CS 4260 and CS 5260
Vanderbilt University

Lecture on Uncertainty (Probabilities)

This lecture assumes that you have

- Read Section 8.1 through 8.3 of ArtInt (though there is some repetition, as well as additional material)

ArtInt: Poole and Mackworth, Artificial Intelligence 2E
at http://artint.info/2e/html/ArtInt2e.html
to include slides at http://artint.info/2e/slides/ch04/lect1.pdf

Douglas H. Fisher

Probability (review of basics)

Finite outcome spaces (worlds): $\Omega = \omega 1, \omega 2, \ldots, \omega N$ with $P(\omega i)$ such that

- $0 <= P(\omega i) <= 1.0$, and
- $\Sigma\ P(\omega i) = 1.0$

Examples:

- Roll of a die: $\Omega = 1, 2, 3, 4, 5, 6$
- Toss of a coin: $\Omega = H, T$

Probability (review of basics)

Assigning probabilities to each ωi can be:

- *subjective* (first principles, aka "educated guess" based on domain knowledge) or
- *objective* (proportion/frequency)
  - [ωi] = number of occurrences of ωi in experiment
  - P(ωi) = [ωi] / ∑[ωj]

Examples

- subjective (fair die, fair coin): P(1) = 1/6, …, P(6) = 1/6 P(H) = ½, P(T) = ½
- objective: 57 H in 100 tosses: P(H) = 57/100, P(T) = 43/100

  These are correct/exact probabilities over the conducted experiments, but only an estimate of probability in general.

Probability (review of basics)

Additivity of probabilities.

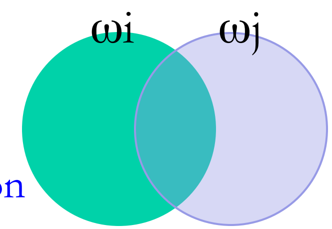
If ωi and ωj are mutually exclusive, then P(ωi or ωj) = P(ωi) + P(ωj)

Example

- In a die roll: P(2 or 5) = P(2) + P(5) = 2/6 = 1/3 (2 and 5 are mutually exclusive)

If ωi and ωj are NOT (necessarily) mutually exclusive,

then P(ωi or ωj) = P(ωi) + P(ωj) – P(ωi and ωj)



Don't double count the intersection

Independence of different outcome spaces or independence of sequence of draws/trials/experiments from same outcome space.

If independent then

P(ωi in trial/space 1 followed by ωj in trial/space 2) = P(ωi) * P(ωj) for all ωi and ωj

Examples

- P(H, H) = P(H)*P(H) = ¼
- P(H, 3) = P(H) * P(3) = ½ * 1/6 = 1/12
  - [H]/([H]+[T]) * ([3]/[1]+[2]+[3]+[4]+[5]+[6])

Probability (review of basics)

Countably infinite outcome spaces: e.g., flip a coin until first head

Ω = H, TH, TTH, TTTH, …..

P(H) = ½, P(TH) = ¼, P(TTH) = 1/8, P(TTTH) = 1/16, ….

probabilities must still sum to 1, since these outcomes are mutually exclusive.

$$\sum_{i=0}^{\text{inf}} 1/(2^{\wedge}(i+1)) = 1$$

Joint outcome spaces and probabilities: W1 ✖ W2

if W1 = H, T then W1 ✖ W1 = H,H H,T T,H T,T

(each with a probability, and probabilities must sum to 1)

if W2 = 1, 2, 3, 4, 5, 6 then W1 ✖ W2 = H,1 H,2 … H,6 T,1 ….. T,6

(each with a probability, and probabilities must sum to 1)

Douglas H. Fisher

Probability (review of basics)

Random "variables": a real-valued function defined over an outcome space:

Ω = H, TH, TTH, TTTH, …..

X1(Ω) =	0	1	2	3	….	(# of T before first H)
X2(Ω)	0.00001	…..		0.0002	…	(muscle fatigue)
X3(Ω)	100.0	…………		95.4	…..	(patience)

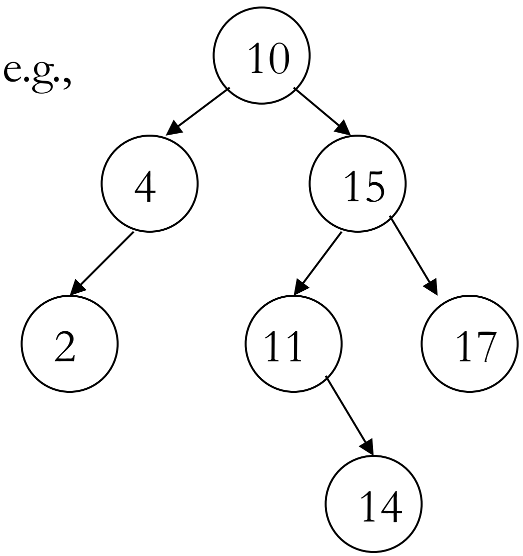
A random variable defines an outcome space. Probabilities can be assigned to random variable values.

After this lecture, I will not use the term "random variable" often. There is considerable confusion around the term. I will refer to variables, outcome spaces (or worlds), and functions defined over outcome spaces.

Probability (review of basics)

Expected value of a random variable (function): EX = ∑ [xi * P(X(**Ω**)=xi)] = ∑ [X(ωi)*P(ωi)]
where xi is a value of the random variable (function) X(**Ω**), and ωi in an outcome in **Ω**

e.g.,



Expected number of examined nodes in successful search of a binary search tree

Ω = (lookup)	10,	4,	15,	2,	11,	17,	14
P(ωi)	0.1	0.1	0.05	0.05	0.2	0.3	0.2
X(ωi)	1	2	2	3	3	3	4

