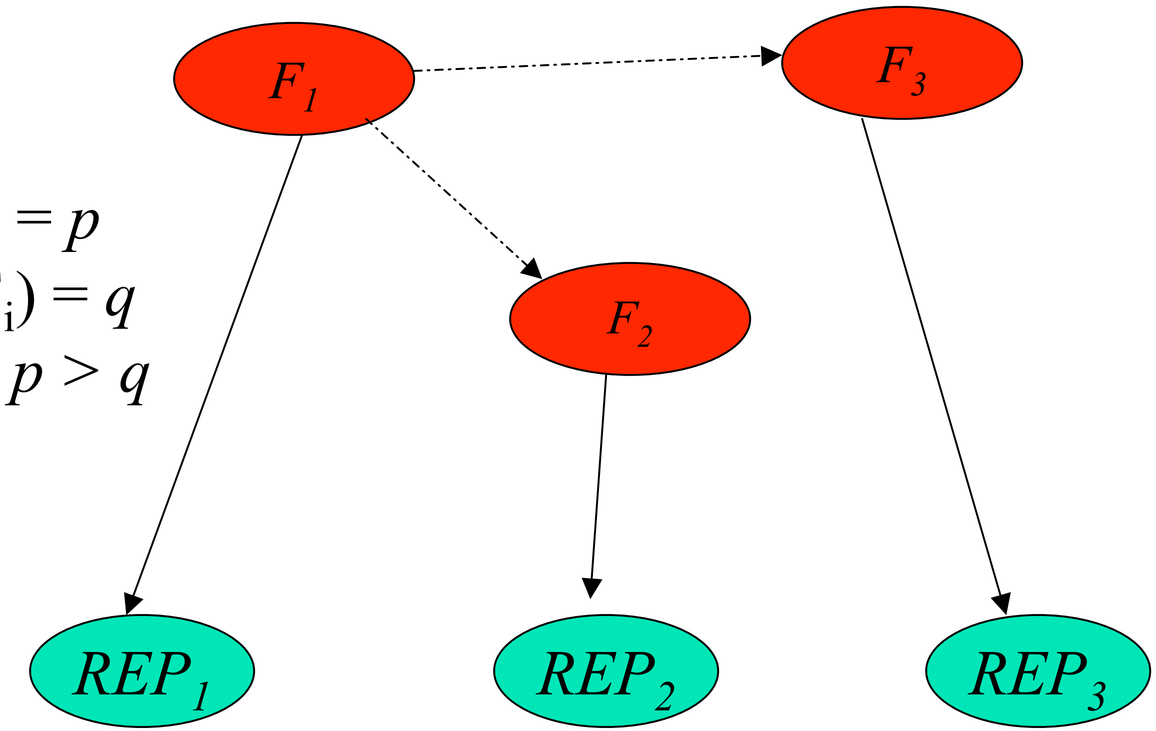
Examine *two* models!

Model I: Fixed Reliability Paradigm: Medical Testing



Model I: Fixed Reliability Paradigm: Medical Testing

$P(\text{REP}_i|\text{F}_i) = p$
$P(\text{REP}_i|\neg\text{F}_i) = q$
for $p > q$



Measuring Reliability

- Let's assume that we get positive reports.

  $p := P(\mathrm{REP}_i|\mathrm{F}_i)$ (for all $i = 1,\ldots, n$)

