PROBABILITY EXAMPLES

John G. Truxal
Department of Technology and Society
State University of New York at Stony Brook

MONOGRAPH SERIES
OF THE
NEW LIBERAL ARTS PROGRAM

The New Liberal Arts (NLA) Program of the Alfred P. Sloan Foundation has the goal of assisting in the introduction of quantitative reasoning and concepts of modern technology within liberal education. The Program is based on the conviction that college graduates should have been introduced to both areas if they are to live in the social mainstream and participate in the resolution of policy issues.

The New Liberal Arts (NLA) Program of the Alfred P. Sloan Foundation has led to significant, new courses and course changes in many colleges and universities. MIT Press and McGraw-Hill are jointly publishing an NLA series of books. These monographs are planned to provide teaching/learning materials for other educational developments.

1. John G. Truxal, Feedback/automation
2. Morton A. Tavel, Information Theory
3. John G. Truxal, Probability Examples

NLA Monograph Series

Co-Directors: John G. Truxal and Marian Visich, Jr.
Production Editor: Carol Galdi

Acknowledgment This material was developed as part of the New Liberal Arts Program, with the monograph possible because of grants from the Alfred P. Sloan Foundation and the Ford Motor Company Educational Fund. The latter grant is for the preparation of teaching modules for basic undergraduate courses.

Contents

INTRODUCTION

There are four primary roads from Long Island westward into Manhattan; the commuter must make a sequence of decisions about which road to take. Overhead signs frequently inform the motorist about travel conditions on the main arteries, so that "rational decisions can be made."

The motorist faces an horrendous problem. If the signs indicate, for example, severe congestion ahead on Northern State Parkway, should he/she switch to the Long Island Expressway? Thousands of other drivers are probably making the switch, and the probability is high that the Expressway will shortly be heavily congested.

We face such "decision problems in probabilistic situations" continually. Should I fly ABC airlines when they announce a bomb threat? Should I buy a sphygmomanometer to take my blood pressure frequently during the day?

As so many of the national studies of U.S. education during the 1980s have emphasized, a primary goal is to develop in students the ability to reason quantitatively. This is especially important in situations governed by *probability*, since in such cases human intuition is often wrong. Indeed, the New York State Education Department requires significant exposure to probability ideas each pre-college year (although this is far less demanding than the Japanese high school requirement of a year of probability).

In recognition of this situation, the Sloan Foundation included quantitative reasoning as one of the key goals of its New Liberal Arts Program. Over the past five years, the Stony Brook center has published a monthly newsletter, *NLA News*, with each issue including an example used in teaching quantitative reasoning.

In this monograph, we have collected a number of these newsletter examples and added mini-case studies which have proved successful in the college classroom. These are divided into two broad categories:

(1) Probability -- basic concepts and simple examples illustrating these ideas

(2) Decision Problems -- where the probabilistic concepts are applied to realistic situations.

The instructor is free to duplicate these examples for his/her classroom use.

SOME KEY FEATURES OF PROBABILITY

The engineer or technologist uses probability in a variety of problems: e.g., assessment of the quality of a manufacturing operation, evaluation of the risk associated with a technological change, or determination of the best decision when options are available. In many cases, only fundamental ideas of probability are needed. The following list includes some of the most important concepts.

(1) **Definition of probability.** When an experiment is repeated many times, the **probability** of success is

$$\frac{\text{Number of successes}}{\text{Number of trials}}$$

Example Of 500 professional basketball games, the home team won 275. Thus, the probability the home team will win a particular game is

$$\frac{275}{500} \quad \text{or} \quad 0.55 \quad \text{or} \quad 55\%$$

This definition requires that we have a large number of trials -- how many to ensure we are close to the actual probability is a basic question of statistics.

Second example The weatherman, after studying weather maps, states the probability of rain tomorrow is 70%. This means that if we considered 100 days when the weather maps were similar to today's, it rained on about 70 of the next days.

(2) **Range of probability.** Every probability has a value from 0 to 1. Zero probability means success almost never occurs. A probability of 1 (or 100%) is almost a certainty.

Example When a woman gives birth, the probability she will have a boy is

$$\frac{104}{204} \quad \text{or} \quad 51\%$$

(104 boys are born for every 100 girls.) The probability of a girl is

$$\frac{100}{204} \quad \text{or} \quad 49\%$$

The probability of a boy *or* a girl is 1 -- a certainty.

(3) **Another definition.** If we do an experiment and all possible outcomes are *equally likely*, the probability of success is

$$\frac{\text{Number of successful outcomes}}{\text{Number of possible outcomes}}$$

Example I roll a pair of fair dice, one red (R) and the other white (W). The number of possible outcomes is 36:

R1 W1	R2 W1	R3 W1	R4 W1	- - -
R1 W2	R2 W2	R3 W2	R4 W2	- - -
R1 W3	R2 W3	R3 W3	R4 W3	
R1 W4	R2 W4	R3 W4	R4 W4	
R1 W5	R2 W5	R3 W5	R4 W5	
R1 W6	R2 W6	R3 W6	R4 W6	

A total of four on the dice can happen in three equally likely ways

R1 W3 R2 W2 R3 W1

so the probability of a four is

$$\frac{3}{36}$$

(4) **Probability of failure.** Suppose one or more outcomes of an experiment are considered "failures," with all others "successes." If the probability of failures is P, the probability of success is

$$1 - P$$

Example A factory for making automobile engines has a 15% failure rate: 15% of the engines made do not work properly. Then we know immediately that 85% of the engines are O.K.

If it costs $500 to manufacture an engine and an additional $300 to repair a faulty engine, we can calculate the average cost of an engine from this factory: For each 100 engines produced,

85 cost $500 15 cost $800

The average cost is the total cost divided by the number of engines:

$$\frac{85 \times 500 + 15 \times 800}{100} \quad \text{or} \quad \$545$$

The cost of faulty products is high.

The average cost can be thought of as the weighted average of the possible costs, each weighted (or multiplied) by its probability:

$$\text{Average cost} = 500 \times \frac{85}{100} + 800 \times \frac{15}{100}$$

(5) **Probability of either of two events.** If there are two outcomes of an experiment, the probability of one *or* the other happening is the sum of the separate probabilities.

$$P\,(a \textit{ or } b) = P\,(a) + P\,(b)$$

if the two ways of succeeding (a and b) are distinct, non-overlapping.

Example A couple wants two children. If for each birth the probability of having a boy is 1/2, what is the probability they will have *at least* one boy? There are four equally likely outcomes:

BG BB GB GG

The probability of having one boy is

$$\frac{2}{4} \quad \text{or } 50\%$$

The probability of having two boys (BB) is

$$\frac{1}{4} \quad \text{or } 25\%$$

The probability of having *at least* one boy is

$$50\% + 25\% \quad \text{or} \quad 75\%$$

[We could obtain this answer by using (4) above, considering two girls as a failure.]

(6) **Calculation of probability from specific example.** Often finding the probability of a certain outcome is easy if we arbitrarily look at a certain number of trials and then use the definition of (1) above.

Example In the same example (5) above, we might look at 100 couples. The first child will be

50	B
50	G

(We assume the two outcomes are equally likely.) Now of the 50 with B first, half or 25 will have a B second, the other half a girl, so we have

50	B	25	BB
		25	BG
50	G	25	GB
		25	GG

Looking at the last column, we see 75 of our 100 couples will have at least one boy, so the probability is

$$\frac{75}{100} \text{ or } 75\%$$

(7) **Probability of two successes in a row.** If we run an experiment two times in a row, the probability of two successes is just the product of the two separate probabilities.

If the two experiments or trials are independent (that is, the probability of the second success in no way depends on the first outcome),

$$P\,(a, \text{then } b) = P\,(a) \times P\,(b)$$

Example Our engine factory turns out 85% good engines. Only 2/3 of the faulty engines can be repaired; the other 1/3 must be discarded. Each engine manufactured costs $500; repair work costs $300 whether it succeeds or not. What is the average cost of a good engine coming from this factory?

To answer this, again we imagine 100 engines being made and use the approach of (6):

85 are O.K.

15 are faulty — $\frac{2}{3}$ or 10 are repaired; $\frac{1}{3}$ or 5 are junked

Now that we have the probabilities, we can find the total cost:

100 engines x $500 per engine + 15 engines x $300/repair

or $54,500

We end up with 95 good engines, so the average cost is

$$\frac{54500}{95} \quad \text{or} \quad \$574 \text{ per good engine}$$

(8) **Odds.** The odds of success means simply the probability of success divided by the probability of failure.

Example Our engine factory produces 85% good engines, 15% faulty engines. The odds of the next engine being good are

$$\frac{0.85}{0.15} \quad \text{or} \quad 5.7/1 \quad \text{(or 5.7 to 1)}$$

Sporting and gambling events are often characterized by odds (or other measures of probability such as the "point spread" in football).

(9) **Value of intuition.** In probability situations, intuition is often of little value. This characteristic is emphasized in several of the following examples in this section of the monograph.

Example of two poker hands:

Hand One	Hand Two
A of spades	10 of hearts
A of hearts	9 of spades
A of diamonds	7 of clubs
A of clubs	4 of diamonds
K of spades	2 of hearts

Are the odds of being dealt the poker hand on the right better than being dealt the one on the left?

Answer: NO -- the odds are the same for any specific five cards. Indeed, the probability of *either* hand is 1/2598960, in spite of our much greater psychological surprise upon being dealt Hand One.

References

There are two excellent paperbacks introducing probability:

Derek Rowntree, *Probability Without Tears*, Charles Scribner's Sons, New York, 1984.

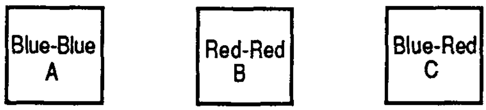
Warren Weaver, *Lady Luck* (The Theory of Probability), Doubleday & Co., Inc., Garden City, NY, 1963.

Then there is a wide selection of introductory textbooks.

The basic ideas of probability typically appear early in a course on quantitative reasoning. I find that it is important to have a few simple class activities which promote student confidence in their ability to handle the subject.

SIMPLE PROBABILITY

Before class, I cut three identical, one-inch squares of white paper. One I color blue on both sides, one red on both sides, and the last blue on one side, red on the other. Thus, I have three squares:



I now hold these three in my hand, mix them up randomly, blindly select one and hold it in my palm to show the class one side only. The side visible to the class and me is Red.

I now argue that we obviously have either B or C, since we can see the Red side. There are only two choices; clearly I am as likely to have picked B as C. I would bet a dollar that the hidden side of the card in my hand is Red, and I look for a student willing to bet a dollar it is Blue.

Is this a fair bet? We repeat the activity 12 times; each time I bet that the hidden color is the same as the color showing. Will I probably end up about even in money?

Answer. The bet gives me an awesome advantage. On the average, I will win 2/3 of the time -- for 12 plays, I expect to win 8 and lose 4, or end up $4 ahead. These 2:1 odds are far more generous than a Las Vegas casino, where the odds may be 55:45, in contrast to my 67:33.

The error in my argument comes from the fact that there are two ways I can show a Red side from square B -- I can be showing either side of the square. Thus, the Red side showing is equally likely to be

Side a of square B
Side b of square B
Side b of square C

2/3 of the time the Red side showing will be from square B, and the other side will also be Red.

The example illustrates beautifully the significance of the phrase "equally likely" in one definition of probability (page 3):

> When all outcomes of an experiment are *equally likely*, the probability of success is the number of ways success occurs divided by the number of possible outcomes.

Other Forms

In an excellent module on Intuition in Statistical Analyses, Economics Professor Alan S. Caniglia (Franklin & Marshall College) uses the example of three identical cups (A, B, and C):

A	containing two blue marbles
B	containing two red marbles
C	containing one red, one blue marble

He blindly picks a cup and one marble from that cup. The question is: What is the color of the other marble in that cup?

Bertrand's Box Paradox depicts three boxes (A, B, and C), each with two drawers:

A	each drawer has a gold coin
B	each drawer has a silver coin
C	one drawer has a gold coin, one has silver

He selects a box at random, then a drawer randomly, and finds a gold coin. What is the probability the other drawer in that box has a gold coin?

