Probabilistic Reasoning and Bayesian Decision Analysis

Introduction

- Reasoning under uncertainty using probability theory
- Dealing with uncertainty is one of the main advantages of an expert system over a simple algorithm in which all the facts must be known to achieve an outcome
- By understanding the advantages and disadvantages of each approach to uncertainty, you will create an expert system that is best for the particular expertise being modeled

Uncertainty

- Uncertainty can be considered as the lack of adequate information to make a decision.
- Uncertainty may prevent us from making the best decision and may even cause a bad decision to be made.
- A number of theories have been devised to deal with uncertainty:
    - classical probability, Bayesian probability, Hartley theory based on classical sets, Shannon theory based on probability, Dempster-Shafer theory, Markov Models, and Zadeh's fuzzy theory.
- Bayesian theory is popular in many diverse areas such as biology, psychology, music, and physics.

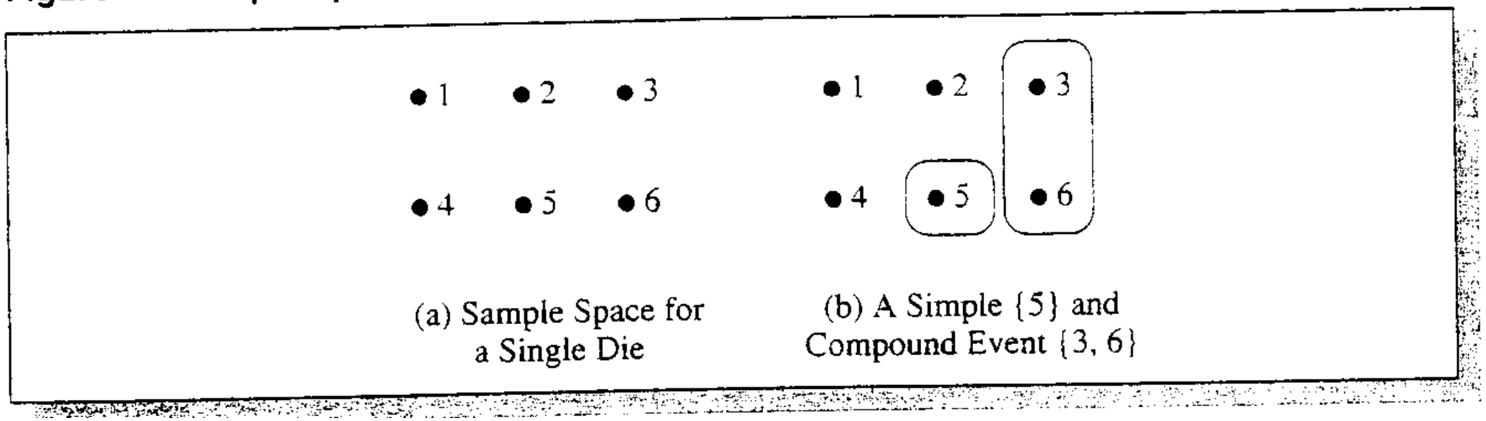
Uncertainty

- The deductive method of reasoning is called **exact reasoning** because it deals with exact facts and the exact conclusions that follow from those facts
  - the conclusion *must* be true
- Many expert systems applications require **inexact reasoning** because the facts or knowledge itself is not known precisely
  - E.g. In medical business or expert systems, there may be uncertain facts, rules, or both
- Classic examples of successful expert systems that dealt with uncertainty are
  - MYCIN for medical diagnosis
  - PROSPECTOR for mineral exploration

Classical probability

- A classical probability considers games such as dice, cards, coins, etc, as ideal systems that do not become worn out
- Sample spaces
  - The result of a trial is a sample point and the set of all possible sample points defines a sample space
  - An event is a subset of the sample space. For example, the event { 1 } occurs if the die comes up with a 1
  - A simple event has only one element and a compound event has more than one

Figure 4.2 Sample Space and Events



Theory of probability

- A formal theory of probability can be made using three axioms:
  - axiom 1: $0 \leq P(E) \leq 1$
    - A certain event is assigned probability 1 and an impossible event is assigned probability 0
  - axiom 2: $\sum_i P(E_i) = 1$
    - This axiom states that the sum of all events that do not affect each other, called mutually exclusive events, is 1. Mutually exclusive events have no sample point in common
    - As a corollary of this axiom:

      $P(E) + P(E') = 1$
    - where E' is the complement of event E
    - This corollary means that the occurrence and nonoccurrence of an event is a mutually exclusive and complete sample space

Theory of probability

- *axiom 3:* $P(E_1 \cup E_2) = P(E_1) + P(E_2)$
  - where E_1 and E_2 are mutually exclusive events.
  - This axiom means that if E_1 and E_2 cannot both occur simultaneously (mutually exclusive events) then the probability of one or the other occurring is the sum of their probabilities
- From these axioms, theorems can be deduced concerning the calculation of probabilities under other situations, such as non-mutually exclusive events, these axioms form the basis for a theory of probability
- These axioms put probability on a sound theoretical basis

Compound probabilities

- The probabilities of compound events can be computed from their sample spaces.
- For example, consider the probability of rolling a die such that the outcome is an even number and a number divisible by three

  A = { 2, 4, 6 }   B = { 3, 6 }

  in the sample space of the die
- the intersection of the sets A and B is:

$$A \cap B = \{ 6 \}$$

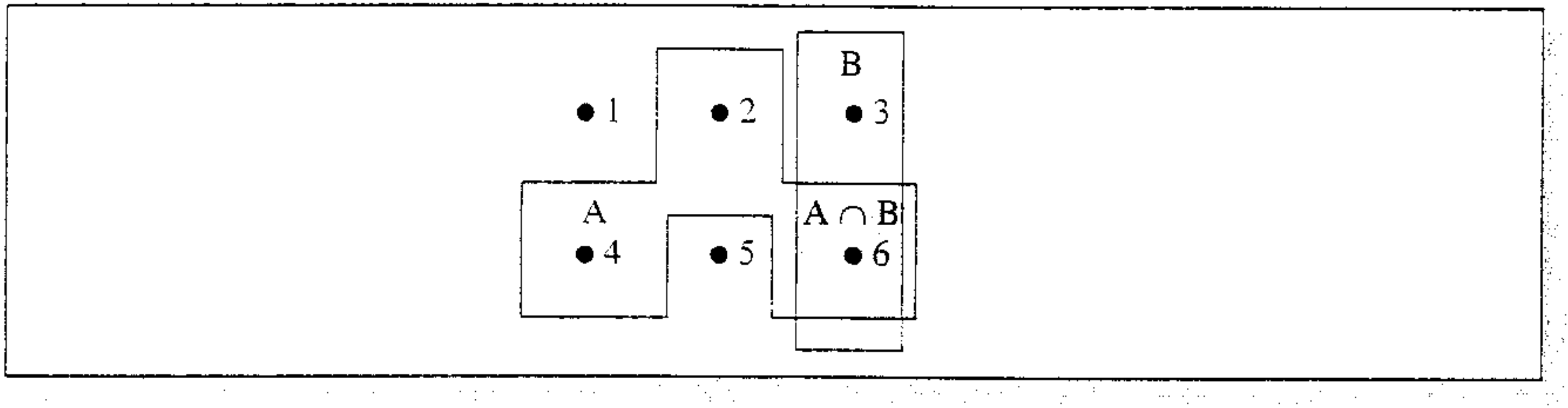
- The compound probability of rolling an even number and a number divisible by three is:

$$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{1}{6}$$

  - where n is the number of elements in the sets and S is the sample space

Compound probabilities

Figure 4.5 Compound Probability of Rolling a Single Die to Give an Even Number and a Number Divisible by Three



Compound probabilities

- Events that do not affect each other in any way are called independent events.
  - For two independent events A and B, the probability is simply the product of the individual probabilities.
  - Events A and B are said to be pairwise independent:

$$P(A \cap B) = P(A)\ P(B)$$

- Two events are called stochastically independent events if and only if the above formula is true
- For three events you might assume that independence is:

$$P(A \cap B \cap C) = P(A)\ P(B)\ P(C)$$

- The formula for the mutual independence of N events requires that 2^N equations be satisfied:

$$P(A^*_1 \cap A^*_2 \ \ldots \cap A^*_N) = P(A^*_1)\ P(A^*_2)\ \ldots\ P(A^*_N)$$

  - where the asterisks mean that every combination of an event and its complement must be satisfied

Compound probabilities

- For three events, the above equation for mutual independence requires that *all* the following equations be satisfied:

$$P(A \cap B \cap C) = P(A)\ P(B)\ P(C)$$
$$P(A \cap B \cap C') = P(A)\ P(B)\ P(C')$$
$$P(A \cap B' \cap C) = P(A)\ P(B')\ P(C)$$
$$P(A \cap B' \cap C') = P(A)\ P(B')\ P(C')$$
$$P(A' \cap B \cap C) = P(A')\ P(B)\ P(C)$$
$$P(A' \cap B \cap C') = P(A')\ P(B)\ P(C')$$
$$P(A' \cap B' \cap C) = P(A')\ P(B')\ P(C)$$
$$P(A' \cap B' \cap C') = P(A')\ P(B')\ P(C')$$

Conditional Probabilities

- Events that are not mutually exclusive influence one another
- Multiplicative Law
  - Conditional Probability
    - The probability of an event A, given that event B occurred:

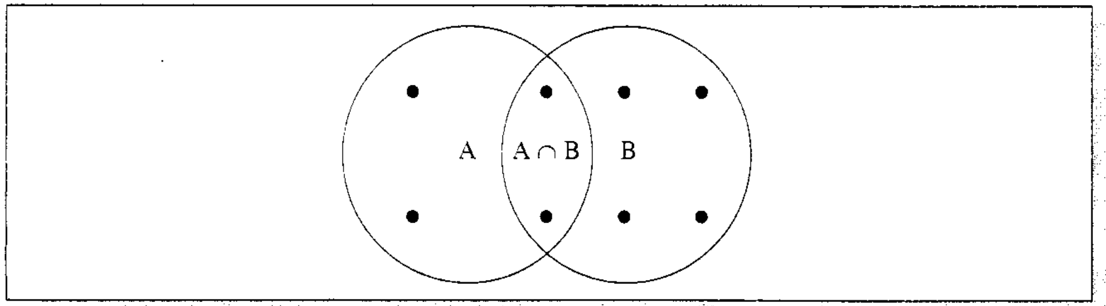
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad \text{for } P(B) \neq 0$$