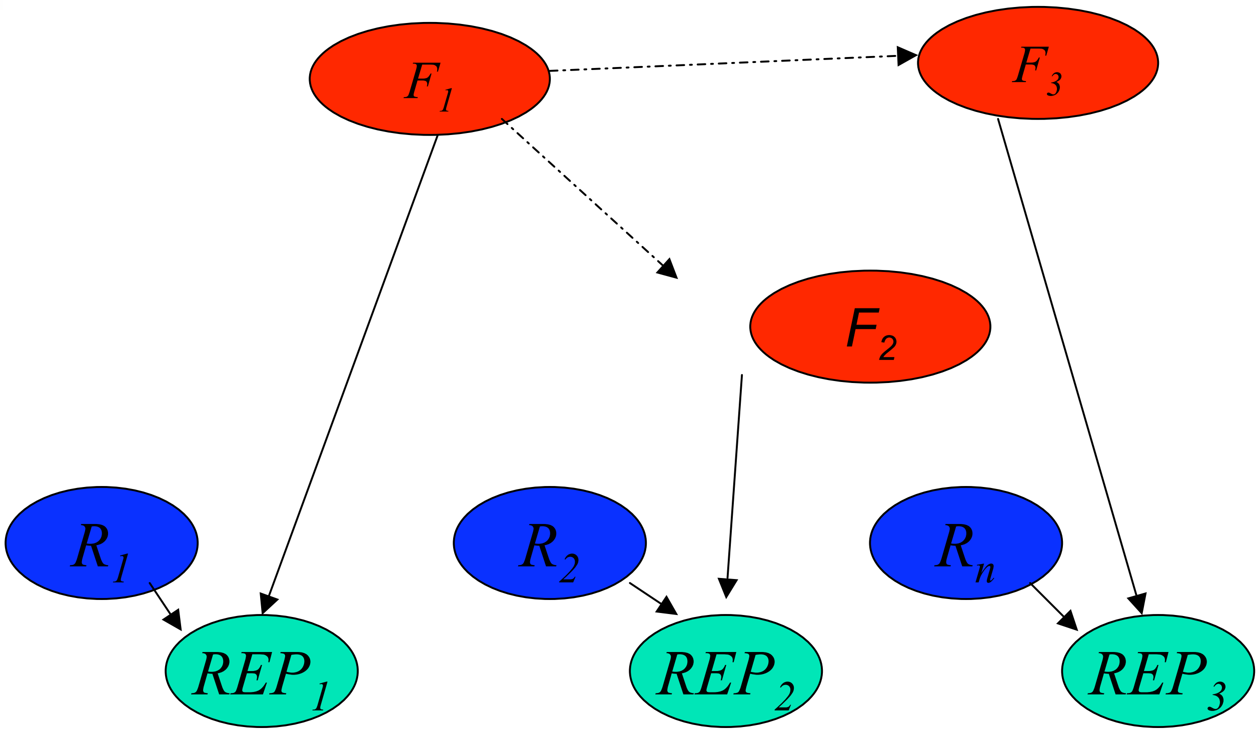
  $q := P(\mathrm{REP}_i|\neg\mathrm{F}_i)$ with $p > q$

  $r := 1 - q/p$

Randomization Full Reliability

0--1

Model II: Variable Reliability, Fixed Randomization Parameter Paradigm: Scientific Instruments



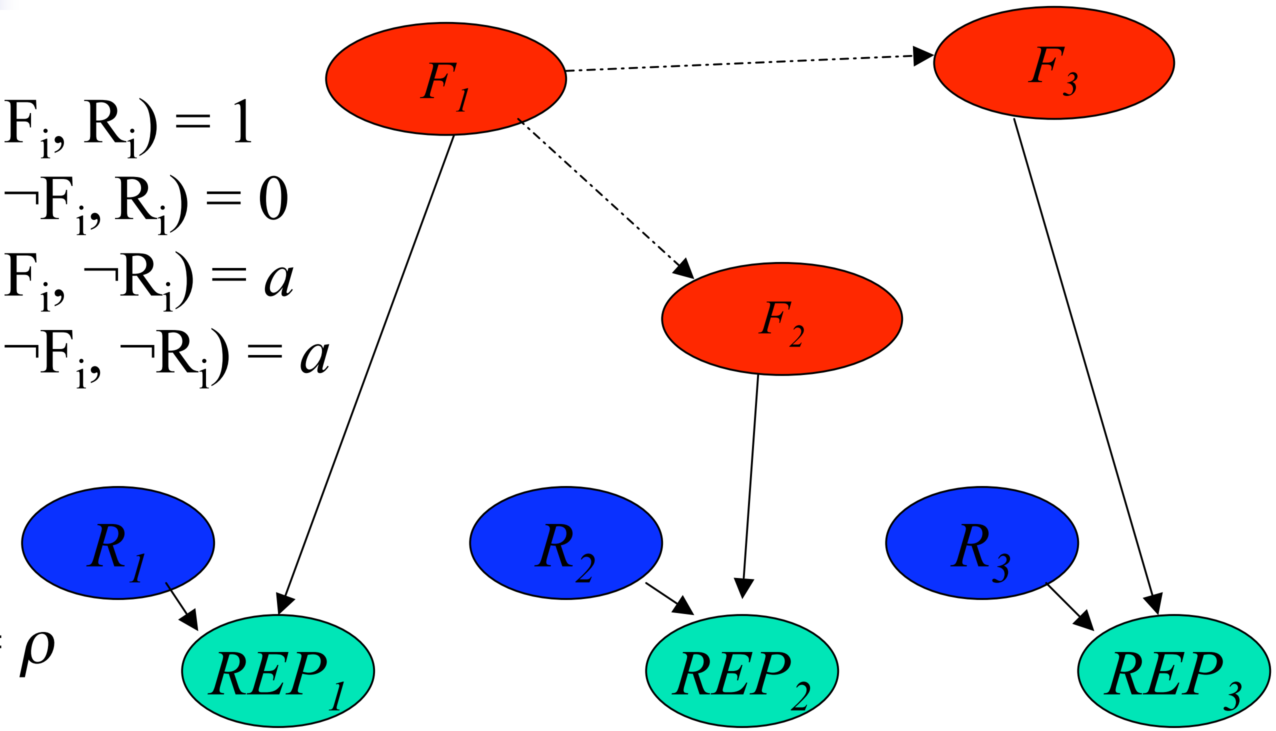
Model II: Variable Reliability, Fixed Randomization Parameter Paradigm: Scientific Instruments

