RN, Chapter 13

Decision Theoretic Agents: Acting Under Uncertainty

Environment is (in)accessible, nondeterministic
... known and modeled using belief networks

Decision Theoretic Agents

- Introduction to Decision Theory [Ch13]
  - Decision Theory 101
  - Probability 101
  - Bayes Theorem
  - Independence and Conditional Independence
  - Intro Belief Nets
  - Dutch Book Theorem
- Belief networks [Ch14]
- Dynamic Belief Networks [Ch15]
- Single Decision [Ch16]
- Sequential Decisions [Ch17]

Need to deal with Uncertainty

- Agent must act based on percepts (+ prior knowledge)
- Relatively easy if agent has COMPLETE knowledge of world
  - ie, if world is Accessible + Known
- Seldom true … even in Wumpus World:
  - Which square has wumpus? … has pit?
  - + Qualification problem, …
- Logical Reasoning can help ---

  eg, infer info about new state, from prior state, other info
  - But#1: … expects world to be Known, Deterministic, Discrete

    State s known, Result(a,s) is single state, …
  - But#2: … while Logic may specify *set* of options,

    need more to *decide which option to follow* (utilities)

Why not use Predicate Calculus?

- Eg: Consider diagnosing toothache:
  1. $\forall p$ Symptom(p, Toothache) $\Rightarrow$ Disease(p, Cavity)
     Wrong – other factors cause toothaches:
  2. $\forall p$ Symptom(p, Toothache) $\Rightarrow$ Disease(p, Cavity) v
     Disease(p, GumDisease) v Disease(p, ImpactedWisdom) v ...
     Too many! Maybe diagnostic:
  3. $\forall$ p Disease(p, Cavity) $\Rightarrow$ Symptom(p, Toothache)
     Wrong – many other factors (on lhs)!

- Difficulties of Building Exhaustive KB
  - *Laziness*: Just too many rules and contingencies
  - *Theoretical Ignorance*: No complete theory for the domain
  - *Practical Ignorance*: Don't have all the (patient) information available

- Probabilities provide way of summarizing uncertainty from
  - laziness
  - Ignorance (general, specific)

Using Probability

- Not everyone with cavity has toothache
  $\neg[\forall p \text{ Disease}(p, \text{Cavity}) \Rightarrow \text{Symptom}(p, \text{Toothache})]$
  but… perhaps *80%* do

- "*80%*" summarizes ...
  - factors required for cavity to cause toothache
  - patient has cavity & toothache (but unrelated)

Remaining *20%* ≡ all other possible causes of toothache

- Meaning:
  - An individual with cavity either has toothache, or not.
  - In 80% of situations where x has Cavity,
    (ie, indistinguishable from this situation based on current knowledge)
    x has toothache

Rational Decision

But#2: Need prob's to decide on action:

- Eg: Getting to airport:
  Leave 45 minutes early? … 90 mins? 2000 mins?
  Might arrive on time, or not.
  Might have to wait, or not.
  Might get ticket, or not.

... logic only specifies set of legal options/results…

- **Rational Decision**: Depends on
  - relative importance of various goals,
  - likelihood they will be achieved,
  - how much they are achieved.
- Need to know “likelihood" of possible occurrence
  ... not just that it is possible

Decision Theory

- **Utility Theory**:
  … gives preference ordering to outcomes of actions.
  (Every state has a degree of usefulness.)
- **Decision Theory =**
  Probability Theory + Utility Theory
- **Expected Utility**: Weight utility of outcome by probability that it occurs.
- **Maximize Expected Utility**:
  Prefer action that produces outcome (state) with maximal expected utility

Decision-Theoretical Agents

function DT-AGENT(*percept*) **returns** an *action*
static: probabilistic beliefs about the state of the world
calculate *updated probabilities* for current state based on available evidence including current *percept* and previous action
calculate *outcome probabilities* for actions, given action descriptions, probabilities of current states
select *action* with highest expected utility given probabilities of outcomes and utility information
return *action*

Terms from Probability Theory

- **Random Variable**:
  Weather $\in$ { Sunny, Rain, Cloudy, Snow }
- **Domain**: Possible values a random variable can take.
  (… finite set, $\Re$, ...)
- Probability distribution:
  mapping from domain to values $\in$ [0, 1]
- P(Weather) = $\langle$ 0.7, 0.2, 0.08, 0.02 $\rangle$

  means
  $$\left\{ \begin{array}{l} P(\text{ Weather } = \text{ Sunny }) = 0.7 \\ P(\text{ Weather } = \text{ Rain }) = 0.2 \\ P(\text{ Weather } = \text{ Cloudy }) = 0.08 \\ P(\text{ Weather } = \text{ Snow }) = 0.02 \end{array} \right\}$$
- Event:
  Each assignment (eg, Weather = Rain) is “event”

Probability as Relative Frequency

- What is probability of *event* E ?
- Over long sequence of experiments, ratio of
  - (# of times *E* occurred)
    number of times *E* occurs in sequence, to
  - (# of trials)
    total number of experiments
- Estimate:
  P(*E*) ≈ (# of times *E* occurred) /(# of trials)

- As (# of trials) $\rightarrow \infty$,
  ratio approaches true probability
  - given std assumptions

Example

- P(Swimmer succeeds)
  - Swimmer S tries...
    - 100 attempts are made to swim 50' in 15 secs
    - Succeeds 20 occasions
  - Estimate: probability that
    swimmer can swim 50' in 15 seconds is:
    - P(Swimmer succeeds) ≈ 20/100 = 0.2

- For probability to be meaningful, must clearly define
  - experiments
  - sample space
  - events