    - The probability P(B) is the a priori or prior probability
      - it is sometimes called the unconditional probability or an absolute probability.
    - The probabilities can be calculated as the ratios of the number of events, n(A) or n(B), to the total number in the sample space n(S)

$$P(A) = \frac{n(A)}{n(S)} = \frac{4}{8}$$

$$P(B) = \frac{n(B)}{n(S)} = \frac{6}{8}$$

Figure 4.6 Sample Space of Two Intersecting Events



Conditional Probabilities

- If event B has occurred, the reduced sample space is just that of B:
  - n(B)=6
  - And only the events in A that are associated with B are considered:

$$P(A \mid B) = \frac{n(A \cap B)}{n(B)} = \frac{2}{6}$$

  - In terms of probabilities:

$$P(A \mid B) = \frac{\frac{n(A \cap B)}{n(S)}}{\frac{n(B)}{n(S)}}$$

$$= \frac{P(A \cap B)}{P(B)} \quad \text{for } P(B) \neq 0$$

Conditional Probabilities

The **Multiplicative Law** of probability for two events is then defined as:

$$P(A \cap B) = P(A \mid B)\ P(B)$$

which is equivalent to the following:

$$P(A \cap B) = P(B \mid A)\ P(A)$$

The Multiplicative Law for three events is the following:

$$P(A \cap B \cap C) = P(A \mid B \cap C)\ P(B \mid C)\ P(C)$$

and the Generalized Multiplicative Law is:

$$\begin{aligned} P(A_1 \cap A_2 \cap \ldots \cap A_N) = {} & P(A_1 \mid A_2 \cap \ldots \cap A_N) \cdot \\ & P(A_2 \mid A_3 \cap \ldots \cap A_N) \cdot \\ & \ldots\ P(A_{N-1} \mid A_N)\ P(A_N) \end{aligned}$$

Conditional Probabilities

Table 4.6 Hypothetical Probabilities of a Disk Crash Within One Year

	Brand X	Not Brand X′	Total of Rows
Crash C	0.6	0.1	0.7
No Crash C′	0.2	0.1	0.3
Total of Columns	0.8	0.2	1.0

- **Interior probabilities** represent the intersections of events.
- The sum of the rows and columns are displayed as *Totals* and are called **marginal probabilities** because they lie on the margin of the table

Table 4.7 Set Interpretation

	X	X′	Total of Rows
C	$C \cap X$	$C \cap X'$	$C = (C \cap X) \cup (C \cap X')$
C′	$C' \cap X$	$C' \cap X'$	$C' = (C' \cap X) \cup (C' \cap X')$
Total of Columns	$X = (C' \cap X) \cup (C \cap X)$	$X' = (C' \cap X') \cup (C \cap X')$	S (Sample Space)

Conditional Probabilities

Figure 4.7 The Sample Space Interpretation of Two Sets

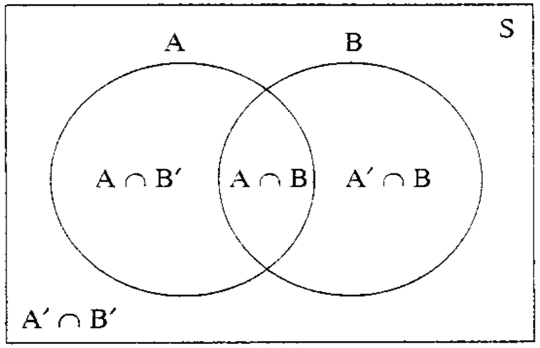


Table 4.8 Probability Interpretation of Two Sets

	X	X′	Total of Rows
C	$P(C \cap X)$	$P(C \cap X')$	$P(C)$
C′	$P(C' \cap X)$	$P(C' \cap X')$	$P(C')$
Total of Columns	$P(X)$	$P(X')$	1.0

Conditional Probabilities

1. The probability of a crash for both Brand X and not Brand X (the sample space) is

   P(C) = 0.7
2. The probability of no crash for the sample space is:

   P(C') = 0.3
3. The probability of using Brand X is:

   P(X) = 0.8
4. The probability of not using Brand X is:

   P(X') = 0.2
5. The probability of a crash and using Brand X is:

   $P(C \cap X) = 0.6$
6. The probability of a crash, given that Brand X is used, is:

$$P(C \mid X) = \frac{P(C \cap X)}{P(X)} = \frac{0.6}{0.8} = 0.75$$

7. The probability of a crash, given that Brand X is not used, is:

$$P(C \mid X') = \frac{P(C \cap X')}{P(X')} = \frac{0.1}{0.2} = 0.50$$

Conditional Probabilities

- The meaning of the intersection, (5), $P(C \cap X)$, is the following:
  - If a disk drive is picked randomly, then 0.6 of the time it will be Brand X and have crashed
- The meaning of the conditional probability (6), $P(C \mid X)$, is very different:
  - If a Brand X drive is picked, then 0.75 of the time it will have crashed
- If any of the following equations are true, then events A and B are **independent:**
  - $P(A \mid B) = P(A)$ or
  - $P(B \mid A) = P(B)$ or
  - $P(A \cap B) = P(A)\, P(B)$

Bayes' Theorem

- The inverse problem is to find the **inverse probability,** which states the probability of an **earlier event** given that a **later one** occurred
  - The solution to this problem is Bayes' Theorem
- Bayesian theory is extensively used today in many applications
- an example: disk drive crashes
  - The inverse question is, suppose you have a drive and don't know its brand, what is the probability that if it crashes, it is Brand X? Non-Brand X?
  - Given that a drive crashed, the probability of it being Brand X can be stated using conditional probability and the results (1), (5):

$$P(X \mid C) = \frac{P(C \cap X)}{P(C)} = \frac{0.6}{0.7} = \frac{6}{7}$$

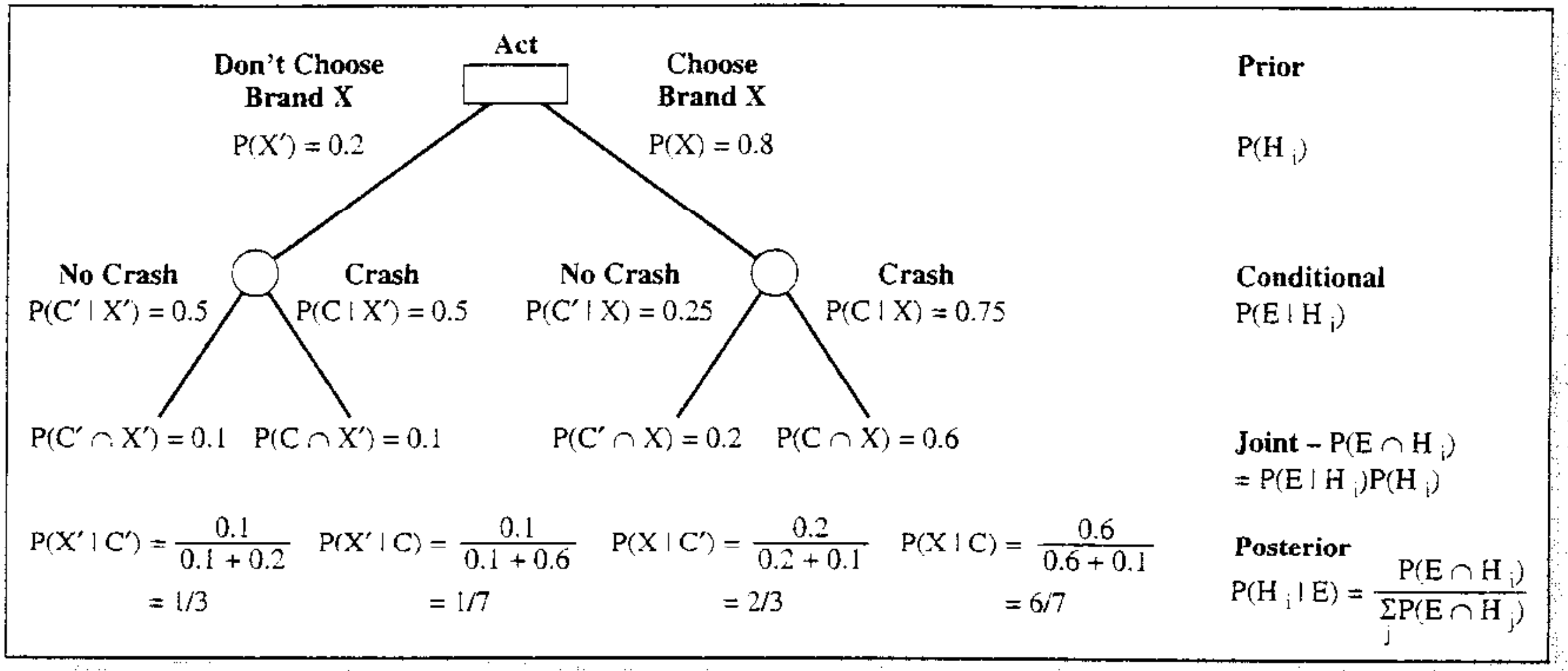
  - Alternatively, using the Multiplicative Law on the numerator, and (1), (3), (6):

$$P(X \mid C) = \frac{P(C \mid X)P(X)}{P(C)} = \frac{(0.75)(0.8)}{0.7} = \frac{0.6}{0.7} = \frac{6}{7}$$

Bayes' Theorem

- The probability P(X | C) is the **inverse** or a posteriori probability, which states that if the drive crashed, it was Brand X

Figure 4.8 Decision Tree for the Disk Crashes



Bayes' Theorem

- The general form of Bayes' Theorem

$$P(H_i \mid E) = \frac{P(E \cap H_i)}{\sum_j P(E \cap H_j)}$$

$$= \frac{P(E \mid H_i)\, P(H_i)}{\sum_j P(E \mid H_j)\, P(H_j)}$$

$$= \frac{P(E \mid H_i)\, P(H_i)}{P(E)}$$

Hypothetical reasoning and backward induction

- Bayes' Theorem is commonly used for decision analysis of business and the social sciences
- Many decision analysis problems can be handled.
  - Decision Analysis Examples
  - Problem Formulation
  - Decision Table and Decision Tree
  - Decision Making with Probabilities
  - Expected Value of Perfect Information
  - Decision Analysis with Sample Information
  - Expected Value of Sample Information

Hypothetical reasoning, backward induction and Decision Analysis

- Managers often must make decisions in environments that are fraught with uncertainty.
- Some Examples
  - A manufacturer introducing a new product into the marketplace
    - What will be the reaction of potential customers?
    - How much should be produced?
    - Should the product be test-marketed?
    - How much advertising is needed?
  - A financial firm investing in securities
    - Which are the market sectors and individual securities with the best prospects?
    - Where is the economy headed?
    - How about interest rates?
    - How should these factors affect the investment decisions?