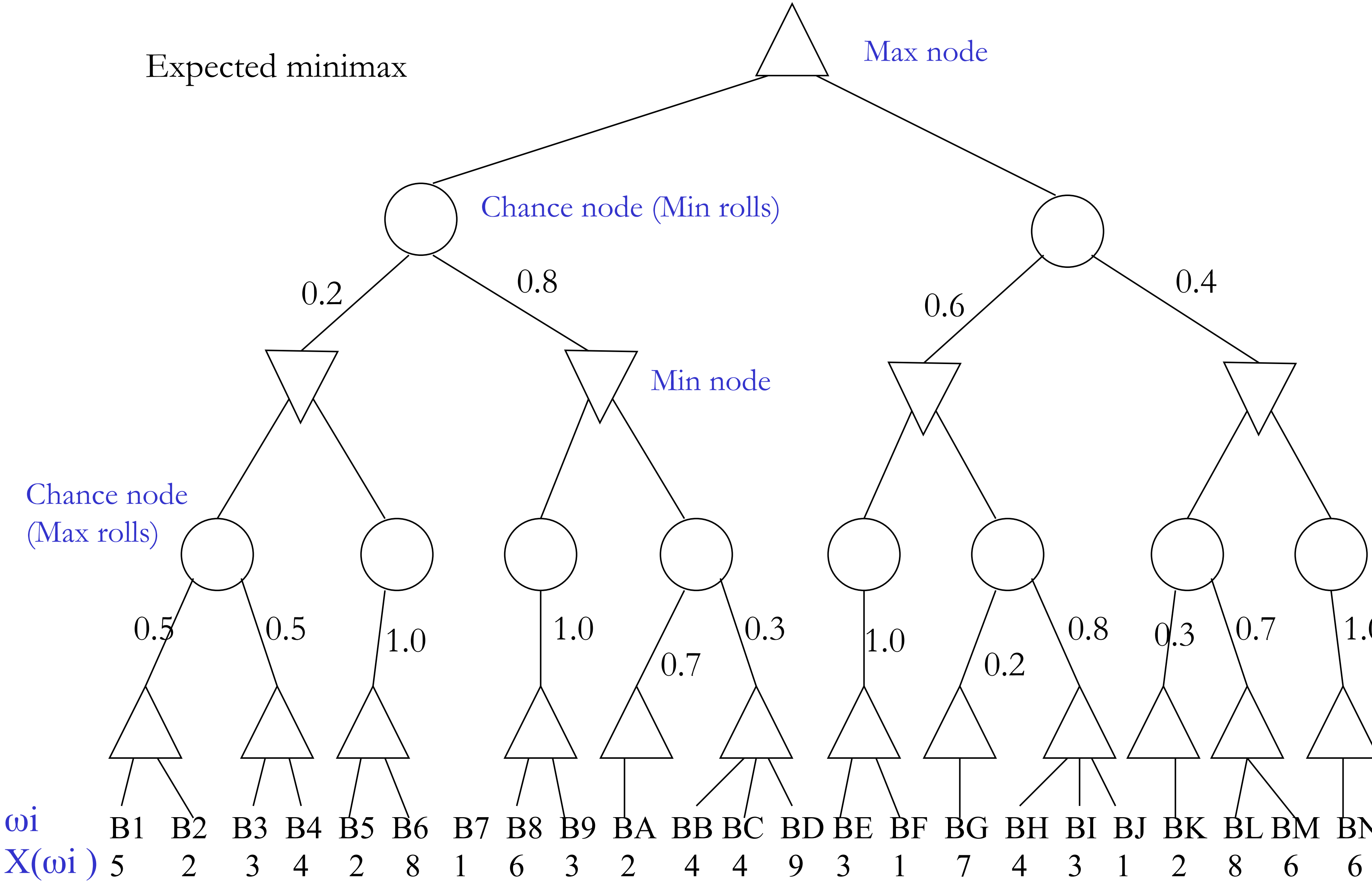
EX = 1(0.1) + 2(0.1+0.05) + 3(0.05+0.2+0.3) + 4(0.2)

Random variables (functions) used to represent cost (space, time), value (goodness, utility), etc.

Probabilistic Games
(e.g., Risk)

$$EX = \sum_{xi} [xi * P(X(\Omega)=xi)] = \sum_{\omega i} [X(\omega i)*P(\omega i)]$$

