Representations for KBS: Uncertainty & Decision Support

6.871 -- Lecture 10

Outline

- A Problem with Mycin
- Brief review of history of uncertainty in AI
- Bayes Theorem
- Some tractable Bayesian situations
- Bayes Nets
- Decision Theory and Rational Choice
- A recurring theme: battling combinatorics through model assumptions

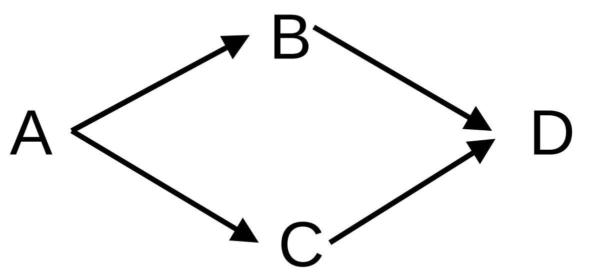
A Problem with Mycin

- Its notion of uncertainty seems broken
  - In Mycin the certainty factor for OR is Max
    - CF (OR A B) = (Max (Cf A) (Cf B))
- Consider
  - Rule-1 IF A then C, certainty factor 1
  - Rule-2 If B then C, certainty factor 1
  - This is logically the same as
    If (Or A B) then C, certainty factor 1

More Problems

- If CF(A) = .8 and CF(B) = .3
  A→C
  B→C
  A or B→C

- IF A → B, A → C, B → D, C → D there will also be a mistake: (why?)

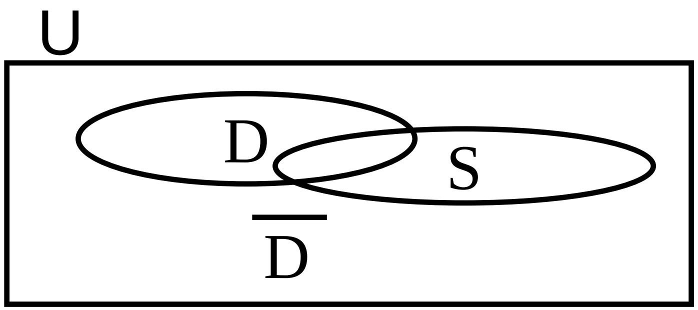


Then CF (C) = .8 + .3 * (1 - .8) = .8 + .06 = .86
CF (OR A B) = (Max .8 .3) = .8 and CF(C) = .8

Some Representations of Uncertainty

- Standard probability
  – too many numbers
- Focus on logical, qualitative
  – reasoning by cases
  – non-monotonic reasoning
- Numerical approaches retried
  – Certainty factors
  – Dempster-Schafer
  – Fuzzy
- Bayes Networks

Background



Conditional Probability of S given D

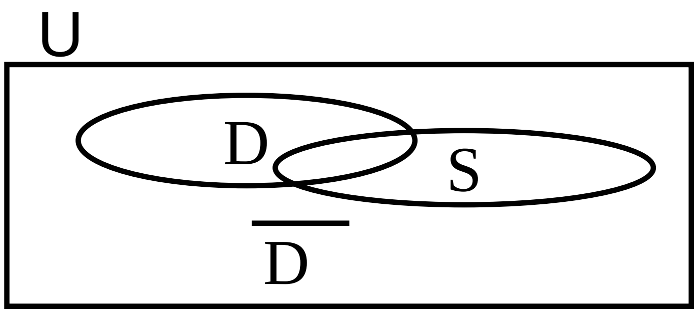
$$P(S \mid D) = \frac{P(S \,\&\, D)}{P(D)}$$