Hypothetical reasoning, backward induction and Decision Analysis

- Problem Formulation
  - A decision problem is characterized by decision alternatives, states of nature, and resulting payoffs.
  - The decision alternatives are the different possible strategies the decision maker can employ.
  - The states of nature refer to future events, not under the control of the decision maker, which may occur. States of nature should be defined so that they are mutually exclusive and collectively exhaustive.

Hypothetical reasoning, backward induction and Decision Analysis

- Payoff Tables
  - The consequence resulting from a specific combination of a decision alternative and a state of nature is a payoff.
  - A table showing payoffs for all combinations of decision alternatives and states of nature is a payoff table.
  - Payoffs can be expressed in terms of profit, cost, time, distance or any other appropriate measure.

States of nature

Economic Outlook

Decision alternatives

	Maintained (S_1)	Improving (S_2)	Getting worse (S_3)
CLP (d1)	4 Millions	4 Millions	-2 Millions
TRACKER FUND (d2)	0 Million	3 Millions	-1 Million
HSBC (d3)	1 Million	5 Millions	-3 Millions

Payoff (Profit)

Price of Oil

	Increasing (S_1)	Stable / Decreasing (S_2)
Product A (d_1)	HK$20	HK$10
Product B (d_2)	HK$18	HK$17

Payoff (Cost)

Traffic Jam

	No (S_1)	Yes (S_2)
Bus (d_1)	1 hr.	2 hr.
MTR (d_2)	45 min.	45 min.

Payoff (Time)

Hypothetical reasoning, backward induction and Decision Analysis

- Expected Value Approach
  - The decision maker generally will have some information about the relative likelihood of the possible states of nature. These are referred to as the prior probabilities.
  - If probabilistic information regarding he states of nature is available, one may use the expected value (EV) approach.
  - Here the expected return for each decision is calculated by summing the products of the payoff under each state of nature and the probability of the respective state of nature occurring.
  - The decision yielding the best expected return is chosen.

Hypothetical reasoning, backward induction and Decision Analysis

- The <u>expected value of a decision alternative</u> is the sum of weighted payoffs for the decision alternative.
- The expected value (EV) of decision alternative d_i is defined as:

$$EV(d_i) = \sum_{j=1}^{N} P(s_j) V_{ij}$$

where: N = the number of states of nature

$P(s_j)$ = the probability of state of nature s_j

V_{ij} = the payoff corresponding to decision alternative d_i and state of nature s_j

Hypothetical reasoning, backward induction and Decision Analysis

Example: Burger Prince

Burger Prince Restaurant is contemplating opening a new restaurant on Main Street. It has three different models, each with a different seating capacity. Burger Prince estimates that the average number of customers per hour will be 80, 100, or 120. The payoff table for the three models is as follows:

Average Number of Customers Per Hour

	$s_1 = 80$	$s_2 = 100$	$s_3 = 120$
Model A	\$10,000	\$15,000	\$14,000
Model B	\$ 8,000	\$18,000	\$12,000
Model C	\$ 6,000	\$16,000	\$21,000
Probability	0.4	0.2	0.4

Hypothetical reasoning, backward induction and Decision Analysis

- Expected Value Approach

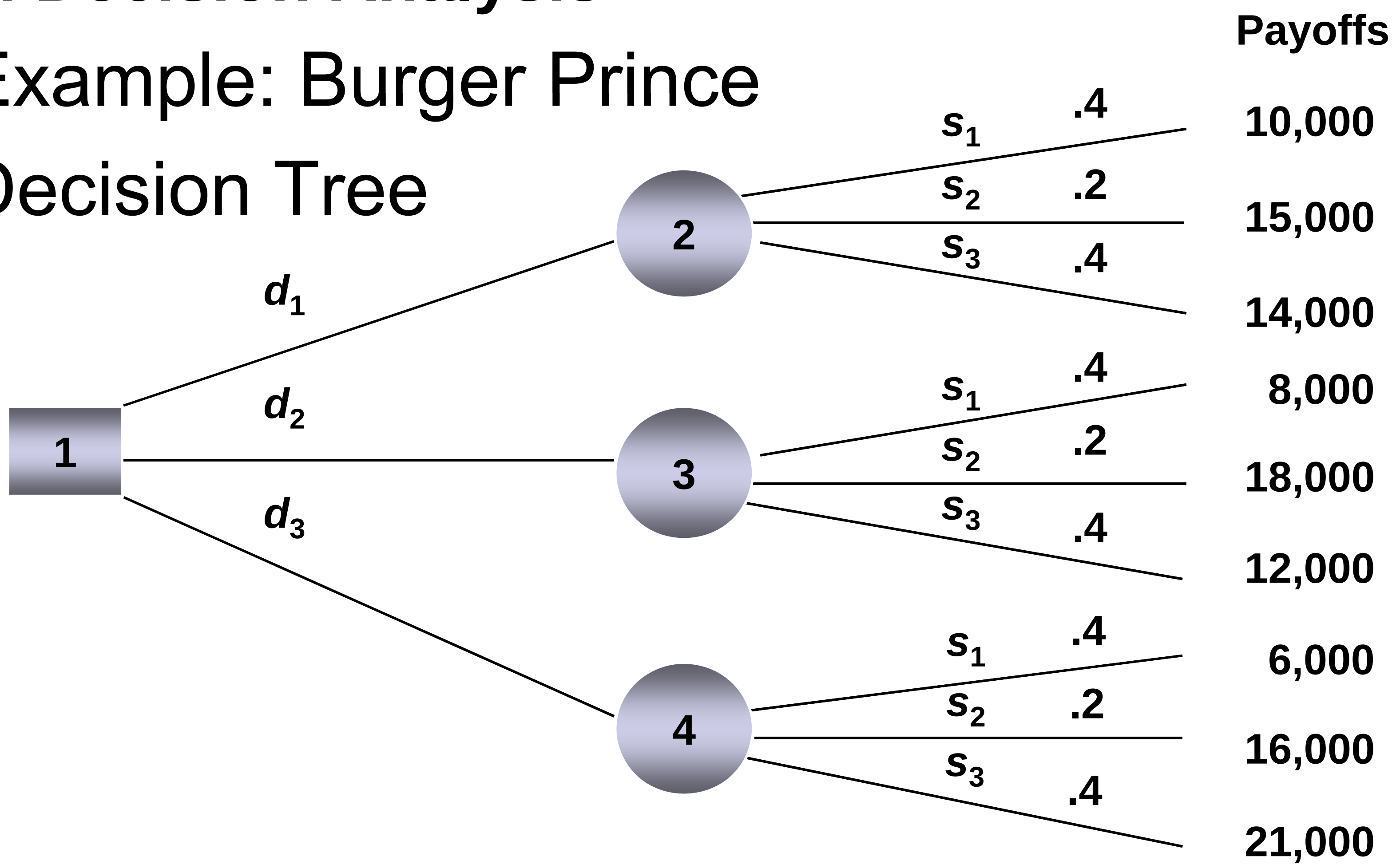
  Calculate the expected value for each decision. The decision tree on the slide 32 can assist in this calculation. Here d_1, d_2, d_3 represent the decision alternatives of models A, B, C, and s_1, s_2, s_3 represent the states of nature of 80, 100, and 120.

Hypothetical reasoning, backward induction and Decision Analysis

- Decision Trees
  - A decision tree is a chronological (sequential) representation of the decision problem.
  - Each decision tree has two types of nodes; round nodes correspond to the states of nature while square nodes correspond to the decision alternatives.
  - The branches leaving each round node represent the different states of nature while the branches leaving each square node represent the different decision alternatives.
  - At the end of each limb of a tree are the payoffs attained from the series of branches making up that limb.