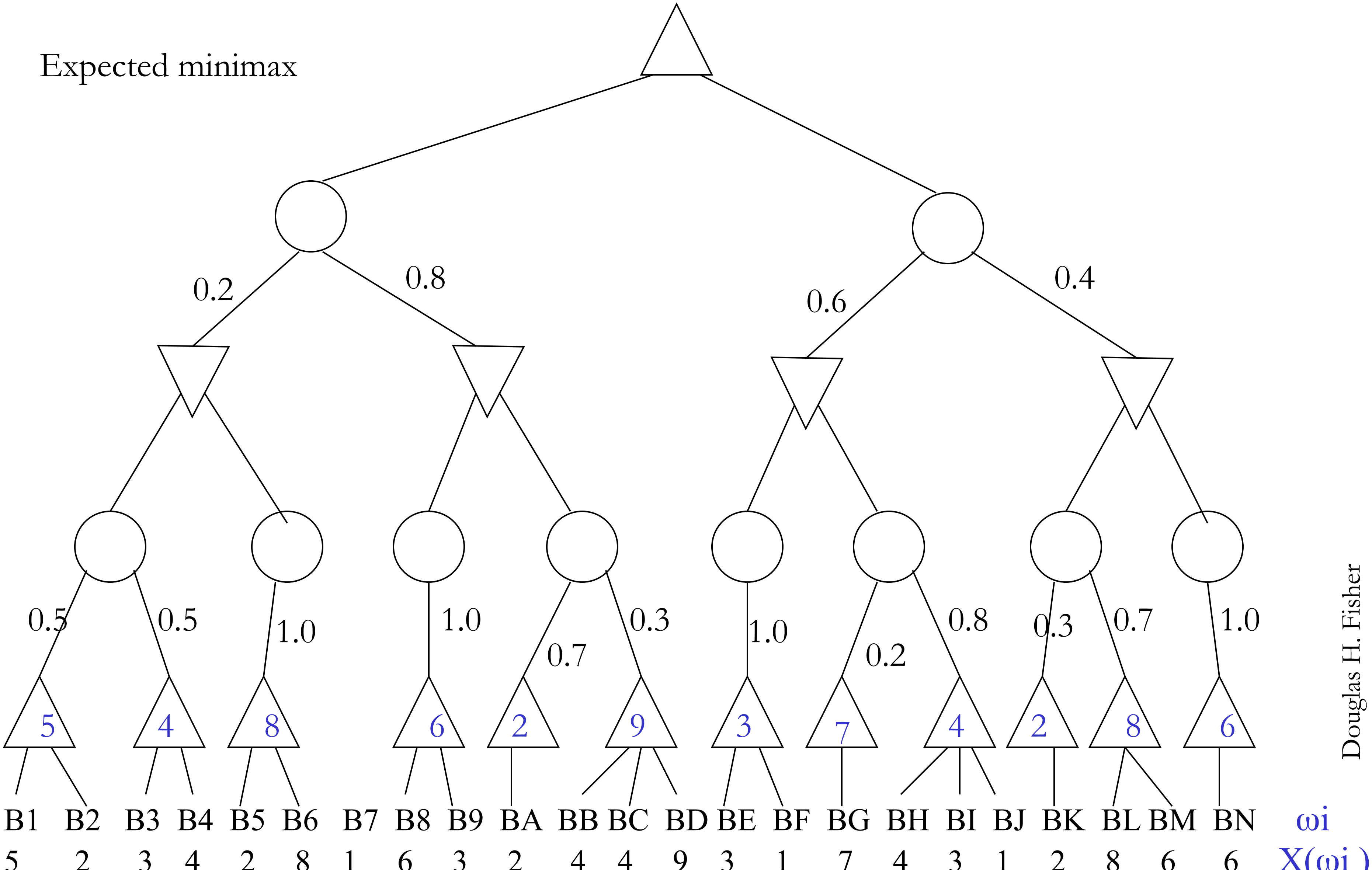
Expected minimax



Douglas H. Fisher

$$EX = \sum_{xi} [xi * P(X(\boldsymbol{\Omega})=xi)] = \sum_{\omega i} [X(\omega i) * P(\omega i)]$$

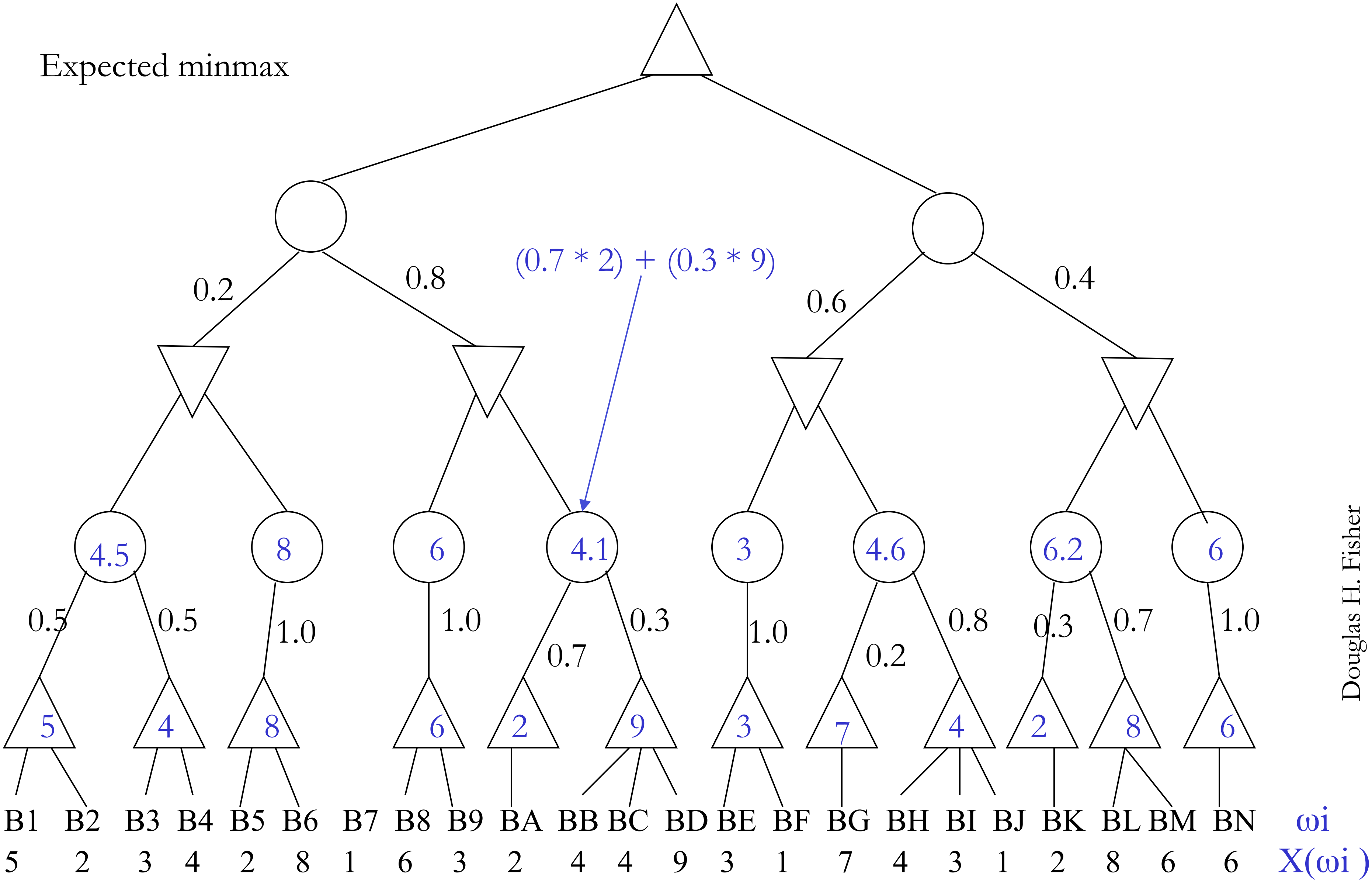
Expected minimax



Douglas H. Fisher

$$EX = \sum_{xi} [xi * P(X(\Omega)=xi)] = \sum_{\omega i} [X(\omega i)*P(\omega i)]$$

