will not be followed, as in Figure 4.12. Likewise at action node D, the best choice between a payoff of -\$50,000 and a payoff of –\$500,000 is –\$50,000 and so this is written at D.

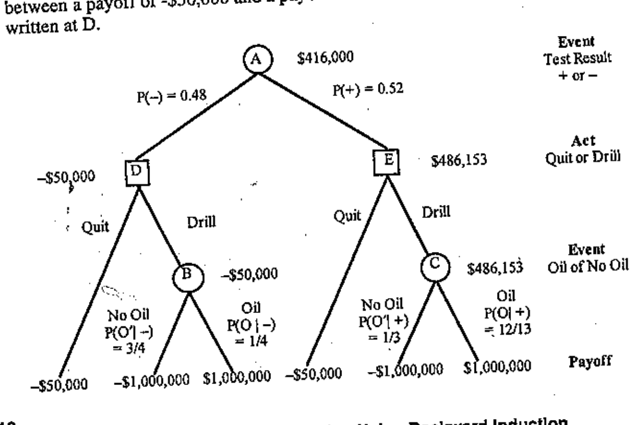


Figure 4.12
Complete Bayesian Decision Tree for Oil Exploration Using Backward Induction

Finally, the expected payoff at the beginning is computed as follows:

```
Expected payoff at node A
$416,000 = ($846,153) (0.52) - ($50,000) (0.48)
```

and this value is written at node A. No path can be pruned at node A because we have already decided to do the seismic test. To decide on a seismic test, the decision tree would have to be extended backward before node A. Notice that in backward induction, we are reasoning back in time so that our actions in the future will be optimal.

The decision tree is an example of hypothetical reasoning or "What if" type situations. By exploring alternate paths of action, we can prune paths that do not lead to optimal payoffs. Some types of expert system tools such as ART, from the Inference Corp., have elaborate mechanisms for hypothetical reasoning and poisoning.

The decision tree in Figure 4.12 shows the optimal strategy for the prospector. If the seismic test is positive, the site should be drilled. If the seismic test is negative, the site should be abandoned. Although this is a very simple example of Bayesian decision making, it does illustrate the type of reasoning involved in dealing with uncertainty. In more complex cases, such as deciding to use a seismic test, the decision trees may grow much larger (Lapin 78).

4.10 TEMPORAL REASONING AND MARKOV CHAINS

Reasoning about events that depend on time is called ***temporal reasoning*** and is something that humans do fairly easily. However, it is difficult to formalize **temporal events** so that a computer can make temporal inferences. Yet expert systems that reason about temporal events such as aircraft traffic control could be

very useful. Expert systems that reason over time have been developed in medicine. These include the VM system (Fagan 1979) for ventilator management of patients on respirators, to help them breathe. Other systems are CASNET for eye glaucoma treatment (Weiss 78) and the digitalis therapy advisor for heart patients (Gorry 78).

Except for VM, the other medical systems mentioned above have a less difficult problem with temporal reasoning compared to an aircraft control system, which must operate in **real time**. Most expert systems cannot operate in real time because of the inference engine design and the large amounts of processing required. An expert system that does a lot of temporal reasoning to explore multiple hypotheses in real time is very difficult to build. Different temporal logics have been developed based on different axioms (Turner 84). Different theories are based on the ways certain questions are answered. Does time have a first and last moment? Is time continuous or discrete? Is there only one past but many possible futures?

Depending on how these questions are answered, many different logics can be developed (McDermott 82; Allen 81). Temporal logic is also useful in conventional programs, such as in the synthesis and the synchronization of processes in concurrent programs (Manna and Wolper 84).

Another approach to temporal reasoning is with probabilities. We can think of a system moving from one state to another as evolving over time. The system can be anything that is probabilistic such as stocks, voters, weather, business, disease, equipment, genetics, and so forth. The system's progression through a sequence of status is called a ***stochastic process*** if it is probabilistic.

It is convenient to represent a stochastic process in the form of a transition matrix. For the simple case of two states, S_1 and S_2, the transition matrix is

$$
\begin{array}{ccc}
 & & \text{Future} \\
 & & \begin{array}{cc} S_1 & S_2 \end{array} \\
\text{Present} & \begin{array}{c} S_1 \\ S_2 \end{array} & \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}
\end{array}
$$

where P_{mn} is the probability of a transition from state m to state n.

As an example, assume that 10 percent of all people who now use Brand X drives will buy another Brand X drive when needed. Also, 60 percent of the people who don't use Brand X now will buy Brand X when they need a new drive (the only good thing about Brand X is its advertising). Over a long period of time, how many people will use Brand X?

The transition matrix, T, is

$$
\begin{array}{ccc}
 & & \begin{array}{cc} X & X' \end{array} \\
T = & \begin{array}{c} X \\ X' \end{array} & \begin{bmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix}
\end{array}
$$

where the sum of each row must add up to 1. A vector whose components are not negative and add up to 1 is called a ***probability vector***. Each row of T is a probability vector. One way of interpreting the transition matrix is in terms of a state diagram, as shown in Figure 4.13. Notice that there is a 0.1 probability of remaining in state X, a 0.4 probability of remaining in state X′, a 0.9 probability of going from X to X′, and a 0.6 probability of going from X′ to X.