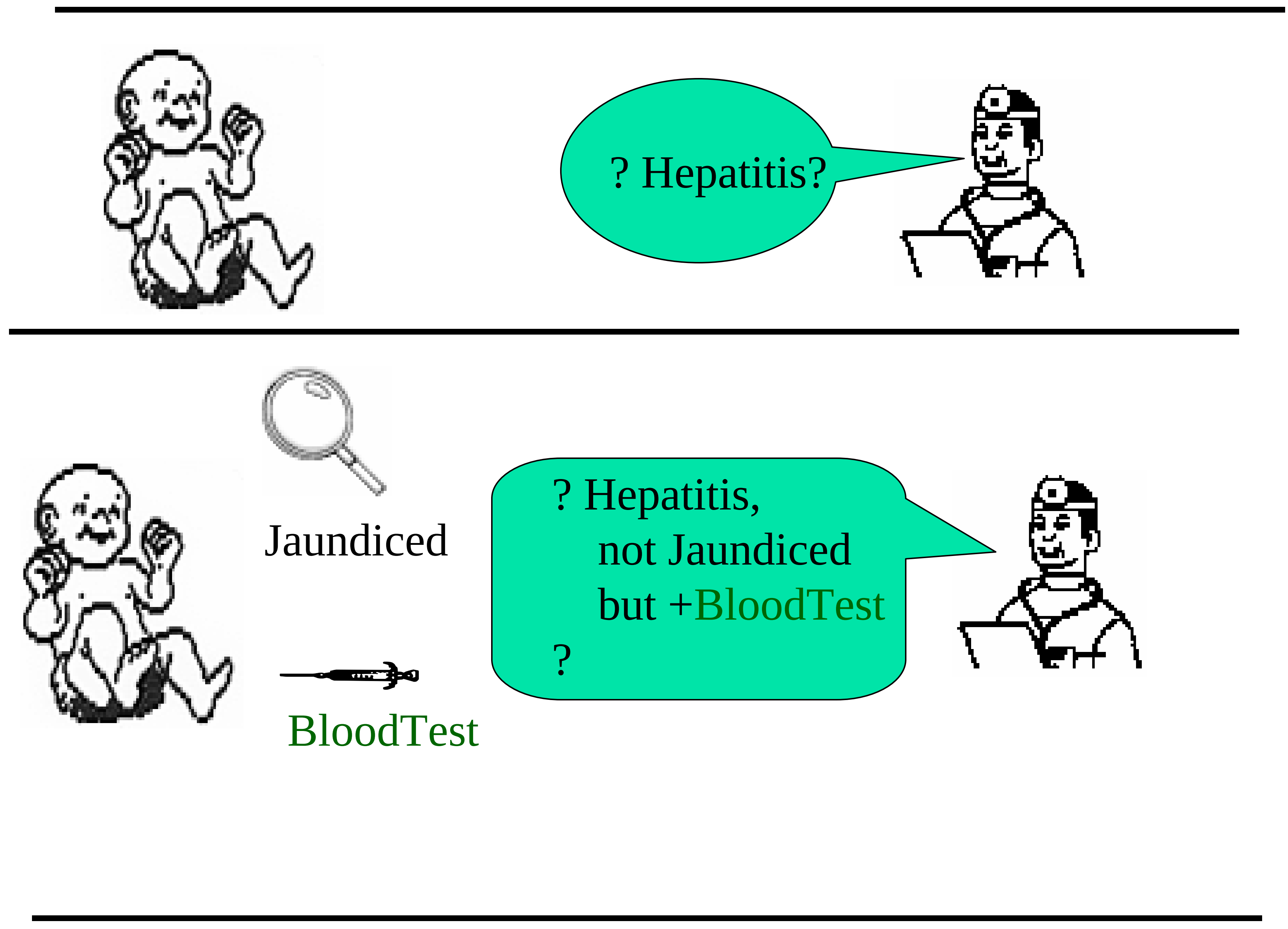
- What is the probability of an *accident* ?

Interpretations of Probability – A can of worms!

- Frequentist
  - $P(\alpha)$ is the frequency of α in the limit
  - Many arguments against this interpretation
    - What is the frequency of the event “it will rain tomorrow” ... “nuclear war tomorrow”?
- Subjective interpretation
  - $P(\alpha)$ is my degree of belief that α will happen
  - Where “degree of belief” means...

    If I say $P(\alpha)=0.8$, then I am willing to bet!!!
- For this class...

  we (mostly) don’t care what camp you are in


? Hepatitis?
Jaundiced
BloodTest
? Hepatitis,
not Jaundiced
but +BloodTest
?

Underlying Task

- *Situation*: Given observations $\{O_1=v_1, \ldots O_k=v_k\}$
  (symptoms, history, test results, …)
  what is best DIAGNOSIS Dx_i for patient?
- *Approach1*: Use set of obs_1 & … & $obs_m \rightarrow Dx_i$ rules

  but… *Need rule for each situation*
  - for each diagnosis Dx_r
  - for each set of possible values v_j for O_j
  - for each subset of obs. $\{O_{x1}, O_{x2}, \ldots\} \subset \{O_j\}$

  Can't use

  If Temp>100 & BP = High & Cough = Yes → DiseaseX

  if only know Temp and BP

- *Seldom Completely Certain*

Underlying Task

- *Situation*: Given observations $\{O_1=v_1, \ldots O_k=v_k\}$

  (symptoms, history, test results, ...)

  what is best DIAGNOSIS Dx_i for patient?

- *Approach 2*: Compute Probabilities of Dx_i

  given observations $\{O_1=v_1, \ldots O_k=v_k\}$

$$P(\text{ Dx} = u \mid O_1 = v_1, \ldots, O_k = v_k)$$

General Events

- **Atomic Event**: "Complete specification"
  Conjunction of assignments to EVERY variable [PossibleWorld]
- **Joint Probability Distribution**:
  Probability of every possible atomic event

n binary variables: 2^n entries
($2^n - 1$ independent values, as sum = 1)
A huge table!

J	B	H	P(j,b,h)
0	0	0	0.03395
0	0	1	0.0095
0	1	0	0.0003
0	1	1	0.1805
1	0	0	0.01455
1	0	1	0.038
1	1	0	0.00045
1	1	1	0.722

H	Hepatitis	
J	Jaundice	
B	(positive) Blood test	

Inference by Enumeration

- Using only joint probability distribution:

- For any proposition φ, add the atomic events where it is true:

  $P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$

J	B	H	P(j,b,h)
0	0	0	0.03395
0	0	1	0.0095
0	1	0	0.0003
0	1	1	0.1805
1	0	0	0.01455
1	0	1	0.038
1	1	0	0.00045
1	1	1	0.722

- P(+j)

  = 0.01455 + 0.038 + 0.00045 + 0.722

  = 0.775

Cost of Marginalization

- Called "marginal"

$$P(X_n) = \sum_{x_1,\ldots,x_{n-1}} P(x_1, \ldots, x_{n-1}, X_n)$$

- To compute marginal distribution P*(X$_n$)*:
  If all binary, 2^{n-1} additions
  - one term for each value of *$x_1, \ldots, x_{n-1}$*

H	Hepatitis
J	Jaundice
B	(positive) Blood test

Inference by Enumeration

- Using only joint probability distribution:

J	B	H	P(j,b,h)
0	0	0	0.03395
0	0	1	0.0095
0	1	0	0.0003
0	1	1	0.1805
1	0	0	0.01455
1	0	1	0.038
1	1	0	0.00045
1	1	1	0.722

- For any proposition φ, add the atomic events where it is true:

$$P(\varphi) = \Sigma_{\omega:\omega \models \varphi} P(\omega)$$

- P(-j v +b)

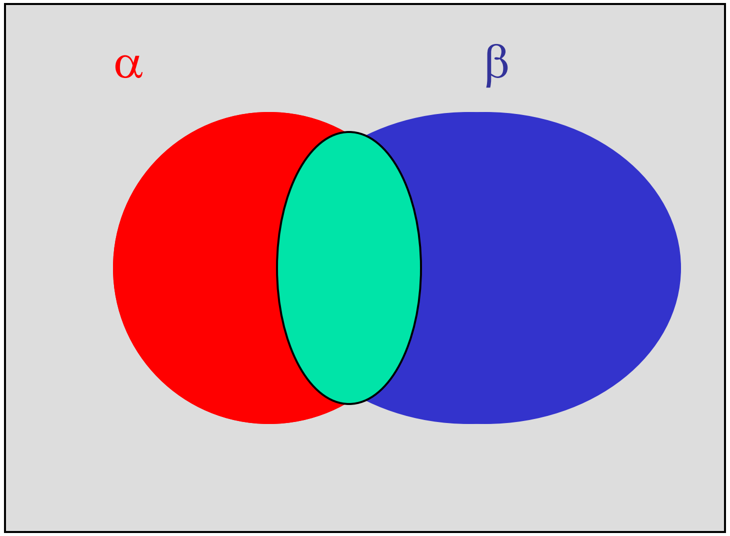
$= .03395 + .0095 + .0003 + .1805 + .00045 + .722 = 0.9467$

Conditional Probabilities

- After learning that β is true,
  how do we feel about α?
- If roll EVEN, what is chance of rolling 2?
- If have hepatitis, what is chance of jaundice?

β α

$$P(\alpha \mid \beta)$$



Conditional Probability

- Conditional Probability:
  P(α | β) = Probability of event α,
  given that event β has happened
- P(Jaundice | Hepatitis) = 0.8

- In gen'l:

$$P(\alpha \mid \beta) = \frac{P(\alpha \;\&\; \beta)}{P(\beta)}$$

$$P(\alpha \;\&\; \beta) = P(\alpha \mid \beta)\; P(\beta)$$

Conditional Probability

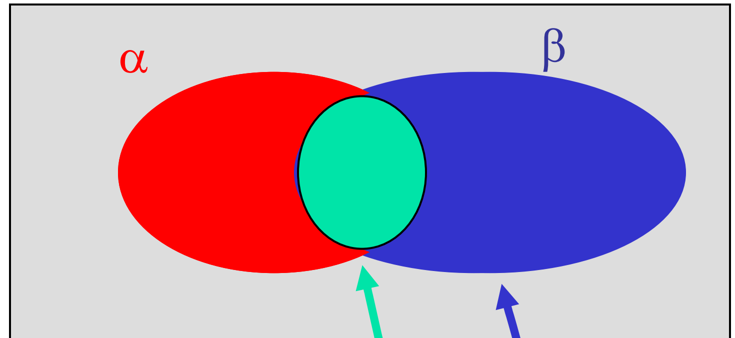
$$P(\alpha \mid \beta) = \frac{P(\alpha \& \beta)}{P(\beta)}$$

$$P(\alpha \& \beta) = P(\alpha \mid \beta)\, P(\beta)$$

- **Unconditional (prior) Probability**:
  - Probability of event before evidence is presented
  - P(Jaundice) = 0.04

    prob that someone (from this population) is jaundiced is 4 in 100
- **Evidence**: Percepts that affects degree of belief in event
- **Conditional (posterior) Probability**:
  - Probability of event after evidence is presented
  - N.b., posterior prob can be COMPLETELY different than prior prob!

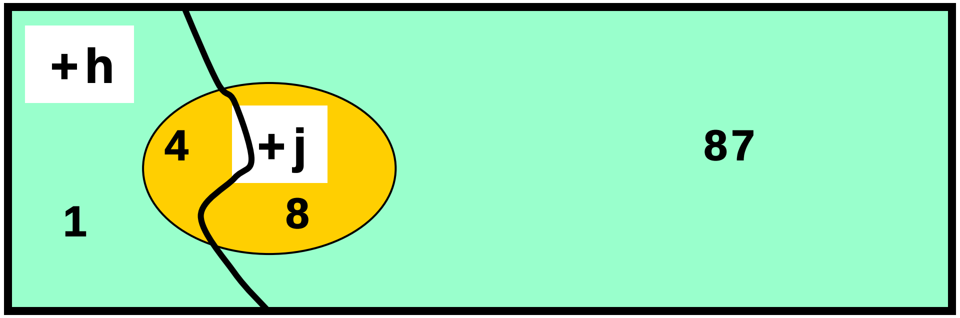
Frequentist: How to Estimate P(α|β)?



$$P(\alpha \mid \beta) = \frac{P(\alpha, \beta)}{P(\beta)} \approx \frac{\dfrac{\#(\alpha \text{ and } B)}{N}}{\dfrac{\#(\beta)}{N}} = \frac{\#(\alpha \text{ and } \beta)}{\#(\beta)}$$

$$P(\text{jaundice} \mid \text{hepatitis}) = \frac{\#(\text{jaundice and hepatitis})}{\#(\text{hepatitis})}$$

Hepatitis Examples



- 100 patients visit clinic
  - 5 have hepatitis
- 12 patients are jaundiced
  - 4 of these have hepatitis
- +h = the event that a randomly selected patient has hepatitis
- +j = the event that a randomly selected patient is jaundiced
- P(+j) = ? P(+j|+h) = ?