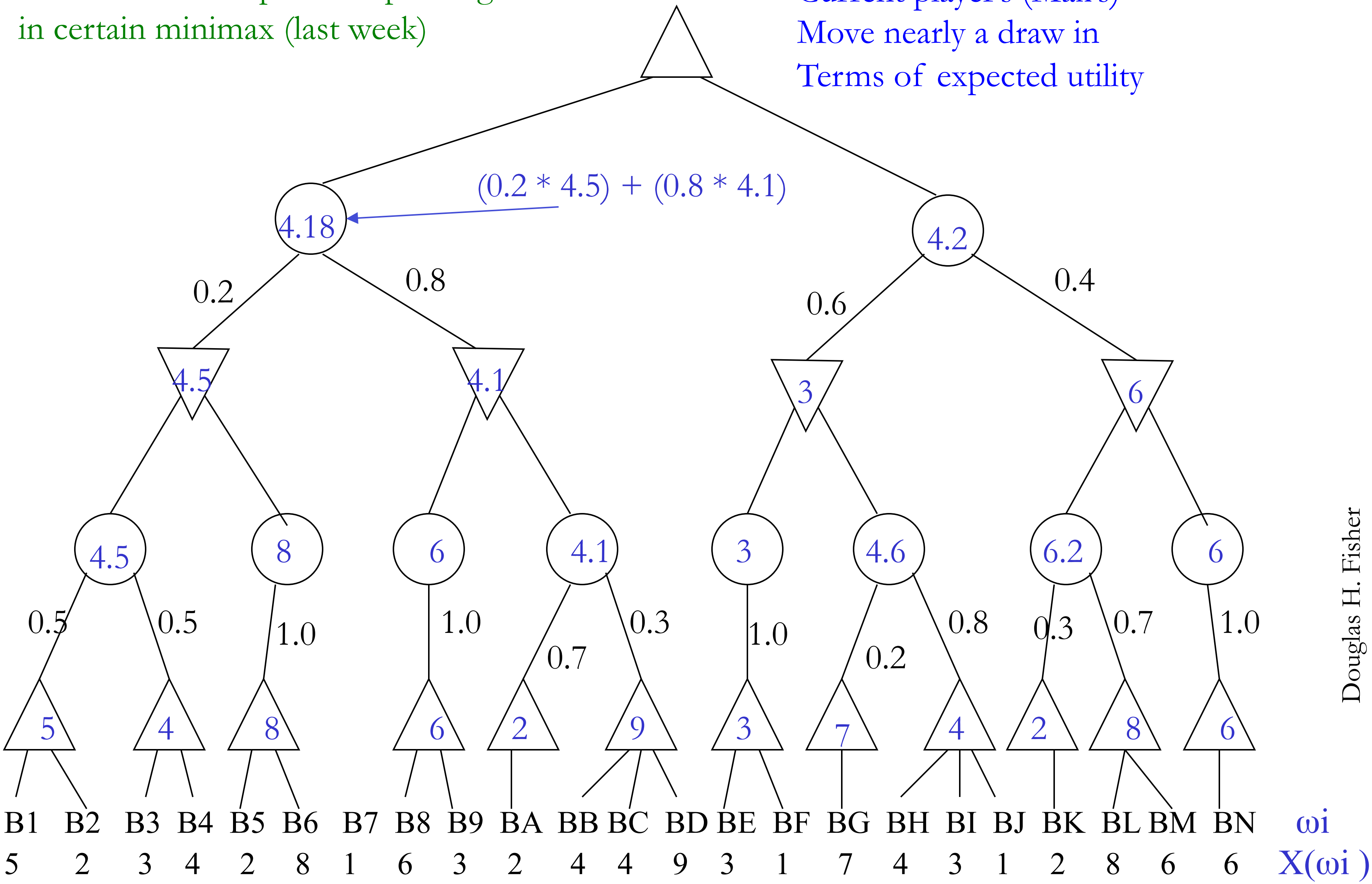
Expected minmax



Douglas H. Fisher

Expected minimax: there is a pruning method akin to alpha beta pruning in certain minimax (last week)

Current player's (Max's) Move nearly a draw in Terms of expected utility



Douglas H. Fisher

Probability (review of basics)

Conditional probability: P(e1 | e2) = P(e1 and e2) / P(e2) where e1 is an event (a draw from an outcome space including the value of a random variable (function), and e2 is a draw from another outcome space or a proceeding/preceding draw from the same outcome space as e1 was drawn from.

e.g., P(flu | sore-throat) = P(flu and sore-throat) / P(sore-throat)

P(battery-dead | car-wont-start) = P(battery-dead and car-wont-start) / P(car-wont-start)

In terms of objective probability assignment

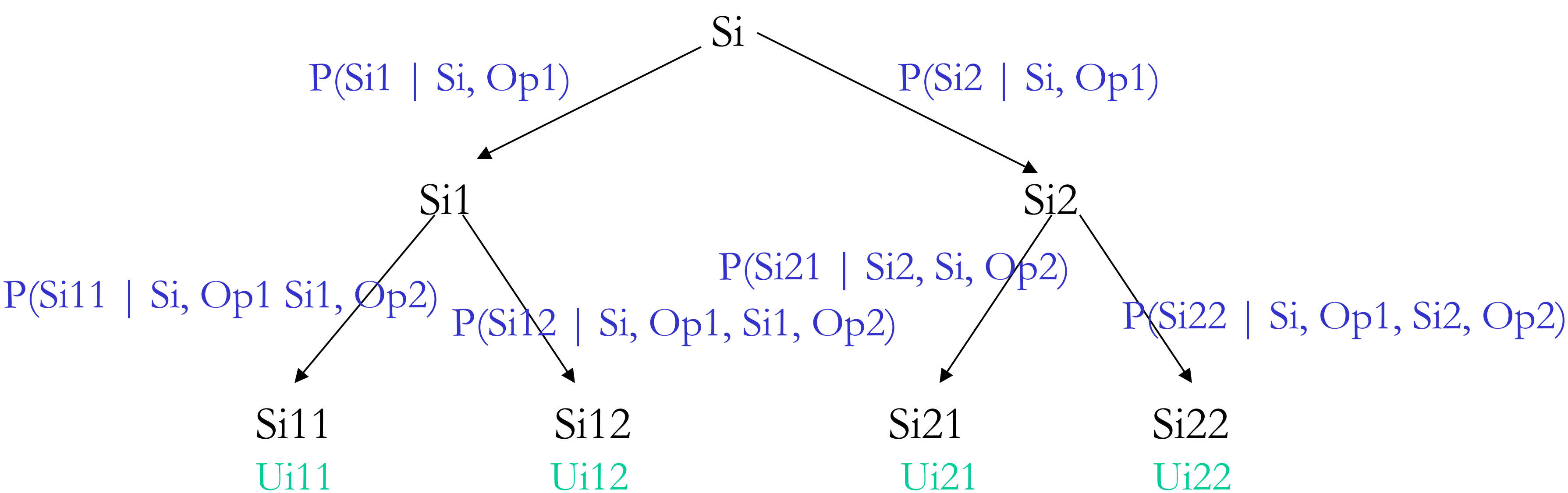
P(e1 | e2) = P(e1 and e2) / P(e2) = ([e1 and e2] / [o]) / ([e2] / [o])
= [e1 and e2] / [e2]