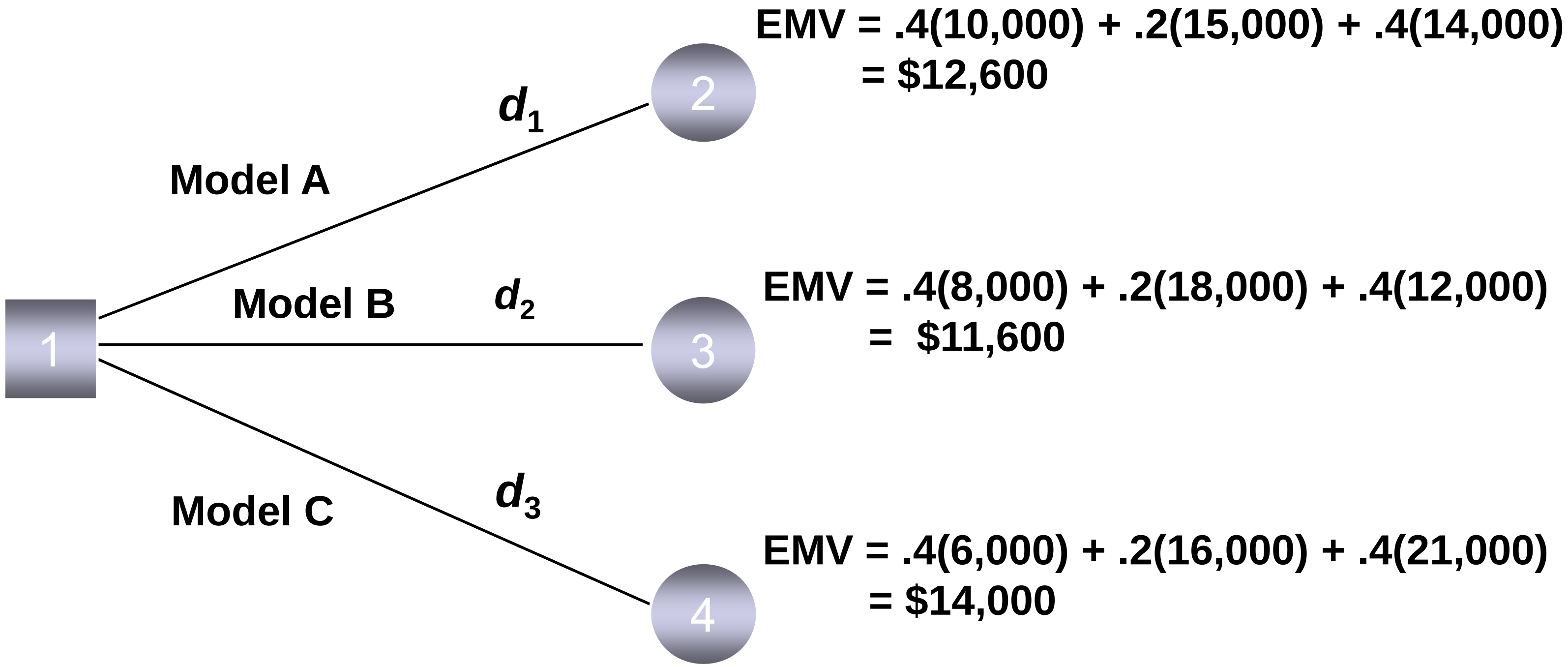
Hypothetical reasoning, backward induction and Decision Analysis

- Example: Burger Prince
- Decision Tree



Hypothetical reasoning, backward induction and Decision Analysis

- **Expected Value For Each Decision**



Choose the model with largest EV, Model C.

Hypothetical reasoning, backward induction and Decision Analysis

- Expected Value of Perfect Information
  - Frequently information is available which can improve the probability estimates for the states of nature.
  - The expected value of perfect information (EVPI) is the increase in the expected profit that would result if one knew with certainty which state of nature would occur.
  - The EVPI provides an upper bound on the expected value of any sample or survey information.

Hypothetical reasoning, backward induction and Decision Analysis

- EVPI Calculation
  - Step 1:

    Determine the optimal return corresponding to each state of nature.
  - Step 2:

    Compute the expected value of these optimal returns.
  - Step 3:

    Subtract the EV of the optimal decision from the amount determined in step (2).

Hypothetical reasoning, backward induction and Decision Analysis

- Example: Burger Prince
- Expected Value of Perfect Information

  Calculate the expected value for the optimum payoff for each state of nature and subtract the EV of the optimal decision.

EVPI= .4(10,000)+.2(18,000)+.4(21,000) - 14,000 = $2,000

The expected value for the optimum payoff for each state of nature
It is also called EPPI (Expected Payoff with perfect information)

EV of the optimal decision
(P.33)

It is also called EV without perfect information

Hypothetical reasoning, backward induction and Decision Analysis

	A	B	C	D	E	F
1	PAYOFF TABLE					
2						
3	**Decision**	**State of Nature**			**Expected**	**Recommended**
4	**Alternative**	s1 = 80	s2 = 100	s3 = 120	**Value**	**Decision**
5	d1 = Model A	10,000	15,000	14,000	12600	
6	d2 = Model B	8,000	18,000	12,000	11600	
7	d3 = Model C	6,000	16,000	21,000	14000	**d3 = Model C**
8	**Probability**	0.4	0.2	0.4		
9		**Maximum Expected Value**			**14000**	
10						
11		**Maximum Payoff**			**EPPI**	**EVPI**
12		10,000	18,000	21,000	**16000**	**2000**

Hypothetical reasoning, backward induction and Decision Analysis

- Knowledge of sample or survey information can be used to revise the probability estimates for the states of nature.
- Prior to obtaining this information, the probability estimates for the states of nature are called <u>prior probabilities</u>.
- With knowledge of <u>conditional probabilities</u> for the outcomes or indicators of the sample or survey information, these prior probabilities can be revised by employing <u>Bayes' Theorem</u>.
- The results of this analysis are called <u>posterior probabilities</u> or <u>branch probabilities</u> for decision trees.

(See the example on Slides 43-44)

Hypothetical reasoning, backward induction and Decision Analysis

- Branch (Posterior) Probabilities Calculation
  - Step 1:

    For each state of nature, multiply the prior probability by its conditional probability for the indicator (outcome)-- this gives the <u>joint probabilities</u> for the states and indicator (outcome).
  - Step 2:

    Sum these joint probabilities over all states -- this gives the <u>marginal probability</u> for the indicator (outcome).
  - Step 3:

    For each state, divide its joint probability by the marginal probability for the indicator (outcome) -- this gives the posterior probability distribution.

(See the example on Slides 43-46)

Hypothetical reasoning, backward induction and Decision Analysis

- The expected value of sample information (EVSI) is the additional expected profit possible through knowledge of the sample or survey information.

Hypothetical reasoning, backward induction and Decision Analysis

- EVSI Calculation
  - Step 1: (See the example on slides 47-49)

    Determine the optimal decision and its expected return for the possible outcomes of the sample or survey using the posterior probabilities for the states of nature.
  - Step 2: (See the example on slide 50)

    Compute the expected value of these optimal returns.
  - Step 3: (See the example on slide 50)

    Subtract the EV of the optimal decision obtained without using the sample information from the amount determined in step (2).

Hypothetical reasoning, backward induction and Decision Analysis

- Efficiency of sample information measure the value of the market research information.
- Is the ratio of EVSI to EVPI.
- As the EVPI provides an upper bound for the EVSI, efficiency is always a number between 0 and 1.

$$\text{Efficency of sample information} = \frac{\text{EVSI}}{\text{EVPI}} * 100\%$$

(See the example on Slides 52)

Hypothetical reasoning, backward induction and Decision Analysis

- Example: Burger Prince

  Burger Prince must decide whether or not to purchase a marketing survey from Stanton Marketing for $1,000. The results of the survey (outcomes) are "favorable" or "unfavorable". The conditional probabilities are:

  P(favorable | 80 customers per hour) = .2

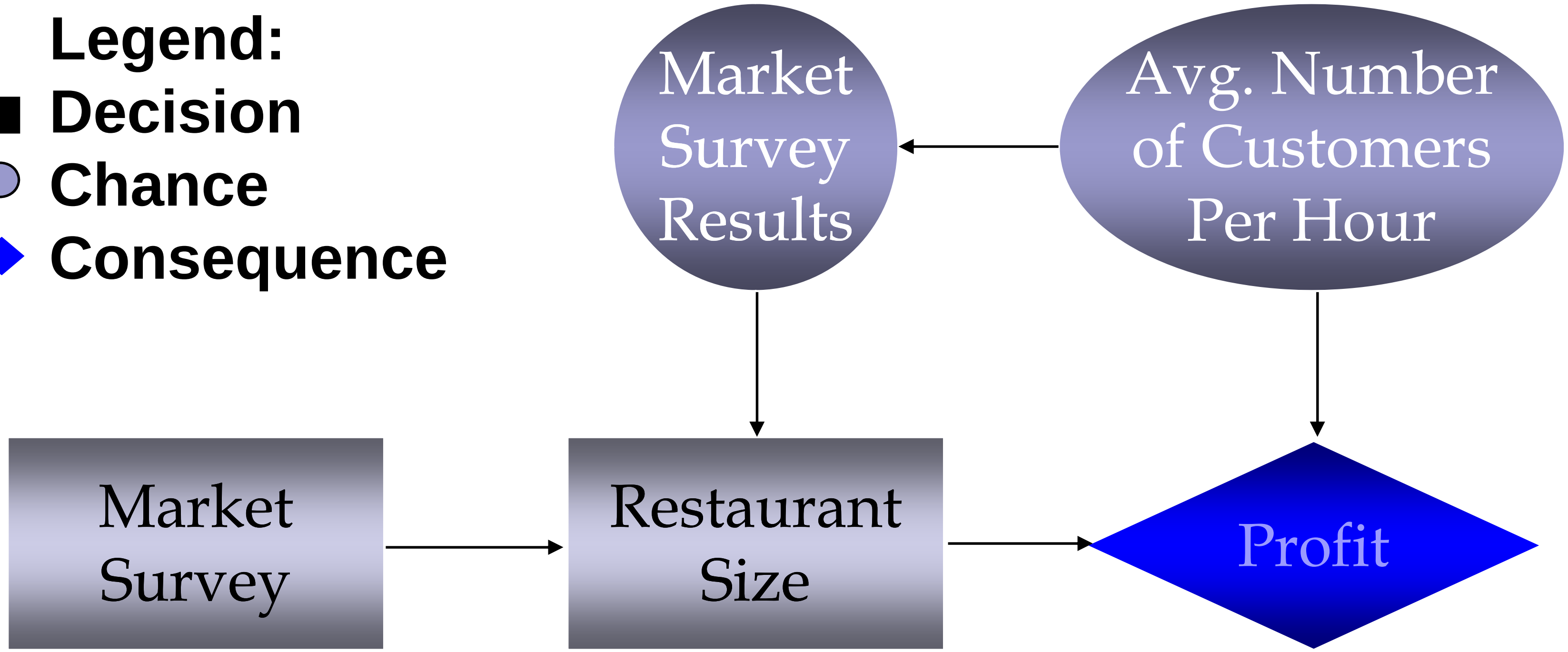
  P(favorable | 100 customers per hour) = .5

  P(favorable | 120 customers per hour) = .9

  Should Burger Prince have the survey performed by Stanton Marketing?