Clearly, it is easy to find many different forms of the same example. Caniglia's form has the advantage that we can then extend this to a fuller discussion of *conditional probability*: we can, for example, visualize three cups:

A	contains 1 Red and 99 Blue marbles
B	contains 2 Blue
C	contains 2 Red

I now blindly pick a cup and then a marble, which turns out to be Red. The cup is either A or C, but the two are clearly not equally likely. If I pick cup A, the probability of my *then* drawing a Red marble is 1/100 (this is the conditional probability -- the probability the marble is Red if we are in cup A). Thus, at the outset of the "game," I will pick a Red marble from A with a probability of

$$\underset{\text{Pick A}}{\frac{1}{3}} \times \underset{\text{Pick Red}}{\frac{1}{100}} = \frac{1}{300}$$

but a Red marble from C with a probability of

$$\underset{\text{Pick C}}{\frac{1}{3}} \times \underset{\text{Pick Red}}{1} = \frac{1}{3}$$

If a second marble is picked from the same cup (without replacing the first red marble), the probability the second marble picked will be Red is

$$\frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{300}} = \frac{100}{101}$$

or very close to 1. *Almost* always, the Red marble picked first is from C; if so, the second marble picked will be Red.

(This problem can also be explained with either the joint frequency table or the probability tree.)

EQUALLY LIKELY

In a problem governed by probability, we have to be very careful if we assume that all outcomes are equally likely. The pitfall can be illustrated by a very simple classroom activity.

The instructor tells the students they are about to hold an important lottery. Each student is to select a two-digit number from 00 through 99 and write that number on a paper kept by the student. The instructor has already picked a two-digit number; if anyone matches, he/she wins (and receives an automatic A on the next quiz, or some such prize).

Before revealing the winning number, the instructor polls the students to determine how often each digit was chosen. For example, for 9 first, each student whose number contains a 9 raises his/her hand (if the number was 99, both hands are raised).

In a class of 40 students (80 digits chosen), a typical distribution is

0 - 2	4 - 11	7 - 12
1 - 4	5 - 9	8 - 9
3 - 12	6 - 10	9 - 11

Very few use 0; the digit 1 is chosen appreciably less than the average. The digits are decidedly *not* equally likely,

MIT/Harvard Professor Chernov found the same phenomenon in a study of the Massachusetts lottery. Each digit is equally likely in the winning numbers. Few players chose 0. Consequently, if you won and had chosen a number with zeros (and ones), you were much more likely to be a lone winner, with a large pay-off.

Of course, once such an analysis is published in the newspapers, you probably should stay away from 0 and 1.

PROBABILITY of AUTO ACCIDENTS

In a class of 80 college freshmen, how many are likely to die in automobile accidents? How many will probably be injured seriously enough to require major hospitalization?

We are really asking the question: Do auto accidents in the U.S. constitute a serious national problem? Are auto accidents sufficiently common that we as individuals should be concerned?

To answer this question, we need some data. Newspaper articles reveal the following facts on automobile accidents:

(1) There are approximately 50,000 deaths/year.

(2) There are about 500,000 serious injuries each year.

(3) There are 3 fatalities for every 100 M (one hundred million) miles driven.

(4) Alcohol is a significant factor in almost half (45%) of the serious or fatal accidents -- hardly surprising since there are over 10 M people in the U.S. with severe alcohol problems.

Probability of dying

The information in either (1) or (3) above is enough to allow us to estimate the probability an individual will die in an auto accident. If we use (1) first, we reason as follows:

About 150 M people ride in cars frequently (i.e., we exclude people who ride never or very rarely). Thus, the probability of death in one year is

$$\frac{50,000}{150,000,000} \text{ or } \frac{1}{3,000}$$

One in every 3000 people riding frequently in cars will be killed each year. A college freshman will be riding in cars for about 60 more years, so the probability he/she dies in an auto accident is

$$\frac{60}{3,000} \text{ or } \frac{1}{50} \text{ or } 0.02$$

Clearly, this is a rough calculation: we have estimated the total number of U.S. riders and the life expectancy.

Next, let's see what we can do with (3) above. Our average college freshman is likely to be mobile throughout life and average 30,000 miles of riding each year. In 60 years, he/she will ride

$$60 \times 30{,}000 \quad \text{or} \quad 1.8 \text{ M miles}$$

Statement (3) says that there are three deaths for every 100 M vehicle rides. Since on the average there are 1.8 passengers, (3) says there are three deaths for 180 M passenger miles, or one death per 60 M passenger miles. The probability the freshman will die is then

$$\frac{1.8 \text{ M}}{60 \text{ M}} \quad \text{or} \quad \frac{1}{33} \quad \text{or} \quad 0.03$$

Comparison of two probabilities

In the light of the rough assumptions we made, it is remarkable the two probabilities are so close: 0.02 and 0.03.

We might look at the *sensitivity* of each answer to the assumptions made. In the first calculation, if we change the 150 M people who ride frequently to 100 M (and this was just a crude estimate), the probability comes out 0.03. In the second calculation, if we change the average mileage to 20,000, we obtain the probability 0.02. So the two answers are really not very far apart in view of the roughness of our assumptions.

The correct probability is probably around 0.02 and 0.03. We might guess an intermediate value

$$0.025 \quad \text{or} \quad \frac{1}{40}$$

Answers to questions

With this estimate of the probability of dying in an auto accident, we can now return to the questions at the start of this unit. In a class of 80 college freshmen, each has a probability of 1/40, so we can expect two will die in auto accidents. Ten times as many, or 20, will be seriously injured.

These answers are, of course, rough estimates. We have not worried about details: e.g., a few might be seriously injured twice, or the risk is greater for males (especially young males) than for females, or the probability depends on economic and educational levels. Our answers, nevertheless, indicate the magnitude of the auto-accident problem: in any group of young adults, we can expect 1/4 to be seriously injured and 1/40 to be killed.

Comment

The start of this unit gave us four data statements, we have used only the first three. The relation of alcohol to auto accidents was not relevant. Actually, in any problem we normally have much more data than are needed. An important part of the solution is to select only those data that are useful.

Additional problem

You are driving down a four-lane highway (two lanes in each direction and no center divider). What fraction of the cars coming toward you can you expect to be driven by an individual with a severe alcohol problem?

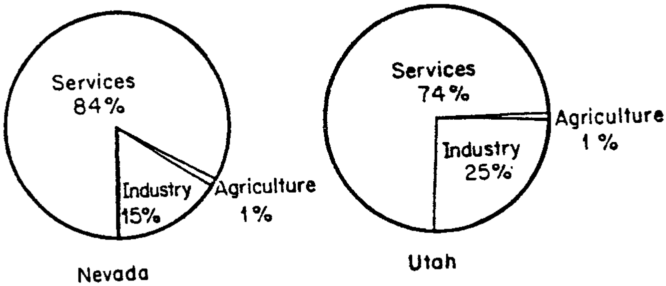
* * * * *

Topics for further discussion might be:

- the recent abandonment of the 55-mph speed limit
- the on-going court battle over mandatory passive restraints
- the Massachusetts and Nebraska voters over-throwing the mandatory seat belt law in the 1986 elections
- the fundamental question of the appropriate role of the federal government in protection of citizens

BIRTH AND DEATH RATES

Nevada and Utah are neighbors and reasonably alike in many ways. The topographies are similar; the types of industry closely match as the following graphs show.



We can compare the two states in a wide variety of ways by using census data. Since we want to compare one state to the other, in each case we use the ratio

$$\frac{\text{Number for Nevada}}{\text{Number for Utah}}$$

Area	1.3 (Nevada is 30% larger than Utah)
Population	0.6 (Nevada has 60% as many people as Utah, or Nevada has 40% fewer people than Utah)
Birth rate	0.7 (For each 1,000 people, Nevada has only 70% as many births as Utah)
Per capita income	1.4
% of high school graduates	1.0
Average years of school	1.0
Male/female ratio	1.0
% people over 65	0.9
% people over 21	1.1
Median age	1.2

These numbers certainly give the impression that we have "sister" states.

There is one number missing above -- the death rate. Even though the age distributions are similar, 40% more people die in Nevada each year for each 1,000 of population. Perhaps there is inadequate health care in Nevada. But census data indicate

Hospital beds per person	1.2
Physicians per person	0.8
Motor vehicle death rate	1.2

so in these three categories the states do not differ in a major way.

In fact, Nevada has a higher mortality rate at every age than Utah. The following graph shows that at age 45 a woman in Nevada has 1.7 times the probability of dying as a woman living in Utah.

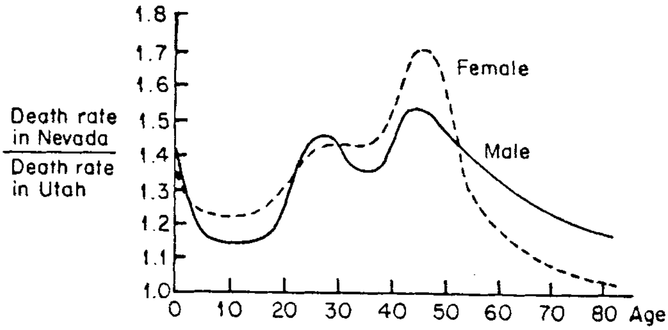


Fig. 2 Increased chance of dying when you live in Nevada rather than Utah. Vertically we plot the ratio

$$\frac{\text{Number of deaths in N for every 1000 people at this age}}{\text{Number of deaths in U for every 1000 people at this age}}$$

The important feature of the graph is the ease with which information is conveyed. Figure 2 is enormously more useful than a page full of numbers. At a glance we can see the most dangerous age for females in Nevada is 45, for males 43.

The question arises as to why the death rate is so much higher in one state than the other. Certain factors are probably important:

(a) The population is much more stable in Utah than in Nevada. 40% of the Utah population were born in that state, less than 10% for Nevada.

(b) The death rate of U. S. males, age 45 to 54, is over twice as great if the males are living alone rather than with a wife. Utah has only 10% in this category, Nevada has 20%.

(c) Nevada has 170% more violent crime than Utah.

(d) Nevada has 25 times as many liquor stores per capita.

(e) Cirrhosis of the liver and respiratory-system cancer are much more prevalent in Nevada than in Utah.

(f) The Mormon religion forbids the use of alcohol and other stimulants.

Students inevitably raise the question of whether the extensive underground nuclear testing in Nevada affects the death rate. The question obviously can not be answered definitively, but we can note that, if the test effects reach the atmosphere, the prevailing winds are toward Utah.

Let's look at the characteristics of the graph. In particular, we note that:

(1) The graph "*stands alone*" -- that is, a reader can understand the graph without reading the accompanying text.

(2) Along the horizontal axis, we show age, so we naturally start from birth (0) and extend out to 80. One year is as important as any other, so the scale is "*linear*" -- each ten years covers the same length.

(3) The vertical scale shows the *ratio* of the two mortality rates. When the graph was first made, I had data for each five years from 0 to 80. All of these data were between 1 and 1.8, so I chose to let the vertical scale cover this same range.

This is a decision which is open to criticism. I could plot the graph for the male mortality ratio as shown in the following figure, where the vertical axis goes from 0 to 1.6. The bottom portion of the plot is now really useless, and the undulations in the ratio with age are not as easy to see. Also this figure does not emphasize as dramatically how poorly Nevada compares to Utah.

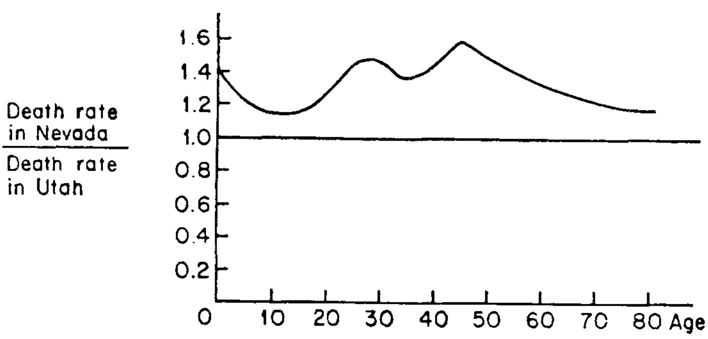


Fig. 3 Alternate plot for male mortality ratio.

* * * * *

This example is taken from the book *Who Shall Live*, by Victor Fuchs (Basic Books, New York, 1974) -- an excellent treatment of the national health care system by a brilliant economist. We have borrowed also from an adaptation by Leon E. Clark, "*Mortality American Style: A Tale of Two States*," May, 1976.

EXPECTED WAIT

My son, Jim, decides to collect baseball cards. He soon becomes discouraged by the large number of duplicates he is buying with his limited amount of money. He comes to me for advice.

"Dad, I know there are 100 different cards. If the company making the cards is honest and puts out an equal number of each card and packages them randomly, how many can I expect to have to buy before I have the complete set?"

How do I answer Jim?

This problem is a good example of the failure of human intuition in analyzing probabilistic situations. Most students will guess 1,000 or 3,000 or some such number; engineering students often guess in the millions (perhaps they are accustomed to larger numbers).