where [o] = [e1 and e2] + [e1 and ~e2] + [~e1 + e2] + [~e1 + ~e2]
= ([e1 and e2] + [e1 and ~e2]) + ([~e1 + e2] + [~e1 + ~e2])
= [e1] + [~e1]
= ([e1 and e2] + [~e1 and e2]) + ([e1 + ~e2] + [~e1 + ~e2])
= [e2] + [~e2]

Douglas H. Fisher

Conditional expectation

$$E(X \mid Y = yj) = \sum_{xi} xi * P(X = xi \mid Y = yj) = \sum_{\omega i} X(\omega i) * P(W = \omega i \mid Y = yj)$$

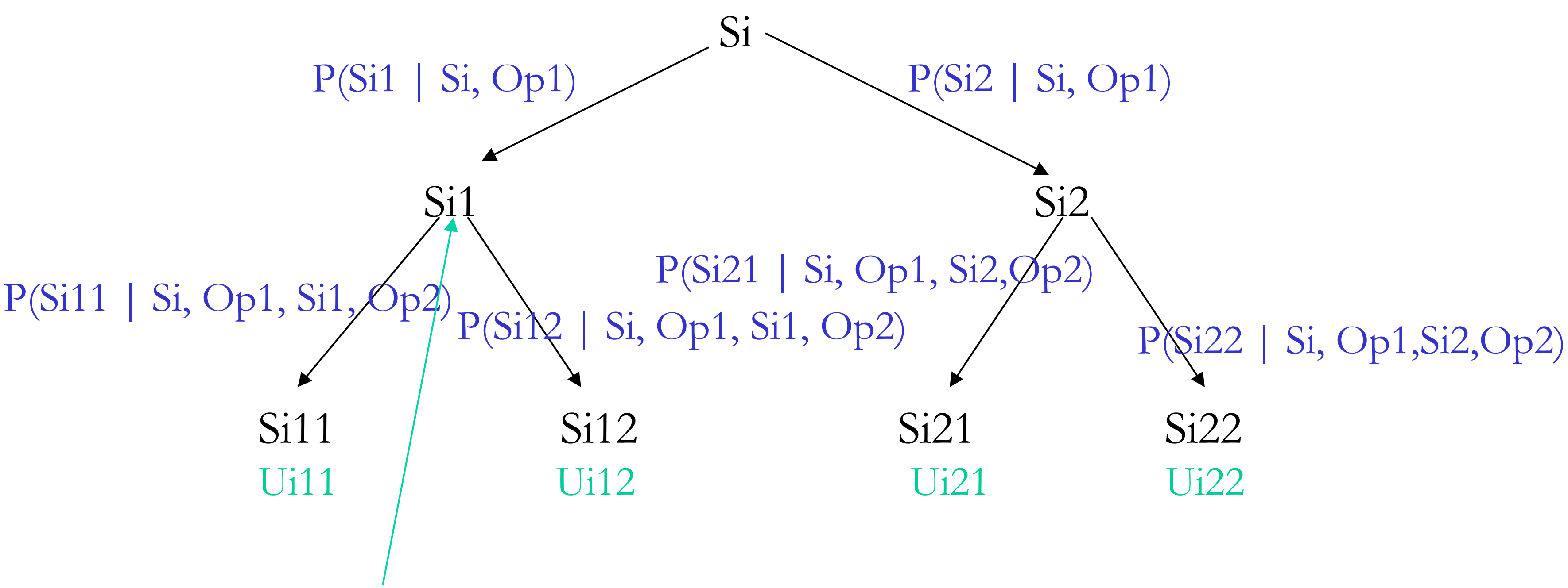


Assume a plan executer in an environment in which operators do not achieve their effects with certainty, but in which some anticipated effects of an operator may not be present after an operator has been applied. Then, with some probability, applying Op1 from state Si will lead to state Si1 and with some probability it will lead to Si2, etc. The U values are utility values of the possible resulting states (e.g., the number of goal conditions satisfied by the state).