Hypothetical reasoning, backward induction and Decision Analysis

Legend:
- **Decision**
- **Chance**
- **Consequence**



Hypothetical reasoning, backward induction and Decision Analysis

P(80 | Favorable)

- Posterior Probabilities

		Favorable		
State	Prior	Conditional	Joint	Posterior
80	.4	.2	.08	.148
100	.2	.5	.10	.185
120	.4	.9	.36	.667
		Total	.54	1.000

P(favorable) = .54

Hypothetical reasoning, backward induction and Decision Analysis

P(80 | Unfavorable)

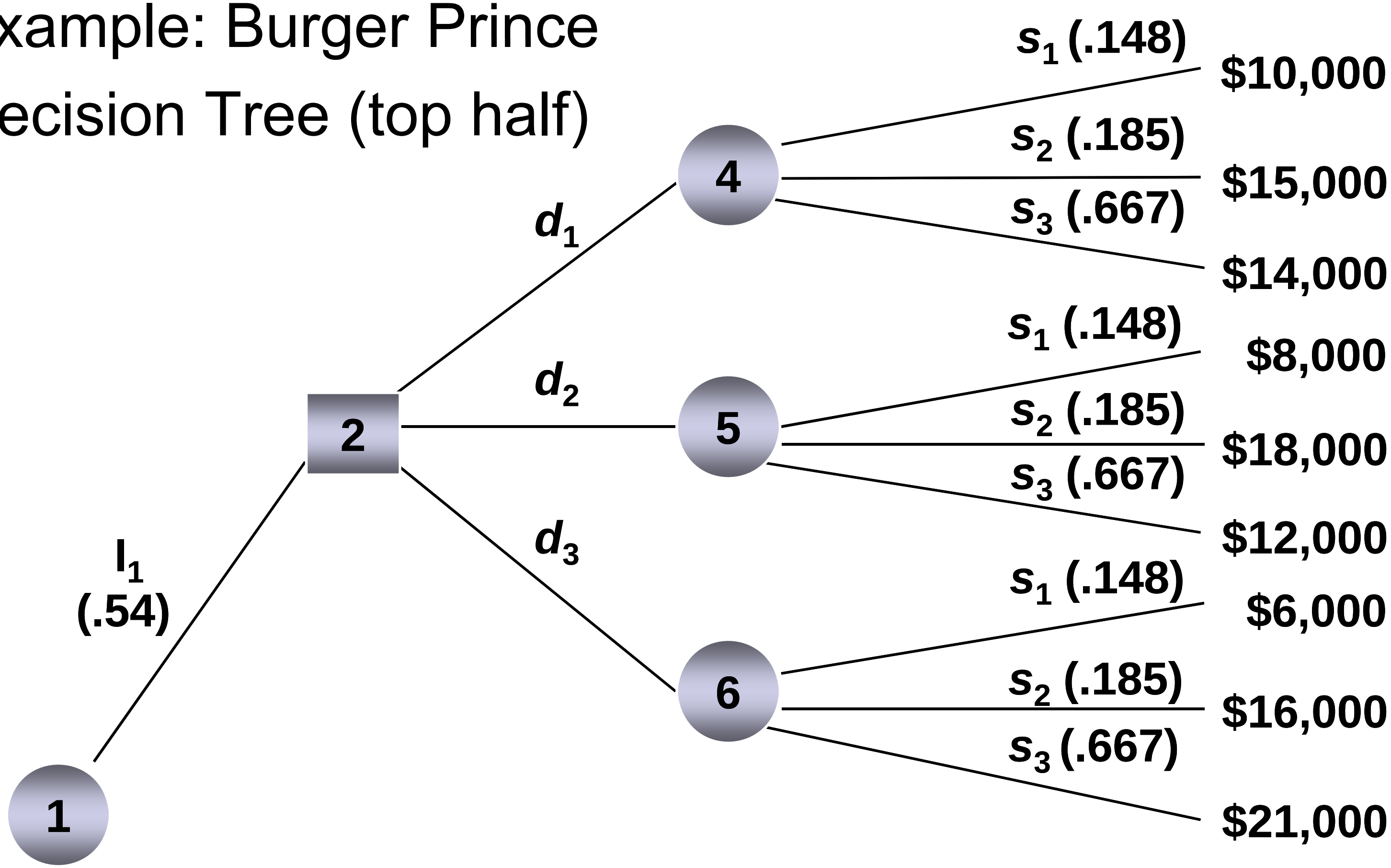
- Posterior Probabilities

State	Prior	Unfavorable Conditional	Joint	Posterior
80	.4	.8	.32	.696
100	.2	.5	.10	.217
120	.4	.1	.04	.087
		Total	.46	1.000

P(unfavorable) = .46

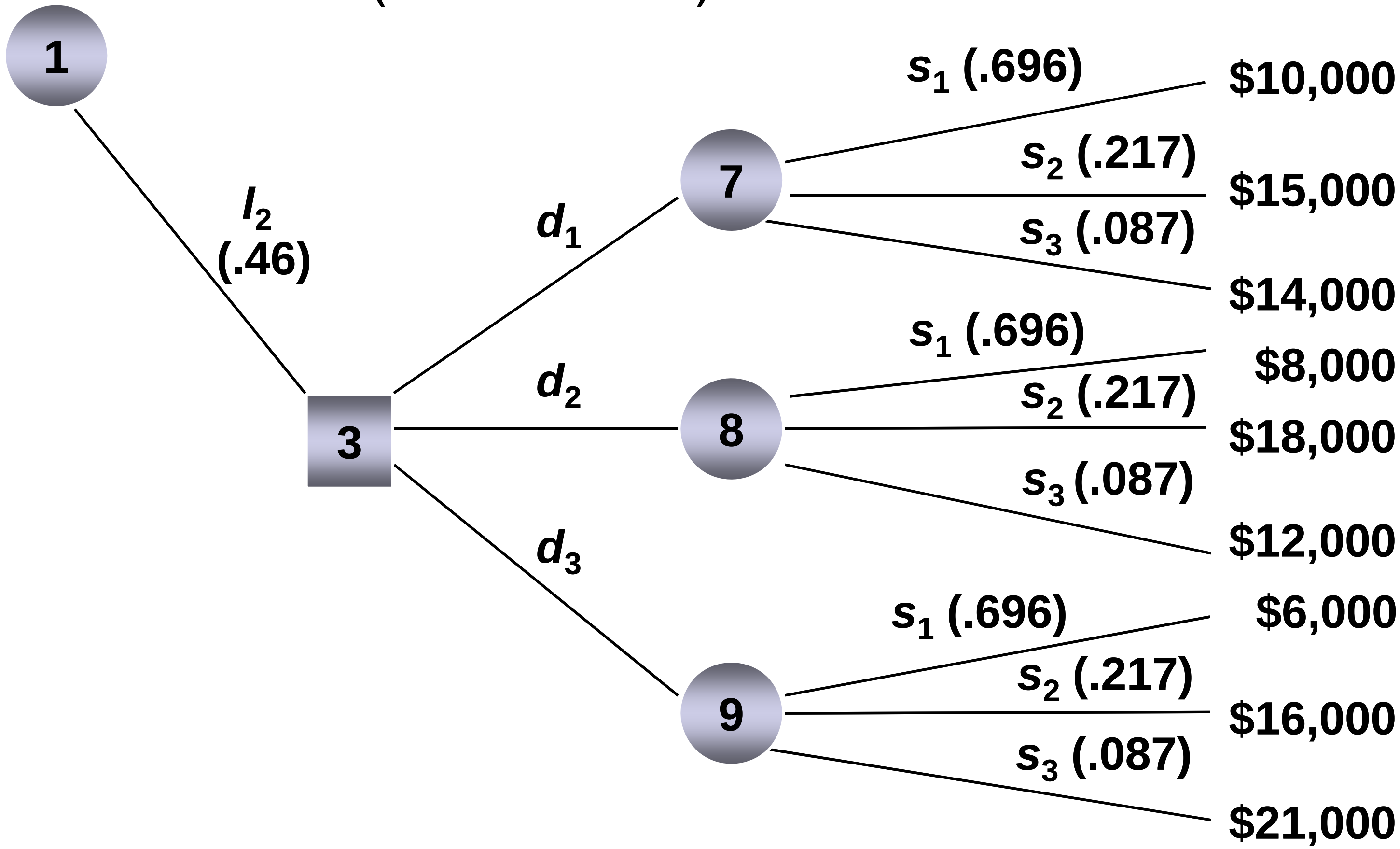
Hypothetical reasoning, backward induction and Decision Analysis

- Example: Burger Prince
- Decision Tree (top half)