The correct answer is 519.

Expected Wait

Before we tackle the problem, we need to understand one important result of probability theory: Suppose the probability of success in an experiment is a number we call p. I run this experiment over and over again, *until* I have a success. On the *average*, success will occur in a number of trials equal to 1/p.

As an example, look at the box describing the rolling of a single die until a four is obtained. The probability p of a four is 1/6 (there are six equally likely faces of the die, and a four on only one face).

Thus, on the average we will have to wait

$$\frac{1}{p} \quad \text{or} \quad \frac{1}{1/6} \quad \text{or} \quad 6$$

rolls for the four to appear. In the box, the average wait is 5.9 rolls -- surprisingly close to 6, since I really should have found 1,000 sequences to estimate the average.

EXPERIMENT

I roll a single die. I roll again and continue until a four appears, and I keep a record of each roll. The first sequence is

5-2-1-1-1-2-5-1-4

That is, the first roll is a five, then a two, and so on, until I finally roll a four on the ninth roll. The sequence then ends.

This game may be unexciting, but I play again and again. After ten sequences, I have

Sequence	Number of rolls
5-2-1-1-1-2-5-1-4	9
3-3-1-6-2-1-3-4	8
4	1*
5-5-2-4	4*
5-1-4	3*
1-4	2*
5-2-2-5-3-2-1-6-2-6-6-5-4	13
2-2-5-5-2-2-6-3-5-4	10
4	1*
6-6-3-6-2-3-3-4	8

I actually did this experiment and obtained the above results. All told, I had to roll 59 times to get the 10 sequences. Thus, on the *average*, each sequence is

$$\frac{59}{10} \quad \text{or} \quad 5.9 \text{ rolls}$$

Half the time (the five sequences with asterisks), I had a four in 4 rolls or less. Thus, if I bet you a dollar I would roll a four in 4 rolls or less, we would have ended up even -- each of us winning 5 times. (This experiment can easily be carried out on a computer, and we can use many more than 59 runs.)

Solution of Our Problem

With this knowledge, we can return to the baseball-card problem. The first card bought will certainly be new. For the second card, the probability it is new is 99/100, so the expected waiting time or number of purchases is

$$\frac{1}{99/100} \text{ or } 100/99.$$

Similarly, the successive expected numbers of purchases are

$$\underset{\text{Card 1}}{\frac{100}{100}} + \underset{\text{Card 2}}{\frac{100}{99}} + \underset{\text{Card 3}}{\frac{100}{98}} + \underset{\text{Card 4}}{\frac{100}{97}} + \cdots + \underset{\text{Card 99}}{\frac{100}{2}} + \underset{\text{Card 100}}{\frac{100}{1}}$$

Finding the sum of these 100 numbers is tedious, but a calculator reveals the answer to be 519.

Other Applications

(1) If I roll a single die, the probability of rolling a four is 1/6. On the average I will need six rolls to obtain a four, by our above property. Over how many rolls will the probability be at least 1/2 that I have a four? In other words, if I start rolling and continue until I obtain a four, I might obtain the sequences shown in the box. If I do this many times, half the time the sequences will be less than what length?

This problem is different from the question of the *average wait*. Once in a long while, I will have to roll the die 100 times before a four appears. Such a very long sequence pulls the average up.

To solve this problem, it is easiest to find the probability of the first four appearing on the first roll, second roll, etc.

First four on	Probability	Cumulative
First roll	1/6 or 0.167	
Second roll	* $\frac{5}{6} \times \frac{1}{6}$ or 0.139	.167 + .139 = .306
Third roll	$\frac{5}{6} \times \frac{5}{6} \times \frac{1}{6}$ or 0.116	.306 + .116 = .422
Fourth roll	$\frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} \times \frac{1}{6}$ or 0.096	.422 + .096 = .518

* Here we are finding the probability that a four appears first on precisely the second roll. The first roll must be any other number (probability of 5/6); the second roll must be a four (probability of 1/6); hence, the answer is

$$\frac{5}{6} \times \frac{1}{6} \text{ or } 0.139$$

Thus, the probability the first four appears on the fourth roll *or earlier* is 0.518 -- more than 1/2. More than half the time I will have a four by the fourth roll. If I am betting with you, I should wager that a four will appear by at least the fourth roll. The 50% probability is passed about 2/3 of the way to the average (by four rather than the average of six rolls).

Actually one of the first probability calculations was made in the 17th century when the question arose: should I bet that a 6-6 appears in the first 24 rolls of a pair of dice? (The answer is no -- the probability is a little less than 1/2.)

(2) The mean time between fires (MTBF) for a residential house in this country is 100 years. Thus, the probability a house will burn down (or suffer major damage from a fire) in the next tear is 1/100. (Such data are the basis for fire insurance premiums.) There is a 50% chance my house will burn in how many years? (The answer is about 2/3 of the average or 65 years.)

(3) Problem (2) gives an average for the U.S.; there are certain houses at unusually high risk. For example, there are sections of Brooklyn where the MTBF is five years. This means that about half of these homes will burn in the next three years. What are the principal reasons for such a low MTBF?

The existence of such high-risk housing means that the risk for my house is actually less than the average. Simple calculations show that I really should not buy fire insurance; it would be better on the average to take the risk. Such a probability calculation is not useful, however, since I must buy insurance because the loss of the uninsured house would be a catastrophe. This interpretation brings us into a complex economic issue: To compare real costs, we must consider the *utility* of money rather than money itself. I can spare the money for insurance, but not the money to replace my home.

Proof: Expected Wait is 1/p

To prove that the expected wait is 1/p, we proceed to list the probabilities of success in trial 1, then 2, then 3, etc.:

One trial	p
Two trials	$(1-p)p$ [failure first, then success]
Three trials	$(1-p)^2p$

etc.

The average number of trials is the fraction (p) occurring in one trial times 1, plus the fraction in two trials times 2, and so on:

$$\text{Average wait} = p \times 1 + (1-p)\, p \times 2 + (1-p)^2\, p \times 3 + \ldots$$
$$= p\,[1 + 2\,(1-p) + 3\,(1-p)^2 + 4\,(1-p)^3 + \ldots]$$

The infinite series inside the bracket has the sum $1/p^2$. We can look this up in a reference book or, if we have introductory calculus, we can note that the bracketed term is the derivative of

$$1 - \{1 + (1-p) + (1-p)^2 + \ldots\}$$

$$= 1 - \frac{1}{1-(1-p)}$$

since the term in braces is just the geometric series.

PROBLEMS per HUNDRED

Auto manufacturers describe their problems with reliability or warranty costs in terms of "Problems Per Hundred:" for every 100 new cars sold, how many significant problems will there be in the first six months? Each car may have more than one problem, so the PPH may be greater than 100.

Three manufacturers have the following results, in each case for a particular product:

Manufacturer	PPH
A	80
B	120
C	360

Let us assume that I know these data. I am about to buy a new car and I must pick from A, B, and C. The above data, however, are not what I really would like to know. I am anxious to buy a car which gives me *no* problems during the first six months.

If I buy one of the A cars, what is the probability I have no significant problem in the first six months? Is this significantly different from the probability if I buy a car from B or C?

Solution

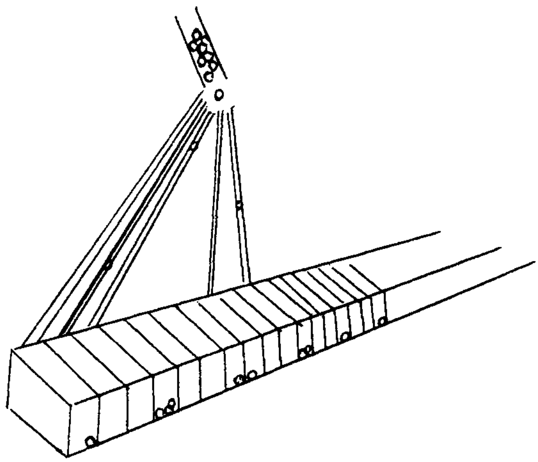
I would very much like to have more information: e.g., if a car has one problem, is it very likely to have more than one? In other terms, what are the relative frequencies of one problem, two problems, three problems, etc.? Maybe, in B's line, one car is a real lemon with 120 problems, and the other 99 cars are problem-free.

Unfortunately, no other data are available, and we have to do the best we can with what we have.

We might visualize a model or analog for this problem. There are 100 bins, each representing a car. For manufacturer A, we have 80 marbles, each representing a problem. The first marble is dropped and randomly falls into one bin. The second marble follows in the same way to one bin picked at random. After 80 marbles fall, what is the probability that a specific bin (say no. 72, corresponding to my car) has *no* marbles?

Let's consider the marbles dropping one at a time. The first marble falls. The probability it is *not* in bin 72 (my car) is 99/100 because there are 99 other bins. Thus, the probability bin 72 has no marbles after one drop is

$$0.99$$



The second marble falls independently of the first -- hence the probability it lands in bin 72 is 1/100, anywhere else 99/100. Thus, after two drops, the probability my bin 72 has no marbles is

$$(0.99)^2$$

If we continue this reasoning for 80 drops, the probability my bin 72 has *no* marbles is

$$(0.99)^{80}$$

A calculator reveals this value is

$$0.45$$

Thus, there is a 45% chance that the A car I buy will have no problems at all -- almost half the buyers experience no problems.

The procedure is exactly the same for finding the probability when I purchase a B or C car. For a B car, 120 marbles must drop. The probability my bin 72 has no marbles after 120 independent drops is

$$(0.99)^{120} \text{ which is } 0.30$$

Similarly, a C car has 360 problems, so the probability my C car has no problems in six months is

$$(0.99)^{360} \text{ which is } 0.027$$

Now if I am primarily interested in buying a car with no problems, I can compare the A, B, C possibilities. My three probabilities are

45%   30%   2.7%

Clearly, buying car C, I am almost sure to have a problem in the first six months. Whether there is a significant preference between A and B is not as clear, particularly if A is more expensive to purchase and/or to maintain.

Is the model valid?

For study of what happens if I buy an A car, I have used a model: 80 marbles dropping randomly into 100 bins to simulate allocation of the 80 problems among 100 cars. Is this model appropriate?

Intuitively I suspect that there is a reasonable upper limit on the number of problems with any one car. When I buy a car, I may have one, two, three, etc. problems, but it is very unlikely that I will have six. Does the model give a very low probability for six problems with my car?

To solve this, let's look at the A car and find the probability the first six marbles end up in bin 72, none of the others ends up there

$$\left(\frac{1}{100}\right)^6 \quad \left(\frac{99}{100}\right)^{74}$$

To end up with six marbles in my bin 72, however, the first can be any of 80, the next any of the remaining 79, and so forth. Thus, there are

$$80 \times 79 \times 78 \times 77 \times 76 \times 75$$

different ways we can pick the sequence of six marbles to fall into bin 72. But all marbles are the same, so the order is inconsequential. Of all these combinations in the last number, 6 x 5 x 4 x 3 x 2 x 1 simply represent the reordering of the six ending up in bin 72.

We now can calculate the probability six marbles end up in my bin

$$\left(\frac{1}{100}\right)^6 \quad \left(\frac{99}{100}\right)^{74} \quad \left(\frac{80 \times 79 \times 78 \times 77 \times 76 \times 75}{6 \times 5 \times 4 \times 3 \times 2 \times 1}\right)$$

A calculator reveals this product is about 0.0002

Thus, about one car in 5000 will have six problems. This is a very small probability which seems reasonable. Perhaps the model is O.K. (If you have studied probability, you will recognize that we are simply using the binomial distribution.)

SAMPLING EXAMPLE

I have two similar, opaque bags. Into one I place 14 red and 6 white poker chips -- we call this the red bag. Into the second bag I place 14 white and 6 red chips -- this is the white bag.

I ask you to take the two bags outside the room, leave one there, and return the other to me. Do I now hold the red or white bag?

I draw one chip from the bag, look at the color, then return this chip and shake the bag. I repeat this sampling and replacement until I have ten samples.

Suppose seven of those samples are red and three are white. What is the probability that this is the red bag?

The most popular guess is 0.7, probably because 70% of the samples I drew were red. A few people insist the probability is 1/2 (the same as when you brought one bag into the room).

The correct answer is 0.97 or almost one. We can state that this is the red bag with a high degree of confidence. Ten samples are enough. This surprising result illustrates the same principles validating political polling: when the samples are randomly chosen, very few samples are needed to obtain an accurate picture.