Douglas H. Fisher

Conditional expectation cont

$$E(X \mid Y = yj) = \sum_{xi} xi \; P(X = xi \mid Y = yj) = \sum_{\omega i} X(\omega i) \; P(W = \omega i \mid Y = yj)$$



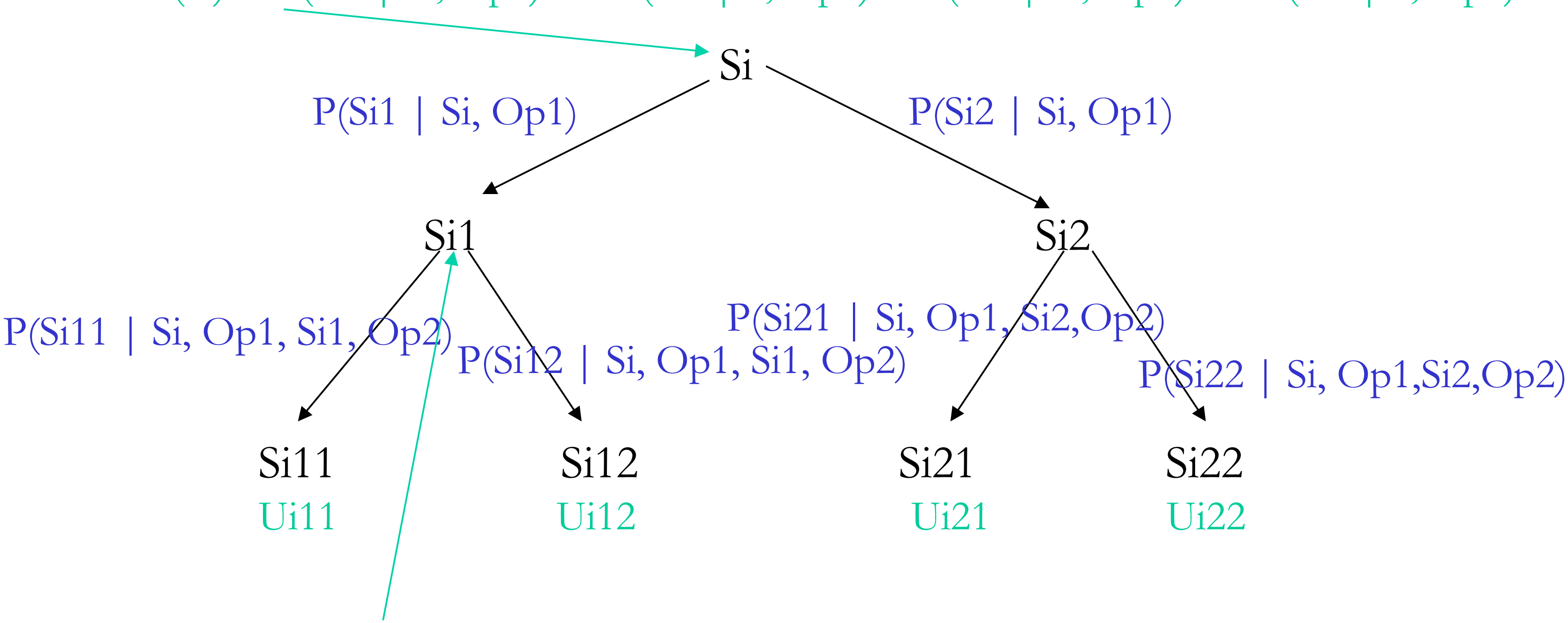
EU(Si1 | Si, Op1) = P(Si11 | Si, Op1, Si1, Op2) * Uill + P(Si12 | Si, Op1, Si1, Op2) * Uil2

Douglas H. Fisher

Conditional expectation

$$E(X \mid Y = yj) = \sum_{xi} xi \; P(X = xi \mid Y = yj) = \sum_{\omega i} X(\omega i) \; P(W = \omega i \mid Y = yj)$$

EU(Si) = P(Si1 | Si, Op1) * EU(Si1 |Si, Op1) + P(Si2 | Si, Op1) * EU(Si2 |Si, Op1)



EU(Si1 | Si, Op1) = P(Si11 | Si, Op1, Si1, Op2) * Uill + P(Si12 | Si, Op1, Si1, Op2) * Uil2

Make sure that you can work out an example like this, and with specific values

Douglas H. Fisher

Probability (review of basics)

Bayes rule:

$P(e1 \mid e2) = P(e1 \text{ and } e2) / P(e2) \rightarrow P(e1|e2)P(e2) = P(e1 \text{ and } e2)$
$P(e2 \mid e1) = P(e1 \text{ and } e2) / P(e1) \rightarrow P(e2|e1)P(e1) = P(e1 \text{ and } e2)$

$\rightarrow$ **$P(e1 \mid e2) = [P(e2 \mid e1) * P(e1)] / P(e2)$**

Consider that a diagnostician may want to estimate **P(Di | Sj)** where Di is a disease, Sj is a symptom. P(Di | Sj) is hard for most experts to accurately estimate, but

$P(Di \mid Sj) = [P(Sj \mid Di) * P(Di)] / P(Sj)$ and P(Sj | Di) and P(Di) is easier for experts to accurately estimate.

P(Sj) is hard to estimate, but it may not be important!

$P(D1 \mid Sj) = [P(Sj \mid D1) * P(D1)] / P(Sj)$
$P(D2 \mid Sj) = [P(Sj \mid D2) * P(D2)] / P(Sj)$
…..
$P(DM \mid Sj) = [P(Sj \mid DM) * P(DM)] / P(Sj)$

which Di (disease) is most probable
note that P(Sj) is constant across choices

Probability (review of basics)

$P(D1 \mid Sj) = [P(Sj \mid D1) * P(D1)] / P(Sj)$
$P(D2 \mid Sj) = [P(Sj \mid D2) * P(D2)] / P(Sj)$
…..
$P(DM \mid Sj) = [P(Sj \mid DM) * P(DM)] / P(Sj)$

→

$P(D1 \mid Sj) \; \alpha \; P(Sj \mid D1) * P(D1)$
$P(D2 \mid Sj) \; \alpha \; P(Sj \mid D2) * P(D2)$
…..
$P(DM \mid Sj) \; \alpha \; P(Sj \mid DM) * P(DM)$

proportional to

answering which Di which is most probable does not require P(Sj)

Recall a similar observation when we discussed the Naïve Bayesian Classifier (Machine Learning)

Probability (review of basics)

Chain rule:

Consider $P(e1 \text{ and } e2) = P(e1 \mid e2)\, P(e2)$

$$P(e1 \text{ and } e2 \text{ and } e3) = P(e1 \mid e2 \text{ and } e3)\, P(e2 \text{ and } e3)$$

$$= P(e1 \mid e2 \text{ and } e3)\, P(e2 \mid e3)\, P(e3)$$

In general:

$P(e1, e2, e3, \ldots, eN)$ read "," as "and"

$$= P(e1 \mid e2, e3, \ldots, eN)\, P(e2 \mid e3, \ldots, eN)\, P(e3 \mid e4, \ldots, eN) \ldots P(e(N-1) \mid eN)\, P(eN)$$

You have seen this before in lecture on Naïve Bayesian learning

Probability (review of basics)

Put the *chain rule and Bayes rule* together:

Bayes rule

$$P(Di \mid S1, S2, \ldots, SN) = (P(S1, S2, \ldots, SN \mid Di) * P(Di)) / P(S1, S2, \ldots SN)$$

Chain rule

$$= P(Di) * \frac{P(S1 \mid Di)}{P(S1)} * \frac{P(S2 \mid Di, S1)}{P(S2 \mid S1)} * \ldots * \frac{P(SN \mid Di, S1, \ldots, S(N-1))}{P(SN \mid S1, \ldots, S(N-1))}$$

$$\alpha \quad P(Di) * P(S1 \mid Di) * P(S2 \mid Di, S1) * \ldots * P(SN \mid Di, S1, \ldots, S(N-1))$$

allows *Bayesian updating (i.e., incremental revision of probability estimate with each new piece of evidence)*

Probability (review of basics)

Where do $P(Sj \mid Di, S1, S2, \ldots, S(j-1))$ come from???

Would need a lot of data for an objective assignment that was accurate.
Experts find it difficult to estimate in a subjective assignment

Independence revisited

Outcome spaces $\Omega 1$ and $\Omega 2$ are independent iff $P(\omega i \text{ and } \omega j) = P(\omega i) * P(\omega j)$ for all ωi in $\Omega 1$ and all ωj in $\Omega 2$

if independent

Alternate definition of independence

$$P(\omega i \text{ and } \omega j) = P(\omega i \mid \omega j)\, P(\omega j) = P(\omega i) * P(\omega j) \leftrightarrow P(\omega i \mid \omega j) = P(\omega i)$$

$$\leftrightarrow P(\omega j \mid \omega i) = P(\omega j)$$

Probability (review of basics)

Conditional independence

Ω1 and Ω2 are conditionally independent given (any known outcome from) Ω3 iff

P(ωi and ωj | ωk) = P(ωi | ωk) * P(ωj | ωk) for all ωi in Ω1, all ωj in Ω2 , and all ωk in Ω3

←→ P(ωi | ωj and ωk) = P(ωi | ωk)   P(ωi and ωj | ωk) = P(ωi | ωj and ωk) P(ωj | ωk)

←→ P(ωj | ωi and ωk) = P(ωj | ωk)

We can also speak of Ω1 and Ω2 as conditionally independent given a particular outcome from Ω3

Probability (review of basics)

Conditional independence cont

If symptoms independent given disease then

$P(Di \mid S1, S2, \ldots SN)$

α

$P(Di) * P(S1 \mid Di) * P(S2 \mid Di, S1) * \ldots * P(SN \mid Di, S1,\ldots,S(N-1))$

if independent $=$

$P(Di) * P(S1 \mid Di) * P(S2 \mid Di) * \ldots * P(SN \mid Di)$

Belief (or Bayesian) Networks

Consider an ordering of variables to factor a joint probability distribution: V1, V2, V3, V4, V5

e.g. $P(v1 \text{ and } v2 \text{ and } \sim v3 \text{ and } v4 \text{ and } \sim v5)$

factorization ordering

$= P(v1) * P(v2 \mid v1) * P(\sim v3 \mid v1,v2) * P(v4 \mid v1,v2,\sim v3) * P(\sim v5 \mid v1,v2,\sim v3,v4)$

Assume the following conditional independencies:

$P(v1)$

V2 independent of V1

$P(v2 \mid v1) = P(v2)$ and $P(v2 \mid \sim v1) = P(v2)$, $P(\sim v2 \mid v1) = P(\sim v2)$, $P(\sim v2 \mid \sim v1) = P(\sim v2)$

$P(\sim v3 \mid v1,v2) = P(\sim v3 \mid v1)$

and $P(\sim v3 \mid v1,\sim v2) = P(\sim v3 \mid v1)$, $P(\sim v3 \mid \sim v1,v2) = P(\sim v3 \mid \sim v1)$, $P(\sim v3 \mid \sim v1,\sim v2) = P(\sim v3 \mid \sim v1)$,

$P(v3 \mid v1,v2) = P(v3 \mid v1)$, $P(v3 \mid v1,\sim v2) = P(v3 \mid v1)$, $P(v3 \mid \sim v1,v2) = P(v3 \mid \sim v1)$,

$P(v3 \mid \sim v1,\sim v2) = P(v3 \mid \sim v1)$ V3 independent of V2 conditioned on V1

$P(v4 \mid v1,v2,\sim v3) = P(v4 \mid v2, \sim v3)$ and ……

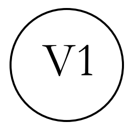
$P(\sim v5 \mid v1,v2,\sim v3,v4) = P(\sim v5 \mid \sim v3)$ and …..

A Belief (or Bayesian) Network is a graphical representation of a joint probability distribution with (conditional) independence relationships made explicit

For a particular factorization ordering (V1, V2, V3, V4, V5), construct a belief network as follows:

V1 a "root"

P(v1), P(~v1)



P(v1) = 0.75
P(~v1) = 0.25 = 1 – P(v1)