$$P(S \,\&\, D) = P(S \mid D) * P(D)$$

Reviewing Bayes Theorem

Symptom S

$$\textit{Diseases(health states)}\ D_i \ \textit{such that} \sum_i P(D_i) = 1$$



$$P(S) = P(S \,\&\, D) + P(S \,\&\, \overline{D})$$

$$P(S) = P(D)P(S \mid D) + P(\overline{D})P(S \mid \overline{D})$$

$$P(S) = \sum_j P(D_j) \times P(S \mid D_j)$$

$$P(D_i \mid S) = \frac{P(S \mid D_i) \times P(D_i)}{P(S)}$$

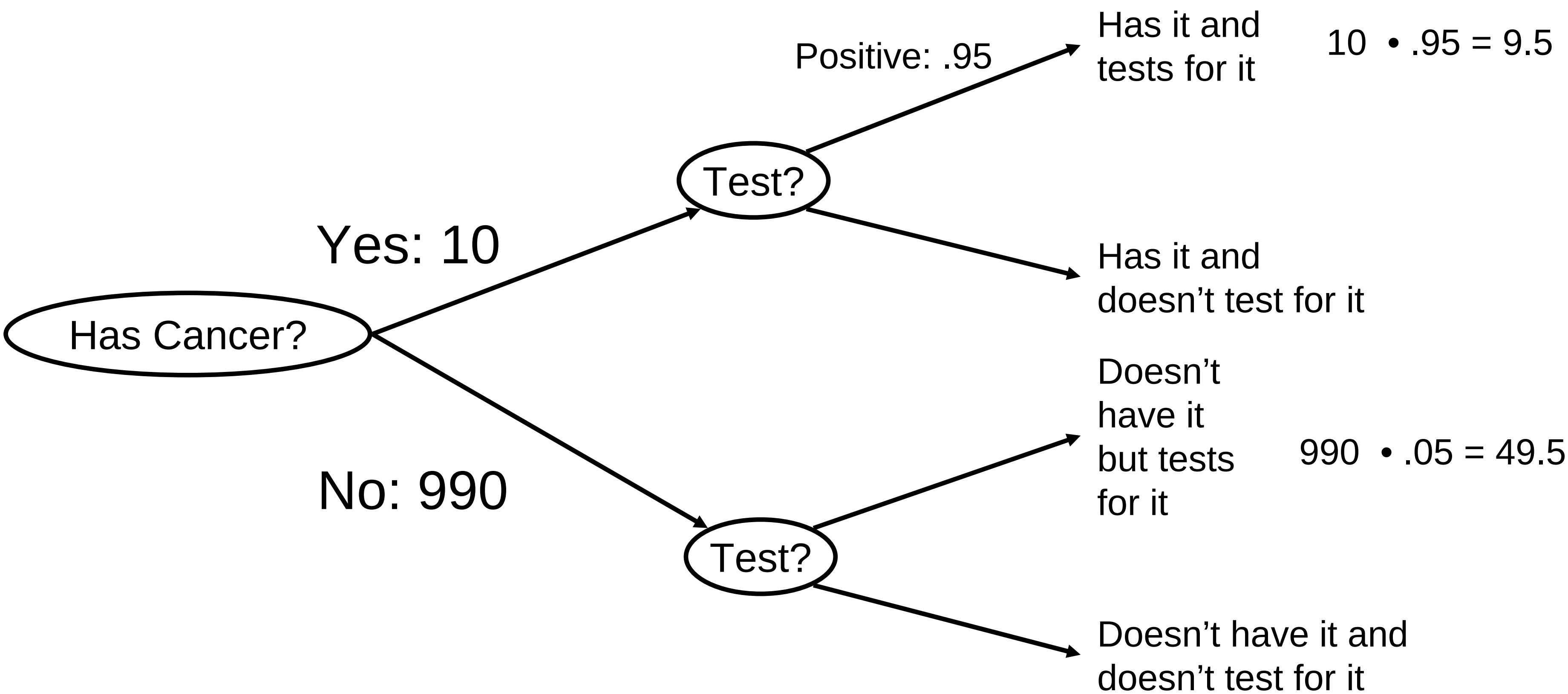
Conditional Probability of S given D

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

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

$$P(D \mid S) = \frac{P(S \mid D)P(D)}{P(S)}$$

Understanding Bayes Theorem



Number that test positive

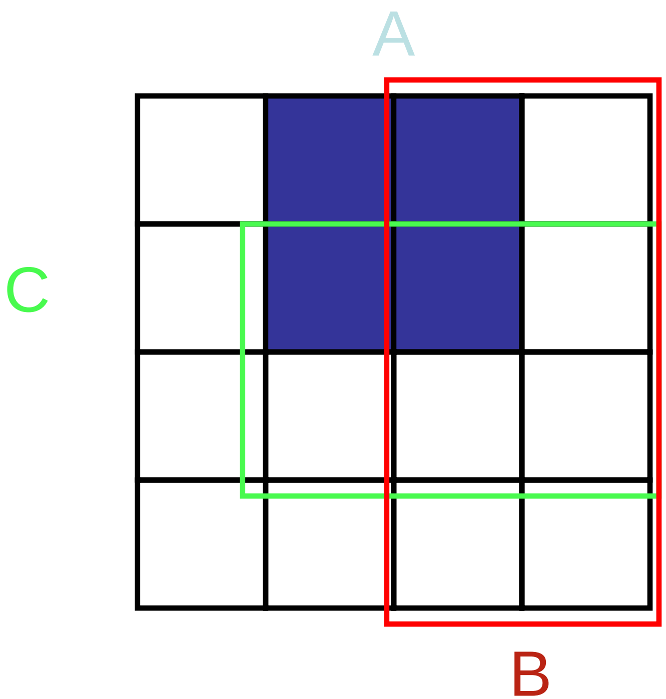
If you test positive your probability of having cancer is?

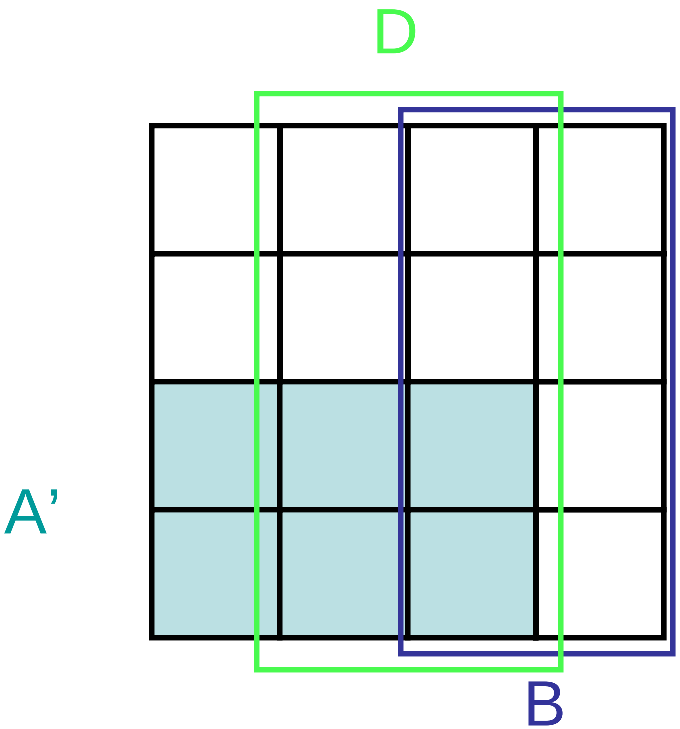
Independence, Conditional Independence

- Independence:
  $P(A\&B) = P(A) \bullet P(B)$
  - A varies the same within B as it does in the universe