Proof

The probability of drawing 7 red, 3 white in the sequence R R R R R R R W W W from the red bag is

$$(0.7)^7 \ (0.3)^3$$

There are

$$\frac{10 \times 9 \times 8}{1 \times 2 \times 3} \quad \text{or} \quad 120$$

ways to arrange three white chips in 10 places -- hence, 120 different ways we can draw seven red and three white. Thus, the probability of drawing seven red and three white from the red bag is

$$120 \ (0.7)^7 \ (0.3)^3$$

The probability of drawing seven red and three white from the white bag is, similarly,

$$120 \ (0.3)^7 \ (0.7)^3$$

Thus, the probability the bag is red is

$$\frac{120 \ (0.7)^7 \ (0.3)^3}{120 \ (0.7)^7 \ (0.3)^3 + 120 \ (0.7)^3 \ (0.3)^7}$$

Algebraically, we can rewrite this as

$$\frac{1}{1 + \left(\frac{0.3}{0.7}\right)^4} \quad \text{or} \quad 0.97$$

Indeed, we can set up the table:

Number of red chips in 10 samples	Probability bag is red
5	0.5
6	0.84
7	0.97
8	0.994
9	0.999
10	0.9998

Sensitivity

The system is quite sensitive to the parameters. For example, if we change the number of red chips in the red bag from 14 to 15 (and the white bag correspondingly), the probability the bag is red if we draw six red chips changes from

0.84 to 0.9

The topic of sampling to find probabilities in decision analysis is discussed in highly understandable and captivating terms in the book, Howard Raiffa, *Decision Analysis*, Addison-Wesley, Reading, Ma, 1968.

ADDITIONAL PROBLEMS ON PROBABILITY

Problem 1 The game of Keno is very popular in Las Vegas casinos and commonly goes on continuously so that customers can gamble while they leave the roulette table to eat. As you sit at lunch, a runner comes by to accept your money and a Keno card you have marked. Each card lists numbers from 1 through 80; you can pick one or several numbers. Then, just as in a lottery, the casino selects randomly 20 numbers. There is an enormous variety of possible bets (you might pick 9 numbers, then be paid a given amount if six are among the 20 chosen). Two simpler bets are

(1) "Mark 1 number." You choose one number, are paid $3 if it is among the 20 randomly selected for this game.

(2) "Mark 2 numbers." You choose two numbers, are paid $12 if both are among the 20.

In each case, what is the expected payoff? Which of the two bets is better from your standpoint? How do your chances compare with roulette, where the expected return to the bettors may be 94 cents per dollar, or with blackjack, where in some casinos the expert bettor actually has a slight edge on the casino?

Problem 2 The following shows one page from a 1939 novel. The astonishing characteristic of this book is that it contains no E, normally the most common letter in written English.

The relative frequency of the 26 letters used in English is given by the table below. In 1000 letters, we expect to find each letter the following number of times:

E	132	R	68	L	34	G	20	V	9
T	104	I	63	F	29	Y	20	K	4
A	82	S	61	C	27	P	20	X	1
O	80	H	53	M	25	W	19	J	1
N	71	D	38	U	24	B	14	Q	1
								Z	1

XXIX

Gadsby was walking back from a visit down in Branton Hills' manufacturing district on a Saturday night. A busy day's traffic had had its noisy run; and with not many folks in sight, His Honor got along without having to stop to grasp a hand, or talk; for a Mayor out of City Hall is a shining mark for any politician. And so, coming to Broadway, a booming bass drum and sounds of singing, told of a small Salvation Army unit carrying on amidst Broadway's night shopping crowds. Gadsby, walking toward that group, saw a young girl, back towards him, just finishing a long, soulful oration, saying: --

" . . . and I can say this to you, for I know what I am talking about; for I was brought up in a pool of liquor!!"

As that army group was starting to march on, with this girl turning towards Gadsby, his Honor had to gasp, astonishingly: --

"Why! Mary Antor!!"

"Oh! If it isn't Mayor Gadsby! I don't run across you much, now-a-days. How is Lady Gadsby holding up during this awful war?"

(a) Show that the probability of one page with 780 letters having no E is approximately $1/10^{48}$ -- that is 1/10 . . . with 48 zeros in the denominator. This number is so near zero that we would be dumbfounded to find such a page (unless, of course, there has been an author with such a goal).

(b) Using the letter frequencies given above, determine the percentage of time a typist uses the left hand (if we consider only the 26 letters). The QWERTY keyboard common on typewriters is very mismatched to the capabilities of the human being. The keyboard was designed this way in the last century to avoid successive letters catching or interfering (in those machines, each letter moved up to hit a carbon ribbon over the paper and thereby make a mark on the paper).

Problem 3 Death rates are measured in deaths per thousand people per year in the population. Some typical figures for various parts of the United States are

District of Columbia	13.9	New York	10.1
Florida	11.5	Mississippi	9.3
Maine	10.2	California	7.7

(a) What conclusions can be drawn from these data? What additional information would be desired before statements could be made about the relative attractiveness of these areas if one is interested in minimizing the risk of death?

(b) The gravediggers of New York City (population about 7.5 million) go on strike for three weeks, and bodies accumulate in undertaking establishments. Estimate the number of corpses awaiting interment at the end of the strike.

Problem 4 A couple of years ago, one California legislator gave a speech in which he argued essentially as follows. "95% of the serious auto accidents in this state last year involved only 2% of the drivers. Let's take away the licenses of this small minority, and we'll cut the accident rate dramatically." What is wrong with this policy?

Problem 5 In New York State, with almost 18 M people, there are 280,000 home break-ins or robberies per year.

(a) What is the probability my home will be broken into in the next year?

(b) What factors might change thus probability significantly?

(c) I live in a development of 100 homes. If we are average N. Y. residents, how many home robberies can we expect in 20 years?

(d) How many years will pass before the total probability of my house being robbed is 50%?

(e) Paris is the worst city in the developed world for home break-ins with 27/1000 each year. Los Angeles is second with 26. New York City has 24, Tokyo 5, and the U. S. 15. What factors might cause this wide range?

Problem 6 You are car-pooling to Stony Brook. One morning, one of your friends bets $10 against your $5 that in the next 20 license plates you see, there will be at least one in which the last two digits are the same. Are you making a good bet? (Actually, your friend should offer $36 to your $5 to have an even bet.)

Problem 7 Zipf's law. "Ulysses," the remarkable novel by James Joyce, contains 260,430 words including 29,899 different words -- an astonishing indication of Joyce's working vocabulary (a bright college graduate may use 12,000 words). An interesting feature of the novel is that the 20 most common words appear the following number of times:

the	14877	in	4884	with	2506	you	1894
of	7786	he	4001	it	2350	her	1775
and	7170	his	3326	was	2125	him	1460
a	6396	that	3082	on	2095	is	1346
to	4907	I	2653	for	1972	all	1311

First, it is startling that anyone would spend the hundreds of hours needed to find such detailed information in the days before the computer age. Second, the common words are all short -- we have to go to #41 to find "their" of five letters, to #92 to find "little", the first with six letters.

Beyond these two trivial observations, Zipf noticed that:

Let N be the number of times the most common words appear. Then the second most common appears about N/2 times, the third N/3 times, and so on -- the r^{th} most common, about N/r times.

Surprisingly, Zipf's law works fairly well for other novels or the front page of *The New York Times*. Zipf postulated that such a language structure corresponds to writing with a minimum effort.

Zipf's law is an example of finding a probability model from one case, then testing the validity, and finally trying to explain why the law holds. After the law was presented, others tried to apply it to different areas -- e.g., the population of cities of the U. S. -- even though there's no reason for cities to obey this relation (but they come close).

(a) What is the probability of the word "the" in Ulysses?"

(b) Take the front page of any convenient newspaper and measure the probability of "the" in that text.

A 1790 DECISION PROBLEM

Smallpox Inoculation

As the father of a five-year old child in England in 1790, you are faced with a difficult decision. Smallpox is a major disease, each year affecting one out of eight people who have not had it previously, and killing one eighth of those who contract it. The disease seems to be most threatening to children from ages five to ten.

The government has authorized inoculation as a "preventive" measure. In this process, a physician scrapes fluid from the pustules of someone with smallpox. This fluid is then rubbed into an arm wound of a healthy child to cause a hopefully mild case of smallpox. This crude immunization process, widely practiced in Europe and America (after it originated in Greece a century earlier), has official government approval. Should your child be inoculated and run the risk of having a severe case of smallpox?

The decision is particularly difficult because you are just beginning to worry less now that your child has survived to the age of five. Half of the children die by the age of four, mostly from influenza, whooping cough, and pneumonia. Life expectancy at birth was 28, but at age five a child could expect to live to age 45. Just when you thought health problems were improving, you're faced with this question of inoculation.

Anxious for a better basis for making this very personal decision, you consult the writings of the most learned scientists of the time. During the last 40 years, there has been a major argument among some of these men about the desirability of inoculations and the associated probability theory.

(1) *Daniel Bernoulli* (1700-1782) made the first mathematical study of the risks and benefits of inoculation. Unfortunately, he had very soft data. Previously several physicians had done small epidemiological studies, but the cause of death was often not known. Working with the available data (one susceptible person in eight contracts the disease each year, and of those who contract one in eight dies), Bernoulli calculated

Mean life of a child aged 5	41 years, 3 months
Mean life of a susceptible child aged 5	39 years, 4 months
Mean life of an inoculated child aged 5	43 years, 9 months
Benefit of inoculation	4 years, 5 months

In these calculations, Bernoulli made two assumptions. First, he said the susceptibility of people to smallpox is constant from age five onward. Actually, a child is most likely to acquire the disease before age ten. Second, he assumed one child in every 200 would die within a month from the smallpox resulting from the inoculation.

Even if you accept Bernoulli's work, your decision is not obvious. You have a five-year-old child. If you order him infected with smallpox, you will *probably* extend the child's life expectancy by more than four years. But there's one chance in 200 the child will die within a few weeks, and a similar chance of serious scarring.

Bernoulli himself argued that inoculation was clearly the only responsible decision a parent could make. After reading Bernoulli, however, you found a later paper on the same topic.

(2) *Jean le Rond d'Alembert* (1717-1783) wrote a series of studies pointing out the weaknesses of the analyses by both the pro-inoculation and anti-inoculation groups. He invited all mathematicians except Bernoulli to comment on his studies.

d'Alembert first pointed out the softness of the data underlying the Bernoulli benefit calculation. The mortality from inoculation was especially questionable. When a doctor inoculated and the patient died, the doctor frequently argued that the patient already had smallpox or died from other causes. Usually, people were inoculated with no attempt at follow-up (so the medical society in Constantinople reported 10,000 inoculations and zero deaths). In one of the few careful studies, Dr. Zabdiel Boylston in Boston reported six deaths from 244 inoculations of young children. To obtain better data, d'Alembert proposed that criminals sentenced to death be given to doctors for experimentation.

d'Alembert also emphasized the great differences between the risk analysis done for the state and that for the parent. The government can accept the risk of dying from the inoculation for the sake of the greater average life expectancy. You, however, may logically wonder whether extending life by 4.4 years 40 years hence is that important a benefit for your child compared to the risk of dying within days.

While d'Alembert ends up favoring inoculation, he argues vehemently that the decision should be left to the individual. d'Alembert's cautions leave you leaning away from inoculation. Having read his criticism of Bernoulli, however, it's only fair you read criticisms of d'Alembert, so you continue your study.

(3) *Benis Diderot* (1713-1784), a mathematician and commentator on literature, science, art, and drama, attacked d'Alembert's understanding of basic probability ideas, then went on to criticize him severely for his paper on inoculation.

Diderot argues against d'Alembert's concern for the right of the parent to make an individual decision. He claims the state has the right to require personal risk and sacrifice just as it does in war.

Now you really are confused. These distinguished mathematicians and scientists seem to disagree violently. Fortunately, you find a recent article by Laplace, the "Newton of France."

(4) *Pierre Simon Laplace* (1749-1827), a professor of mathematics who developed the calculus of probabilities, emphasized the importance of statistical measurements. After studying the inoculation problem, he became a warm supporter of inoculations and again calculated the expected gain in life expectancy. Unfortunately, he made a fundamental error.

To find the increase in expected life from elimination of smallpox, he used public health data. He considered each age group separately. For example, for 1000 people aged 20, suppose for one year

40 die from smallpox	100 will die in all
60 die from other diseases	

Laplace assumed most of the 40 above would be saved by inoculation; he overlooked the fact that several of these would die from other diseases.

You now have all the information known and must make your decision.