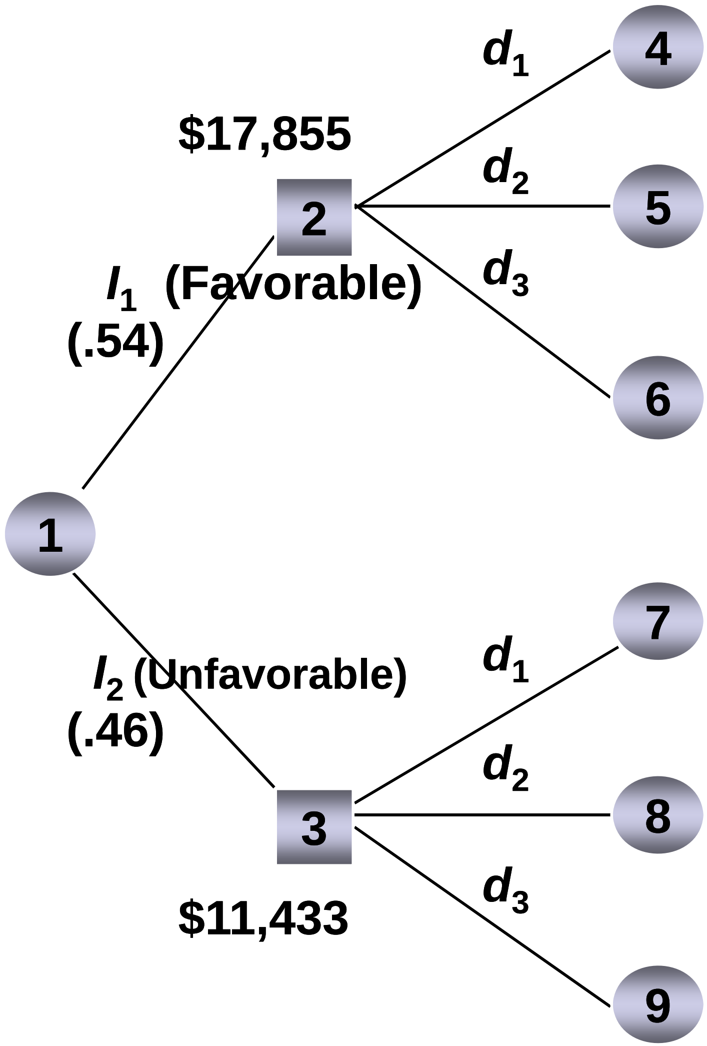


Hypothetical reasoning, backward induction and Decision Analysis

- Decision Tree (bottom half)



Hypothetical reasoning, backward induction and Decision Analysis



EMV = .148(10,000) + .185(15,000) + .667(14,000) = $13,593

EMV = .148 (8,000) + .185(18,000) + .667(12,000) = $12,518

EMV = .148(6,000) + .185(16,000) +.667(21,000) = $17,855 (Model C)

EMV = .696(10,000) + .217(15,000) +.087(14,000) = $11,433 (Model A)

EMV = .696(8,000) + .217(18,000) + .087(12,000) = $10,554

EMV = .696(6,000) + .217(16,000) +.087(21,000) = $9,475

Hypothetical reasoning, backward induction and Decision Analysis

- Expected Value of Sample Information

If the outcome of the survey is "favorable" choose Model C. If it is unfavorable, choose model A.

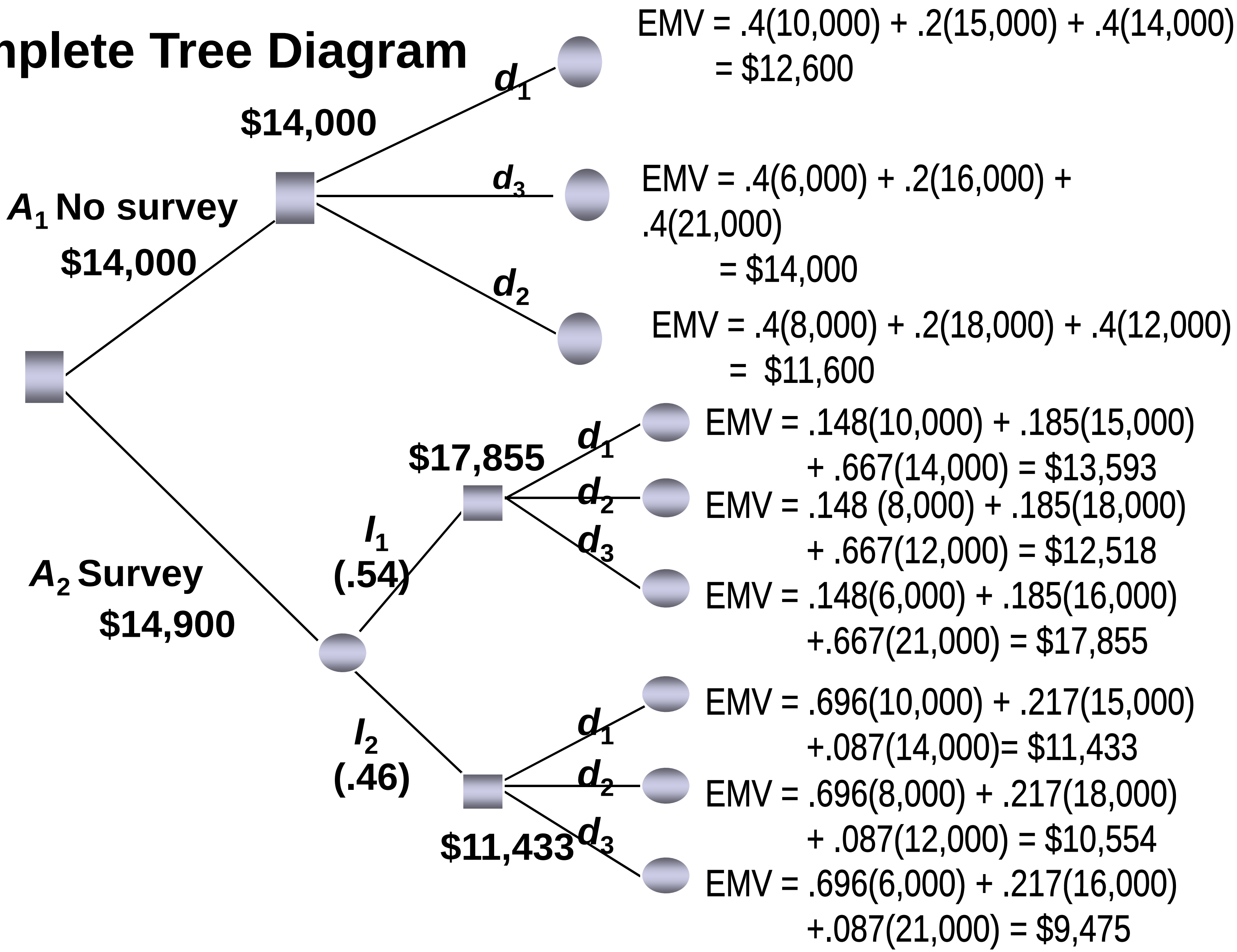
EVSI = .54($17,855) + .46($11,433) - $14,000 = $900.88

The expected value of these optimum returns

The expected value of the optimal decision obtained without using information (P.33)

Since this is less than the cost of the survey, the survey should not be purchased.

A complete Tree Diagram



Hypothetical reasoning, backward induction and Decision Analysis

- Example: Burger Prince
- Efficiency of Sample Information:

  The efficiency of the survey:

  $$\text{EVSI/EVPI} = (\$900.88)/(\$2000) = .4504$$

Hypothetical reasoning, backward induction and Decision Analysis

- Another example of Bayesian decision making under uncertainty,
  - The problem of oil exploration: decide what the chances are of finding oil
    - If there is no evidence either for or against oil, we assign the subjective prior probabilities for oil, O:
      - P(O) = P(O’) = 0.5
    - If we believe that there is better than a 50-50 chance of finding oil, we may set the following:
      - P(O) = 0.6 P(O') = 0. 4

Hypothetical reasoning, backward induction and Decision Analysis

- Suppose the prior probabilities for oil, O, are

  $P(O) = 0.6 \qquad\qquad P(O') = 0.4$

- Assume that the past history of the seismic test has given the following conditional probabilities, where + means a positive outcome and - is a negative outcome

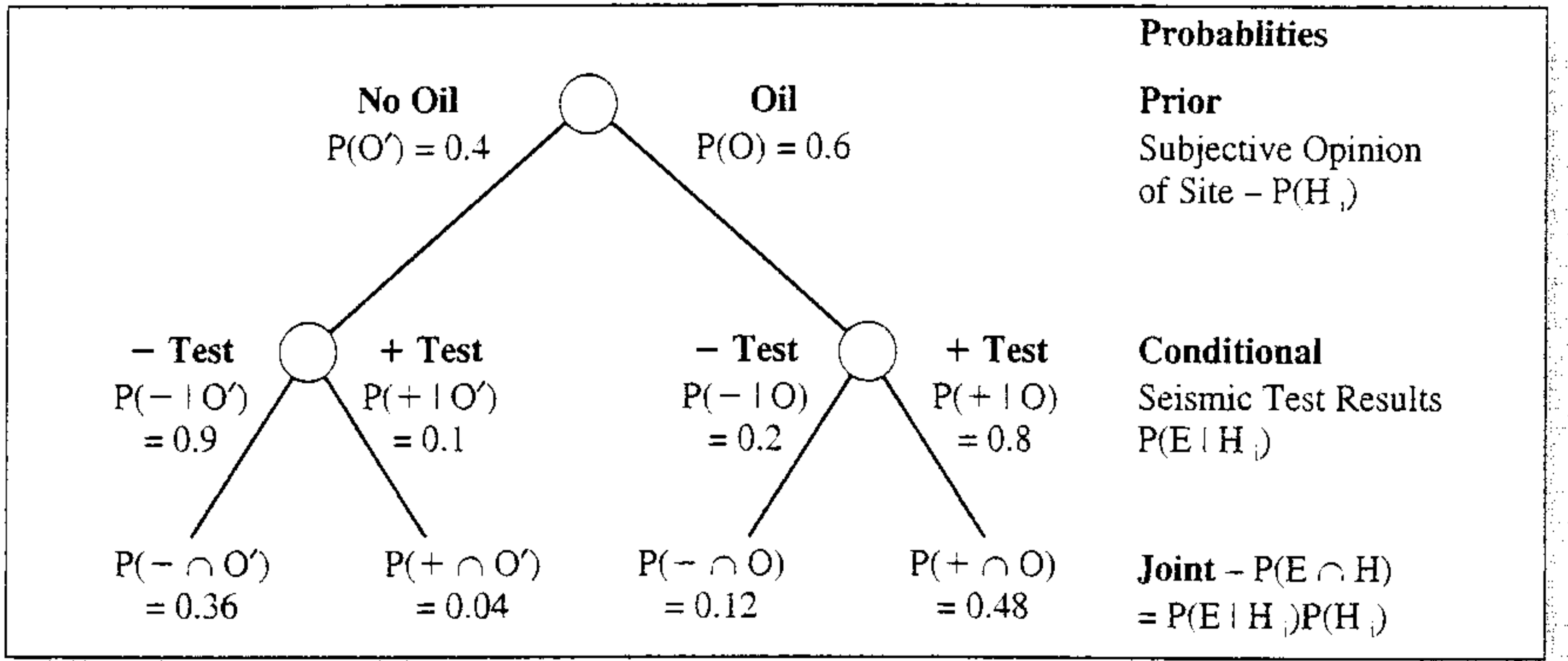
$$P(+ \mid O) = 0.8 \qquad\qquad P(- \mid O) = 0.2 \text{ (false } -\text{)}$$
$$P(+ \mid O') = 0.1 \text{ (false } +\text{)} \qquad P(- \mid O') = 0.9$$