- Conditional independence within C
  $P(A\&B|C) = P(A|C) \bullet P(B|C)$
  - When we restrict attention to C, A and B are independent

Examples





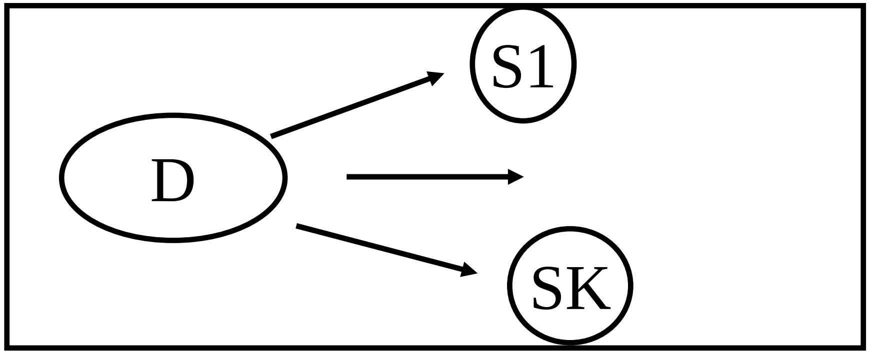
A and B are independent

A and B are conditionally *dependent*, given C

A' and B are dependent

A' and B are *conditionally independent*, given C.

Naïve Bayes Model



- Single disease, multiple symptoms
- N symptoms means how many probabilities?
- Assume symptoms conditionally independent
  – now P(S1,S2|D) = P(S1|D) * P(S2|D)
- Now?

Sequential Bayesian Inference

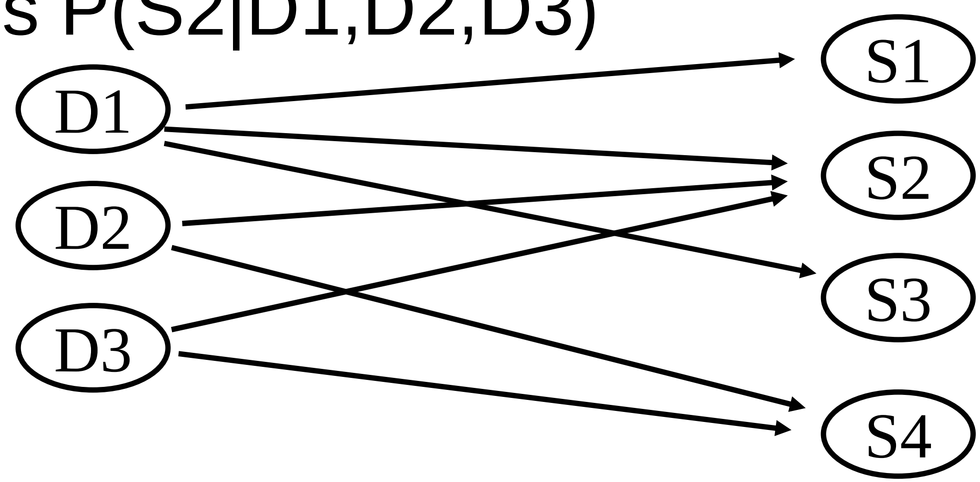
- Consider symptoms one by one
  - Prior probabilities P(Di)
  - Observe symptom Sj
  - Updates priors using Bayes Rule:

$$P(D_i) = \frac{P(S_j \mid D_i) \times P(D_i)}{P(S_j)}$$

  - Repeat for other symptoms using the resulting posterior as the new prior
- If symptoms are conditionally independent, same as doing it all at once
- Allows choice of what symptom to observe (test to perform) next in terms of cost/benefit.

Bipartite Graphs

- Multiple symptoms, multiple diseases
- Diseases are probabilistically independent
- Symptoms are conditionally independent
- Symptom probabilities depend only the diseases causing them
- Symptoms with multiple causes require joint probabilities P(S2|D1,D2,D3)



Noisy OR

Another element in the modeling vocabulary

Assumption: only 1 disease is present *at a time*

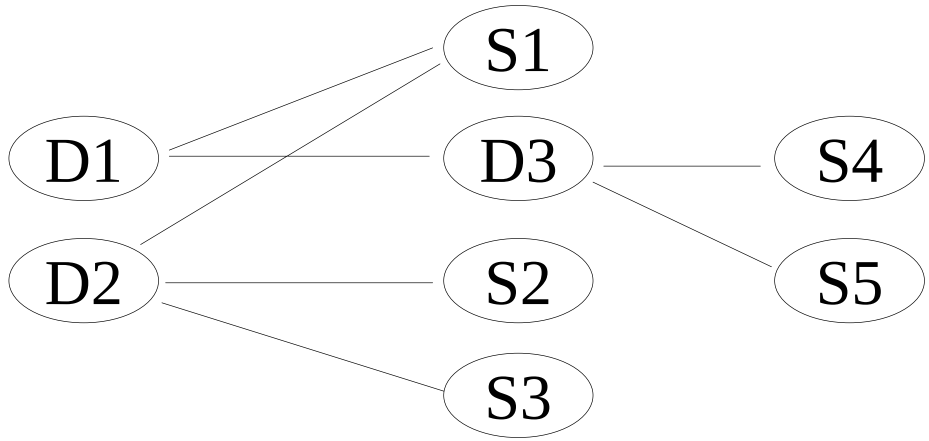
- Probability that all diseases cause the symptom is just the probability that at least 1 does
- Therefore: Symptom is absent only if no disease caused it.