Historical Epilog

In 1796, the English medical student, Edward Jenner, first vaccinated for smallpox. He found that people gained immunity from smallpox if they were purposely given cowpox -- a mild disease common among cows, a disease in which there were only very localized pustules. In modern times, physicians have used the vacinia virus (closely related to cowpox) -- hence the name vaccination.

Jenner's excellent results were published before 1800. Fifteen years later, vaccination was available throughout Europe, but the original inoculations continued. In 1840, the British Parliament finally forbade inoculations.

None of the probability studies in the 18th century considered two important effects of inoculation:

(1) The inoculation also transmitted other diseases (syphilis turned out to be a major problem and the reason Parliament acted). The omission of this aspect was not surprising, since physicians had a strong conviction that the human body could have only one disease at a time. Indeed, the prestigious Commissioners of the Paris Faculty of Medicine made the strong statement, "germs of two different diseases rarely exist together in the same body without one destroying the other."

(2) No mathematician considered the possible growth of the smallpox epidemic through the inoculation program. The inoculated person infects others. Today, our epidemic models show how faster initial spread of the disease can result in a much more serious epidemic. It is remarkable that this feature was not even mentioned by the scientists -- another commentary about the extreme difficulty of thinking in the unfamiliar terms of probabilities.

By 1971, smallpox was so rare that the U.S. stopped the routine vaccination of children and travelers (more children were killed in auto accidents on the way to a doctor for vaccination than were saved by the procedure). In 1979, the World Health Organization announced that smallpox was a disease of the past -- an awesome accomplishment of technology considering that in 1800 essentially everyone caught the disease during their lifetime.

* * * * *

Ref. I. Todhunter, "History of the Theory of Probability," MacMillan and Company, London, 1865.

The smallpox vaccination story is discussed at greater length in the NLA monograph, *Vaccines*, by Newton Copp.

SCREENING OF A POPULATION USING AN IMPERFECT DIAGNOSTIC TEST

Detecting drug-users among college and professional athletes, pilots and other airline workers, government employees, teachers, and many other groups has become a topic of national concern. Although there are important political, economic, and civil liberties issues involved in any mandatory testing program, we ignore these in order to focus on one quantitative aspect of mass screening.

Let us consider a large population in which the *prevalence* of the condition to be detected is one percent. Each member of this population is subjected to a diagnostic test designed to detect the condition. For definiteness, we can consider a urine test to detect the presence of certain drugs. (Other possibilities are various blood tests for the detection of infections or functional disorders, x-ray tests for the detection of abnormal growths, skin tests for the detection of allergic reactions, etc.).

Were such diagnostic tests *perfect*, all would be simple. Those persons with positive test results would be sure to have the condition, those with negative test results would be sure not to have the condition. Such a situation would be characterized by the following joint frequency table in which each person in the population (assumed to be of size 10,000 for this example) is classified in two ways: (1) has the condition (is one of the one percent or 100 persons who are drug users) or is free of the condition (is one of the 99 percent or 9,900 persons who are not drug users), determining the rows of the table, and (2) tests positive (urine sample judged as showing presence of drugs) or tests negative (urine sample judged free of drugs), determining the columns of the table.

Test Result

		Test Positive	Test Negative	Row Total
State of Nature	Has Condition	100	0	100
	Free of Condition	0	9,900	9,900
	Column Total	100	9,900	10,000

Note the two cells with entries of zero, a consequence of the fact that for a perfect test, there are neither false negatives nor false positives among those tested. For such a test, a person receiving notification of a positive test result has good reason to be worried: he or she is certain to have the condition.

Note also how entries in the joint frequency table are determined as soon as one specifies the total population size. For example, if the size is 100 instead of 10,000 then the joint frequency table for a perfect diagnostic test becomes:

Test Result

		Test Positive	Test Negative	Row Total
State of Nature	Has Condition	1	0	1
	Free of Condition	0	99	99
	Column Total	1	99	100

Unfortunately, almost all diagnostic tests are *imperfect*. Someone who has the condition (is a drug user) can go undetected and yield a negative test result (urine sample incorrectly judged free of the drugs), a *false negative*. Also, someone who is free of the condition (is not a drug-user) can somehow yield a positive test result (urine sample incorrectly judged as showing the presence of drugs), a *false positive*. For an imperfect diagnostic test, the false-negative and false-positive rates are not both zero. Equivalently, the true-negative and true-positive rates are not both 100 percent. Some technical jargon is often introduced as a measure of the quality of a test:

> The true-negative rate of a test is called its *specificity*; the true-positive rate of a test is called its *sensitivity*.

A highly specific test (i.e., a test with specificity near 100 percent) is very good at screening out persons who do not have the condition. A highly sensitive test (i.e., a test with sensitivity near 100 percent) is very good at detecting persons who have the condition. Tests with low specificity produce a high proportion of false positives; tests with low sensitivity produce a high proportion of false negatives. Sensitivity and specificity of a test are independent measures of quality -- they need not both be high or both low; either can be high and the other low. Sensitivity applies to persons *with* the condition to be detected; specificity to those *without* the condition.

To continue, let us assume that the (imperfect) test we are administering to the population has specificity 98 percent and sensitivity 90 percent. Put differently, the false-positive rate is two percent and the false-negative rate is ten percent. Recall that we are also supposing that the prevalence of the condition in the entire population is one percent. Now we are able to state the problem:

> You are a member of this population and the (imperfect) diagnostic test is taken by you. You subsequently are informed that your test result is positive. How worried should you be about having the condition? Put differently but equivalently, what proportion of all those persons in the population who have positive test results actually have the condition, i.e., are true positives rather than false positives?

As we have already noted, were the test perfect you would have good cause for worry. For in that case, all or 100 percent of those persons in the population who have positive test results would actually have the condition. But our test is not perfect; false positives are possible. If you happen to be one of the false positives, you have no need to worry. (We are worrying here only about whether we do or do not have the condition, not about what the undesirable consequences might be of a positive test result even if it turns out to be a false positive.)

We proceed by constructing a joint frequency table, making use of our assumed data to determine the entries in its various cells. For ease in referring to the cells, each is numbered.

		Test Result		
		Test Positive	Test Negative	Row Total
State of Nature	Has Condition	1) 90	2) False Negative 10	3) 100
	Free of Condition	4) False Positive 198	5) 9,702	6) 9,900
	Column Total	7) 288	8) 9,712	9) 10,000

Let us start with cell number 9 by assuming a population of 10,000 persons. We know that one percent of this population has the condition. Thus, the entry in cell number 3 is one percent of 10,000 or 100. It follows immediately that the other 9,900 persons in the population belong in cell number 6. Since the false-positive rate is two percent, we know that of all 9,900 persons free of the condition, two percent or 198 yield false-positive tests. Hence, 198 is the entry in cell number 4. Since the entries in cell numbers 4 and 5 must total 9,900, a subtraction produces the entry in cell number 5. We also know the false-negative rate is 10 percent. That is, of the 100 persons with the condition, ten percent or 10 persons yield negative tests and so appear in cell number 2. The remaining 90 persons yield true-positive tests and so are recorded in cell number 1. The entries in cells number 7 and 8 are obtained by adding to get column totals: 90 + 198 = 288 in cell number 7 and 10 + 9,702 = 9,712 in cell number 8. Finally, we are encouraged by checking that the entries in cell numbers 7 and 8 do indeed, as required, sum to 10,000. This completes the joint frequency table for our imperfect diagnostic test.

Now we turn to the problem of determining the proportion of all positive test results that are actually true positives. There are 288 positive test results all together (cell number 7). Of these, 90 are associated with persons having the condition (cell number 1). Hence, the answer to our problem is the ratio 90/288 or approximately 0.31. Somewhat fewer than one-third of all those receiving notices of positive test results actually have the condition. About two of every three persons receiving notices of positive test results are free of the condition, have received false alarms, and have no need to worry at all.

Mass screening of populations, whether to detect drug use or some disease or abnormality, is a public policy option worthy of careful analysis. Costs and benefits need to be considered. We have presented some simple concepts (sensitivity and specificity) that serve to measure the quality of a diagnostic test and have illustrated how one can use joint frequency tables to analyze some consequences of mass screening.

Finally, it is only fair to mention that we have intentionally suppressed the mathematical foundations underpinning the use of joint frequency tables in problems of the sort posed here. Basic probability theory, particularly Bayes' formula, is implicit in the method. For further reading and explanation, the following reference can be consulted:

Weinstein, Milton C. and Fineberg, Harvey V., *Clinical Decision Analysis*, W. B. Saunders Co., Philadelphia, 1980, especially Chapter Four ("The Use of Diagnostic Information to Revise Probabilities"), pp. 75-130.

Exercises

1. (From Weinstein and Fineberg, *Clinical Decision Analysis*, page 95.)

"Physical abuse of children by their parents is a serious public health problem. The potential damage caused by allowing a case of child abuse to go undetected is great, but the costs of falsely accusing a parent are also high. As the pediatrician responsible for a school health program, should you institute a screening program of physical examinations to detect abused children? Clearly your response will depend to a large extent on the fraction of children with positive test results who are actually being abused.

"The experience of school officials indicates that a careful physical examination will detect 95 percent of battered children (i.e., a false-negative rate of 5 percent), with a false-positive rate of only 10 percent. The best information suggests that 3 percent of school children in an average American city are being abused by their parents."

Suppose a child is examined and found positive, i.e., is diagnosed as having been abused. How likely is it that such a child is, in fact, abused?

2. In the example worked out in the text, the prevalence of the condition to be detected in the population was assumed to be one percent. Suppose instead that it is an even more rare condition, say with prevalence 0.1 percent. Assuming a diagnostic test with specificity 98 percent and sensitivity 90 percent, as in our example, construct a new joint frequency table and determine the proportion of all those with positive test results who actually have the condition. (Before starting your work, what do you think will be the effect of lowering the prevalence of the condition? Should it cause those receiving notice of the positive test results to worry less or worry more about actually having the condition?)

3. "When AIDS Tests Are Wrong," an editorial from the September 5, 1987 issue of *The New York Times*, is reprinted below. In analyzing the Army's test procedure, the editorial writer assumes among women blood donors a prevalence of AIDS of 1 in 10,000, a test sensitivity of 100 percent (i.e., a false negative rate of 0 percent), a false positive rate of 0.005 percent, and concludes that a third of all positive test results will be false positives. Verify.

When AIDS Tests Are Wrong

What should an enlightened society do about a health test that can do as much harm as good? The blood tests now used to screen for exposure to the AIDS virus are highly accurate. But in any widespread testing program, the tests are likely to give as many false positives as true.

No senior policy maker in the Administration seems to understand that paradox, or the terrible price of pressing ahead with widespread testing. For every true case detected, as many other people may be falsely branded, exposing them to discrimination and loss of jobs and housing.

AIDS tests have proved very effective so far, especially in screening blood donations and keeping the blood supply free of virus. But when even a highly accurate test is applied to a population at low risk for AIDS, the number of true positives is so small that it doesn't differ much from the number of false positives. The false positives can thus amount to a significant, even overwhelming, share of the total number found.

AIDS testing is now done in two stages. The first test, called ELISA, can give up to 7 percent false positives. To compensate for errors or sloppiness, this test is usually done twice. But false positives can still occur. The blood contains proteins that happen to mimic the antibodies to the AIDS virus or otherwise confuse the test reagents.

A blood sample that still tests positive is then retested with the Western blot test. The change of it giving false positive readings on both tests is very much less. The Army, which tests thousands of recruits each year with carefully standardized equipment, has probably lowered the joint false positive rate to 0.005 percent, or 1 in 20,000. By the standards of most medical tests, that's a fine achievement. For screening high-risk populations like gay men or addicts, the false positives are a tiny fraction of the true positives identified by the test.

But consider what happens when the Army's test is applied to a population at low risk. Assume, as is typical of women blood donors, that only 1 in 10,000 carries the AIDS virus. Thus among 100,000 women, 10 have the virus and 99,990 don't. The Army's AIDS test procedure, assuming it is perfectly sensitive, will pick up all 10 virus carriers. But among the 99,990 uninfected women, the 1-in-20,000 false positive rate will indicate 5 are carriers. Thus of a total of 15 positive results, a third will be false.

These estimates, published recently in the New England Journal of Medicine, were developed by Klemens Meyer and Stephen Pauker of the New England Medical Center. They note that in any test procedure less accurate than the Army's, the ratio of false to true positives rises alarmingly. If the joint false positive rate rises to only 1 in 1,000, 10 people will be wrongly identified as AIDS carriers for every one true infection found.