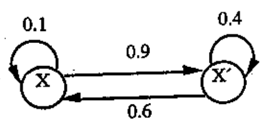
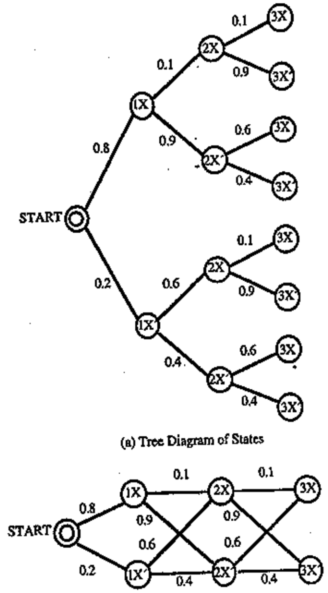


Figure 4.13 State Diagram Interpretation of a Transition Matrix

Suppose initially that 80 percent of the people use Brand X. Figure 4.14 (a) shows a probability tree for a few state transitions, where the states are labeled by state number and drive owned. Notice how the tree is starting to grow. If there were ten transitions, the tree would have $2^{10} = 1024$ branches. An alternate way of drawing the tree, as a lattice, is shown in Figure 4.14 (b). The advantage of the lattice representation is that it doesn't need as many links connecting the states.



(a) Tree Diagram of States

(b) Lattice Diagram of States

Figure 4.14 Tree and Lattice Diagrams of States Evolving over Time

The probability of the system being in a certain state can be represented as a row matrix called the ***state matrix***.

$$S = [\ P_1 \quad P_2 \quad \ldots \quad P_N\ ]$$

where $P_1 + P_2 + \ldots + P_N = 1$

Initially, with 80 percent of the people owning Brand X, the state matrix is

$$S_1 = [\ 0.8 \quad 0.2\ ]$$

As time goes on, these numbers will change depending on which drives people buy.

In order to calculate the number of people in state 2 having Brand X and not Brand X, just multiply the state matrix times the transition matrix using the ordinary laws of matrix multiplication:

$$S_2 = S_1\ T$$

which gives

$$S_2 = [\ 0.8 \quad 0.2\ ] \begin{bmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix}$$

$$= [\ (0.8)(0.1)+(0.2)(0.6) \quad (0.8)(0.9)+(0.2)(0.4)\ ]$$

$$= [\ 0.2 \quad 0.8\ ]$$

Multiplying this second state by the transition matrix gives

$$S_3 = S_2\ T$$

$$= [\ 0.2 \quad 0.8\ ] \begin{bmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix}$$

$$S_3 = [\ 0.5 \quad 0.5\ ]$$

Multiplying this third state by the transition matrix gives

$$S_4 = S_3\ T$$

$$= [\ 0.5 \quad 0.5\ ] \begin{bmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix}$$

$$S_4 = [\ 0.35 \quad 0.65\ ]$$

The next states are

$$S_5 = [\ 0.425 \quad 0.575\ ]$$

$$S_6 = [\ 0.3875 \quad 0.6125\ ]$$

$$S_7 = [\ 0.40625 \quad 0.59375\ ]$$

$$S_8 = [\ 0.396875 \quad 0.602125\ ]$$

Notice that the states are converging on

$$[\ 0.4 \quad 0.6\ ]$$

which is called a ***steady-state matrix***. The system is said to be in equilibrium when it is in the steady state because it does not change. It's interesting that the steady-state values do not depend on the initial state. If any initial probability vector had been used, the steady-state values would be the same.

A probability vector S is a steady-state matrix for the transition matrix T if

$$(1)\quad S = S\ T$$

If T is a **regular transition matrix**, which has some power with only positive elements, then a unique steady-state S exists (Kemeny 59). The fact that the transition matrix elements are positive means that, at some time, it is possible to be in any state no matter what the initial state had been. That is, every state is potentially accessible.

A **Markov chain process** is defined as having the following characteristics:

(1) A finite number of possible states.
(2) The process can be in one and only one state at a time.
(3) The process moves or **steps** successively from one state to another over time.
(4) The probability of a move depends only on the immediately preceding state.

For example, given a finite set of states {A,B,C,D,E,F,G,H}, then if the next state that the process goes to after H is I, the conditional probability is the following.

$$P(I \mid H) = P(I \mid H \cap G \cap F \cap E \cap D \cap C \cap B \cap A)$$

Notice how the lattice diagram of Figure 4.14 (b) resembles a chain.

The disk drive case is a Markov chain process and the steady-state matrix can be found by applying equation (1). Assume some arbitrary vector S with components X and Y and apply equation (1) as follows:

$$[\ X \quad Y\ ] \begin{bmatrix} 0.1 & 0.9 \\ 0.6 & 0.4 \end{bmatrix} = [\ X \quad Y\ ]$$

Multiplying the left side and setting its elements equal to the corresponding elements on the right side gives

$$0.1\ X + 0.6\ Y = X$$
$$0.9\ X + 0.4\ Y = Y$$

which is a dependent system of equations. Solving for X in terms of Y gives

$$X = \frac{0.6}{0.9}\ Y = \frac{2}{3}\ Y$$

To completely solve for X and Y, we'll make use of the fact that the sum of the probabilities equals 1. That is

$$X + Y = 1$$

and so

$$X = 1 - Y = \frac{2}{3}\ Y$$

and therefore

$$X = \frac{2}{5} \qquad Y = \frac{3}{5}$$

so the steady-state matrix is

$$[\ 0.4 \quad 0.6\ ]$$

which is what our trial values were indeed converging to.

4.11 THE ODDS OF BELIEF

So far we have been concerned with probabilities as measures of repeatable events of ideal systems. However, humans are experts at calculating the probabilities of many nonrepeatable events such as medical diagnoses and mineral exploration because each patient and site is unique. In order to make expert systems in areas like this, we must expand the scope of events to deal with propositions, which 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

$$P(A \mid B)$$

is not necessarily a probability in the classic sense if the events and propositions cannot be repeated or have a mathematical basis. Instead, $P(A \mid B)$ can be interpreted as the **degree of belief** that A is true, given B.