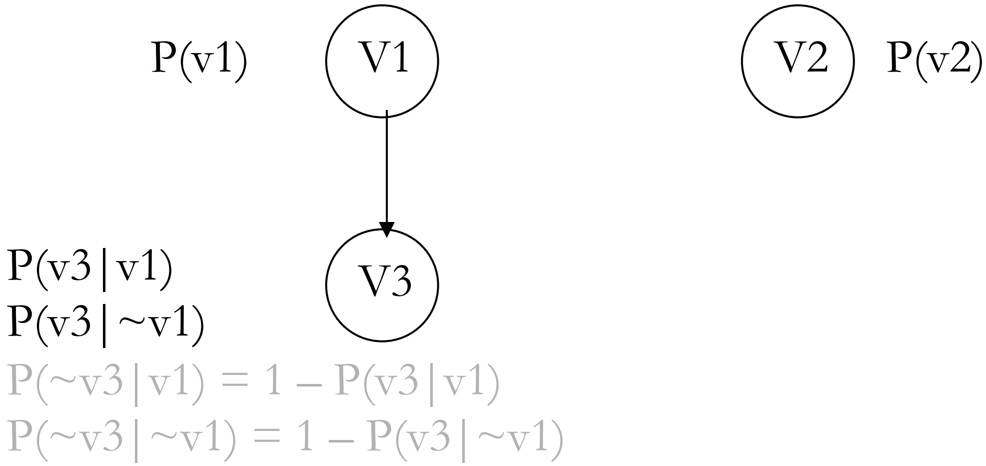
V2 is second variable in ordering. If V2 independent of a subset of its predecessors (possibly the empty set) in ordering conditioned on a disjoint subset of predecessors (including possibly all its predecessors), then the latter subset is its parents, else V2 is a "root"

Since P(v2|v1) = P(v2) ….

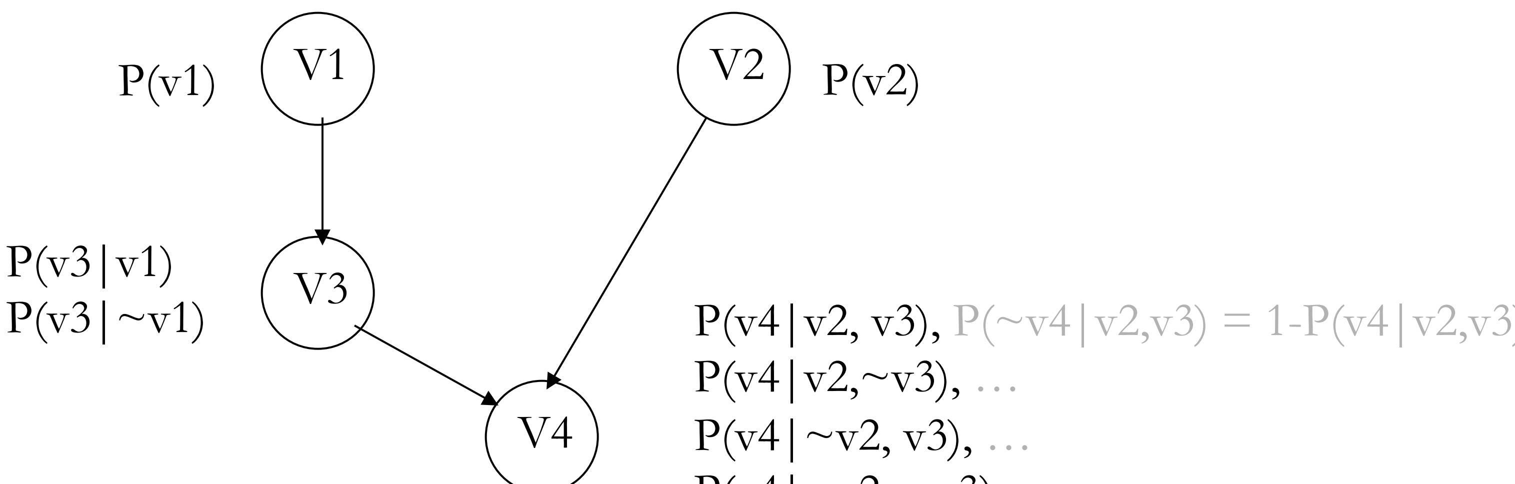
P(v1)

P(v2)

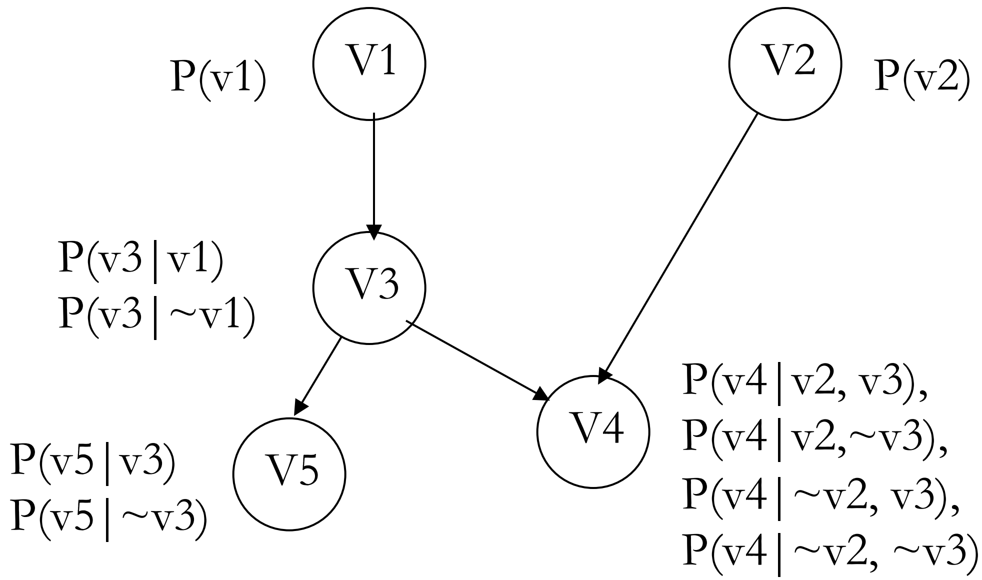
V3 is third variable in ordering. Since P(v3|v1,v2) = P(v3|v1), …:



Since P(v4| v1, v2, v3) = P(v4 | v2, v3), …



Since $P(v5 | v1, v2, v3, v4) = P(v5 | v3), \ldots$:



Components of a belief network: a topology (graph) that qualitatively indicates displays the conditional independencies, and probability tables at each node

Semantics of graphical component: for each variable, V, V is independent of all of its non-descendants conditioned on its parents

More on belief (Bayesian) networks

- constructing BNs and
- inference with BNs

next time