These figures must give serious pause to people who advocate AIDS testing among low-risk populations, like marriage license applicants or hospital patients in low-risk areas. Such a policy puts government in the position of urging or compelling citizens to be tested, when wrongly informing many who test positive that they are infected. The certain harm thus done outweighs the uncertain benefit of identifying a few more AIDS carriers.

No honorable government can bear such a moral burden. If low-risk populations must be tested, the false positive rate must be pushed far lower. Until that occurs, the Administration needs to quench its ambitions for wider AIDS testing.

Professors Andrew Zanella and Newton Copp (Claremont McKenna College) have developed two contrasting case studies -- the following on pertussis vaccine and one on the biological effects of ionizing radiation. Both of these technologies provide benefits in our society but not without also imposing some risks; yet a common perception is that nuclear power is "bad" and vaccines are "good". As others have reported, the acceptability of a risk depends upon whether the technology is perceived to be a necessity or a luxury.

PERTUSSIS VACCINE: A DECISION PROBLEM

Pertussis and Its Prevention

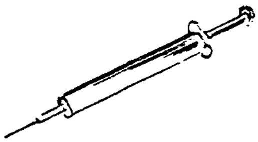
The bacterium *Bordatella pertussis* causes a severe respiratory infection characterized by persistent violent coughing (hence the term "whooping cough"). Complications can develop and include convulsions, brain disorders (categorized under the general term "encephalopathy"), and retardation of development. People are at greatest risk both of acquiring the disease and developing the worst outcomes when they are between the ages of 1 and 5 years.

Prior to 1949, the number of cases of pertussis in the U. S. exceeded 200,000 per year with about 7,000 of these proving fatal. The first vaccine against pertussis was produced and distributed in 1949. This vaccine is made from entire, dead pertussis bacteria and is typically administered in five doses per person (3 doses before the age of one year and 2 more doses by the age of five years).

Pertussis vaccine is one of the components in "DPT" (diphtheria, pertussis, tetanus) injections commonly required of children in the U. S. before admission to school. Pertussis vaccine has thus been a significant component of a nationwide vaccination program that has resulted in decreased incidence of pertussis; the number of pertussis cases has averaged about 1,800 per year over the last 10 years with approximately 10 fatalities each year. (See reference 3.)

Pertussis vaccine has been associated with a variety of unfortunate side effects, although the cause-effect link between the vaccine and alleged serious side effects remains controversial. Single doses may produce minor reactions in recipients, such as fever and "excessive somnolence". These minor reactions rarely persist for more than 48 hours after administration of the dose. More severe reactions can occur and include convulsions, "collapse", and prolonged high-pitched crying. As disturbing as these more severe reactions are, however, they do not lead to permanent disabilities. The third and most serious class of reputed reactions to pertussis vaccine includes encephalopathy leading to permanent brain damage or death. (Reports that pertussis vaccine causes sudden infant death syndrome (i.e., SIDS) are false. See citation in reference 2.)

Objections to Pertussis Vaccine



Public reaction to the side effects of pertussis vaccine led to sharp decreases in public acceptance of the vaccine in England, Sweden and Japan in the late 1970's. An organization of parents (Dissatisfied Parents Together) is currently raising questions about the application of pertussis vaccine in the U. S. and how victims should be compensated. The increasing number of lawsuits for compensation of pertussis vaccine-related injuries, from one in 1978 to 73 in 1984 and 219 in 1985, illustrates increasing public concern about the risks associated with this vaccine.

Questions

1. Would you choose pertussis vaccine for your children?

2. If you were a pediatrician, would you administer pertussis vaccine to your patients?

3. Should the state or federal government require vaccination against pertussis for admission to school?

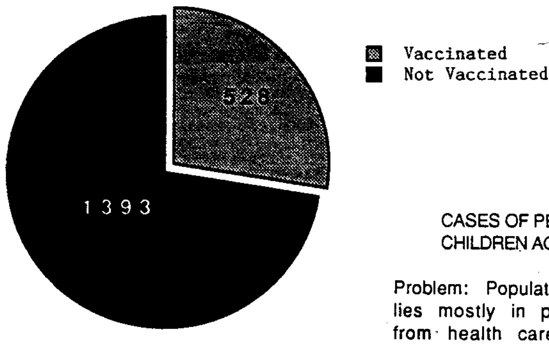
4. If you were the CEO of a major pharmaceutical firm that makes vaccines, would you direct your firm's efforts to the manufacture and distribution of a new pertussis vaccine?

Data

Most published evaluations of the risks associated with pertussis vaccine rely on estimates of the number of side effects because of difficulties in making accurate counts of vaccine-related injuries; the worst outcomes are rare and the reporting of less severe effects is inconsistent. (See reference 1 for an experimental survey of pertussis vaccine and its side effects.) The following data have been established according to published estimates of risk (e.g., references 1, 2 and 4). These data can be used to determine risks per vaccinee that is 5 years old or younger.

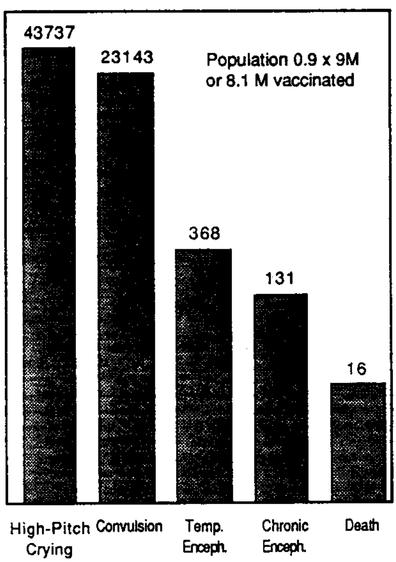
Assume

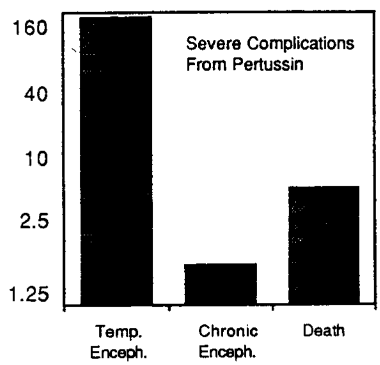
The total population of children 5 years old or younger in the U. S. is nine million. 90% of children aged 5 years or younger have received the vaccine. The vaccine fails to confer immunity against pertussis 10% of the time.



CASES OF PERTUSSIS IN CHILDREN AGE 5 OR LESS

Problem: Population not vaccinated lies mostly in people separated from health care system; cases likely to be under-reported, perhaps by factor of 10.





VACCINE-EFFECTS DATA FOR 9 MILLION CHILDREN THROUGH AGE 5 (90% VACCINATED; 10% OF VACCINATED NOT IMMUNE)

CDC REPORTS 1983

A Decision Tree

In order to expose steps in the reasoning process, we construct a decision tree (page 49) from the data provided above. To decide whether to take the vaccine, it is helpful to consider the *risk per vaccinee*, rather than the more commonly reported risk per dose. This is reasonable in the case of severe reactions because the vaccination series is usually stopped after evidence of one severe reaction. Risks and probabilities needed for the decision tree are calculated first.

Risk of Chronic Brain Damage from Pertussis Vaccine

A total of 67,395 serious adverse effects are expected when 90% of the five-year old or younger cohort receives pertussis vaccine. The risk of experiencing some adverse effect is thus

$$\left.\begin{array}{l}\text{67,395 serious adverse effects}\\ \text{8.1 M children vaccinated}\end{array}\right\} \quad \frac{67395}{8.1\text{ M}} \quad \text{or} \quad \text{8320 per million}$$

Of the 67,395 cases of serious adverse effect from the vaccine, 131 result in chronic brain damage. The risk of experiencing this particular side effect among those people who show at least one adverse reaction is thus

$$\left.\begin{array}{l}\text{131 cases of chronic enceph.}\\ \text{67,395 total serious effects}\end{array}\right\} \quad \frac{131}{67395} \quad \text{or} \quad \text{1944 per million}$$

The total risk of vaccine-associated chronic brain damage for a person at the decision node of the tree is

$$\left.\begin{array}{l}\text{Total probability of this outcome}\\ \text{is product of two probabilities on path}\end{array}\right\} \quad \frac{8320}{1\text{ M}} \quad \text{x} \quad \frac{1944}{1\text{ M}} \quad \text{or} \quad \text{16 per million}$$

This estimate of risk would be less by approximately a factor of ten if the incidence of chronic brain damage was assumed to be 1:510,000 vaccines (see reference 1) rather than 1:310,000 doses (see reference 2) as assumed here.

Risk of Chronic Brain Damage from Pertussis in Non-vaccinated Children

In 1983, 1786 (of a total of 2463) pertussis cases occurred in people *not* vaccinated for pertussis. Because 78% of the pertussis cases in that period occurred in children five years old or younger, the actual number of cases in the cohort of interest was 1393 (1786 x .78). The probability that a non-vaccinated child will contract pertussis (under 1983 conditions of vaccination) is thus

$$\left.\begin{array}{r}\text{1393 cases in non-vaccinated population}\\ \text{0.1 x 9M or 0.9M children not vaccinated}\end{array}\right\} \frac{1393}{0.9\ M} \text{ or 1548 per million}$$

Two of the children with pertussis in 1983 developed chronic brain damage. Without trying to decide whether these cases involved vaccinated or non-vaccinated children, the risk of chronic brain damage from pertussis becomes

$$\left.\begin{array}{r}\text{2 children had chronic enceph.}\\ \text{1921 cases of pertussis in children}\end{array}\right\} \frac{2}{1921} \text{ or 1041 per million}$$

The total risk of chronic brain damage from pertussis in non-vaccinated children thus becomes

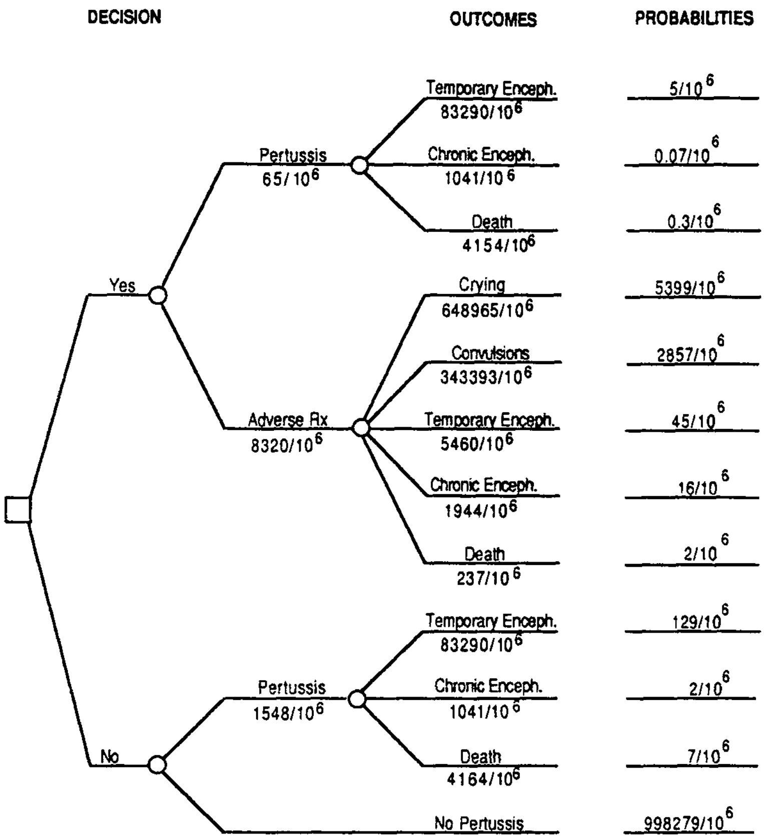
$$\left.\begin{array}{l}\text{Total probability of the outcome}\\ \text{is product of two probabilities on path}\end{array}\right\} \frac{1548}{1\ M} \times \frac{1041}{1\ M} \text{ or 2 per million}$$

Discussion

Inspection of the completed decision tree leads to a very interesting discussion. Regardless of a person's decision, the risk of chronic brain damage is low. The risk of brain damage, however, seems to be 8 times greater for people who elect the vaccine than for people who do not elect the vaccine. This leads some students to recommend avoidance of the vaccine. Other students look at the maximum penalty, death, and see that the risk of death is lower for people who take the vaccine. Which strategy is better?

This question also leads to discussion of what is meant by one strategy being "better" than another. Then there is the question of the estimates on which the risk calculations are based. Published estimates of vaccine-caused chronic brain damage differ by at least a factor of ten. The more pessimistic estimate has been used here. Changing to a more optimistic estimate changes the risk factor and the students' decisions.