H	Hepatitis
J	Jaundice
B	(positive) Blood test

Inference by Enumeration

- Using only joint probability distribution:

J	B	H	P(j,b,h)
0	0	0	0.03395
0	0	1	0.0095
0	1	0	0.0003
0	1	1	0.1805
1	0	0	0.01455
1	0	1	0.038
1	1	0	0.00045
1	1	1	0.722

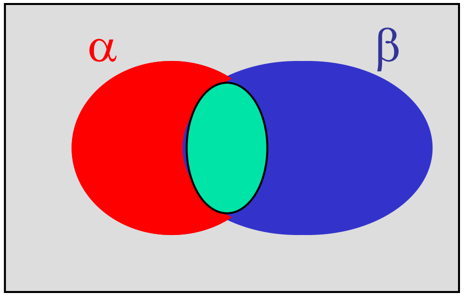
- Can compute *conditional probabilities*:

$$P(-b \mid +j) = \frac{P(-b \wedge +j)}{P(+j)} = \frac{0.01455 + 0.038}{0.01455 + 0.038 + 0.00045 + 0.722}$$

$$\approx 0.0678$$

Useful Rule #1: The chain rule

- $P(\alpha, \beta) = P(\alpha)\, P(\beta | \alpha)$



- More generally:

$$P(\alpha_1, \ldots, \alpha_k) = P(\alpha_1)\, P(\alpha_2|\alpha_1) \cdots P(\alpha_k|\alpha_1, \ldots, \alpha_{k-1})$$

  - ... any order ...

    $P(\alpha_1, \ldots, \alpha_k) = P(\alpha_3)\, P(\alpha_7|\alpha_3)\, P(\alpha_{14}|\alpha_3, \alpha_7) \cdots$

Useful Rule #2. Bayes rule

- $$P(\alpha \mid \beta) = \frac{P(\beta \mid \alpha) P(\alpha)}{P(\beta)}$$

- More generally, external event γ:

$$P(\alpha \mid \beta \cap \gamma) = \frac{P(\beta \mid \alpha \cap \gamma) P(\alpha \mid \gamma)}{P(\beta \mid \gamma)}$$

Bayes' Rule and its Use

- **Diagnosis** typically involves computing P(Hypothesis | Symptoms)

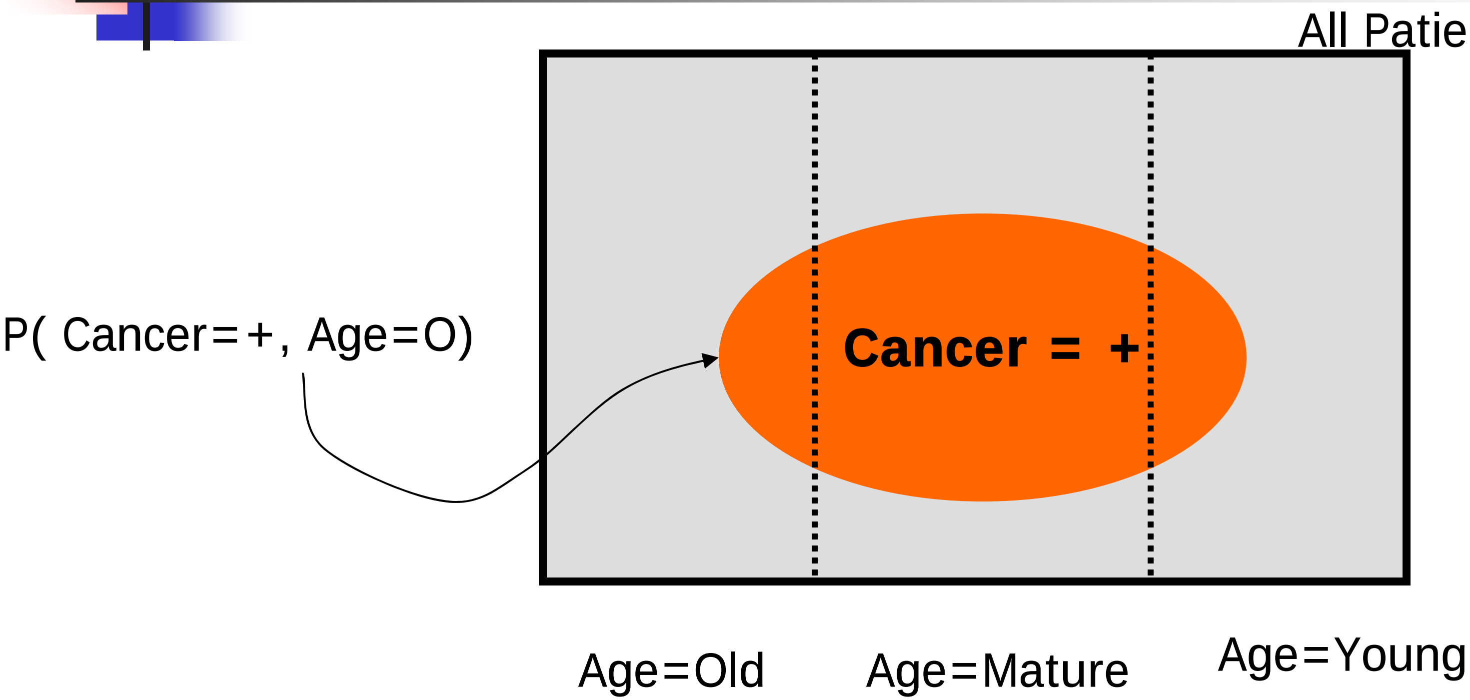
  What is P(Meningitis | StiffNeck) ?
  ≡ prob that patient A has meningitis, given that A has stiff neck?
- Typically have . . .
  - Prior prob of meningitis P(M) = 1/50,000
  - Prior prob of having a stiff neck P(SN) = 1/20
  - Prob that meningitis causes a stiff neck P(SN | M) = 1/2
- Bayes' Rule:

$$P(M \mid SN) = \frac{P(SN \mid M)\; P(M)}{P(SN)}$$

- Eg: P(M | SN) = P(SN | M) P(M) / P(SN) = 0.5 ×0.00002/0.05 = 0.0002

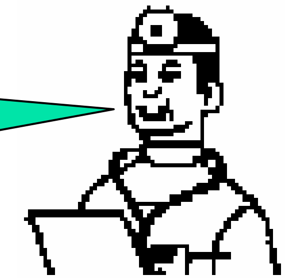
- Only 1 in 5000 stiff necks have meningitis...
  even though SN is major symptom of M...

Factoids



$$P(+c) = \sum_a P(+c, A = a)$$

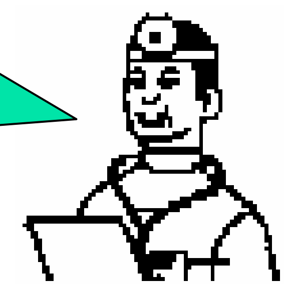
? Hepatitis?