- The Addition Law is then used to calculate the total probability of a + and a - test.

$$P(+) = P(+ \cap O) + P(+ \cap O') = 0.48 + 0.04 = 0.52$$
$$P(-) = P(- \cap O) + P(- \cap O') = 0.12 + 0.36 = 0.48$$

Hypothetical reasoning, backward induction and Decision Analysis

Figure 4.9 Initial Probability Tree for Oil Exploration



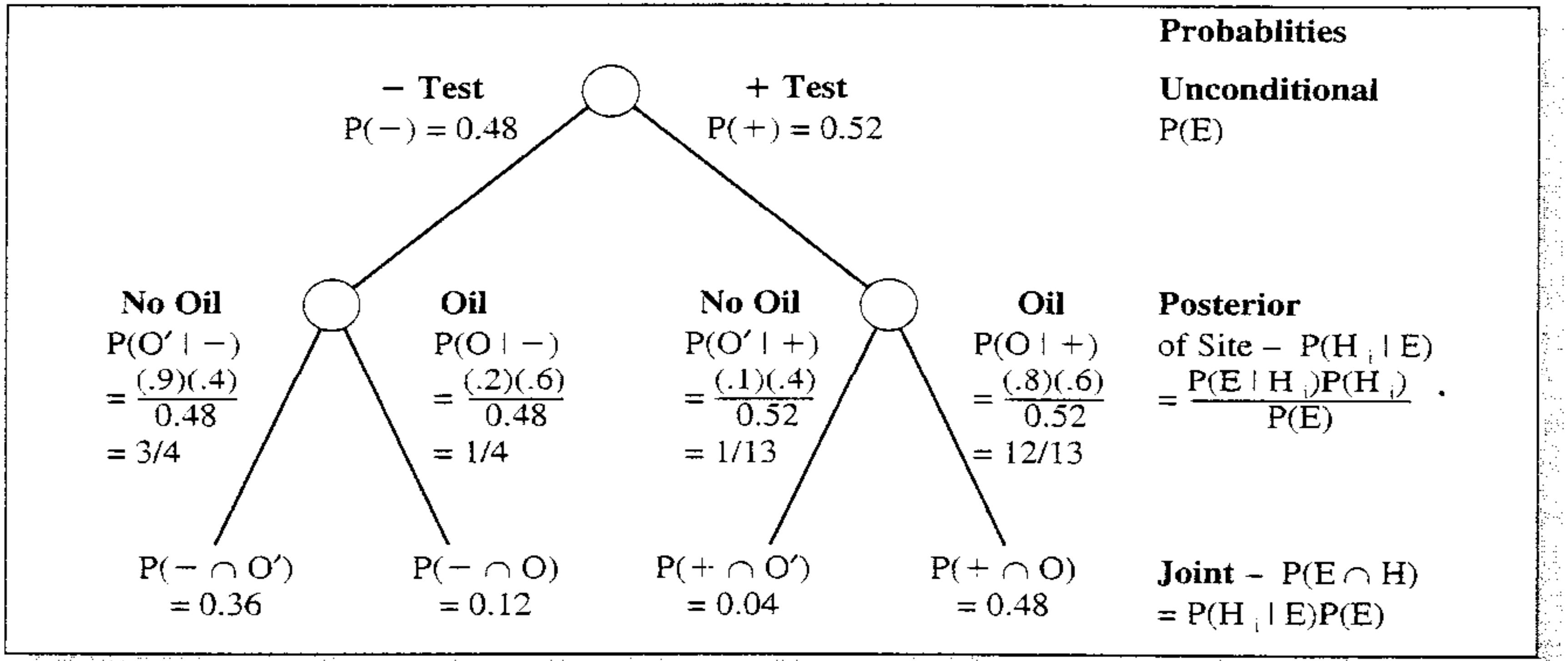
Hypothetical reasoning, backward induction and Decision Analysis

- The P(+) and P(-) are unconditional probabilities that can now be used to calculate the posterior probabilities at the site, as shown in Fig. 4.10.
  - For example, the P(O'| -) is the posterior probability for no oil at the site based on a negative test.
  - The joint probabilities are then computed.
    - The joint probabilities of Fig. 4.10 are the same as in Fig. 4.9.
  - The revision probabilities is necessary to give good results when experimental information, such as the seismic test results, occurs after the initial probability estimates (or guess).
  - Payoff: $1,000,000 (if there is oil)
  - Costs: $200,000 (Drilling expense)

    $50,000 (Seismic survey cost)

Hypothetical reasoning, backward induction and Decision Analysis

Figure 4.10 Revised Probability Tree for Oil Exploration



Hypothetical reasoning, backward induction and Decision Analysis

P(O | +)

- Posterior Probabilities

		Test result +		
State	Prior	Conditional	Joint	Posterior
O(Oil)	0.6	0.8	0.48	0.923 = 12/13
O'(No Oil)	0.4	0.1	0.04	0.077 = 1/13
		Total	0.52	1.000

P(+) = 0.52

Hypothetical reasoning, backward induction and Decision Analysis

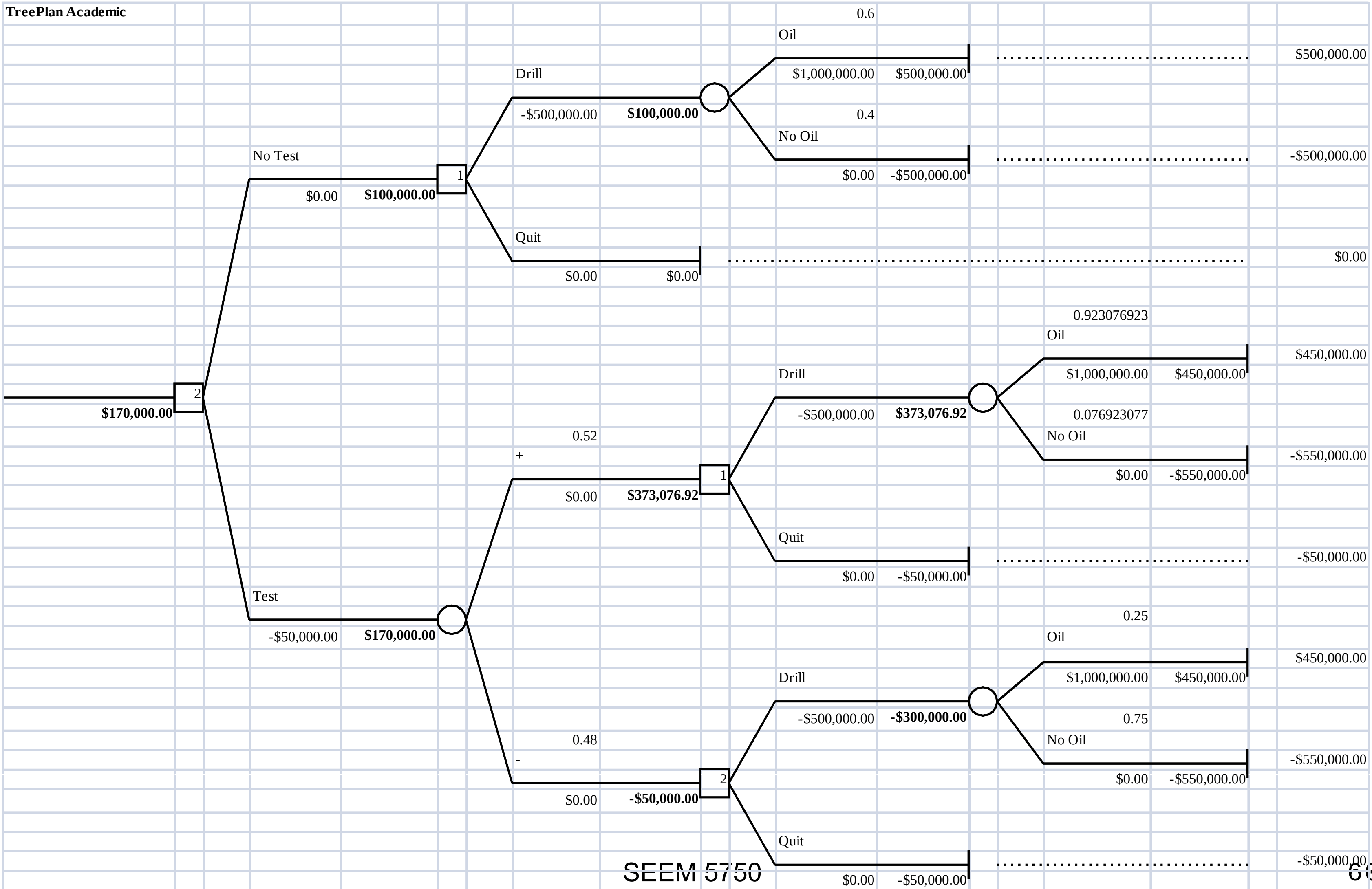
P(O | -)

- Posterior Probabilities

		Test result -		
State	Prior	Conditional	Joint	Posterior
O(Oil)	0.6	0.2	0.12	0.25 = 1/4
O'(No Oil)	0.4	0.9	0.36	0.75 = 3/4
		Total	0.48	1.000