$$1 - P(S2|D1,D2,D3) = (1 - P(S2|D1)) * (1 - P(S2|D2)) * (1 - P(S2|D3))$$

- Reduces probability table size: if n diseases and k symptoms, from k2^n to nk

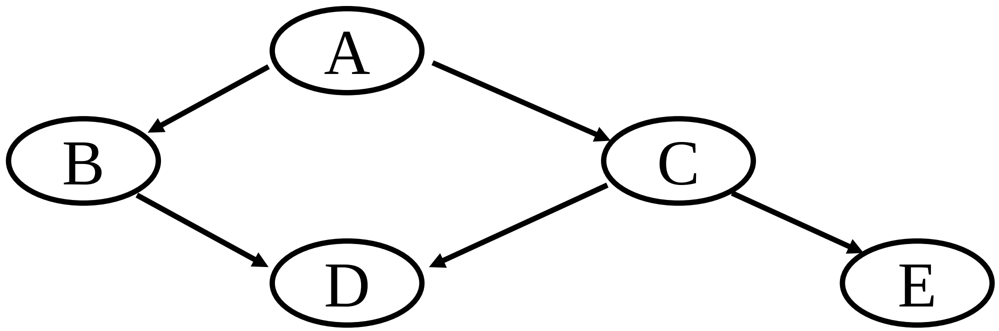
Polytrees

- What if diseases *do* cause or influence each other?



- Are there still well behaved versions?
- Polytrees: At most one path between any two nodes
  - Don't have to worry about "double-counting"
- Efficient sequential updating is still possible

Bayes Nets



- Directed Acyclic Graphs
- *Absence of link → conditional independence*
- P(X1,...,Xn) = Product P(Xi|{parents (Xi)})
- Specify joint probability tables over parents for each node

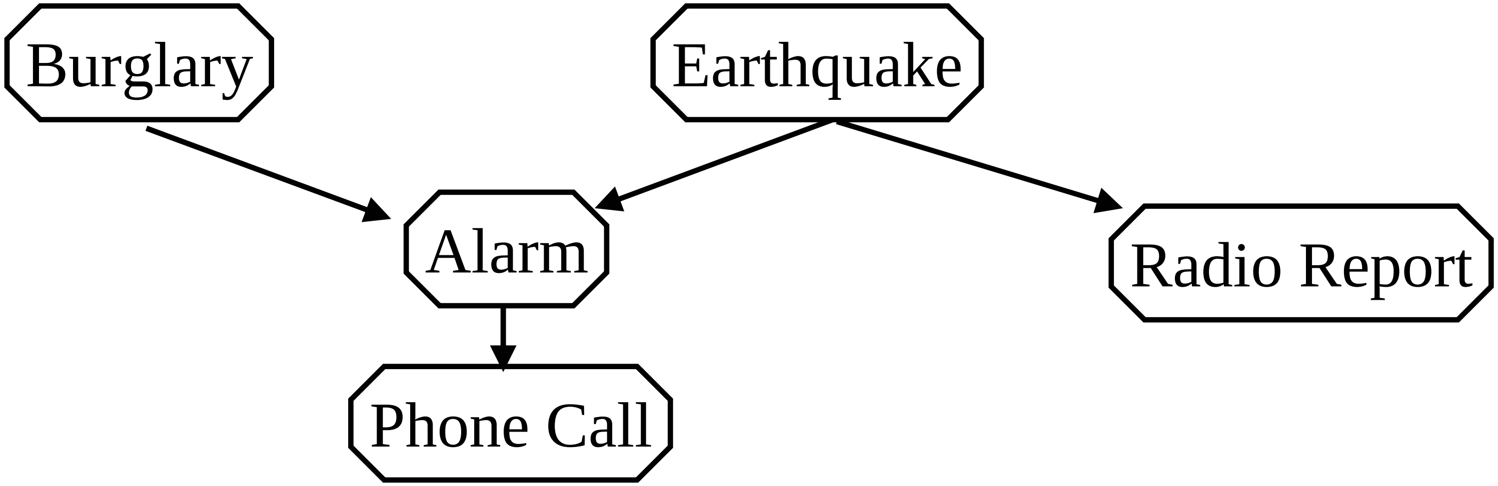
Probability A,B,C,D,E all true:

P(A,B,C,D,E) = P(A) * P(B|A) * P(C|A) * P(D|B,C) * P(E|C)

Probability A,C,D true; B,E false:

P(A,B’,C,D,E’) = P(A) * P(B’|A) * P(C|A) * P(D|B’,C) * P(E’|C)

Example



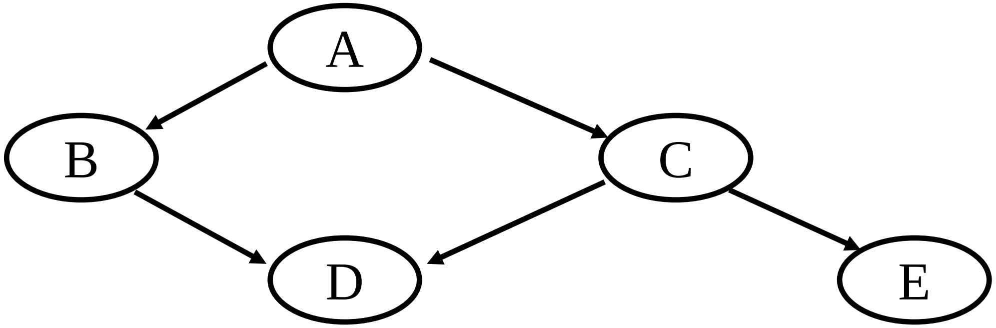
P(Call|Alarm)	t	f
t	.9	.01
f	.1	.99

P(RadioReport|Earthquake)	t	f
t	1	0
f	0	1

P(Alarm|B,E)	t,t	t,f	f,t	f,f
t	.8	.99	.6	.01
f	.2	.01	.4	.99

16 vs. 32 probabilites

Computing with Partial Information



- Probability that A true and E false:

$$P(A,\overline{E}) = \sum_{B,C,D} P(A,B,C,D,\overline{E})$$

$$= \sum_{B,C,D} P(A)P(B \mid A)P(C \mid A)P(D \mid B,C)P(\overline{E} \mid C)$$

$$= P(A)\sum_{C} P(C \mid A)P(\overline{E} \mid C)\sum_{B} P(B \mid A)\sum_{D} P(D \mid B,C)$$

- Graph separators (e.g. C) correspond to factorizations
- General problem of finding separators is NP-hard

Normally have to do 2^3 computations of the entire formula.
By factoring can do 2^3 computations of last term, 2 of second 2, 2 of first
Sum over c doesn’t change when D changes, etc.

Odds Likelihood Formulation

- Define *odds* as $$O(D) = \frac{P(D)}{P(\overline{D})} = \frac{P(D)}{1 - P(D)}$$
- Define *likelihood* as: $$L(S \mid D) = \frac{P(S \mid D)}{P(S \mid \overline{D})}$$

Derive complementary instances of Bayes Rule:

$$P(D \mid S) = \frac{P(D)P(S \mid D)}{P(S)} \qquad P(\overline{D} \mid S) = \frac{P(\overline{D})P(S \mid \overline{D})}{P(S)}$$

$$\frac{P(D \mid S)}{P(\overline{D} \mid S)} = \frac{P(D)P(S \mid D)}{P(\overline{D})P(S \mid \overline{D})}$$

Bayes Rule is Then: $$O(D \mid S) = O(D)L(S \mid D)$$

In Logarithmic Form: Log Odds = Log Odds + Log Likelihood

Decision Making

- So far: how to use evidence to evaluate a situation.
  - In many cases, this is only the beginning
- Want to take actions to improve the situation
- Which action?
  - The one most likely to leave us in the best condition
- Decision analysis helps us calculate which action that is

A Decision Making Problem

Two types of Urns: U1 and U2 (80% are U1)

U1 contains 4 red balls and 6 black balls

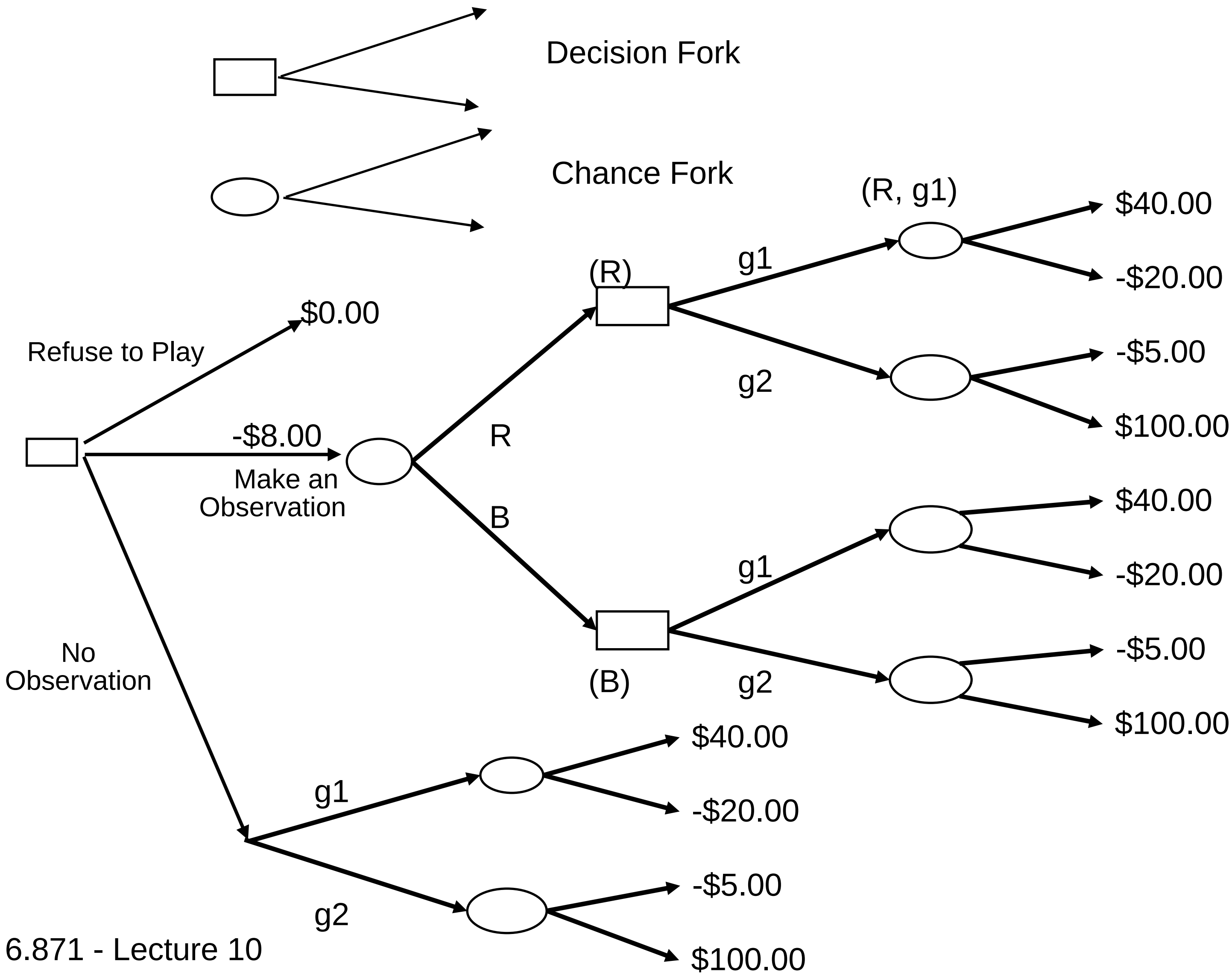
U2 contains nine red balls and one black ball