Jaundiced

BloodTest

? Hepatitis,
not Jaundiced
but +BloodTest
?



What is P(+h | – j, +b) ?

Combining Evidence

- P(+h | -j, +b) ≡
  prob of +Hep, given { -Jaun, +BloodTest }?

- Bayesian Update:

$$
\begin{aligned}
P(+\mathrm{h}|\{\}) &= P(+\mathrm{h}) \\
P(+\mathrm{h}|\text{-j}) &= P(+\mathrm{h}|\{\}) \frac{P(-j|+h)}{P(-j)} \\
P(+\mathrm{h}|\text{-j}, +\mathrm{b}) &= P(+\mathrm{h}|\text{-j}) \frac{P(+b\,|\,+h,-j)}{P(+b|-j)}
\end{aligned}
$$

Each time new evidence is observed (-j, +b, ...) ,
belief in unknown (+h)
is multiplied by factor that depends on new evidence.

(Note: independent of order of observations)

Using Independence

- Needs 3rd order information: P(+b | +h, –j)
  ...Not always available...
- But *sometimes*, NOT NEEDED!
  P(+b | +h, – j) = P(+b | +h)
  (Prob of symptom2, given disease and symptom1
  ≡ Prob of symptom2, given disease)
- If so. . .

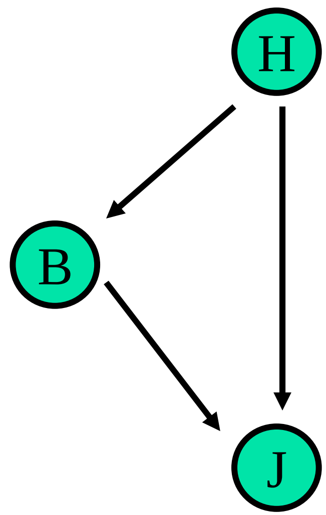
$$P(+\text{h}|\text{-j},+\text{b}) \;=\; P(+\text{h})\frac{P(-j|+h)}{P(-j)}\frac{P(+b\,|+h)}{P(+b|-j)}$$

- ASSUMPTION is NOT ALWAYS TRUE!
  But when it is, just need 2nd order statistics!

- Even better:
  * Denominator is P(–j) P(+b | –j) = P(–j, +b)
  * Independent of H; just normalizing term!

Simple Belief Net

P(H=1)	P(H=0)
0.05	0.95



h	P(B=1 | H=h)	P(B=0 | H=h)
1	0.95	0.05
0	0.03	0.97

h	b	P(J=1|h,b)	P(J=0|h,b)
1	1	0.8	0.2
1	0	0.8	0.2
0	1	0.3	0.7
0	0	0.3	0.7

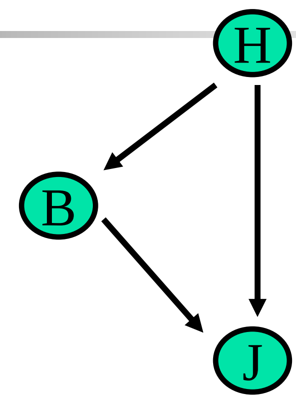
- Node ~ Variable

  Link ~ "Causal dependency"
- "CPTable" ~ P(child | parents)

Encoding Causal Links

P(H=1)
0.05

h	P(B=1 | H=h)
1	0.95
0	0.03



h	b	P(J=1|h , b)
1	1	0.8
1	0	0.8
0	1	0.3
0	0	0.3

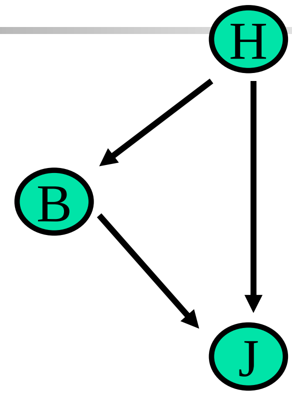
- $P(J \mid H, B=0) \;=\; P(J \mid H, B=1) \;\; \forall\; J,\; H$!

  $\Rightarrow$ **$P(J \mid H, B) = P(J \mid H)$**
- *J* is INDEPENDENT of *B*, once we know *H*
- Don't need $B \rightarrow J$ arc!

Encoding Causal Links

P(H=1)
0.05

h	P(B=1 | H=h)
1	0.95
0	0.03



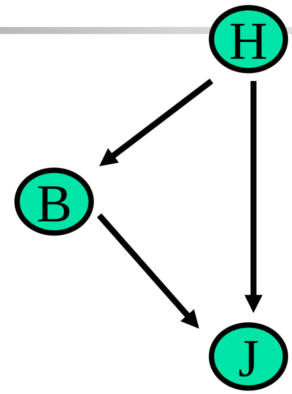
h		P(J=1|h)
1		0.8
1		
0		0.3
0		

- $P(J \mid H, B=0) \;=\; P(J \mid H, B=1) \quad \forall\, J,\; H\,!$

  $\Rightarrow$ **$P(J \mid H, B) = P(J \mid H)$**
- *J* is <u>INDEPENDENT</u> of *B*, once we know *H*
- Don’t need $B \rightarrow J$ arc!

Encoding Causal Links

h	P(B=1 | H=h)
1	0.95
0	0.03



P(H=1)
0.05

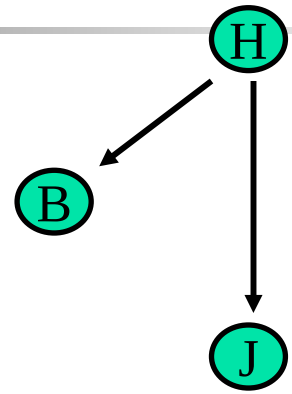
h	P(J=1|h)
1	0.8
0	0.3

- $P(J \mid H, B{=}0) \;=\; P(J \mid H, B{=}1) \;\; \forall\, J,\; H\,!$

  $\Rightarrow \quad \mathbf{P(J \mid H, B) = P(J \mid H)}$
- *J* is <u>INDEPENDENT</u> of *B*, once we know *H*
- Don't need $B \rightarrow J$ arc!