A second issue concerns the effect on risk estimates of assuming 90% compliance with the vaccine program. In inner city areas, compliance has been estimated to be as low as 35%. How does this change the risk calculations?



For the sake of simplicity, only severe outcomes of pertussis and vaccination are shown. A complete decision tree would include all possible outcomes.

Other sources of uncertainty exist. Officials at the CDC estimate, for example, that the number of pertussis cases reported to them may be only 10-15% of the actual figure because of a lack of uniform diagnostic procedures for pertussis and inconsistent reporting of pertussis to the CDC by physicians. The number of adverse reactions to the vaccine is also difficult to estimate both because of a delay between receiving a vaccine dose and the onset of the effect and because a symptom reputed to be caused by the vaccine may have been the result of a condition in existence prior to vaccination; publicity of the risks of pertussis vaccine has undoubtedly increased attribution of symptoms to the vaccine.

*   *   *   *   *

References

(1) Cody, C. L., L. J. Baraff, J. D. Cherry, S. M.. Marcy, and C. R. Manclark. 1981. Nature and rates of adverse reactions associated with DTP and DT immunizations in infants and children. Pediatrics 68:650-660.
(2) Hinman, A. R. and J. P. Koplan. 984. Pertussis and pertussis vaccine; reanalysis of benefits, risks, and costs. JAMA 251:3109-3113.
(3) Immunization Practices Advisory Committee. 1985. Diptheria, tetanus, and pertussis: Guidelines for vaccine prophylaxis and other preventive measures. Morbidity and Mortality Weekly Report 34:405-441, 419-426.
(4) New Vaccine Development: Establishing Priorities, Vol. I. Diseases of Importance in the United States - National Academy Press, 1985.
(5) Vaccine Supply and Innovation - Division of Health Promotion and Disease Prevention, Institute of Medicine, National Academy Press, 1985.

NON-ZERO SUM SITUATIONS

We frequently encounter problems with two characteristics:

(1) Non-zero sum. In a two-party competition, what one person gains may not equal the losses for the other. Indeed, in many cases, both parties may gain from a strategy change.

(2) Counter-intuitive. The best solution leads to results opposite to our intuitive expectation.

Both these features often arise in systems controlled by government policy; then it is obviously important to do enough modeling and system analysis to be able to anticipate such characteristics. Many engineers believe that Jay Forrester's principal contribution in the 1960's was to point out these important features through detailed computer models of a business, an urban region, and the global system.

An example illustrates a way to introduce these ideas within an introductory course in modeling and problem-solving.

Problem of Ice-Cream Vendors

Each of two men has an ice cream wagon which he brings each morning to a mile-long beach. During the day, customers or sun-bathers are distributed uniformly along the beach. When a vacationer wants ice cream, he goes to the nearest vendor, since there is no real difference between the two. Because the desire for ice cream varies widely, the percentage of people buying ice cream falls off as the distance from the nearest vendor increases.

When the first vendor arrives in the morning, where should he locate? Where should the second man then settle with his wagon? We assume that there is no cooperation between the two: each man doesn't trust the other, and each acts solely to maximize his own sales.

A little thought shows that the two vendors should locate next to one another exactly in the middle of the beach (Fig. 1). If A, the first to arrive, makes any other choice, B will then settle just on the side of A toward the longer expanse of beach (Fig. 2). If *A* picks the middle and *B* locates anywhere but beside him, *A* will be serving more than half the customers (Fig. 3).

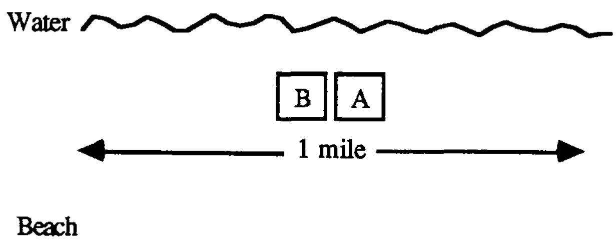


Fig. 1 Optimum location of ice cream wagons A and B in a strictly competitive situation.

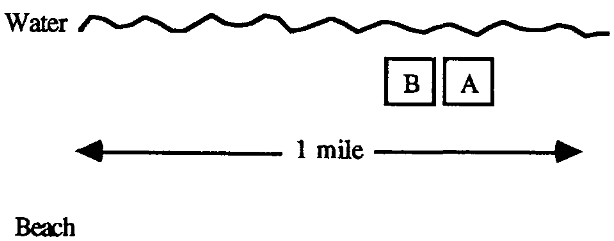


Fig. 2 Final location if A selects any point other than the center, B then optimizes.

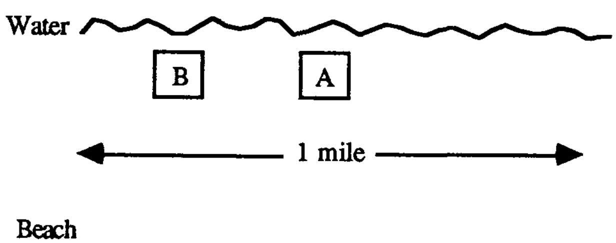


Fig. 3 Final location if A chooses the middle and B makes a non-optimum choice.

The game or problem leads to a ridiculous solution, and certainly one which does not give the best possible service to the customers or profit to the two competitors. From the standpoint of both parties, the optimum solution is shown in Fig. 4. Each vendor is in the middle of "his" half of the beach. The customers have shorter distances for ice cream on the average, yet each vendor serves half the people.

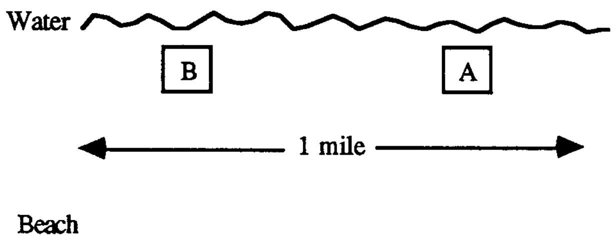
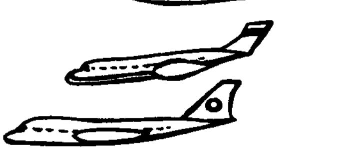


Fig. 4 Solution best for the customers and the competitors.

Why is the solution of Fig. 4 so difficult to achieve in a competitive situation? It requires honest cooperation between the two competing forces -- a cooperation which has to be based on mutual trust. When *A* arrives in the morning, he takes his possibly bad position in confidence that *B* will not try to take advantage of his trust. During the day, both resist the temptation to move gradually toward the center of the beach. This is a non-zero-sum situation: both benefit by disdaining what looks like an advantageous possibility.



Other Examples

This rather trivial example has counterparts in the real world. One of the most obvious is the fierce competition among airlines for passengers on the lucrative New York-Los Angeles trip. In an attempt to attract a large fraction of the customers by having flights at convenient times, each airline runs more flights than necessary (on the average, less than half the seats are filled) and the three schedules tend to be startlingly similar -- with three flights taking off almost concurrently, then a gap of several hours before the next three.

Part of the problem here has been the federal anti-trust laws and the concept that the public is best protected by complete competition among firms. It is only in the past few years that the government has allowed the three airlines to talk together in an attempt to agree on schedules which would improve both service and profits.

In a broader sense, non-zero-sum decisions dominate our society, as G. Hardin has pointed out in his classic article on the tragedy of the commons.* An extreme example illustrates the point. There are two men, each pays $10 per month for rubbish collection from his home. From a selfish standpoint, each would benefit if he simply threw the day's rubbish out the car window while driving to work in the morning. The damage to the environment would be negligible, each man would save $10 monthly, and the quality of life in the area would not deteriorate in any real way if just the two of them followed this practice.

The problem is obvious. If everyone, or even a significant fraction of the populace, acts in this same selfish way, we all will shortly be surrounded by rubbish. In order to find a social optimum -- a decision or strategy which results in the best total picture -- we each must make decisions which are far from optimum for ourselves.

As we consider different socio-technological problems, we find again and again this non-zero-sum characteristic. The total system only works if the individual is willing or can be forced to relegate personal benefit to a secondary position compared to the total, social welfare. Just as in the ice-cream-vendor problem, success depends on individual trust and confidence in others. Whether the concern is obeying traffic regulations or conserving energy, we must design total systems in which the individual is willing to make decisions which are obviously not the best for him/her personally.

*The term was used because he illustrated the concept with the example of a community with a limited grazing acreage (a commons) and each man unwilling to limit his small use of that commons for his cattle. (Garrett Hardin, "The Tragedy of the Commons", *Science*, Dec. 13, 1968, pg. 1243-1248.)

DECISION PROBLEMS WITHOUT ANSWERS

There is a betting game which is so contrary to human intuition that anyone familiar with the game is almost guaranteed large, quick winnings.

The game

If I flip a coin three successive times, there are eight possible outcomes: heads, heads, heads; heads, tails, heads; etc.

HHH HTH THH TTH HHT HTT THT TTT

They are all equally likely, so each has a probability of 1 in 8 or 1/8. If we carry out the three flips 8000 times and carefully record what happens, we will find HHH appears about 1000 times, TTH about 1000 times, and so on.

To play our game, you and I will each select a triplet, then flip a coin until one of the triplets appears. If my triplet appears first, I win. For example, if you select HTH, I might pick HHT. We would then start flipping and recording the results; as soon as either your HTH or my HHT appears, we have a winner and the game is over. As we flip, we might (as an example) obtain the sequence

H T T T T [H T H]

Then we stop; you have won (the last three are your HTH).

I'll let you select your triplet first, since I thought up this game. The astonishing feature is that I can always play with at least 2:1 odds in my favor if I use the following strategy:

(a)	If you select a triplet,	xx-	(that is, the first two the same)
	I pick	yxx	(the last two the same as your first two, the first different)
(b)	If you select a triplet,	xy-	(the first two different)
	I pick	xxy	(the last two the same as your first two, my first the same as my second)

Thus, if you choose HTH, we are in Case (b) above -- your first two are different. Then I select HHT, and I will win 2/3 of the time.

You become suspicious and pick HHT: then I simply choose THH and now win 3/4 of the time. You then select THH, I pick TTH and win 2/3 of the time. No matter what you choose, I can win with overwhelming odds (2:1 or higher odds are very unusual in gambling, where casinos operate with an edge of only a few percent.)

After a while you want me to pick first. Then we can alternate the first choice; if you don't know the strategy, I'll win half the time when I choose first, 2/3 or more of the time when you choose first -- still fantastic odds for me.

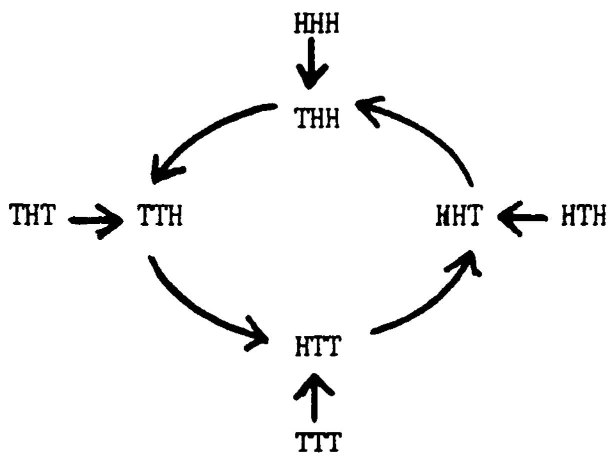
Non-Transitive feature

The remarkable feature of this game is the *non-transitive* characteristic: in simple terms, *no matter what you select*, I can then pick something better.

If we use the symbol

$$A \rightarrow B$$

to mean A loses to B (that is, if you pick A, I should pick B), we can show this non-transitive feature by the following diagram



No matter which triplet you select, there is an arrow leading away toward the triplet I should pick.

Investment example

I have $50,000 to invest; I will need the money in five years, when I plan to buy a house. There are three different ways I can invest: real estate, stocks, or certificates of deposit.

Unfortunately, economists are unable to tell me what is going to happen to the economy over the next few years. They do estimate that, if I invest in real estate, there are roughly three possibilities:

I may earn $40,000 if prices continue to rise.
I may earn 0 -- just receive my money back.
I may lose $10,000 if prices have peaked.

Furthermore, they tell me that the best guess is that these three possibilities are equally probable: each has a probability of 1/3.

The economists analyze the other kinds of investments in the same way, and I have the following table from which I have to make a decision:

Real estate	Stocks	Certificates	
40,000	30,000	15,000	In each column,
0	20,000	10,000	probability of each
-10,000	-20,000	5,000	outcome is 1/3

Notice that this is the best information I have on which to base a decision. I ask: if these data are correct, what is the best decision?