Urn selected at random; you are to guess type.

Courses of action:

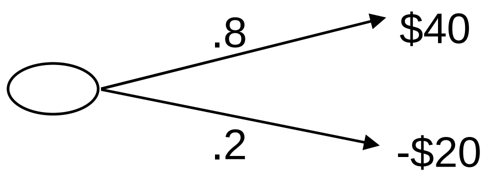
Refuse to play	No payoff, no cost
Guess it is of type 1	\$40 if right, -\$20 if wrong
Guess it is of type 2	\$100 if right, -\$5 if wrong
Sample a ball	\$8 for the right to sample

Decision Flow Diagrams



Expected Monetary Value

- Suppose there are several possible outcomes
  - Each has a monetary payoff or penalty
  - Each has a probability
- The Expected Monetary Value is the sum of the products of the monetary payoffs times their corresponding probabilities.

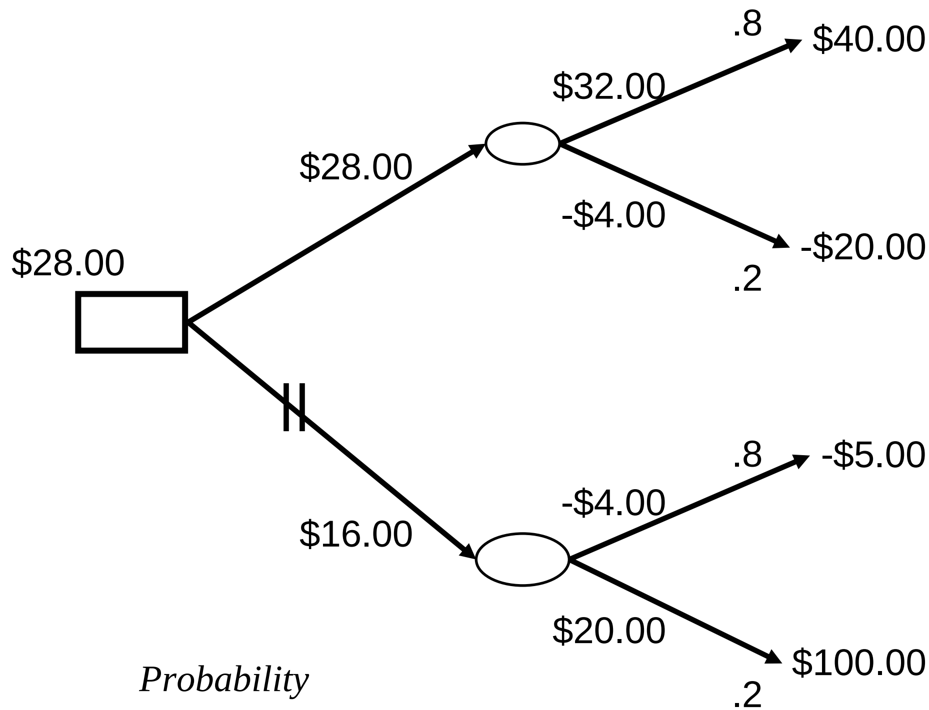


EMV = .8 · \$40 + .2 · -\$20 = \$32 + (-\$4) = \$28

- EMV is a normative notion of what a person who has no other biases (risk aversion, e.g.) should be willing to accept in exchange for the situation. You should be indifferent to the choice of $28 or playing the game.
- Most people have some extra biases; incorporate them in the form of a utility function applied to the calculated value.
- A rational person should choose the course of action with highest EMV.

Averaging Out and Folding Back

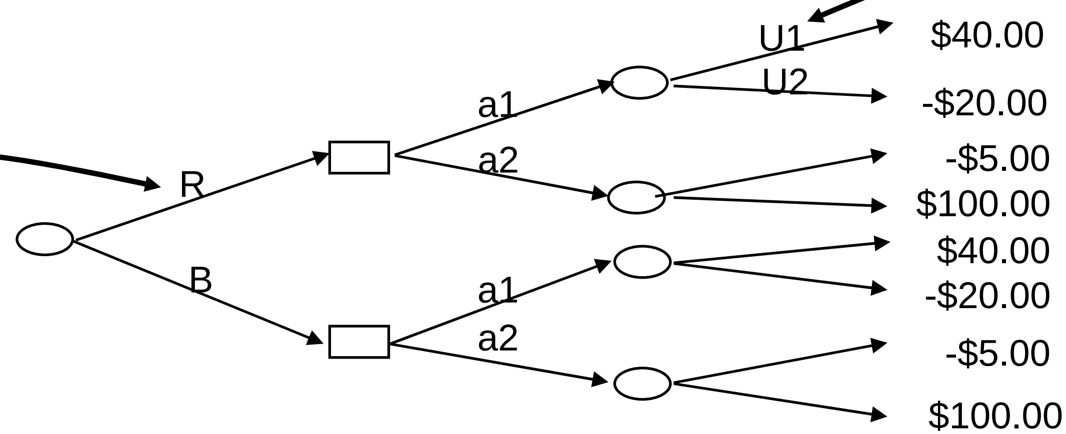
- EMV of chance node is probability weighted sum over all branches
- EMV of decision node is *max* over all branches