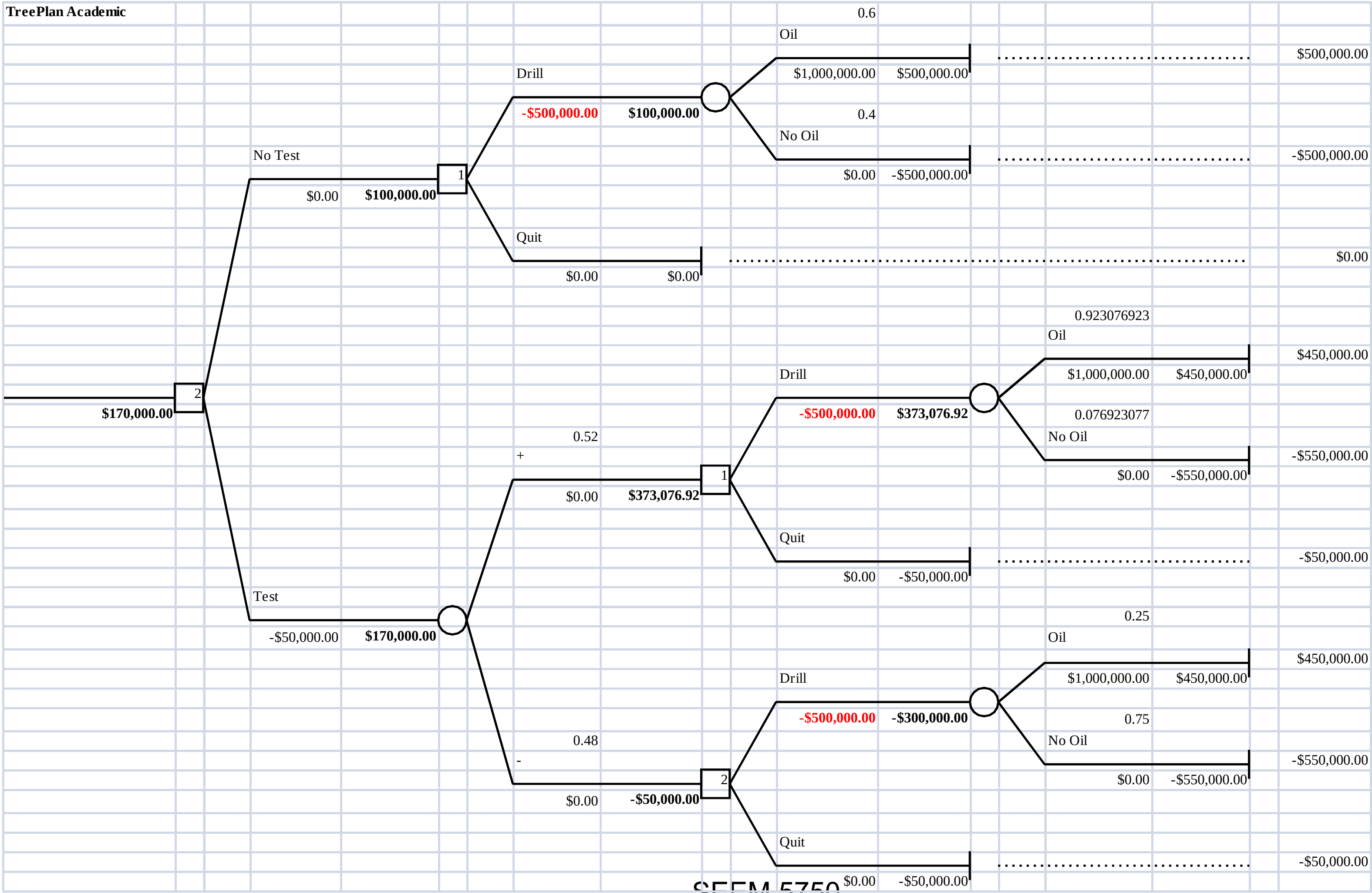
P(-) = 0.48

Hypothetical reasoning, backward induction and Decision Analysis



Hypothetical reasoning, backward induction and Decision Analysis

- $500,000 (Drilling expense)



The odds of belief

- Propositions are statements that are true or false
  - For example, an event may be:
    - "The patient is covered with red spots“
  - And the proposition is:
    - "The patient has measles"
  - Given that A is a proposition, the conditional probability
    - P(A|B) **is not necessarily a probability in the classical sense** if the events and propositions cannot be repeated or have a mathematical basis.
    - P(A|B) can be interpreted as the **degree of belief** that A is true, given B
  - If P(A|B) = 1, then we believe that A is certainly true.
  - If P(A|B) = 0, then we believe A is certainly false

The odds of belief

- Conditional probability
  - is referred to as the likelihood, as in P(H|E), which expresses the likelihood of a hypothesis, H, based on some evidence, E
    - Hypothesis is used for some proposition whose truth or falseness is not known for sure on the basis of some evidence
- Likelihood
  - degree of belief in non-repeatable events
- For example,
  - Suppose you are 95% sure that your car will start the next time.
  - One way of interpreting this likelihood is in terms of the odds of a bet
  - The odds on A against B given some event C is: $\text{odds} = \dfrac{P(A \mid C)}{P(B \mid C)}$
  - If B = A’ $$\text{odds} = \frac{P(A \mid C)}{P(A' \mid C)} = \frac{P(A \mid C)}{1 - P(A \mid C)}$$

The odds of belief

- defining:

  P = P(A|C)
- gives: $$\text{odds} = \frac{P}{1 - P} \text{ and } P = \frac{\text{odds}}{1 + \text{odds}}$$
- In terms of gambling odds, P as wins and 1 - P as losses: $$\text{odds} = \frac{\text{wins}}{\text{losses}}$$
- The likelihood of P = 95% is thus equivalent to:

  $$\text{odds} = \frac{.95}{1 - .95} = 19 \text{ to } 1$$

  that you believe the car will start

The odds of belief

- Probabilities are generally used with deductive problems in which a number of different events, E_i, may occur given the same hypothesis
  - For example, given that a die rolls an even number, there are three possible events:
    - P(2|even)
    - P(4|even)
    - P(6|even)

Sufficiency and necessity

Bayes' Theorem is:

$$(1) \quad P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)}$$

and for the negation of H becomes:

$$(2) \quad P(H' \mid E) = \frac{P(E \mid H')P(H')}{P(E)}$$

Dividing (1) by (2) gives:

$$(3) \quad \frac{P(H \mid E)}{P(H' \mid E)} = \frac{P(E \mid H)P(H)}{P(E \mid H')P(H')}$$

Defining the prior odds on H as:

$$O(H) = \frac{P(H)}{P(H')}$$

Sufficiency and necessity

and the posterior odds as:

$$O(H \mid E) = \frac{P(H \mid E)}{P(H' \mid E)}$$

and finally defining the likelihood ratio:

$$(4) \quad LS = \frac{P(E \mid H)}{P(E \mid H')}$$

then (3) becomes:

$$(5) \quad O(H \mid E) = LS \; O(H)$$

- Equation (5) is known as the odds-likelihood form of Bayes' Theorem

Sufficiency and necessity

- The factor LS is also called the likelihood of sufficiency
  - Equation (5) can be used to solve for LS as follows:

$$(6)\quad LS = \frac{O(H \mid E)}{O(H)} = \frac{\dfrac{P(H \mid E)}{P(H' \mid E)}}{\dfrac{P(H)}{P(H')}}$$

  - Now P(H) /P(H') is some constant, C, and so Equation (6) becomes

$$LS = \frac{\frac{1}{0}}{C} = \infty$$

  - If E is logically sufficient for concluding H, then
    - P(H|E) = 1 and P(H'|E) = 0
  - if LS = ∞ , then the evidence E is logically sufficient for concluding that H is true

Sufficiency and necessity

- Equation (4) also shows in this case that H is sufficient for E

Table 4.10 Relationship Among Likelihood Ratio, Hypothesis, and Evidence

LS	Effect on Hypothesis
0	H is false when E is true or E′ is necessary for concluding H
small ($0 < LS << 1$)	E is unfavorable for concluding H
1	E has no effect on belief of H
large ($1 << LS$)	E is favorable for concluding H
∞	E is logically sufficient for H or Observing E means H must be true

Sufficiency and necessity

- The likelihood of necessity, LN, is defined similarly to LS as:

$$(7)\quad LN = \frac{O(H \mid E')}{O(H)} = \frac{P(E' \mid H)}{P(E' \mid H')} = \frac{\dfrac{P(H \mid E')}{P(H' \mid E')}}{\dfrac{P(H)}{P(H')}}$$

$$(8)\quad O(H \mid E') = LN\ O(H)$$

  - If LN = 0, then P(H|E') = 0
    - This means that H must be false when E' is true
    - If E is not present then H is false, which means that E is necessary for H
- The LS factor shows how much the **posterior odds** are changed when the evidence is present
- The LN factor shows how much the **posterior odds** are changed when the evidence is absent.

Sufficiency and necessity

Table 4.11 Relationship Among Likelihood of Necessity, Hypothesis, and Evidence

LN	Effect on Hypothesis
0	H is false when E is absent or E is necessary for H
small ($0 < LN << 1$)	Absence of E is unfavorable for concluding H
1	Absence of E has no effect on H
large ($1 << LN$)	Absence of E is favorable for H
∞	Absence of E is logically sufficient for H

Sufficiency and necessity

- As an example, in the PROSPECTOR expert system

  IF there are quartz-sulfide veinlets

  THEN there is a favorable alteration for the potassic zone

  - The LS and LN values for this rule are:

    LS = 300

    LN = 0.2

  - which means that observation of quartz-sulfide veinlets is very favorable while not observing the veinlets is mildly unfavorable
  - If LN were < < I, then the absence of quartz-sulfide veinlets would strongly suggest the hypothesis is false
    - An example is the rule:

      IF glassy limonite

      THEN best mineralization favorability

    - with:

      LS = 1000000

      LN = 0.01