Sufficient Belief Net

h	P(B=1 | H=h)
1	0.95
0	0.03



P(H=1)
0.05

h	P(J=1|h)
1	0.8
0	0.3

- Requires: *P(H=1)* known
  P(J=1 | H=1) known
  P(B=1 | H=1) known

  (Only 5 parameters, not 7)

Hence: $P(\text{H=1} \mid \text{B=1, J=0}) = \frac{1}{\alpha} P(\text{H=1})\, P(\text{B=1} \mid \text{H=1})\, \text{P(J=0} \mid \text{B=1,H=1)}$ ~~P(J=0 |B=1,H=1)~~ → $P(\text{J=0} \mid \text{H=1})$

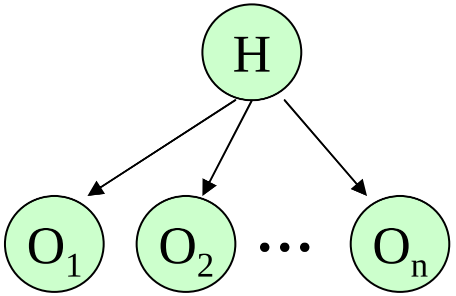
"Naïve Bayes"

- Classification Task:

  Given $\{ O_1 = v_1, \ldots, O_n = v_n \}$

  Find h_i that maximizes $P(H = h_i \mid O_1 = v_1, \ldots, O_n = v_n)$

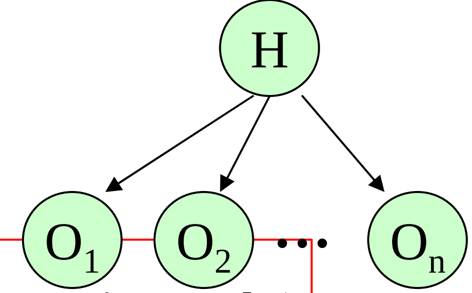
- Using
  - $P(H = h_i)$
  - $P(O_j = v_j \mid H = h_i)$
  - *Independent:* $P(O_j \mid H, O_k, \ldots) = P(O_j \mid H)$



$$P(H = h_i \mid O_1 = v_1 \ldots, O_n = v_n) = \frac{1}{\alpha} P(H = h_i) \prod_j P(O_j = v_j \mid H = h_i)$$

- Find $\operatorname{argmax} \{h_i\}$

Naïve Bayes (con't)



$$P(H = h_i \mid O_1 = v_1, \ldots, O_n = v_n) = \frac{1}{\alpha} P(H = h_i) \prod_j P(O_j = v_j \mid H = h_i)$$

- *Normalizing term* $\alpha = P(O_1 = v_1, \ldots, O_n = v_n) = \sum_i P(H = h_i) \prod_j P(O_j = v_j \mid H = h_i)$

  (No need to compute, as same for all h_i)

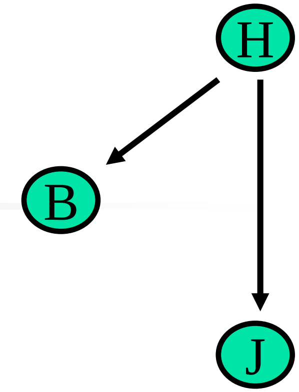
- Easy to use for Classification
- Can use even if some v_js not specified
- If k Dx's and n O_is,

  requires only k priors, $n * k$ pairwise-conditionals

  (Not 2^{n+k}… relatively easy to learn)

n	1+2n	$2^{n+1} - 1$
10	21	2,047
30	61	2,147,438,647

"Factoring"



- B *does* depend on J:

  If J=1, then likely that H=1 $\Rightarrow$ *B =1*

- *but…* *ONLY THROUGH* H:
  - If know H=1, then likely that B=1
  - … doesn't matter whether J=1 or J=0 !

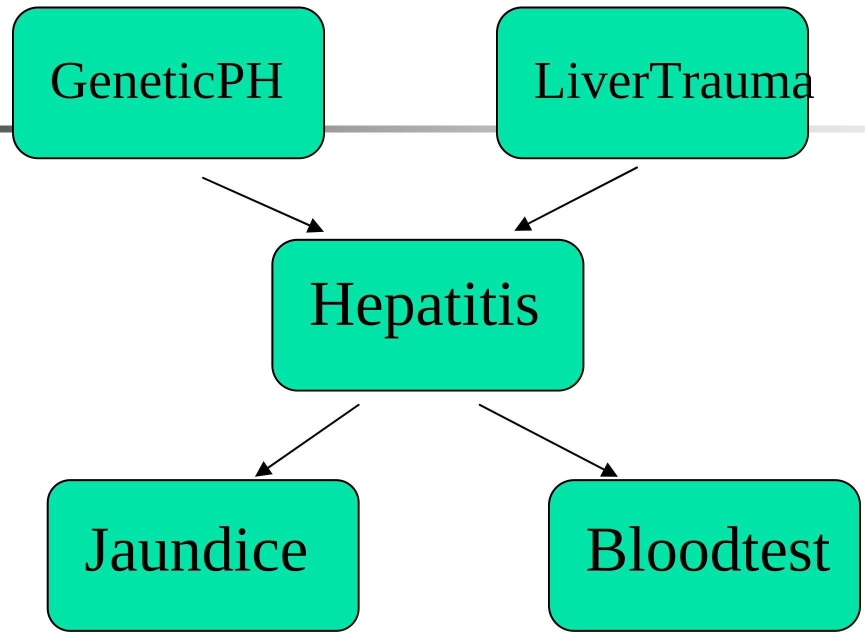
  $\Rightarrow$ **P(J=0 | B=1, H=1) = P(J=0 | H=1)**

N.b., B *and* J ARE *correlated* a priori *P(J | B) ≠ P(J)*
GIVEN H, *they become uncorrelated* *P(J | B, H) = P(J | H)*

Bigger Networks

P(H=1)
0.32

P(I=1)
0.20



g	lt	P(H=1|g ,lt)
1	1	0.82
1	0	0.10
0	1	0.45
0	0	0.04

h	P(J=1| h)
1	0.8
0	0.3

h	P(B=1| h)
1	0.98
0	0.01

- Intuition: Show *CAUSAL* connections:
  GeneticPH CAUSES Hepatitis; Hepatitis CAUSES Jaundice
- If GeneticPH, then expect Jaundice: GeneticPH ⇒ Hepatitis ⇒ Jaundice

But only via Hepatitis: GeneticPH and not Hepatitis $\not\Rightarrow$ Jaundice

$$P(\text{J} \mid \text{G}) \neq P(\text{J}) \quad \text{but}$$
$$P(\text{J} \mid \text{G,H}) = P(\text{J} \mid \text{H})$$