		Action		
State	**A1**	**A2**	**A3**	*Probability*
U1	40	-5	0	.8
U2	-20	100	0	.2
EMV	28	16	0	1

The Effect of Observation

Bayes theorem used to calculate probabilities at chance nodes following decision nodes that provide relevant evidence.



P(R) = P(R|U1) • P(U1) + P(R|U2) • P(U2)

P(U1|R) = P(R|U1) • P(U1) / P(R)

		Action		
State	**A1**	**A2**	**A3**	*Probability*
U1	40	-5	0	.8
U2	-20	100	0	.2
EMV	28	16	0	1

P(r|u1)= .4 P(U1)=.8 P(R|u2)=.9 P(u2)=.2 ➔ P(r)=.5

P(U1|r)= .4 * .8 / .5 = .64

Calculating the Updated Probabilities

Initial Probabilities

P(Outcome|State)	State	
Outcome	U1	U2
Red	.4	.9
Black	.6	.1
	.8	.2

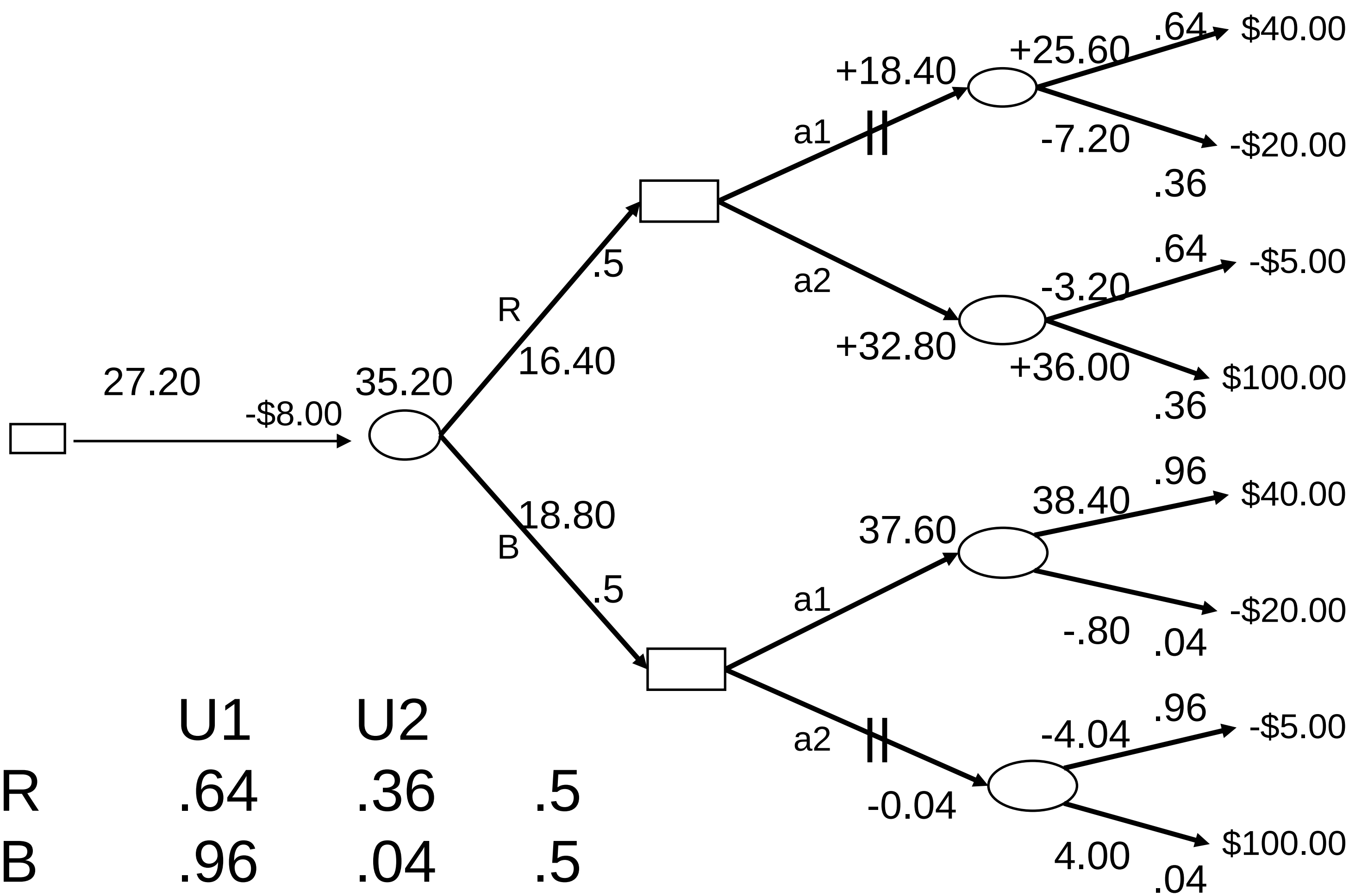
Joint (chain rule)