$P(\text{REP}_i|\text{F}_i, \text{R}_i) = 1$

$P(\text{REP}_i|\neg\text{F}_i, \text{R}_i) = 0$

$P(\text{REP}_i|\text{F}_i, \neg\text{R}_i) = a$

$P(\text{REP}_i|\neg\text{F}_i, \neg\text{R}_i) = a$

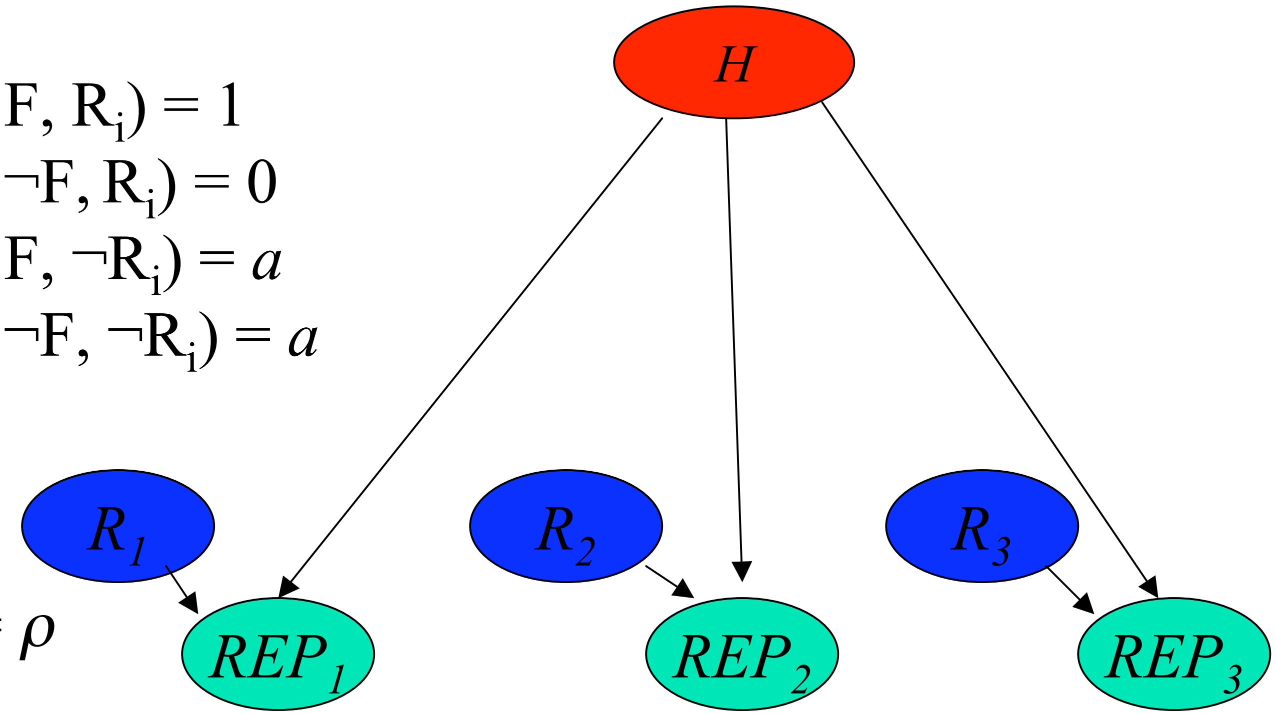


$P(\text{R}_i) = \rho$

Model IIa: Testing One Hypothesis

$P(\mathrm{REP_i}|\mathrm{F}, \mathrm{R_i}) = 1$
$P(\mathrm{REP_i}|\neg\mathrm{F}, \mathrm{R_i}) = 0$
$P(\mathrm{REP_i}|\mathrm{F}, \neg\mathrm{R_i}) = a$
$P(\mathrm{REP_i}|\neg\mathrm{F}, \neg\mathrm{R_i}) = a$

$P(\mathrm{R_i}) = \rho$



Taking Stock

- The Parental Markov Condition is part of the definition of a Bayesian Network.
- *The d*-separation criterion helps us to identify all conditional independences in a Bayesian Network.
- We constructed two basic models of partially reliable information sources:

  - reliability endogenous (medical testing)

  - reliability exogenous (scientific instruments)

The Plan for Lecture II

I. Textbook Bayesianism
II. The Variety-of-Evidence Thesis
III. Testimony
IV. Coherence, Unification and Simplicity
V. What is a Scientific Theory?
VI. The Stability of Normal Science
VII. Voting and Group Decision Making