Factored Distribution

- Symptoms *independent,* given Disease

H	Hepatitis
J	Jaundice
B	(positive) Blood test

$P(\text{ B} \mid \text{J })\ \neq\ P(\text{ B })$ but

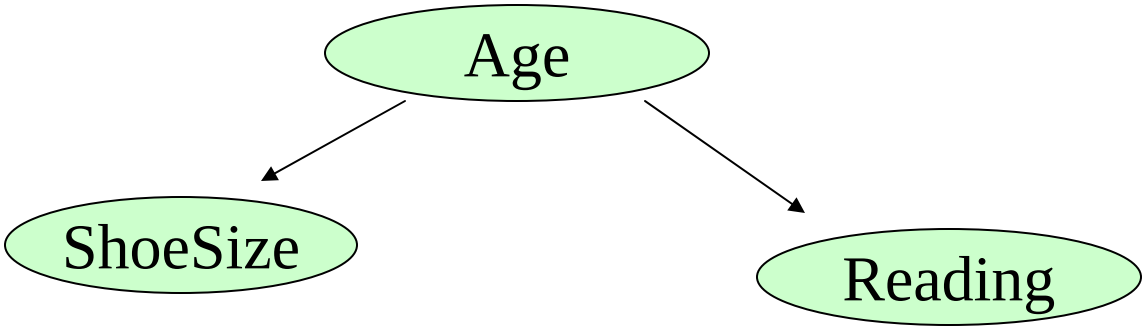
$P(\text{ B} \mid \text{J,H })\ =\ P(\text{ B} \mid \text{H })$

- **ReadingAbility** and **ShoeSize** are dependent,

$P(\text{ReadAbility} \mid \text{ShoeSize }) \neq P(\text{ReadAbility })$

but become independent, given Age

$P(\text{ReadAbility} \mid \text{ShoeSize, Age }) = P(\text{ReadAbility} \mid \text{Age})$



Important concept: (a) Independence

- Coin tosses:
  - T_1: the first toss is a head; T_2: the second toss is a tail
  - $P(T_2 \mid T_1) = P(T_2)$
- α and β ***independent*** iff $P(\beta|\alpha)=P(\beta)$
  - $P \models (\alpha \perp \beta)$
  - ... dist P entails α indep of β
- **Proposition:** α and β *independent* if and only if $P(\alpha \Leftrightarrow \beta) = P(\alpha)\, P(\beta)$

Independence

- Events α and β are independent *iff*
  - $P(\alpha, \beta) = P(\alpha) P(\beta)$
  - $P(\alpha \mid \beta) = P(\alpha)$
  - $P(\alpha \vee \beta) = 1 - (1 - P(\alpha)) (1 - P(\beta))$

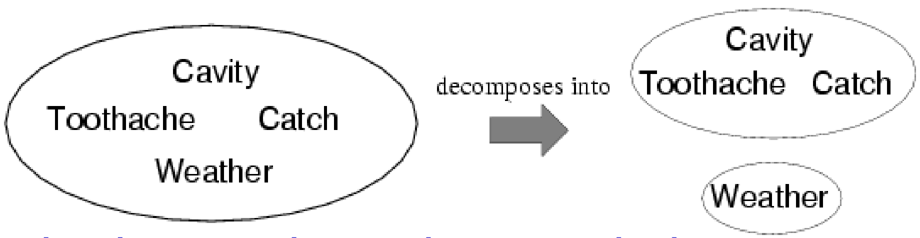
- Variables independent

  $\Leftrightarrow$ independent for all values

$\forall a, b \quad P(A = a, B = b) = P(A = a) \ P(B = b)$

Independence

- *A* and *B* are independent iff
  P(A|B) = P(A) or P(B|A) = P(B) or P(A, B) = P(A) P(B)



P(*Toothache, Catch, Cavity, Weather*)
= P(*Toothache, Catch, Cavity*) P(*Weather*)

- 16 entries reduced to 9;
  for *n* independent biased coins, $O(2^n) \rightarrow O(n)$

- Absolute independence powerful… but rare

- Dentistry is a large field with hundreds of variables, none of which are independent.
  … What to do?

Important concept: (b) Conditional independence

- Independence is rarely true, but conditionally...
  - Shoe size is NOT independent of ReadingAbility
  - But given AGE...

- α and β ***conditionally independent*** given γ if $P(\beta \mid \alpha \Leftrightarrow \gamma) = P(\beta \mid \gamma)$
  - $P \models (\alpha \perp \beta \mid \gamma)$

Proposition: $P \;{}_{\mathrm{T}}\!\geqq (\alpha \perp \beta \mid \gamma)$ if and only if

$P(\alpha \Leftrightarrow \beta \mid \gamma) = P(\alpha \mid \gamma)\, P(\beta \mid \gamma)$

Conditional Independence

- **P**(*Hep, Jaun, BT*) has $2^3 - 1 = 7$ ind. entries
- Given +Hep, Jaun doesn't depend on blood test :
  (1) P (Jaun | +h, BT) = P(Juan | +h)
- The same independence holds given -Hep:
  (2) P(Jaun | -h, BT) = P(Juan| -h)
- *Jaun* is conditionally independent of *BT* given *Hep*:
  P(Jaun | H, BT) = P(Juan| H)
- Equivalent statements:
  P(*Jaun* | *BT, Hep*) = P(*Jaun* | *Hep*)
  P(*Jaun* , *BT* | *Hep*) = P(*Jaun* | *Hep*) P(*BT* | *Hep*)

Conditional Independence

- Events E_1 and E_2 are conditionally independent given E iff

$$P(E_1 \mid E, E_2) = P(E_1 \mid E)$$

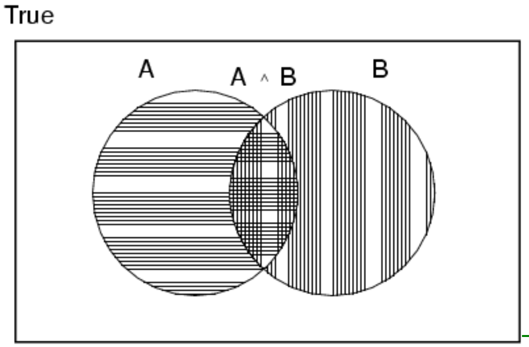
- Given E, knowing E_2 does not change the probability of E_1
- Equivalent formulations:

$$P(E_1, E_2 \mid E) = P(E_1 \mid E)\, P(E_2 \mid E)$$

$$P(E_2 \mid E, E_1) = P(E_2 \mid E)$$

Basic concepts for random variables

- Atomic outcome: assignment $x_1,\ldots,x_n$ to $X_1,\ldots,X_n$
- Conditional probability: $P(X,Y) = P(X)\ P(Y|X)$
- Bayes rule: $P(X|Y) = P(Y|X)\ P(X)\ /\ P(Y)$
- Chain rule:
  $P(X_1,\ldots,X_n) =$
  $P(X_1)\ P(X_2|X_1)\ldots\ P(X_k|X_1,\ldots,X_{k-1})\ \ldots\ P(X_n|X_1,\ \ldots,\ X_{n-1})$