P(Outcome & State)	State		Marginal Probability
Outcome	U1	U2	of Outcome
Red	.4 • .8 = .32	.9 • .2 = .18	.50
Black	.6 • .8 = .48	.1 • .2 = .02	.50

Updated Probabilities

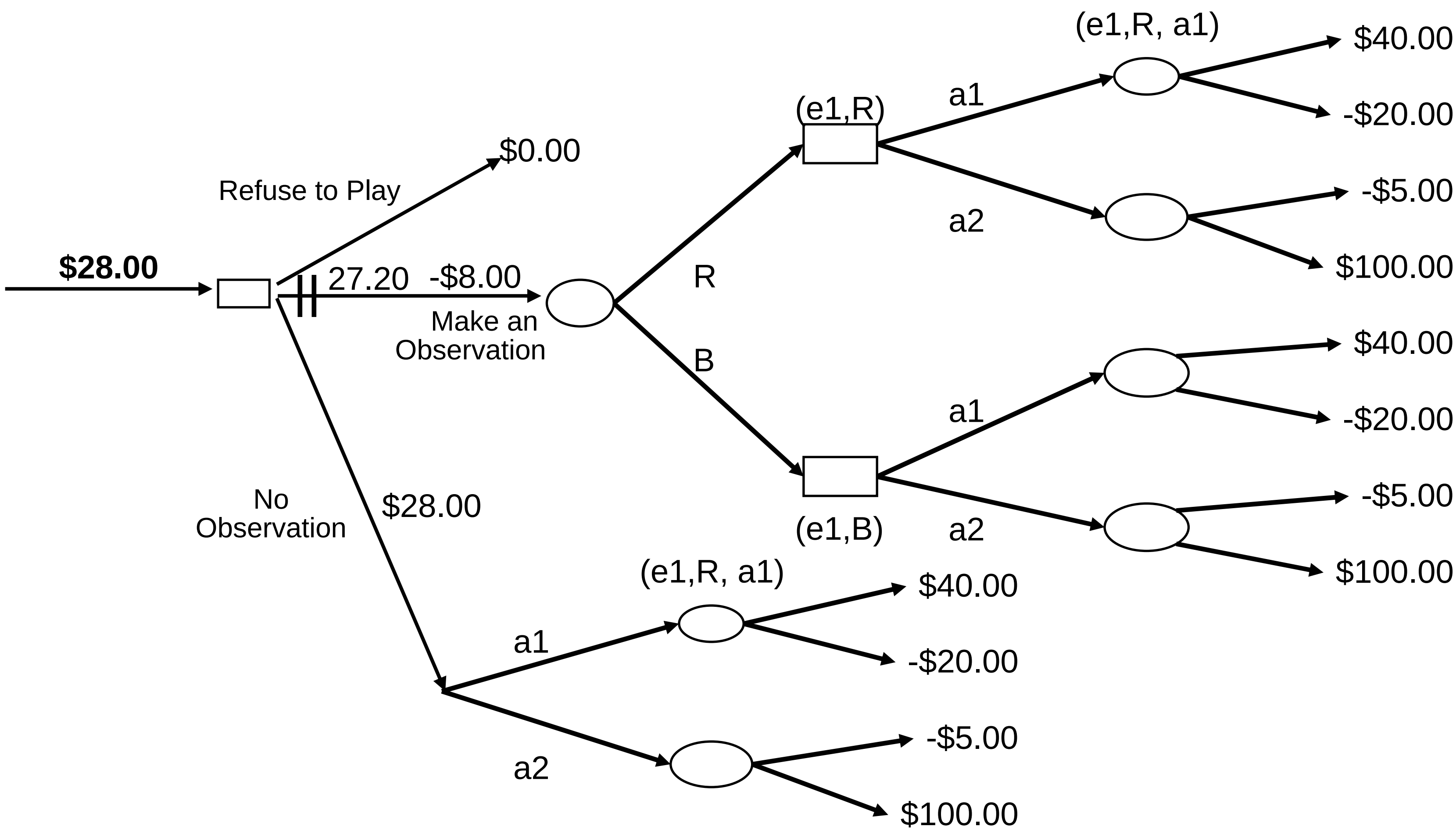
P(State |Outcome)	State	
Outcome	U1	U2
Red	.64	.36
Black	.96	.04

Illustrating Evaluation



	U1	U2	
R	.64	.36	.5
B	.96	.04	.5

Final Value of Decision Flow Diagram



Maximum Entropy

.1 .2 .5 .2

Several Competing Hypotheses
Each with a Probability rating.

- Suppose there are several tests you can make.
  - Each test can change the probability of some (or all) of the hypotheses (using Bayes Theorem).
  - Each outcome of the test has a probability.
  - We're only interested in gathering information at this point
  - Which test should you make?
- Entropy = Sum $-2 \cdot P(i) \cdot \text{Log } P(i)$, a standard measure
- Intuition
  - For .1, .2, .5, .2 = 1.06
  - For .99, .003, .003, .004 = .058

Maximum Entropy

- For each outcome of a test calculate the change in entropy.
  - Weigh this by the probability of that outcome.
  - Sum these to get an expected change of entropy for the test.
- Chose that test which has the greatest expected change in entropy.
  - Choosing test most likely to provide the most information.
- Tests have different costs (sometimes quite drastic ones like life and death).
- Normalize the benefits by the costs and then make choice.

Summary

- Several approaches to uncertainty in AI
- Bayes theorem, nets a current favorite
- Some tractable Bayesian situations
- A recurring theme: battling combinatorics through model assumptions
- Decision theory and rational choice