Probability Theory

- Axioms:

$0 \leq P(A) \leq 1$

$P(\text{ True }) = 1, \quad P(\text{ False }) = 0$

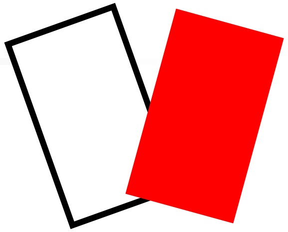
$P(A \vee B) = P(A) + P(B) - P(A \,\&\, B)$

$P(A) + P(\neg A) = 1$

- Not arbitrary:
  - If Agent1 use probabilities that violate axioms, then

    ∃ betting strategy s.t.
    Agent1 guaranteed to lose $
  - “Dutch book”

The Three-Card Problem



- Three cards
  - RR = red on both sides
  - WW = white on both sides
  - RW = red on one side, white on the other
- Draw single card randomly and toss it into the air.
- What is the probability ...

  a. ... of drawing red-red? P(D_RR)

  b. ... that the drawn cards lands white side up? P(W_up)

  c. ... that the red-red card was not drawn,
  assuming that the drawn card lands red side up.
  P(not-D_RR | R_up)

Fair Bets

B believes

- P(D_RR) = 1/3
- P(W_up) = 1/2
- P(not-D_RR | R_up) = ~~1/3~~ 1/2

- A bet is fair to an individual B if,
  - according to B's probability assessment,
  - the bet will break even in the long run.
- B thinks these 3 bets are fair :

Bet **(a)** : Win $4.20 if D_RR;
lose $2.10 otherwise. [B believes P(D_RR)=1/3]

Bet **(b)**: Win $2.00 if W_up;
lose $2.00 otherwise. [B believes P(W_up)=1/2]

Bet **(c)**: Win $4.00 if R_up and not D_RR;
lose $4.00 if R_up and D_RR;
win $0 if not-R_up.

[B believes P(not-D_RR | R_up)=1/2]

Possible Outcomes

1. W_up & not-D_RR: Some card other than RR is drawn, which lands *white* side up.
2. R_up & not-D_RR: Some card other than RR is drawn, which lands *red* side up.
3. R_up & D_RR: RR is drawn, which l... (of course) red sid...

(a): Win \$4.20 if D_RR;
lose \$2.10 otherwise.

(b): Win \$2.00 if W_up;
lose \$2.00...

...t D_RR;

...RR;

B is always guaranteed to lose money...

- whichever card is drawn, &
- however it lands !

			3
(a)	– 2.10	– 2.10	+ 4.20
(b)	+2.00	– 2.00	– 2.00
(c)	± 0.00	+ 4.00	– 4.00

The Dutch Book Theorem

- Spse B accepts any bet it thinks is fair. Then...
- a Dutch book can be made against B

  iff

  B's assessment of probability violates Bayesian axiomatization.

- Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations.
- Rank the following by probability
(1 = most probable; 8 = least probable)
  a. Linda is a teacher in elementary school.
  b. Linda works in a bookstore and takes yoga classes.
  c. Linda is an active feminist.
  d. Linda is psychiatric social worker.
  e. Linda is a member of the League of Women Voters.
  f. Linda is a bank teller.
  g. Linda is an insurance salesperson.
  h. Linda is a bank teller and is an active feminist.

Summary

- Decision Theory 101
- Probability 101
  - Terms; Frequentist vs Subjective
  - Independence, Conditional Independence
  - Bayes Theorem
- Intro Belief Nets
- Axioms and properties
  - Dutch Book Theorem

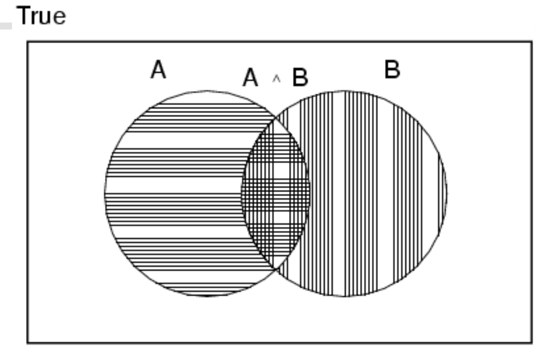
Event Spaces

- Outcome space Ω
- Measurable events S
  - Each $\alpha \in S$ is a subset of Ω
  - S must contain
    - Empty event ϕ
    - Trivial event Ω
  - S closed under
    - Union: $\alpha \cup \beta \in S$
    - Complement: $\alpha \in S$, then $\Omega - \alpha$ also in S

Eg,

- $\Omega = \{1,2,3,4,5\}$
- $S = 2^{\Omega}$
- $\alpha = \{1,2\}$
- $\{1,2\} \in S$ & $\{2,3\} \in S$ $\Rightarrow \{1,2,3\} \in S$
- When $|\Omega| = \infty$, need other tricks

Probability Distribution P over (Ω, S)



- $P(\alpha) \geq 0$
- $P(\Omega) = 1$
- If $\alpha \cup \beta = \phi$, then $P(\alpha \cup \beta) = P(\alpha) + P(\beta)$

- From here, you can prove a lot:
  - $P(\phi) = 0$
  - $P(\alpha \cup \beta) = P(\alpha) + P(\beta) - P(\alpha \cap \beta)$

Random Variable

- Events are complicated – we think about attributes
  - Age, Grade, HairColor
- Random variables formalize attributes:
  - Grade=A shorthand for event $\{\omega \in \Omega: f_{Grade}(\omega) = A\}$
- Properties of random vars, X:
  - Val(X) = possible values of random var X
  - For discrete (categorical): $\sum_{i=1...|Val(X)|} P(X=x_i) = 1$
  - For continuous: $\int_x p(X=x)\, dx = 1$