Best expected earnings

I might decide to invest to receive the largest possible, *expected* earnings. This term, *expected* value, has a precise meaning. If I buy stocks,

- there is one chance in three of earning 30,000
- one chance in three of earning 20,000
- one chance in three of losing 20,000

Hence, with stocks I can expect (on the *average*, I will receive)

$$\frac{1}{3}\ (30{,}000) \quad + \frac{1}{3}\ (20{,}000) \quad + \frac{1}{3}\ (-20{,}000)$$

or $10,000. Notice I will never earn $10,000 (I earn 30,000 or 20,000, or lose 20,000), but, considering all possible outcomes, my expected return is $10,000. (The average here is calculated as described on page 4.)

When I find the expected earnings for each option (real estate, stocks, and certificates), the answer is always the same: $10,000. Consequently, there is no way to choose one on the basis of how much I can expect to earn.

Best future satisfaction

Since I can't choose on the basis of expected earnings, let me try a different approach. Five years from now when I cash in my investment, I will know which of the three possibilities for real estate (for example) has actually happened. If I had bought real estate and stocks have done better over the five years, I will be very unhappy. I will choose today that investment which five years from now is most likely to have been the best choice.

To decide how to do this, I first simplify the problem: I compare real estate to stocks. (Later I'll compare the winner to certificates.) Three different things may happen to real estate in five years (+40,000; 0; -10,000). *Each* of these may be accompanied by each of the three things that might happen to stocks (+30,000; +20,000; -20,000). Hence, five years from now, there are nine different, possible comparisons which I will be making -- each equally likely:

Real estate	**40**	**40**	**40**	0	0	**0**	-10	-10	**-10**
Stocks	**30**	**20**	**-20**	30	20	**-20**	30	20	**-20**

In five of the nine situations (those boxed), real estate will have been a better investment than stocks. Clearly, real estate is in this sense better than stocks: with a fairly high probability of 5/9, I will be happy five years from now that I bought real estate. So I eliminate stocks as an option.

Now I compare real estate to certificates:

Certificates	15	**15**	**15**	10	**10**	**10**	5	**5**	**5**
Real estate	40	**0**	**-10**	40	**0**	**-10**	40	**0**	**-10**

Six of nine pairs of outcomes favor certificates. Obviously, I should invest in certificates.

Certificates beat real estate which beats stocks.

Clearly, certificates are better than stocks -- I wonder how much. So again, I make a comparison of just those two:

Stocks	**30**	**30**	**30**	**20**	**20**	**20**	-20	-20	-20
Certificates	**15**	**10**	**5**	**15**	**10**	**5**	15	10	5

Alas! Stocks are *much* better than certificates.

Comments

The above problem is a system which is given the name "non-transitive." A is better than B, B is better than C, but C is better than A. This kind of situation boggles the mind.

We are comfortable saying:

Jack is heavier than Bill.
Pete is heavier than Jack.
Therefore, Pete is certainly heavier than Bill.

This is a transitive (and logical) situation. But to say

A is better than B is better than C is better than A goes against human intuition.

Both of the examples above (game and investment) are probabilistic: the future can not be predicted; we can only assign probabilities to each of the possible outcomes. The possibility of non-transitive behavior deeply troubles the political analyst, who recognizes that the future can only be described in terms of probabilities, but who argues that quantitative analysis can help guide better decision-making. Many of the decisions faced by political leaders (or by individuals) are described by non-transitive systems -- so that, no matter what decision is made, we could have done better.

This same kind of curious, non-transitive behavior was first noticed in the 18th century by the Marquis de Condorcet, who observed that, if there are three political candidates (A, B, C), voters may prefer B to A, C to B, but A to C -- the "paradox of voting."

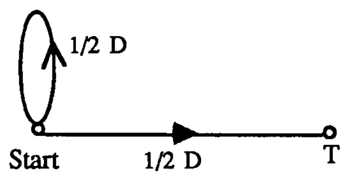
We find non-transitive, decision problems in games, system analysis, political science, sociology, and economics.

Optional Postscript

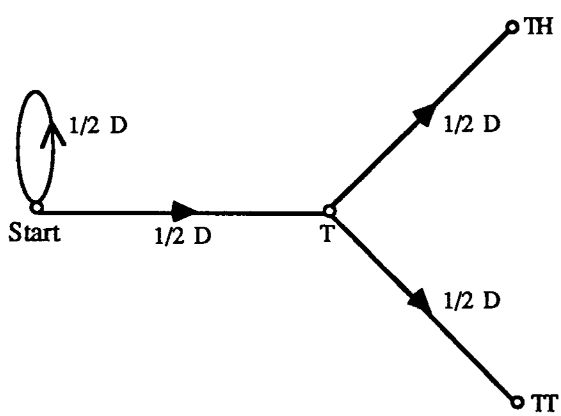
In the above discussion we have simply stated (and not proven) the probabilities of various outcomes. For example, in the coin-triplet game, we stated that TTH would win over THT two thirds of the time.

There are two ways we might prove this statement. First, we can simply play the game many times (probably on a computer to save time). In 3000 games, we will find that TTH wins about 2000 times.

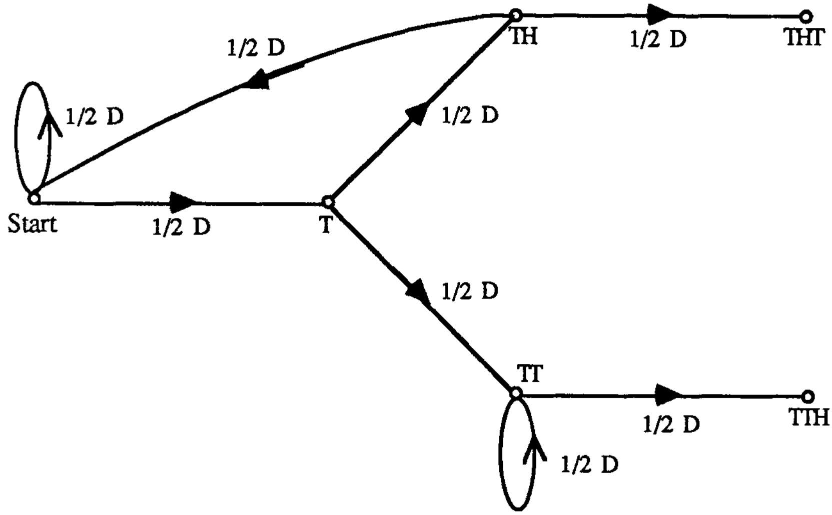
We can also prove the statement mathematically. We first draw a diagram to show the various ways to reach both TTH and THT. From the start, we may flip a tail (moving us to T with a probability of 1/2; D means we have gone through one flip). We might also flip a head, which returns us to the start since we have not made progress toward either TTH or THT.



If we are at T, the next flip can give us either TH or TT for the last two flips:



Following this same line of reasoning, we complete the diagram:



This diagram describes the game; when we begin at the start node, we eventually end up at either THT or TTH.

What are the relative chances of reaching THT or TTH. In both cases, we must reach the T node. Once we are at T, there is a probability of 1/2 of moving to TT. If we reach TT, however, we inevitably reach TTH eventually. Hence, from T there is a probability of 1/2 of reaching TTH.

From T, however, we have only a probability of 1/4 of reaching THT (1/4 of the time we will end up back at the start with two heads in a row). Thus, from T

P of reaching TTH is 1/2
P of reaching THT is 1/4

Two thirds of the time, TTH wins; one third of the time, THT.

Similarly, we can prove the odds stated earlier for all pairs of choices in the triplet game.

(Walter Penney, "Penney-Ante," *Journal of Recreational Mathematics*, 1969, p. 241.)

SIMPSON'S PARADOX

Drug Testing

There are two drugs (called A and B) to be evaluated, and the Food and Drug Administration (FDA) asks Harvard and Stanford medical Schools to carry out the tests. Harvard reports the results

	Number of people tested	Number helped	% helped
Drug A	200	50	25%
Drug B	10	5	50%

So Harvard reports to the FDA that drug B is twice as effective as drug A.

At Stanford there are also 210 patients available, with the following results

	Number of people tested	Number helped	% helped
Drug A	10	1	10%
Drug B	200	40	20%

So Stanford likewise reports that drug B is twice as effective as drug A. The FDA then lists drug B as recommended, drug A as not recommended.

A staff member at FDA is preparing a presentation to a Congressional Committee to explain why Drug B is recommended. To simplify the report, he/she combines the results from the two schools. To his/her dismay, he/she comes out with

	Number of people tested	Number helped
Drug A	210	51
Drug B	210	45

and notices that the not-recommended drug A is obviously better!

SIMPSON'S PARADOX

The Death Penalty

In 1978, Warren McClesky, a black man, was convicted of killing a white police officer and was sentenced to death in Georgia. In an appeal before the U. S. Supreme Court, lawyers for McClesky argued that the imposition of the death penalty in Georgia was racially biased. They presented statistical models showing that defendants accused of murdering whites in Georgia were four times more likely to be sentenced to death than were defendants accused of murdering blacks and argued that race-of-the-victim discrimination played a key role in the decision to sentence McClesky to death.

The Supreme Court, in a 5-4 decision, allowed McClesky's sentence to stand. A full discussion of the legal and statistical issues involved in the case would require hundreds of pages of text. Indeed, many of these issues are the subject of continuing debate. For a more detailed summary, references are available from the author. The purpose of this article is to examine the major statistical issue of the case and to consider a widespread phenomenon known as Simpson's paradox.

There is a pattern in the Georgia data that can be found in the death penalty data from several states. I will focus on data from Florida that make the point clearly. These data, taken from a *Stanford Law Review* article by Samuel Gross and Robert Mauro (Gross and Mauro, 1984), cover homicides in Florida during 1976-1980 in which the victim did not know the suspect. There were 724 such cases in which the suspect was either white or black and the victim was either white or black (cases involving other racial groups are excluded from this analysis). Of these cases, 71 (9.8%) resulted in the suspect being sentenced to death. Table 1 provides a breakdown by race of the suspect.

Table 1

		Death Sentence?			
		Yes	No	Total	% Yes
Race of Suspect	White	39	308	347	11.2%
	Black	32	345	377	8.5%
		—	—	—	
	Total	71	653	724	9.8%

Note that white suspects are more likely to be sentenced to death (11.2%) than are black suspects (8.5%).

The racial factor that has the greater effect on the disposition of these cases is not the suspect's race, but the victim's race. Consider a breakdown of the cases in Table 1 into two parts: cases involving white victims and cases involving black victims. Of the 724 cases in Table 1, 468 involve white victims. In 318 of these cases the suspect was white, while in 150 cases the suspect was black. Black suspects were more likely to be sentenced to death than were white suspects when the victim was white -- see Table 2.

Table 2: White Victims

		Death Sentence?			
		Yes	No	Total	% Yes
Race of Suspect	White	39	279	318	12.3%
	Black	29	121	150	19.3%
	Total	68	400	468	14.5%

Now consider the 256 cases involving black victims. These are summarized in Table 3. Black suspects were more likely to be sentenced to death than were white suspects when the victim was black.

Table 3: Black Victims

		Death Sentence?			
		Yes	No	Total	% Yes
Race of Suspect	White	0	29	29	0.0%
	Black	3	224	227	1.3%
	Total	3	253	256	1.2%

Comparing these three tables we see an example of Simpson's paradox: blacks are more likely (than whites) to be sentenced to death when the victim is white and blacks are more likely to be sentenced to death when the victim is black. These are the only two settings present and, thus, we might think that blacks would be more likely to be sentenced to death overall (combining the two settings). However, this is not so. Table 1 shows that blacks are *less* likely to be sentenced to death in the aggregate. (Note that combining the numbers in Tables 2 and 3 yields Table 1 -- no cases have been added or deleted.)

Compare the "% Yes" values in the tables. In Tables 2 and 3, the percentage is higher for black suspects than for white suspects, but this inequality is reversed when the data are combined into Table 1! How can this be?

The two "% Yes" values for black suspects are 19.3% and 1.3%, which are greater than the "% Yes" values for white suspects of 12.3% and 0%. If we took simple averages, the aggregate percentage for blacks would exceed that for whites. But we do not take simple averages. Most of the black suspects were accused of killing blacks, while the overwhelming majority of white suspects were accused of killing whites. Moreover, the death penalty was imposed far more often when the victim was white than when the victim was black. These two factors -- (1) whites tend to murder whites while blacks tend to murder blacks and (2) cases with white victims are more likely to result in a death sentence than are cases with black victims -- combine to produce Simpson's paradox. We would say that white suspects are more likely to be sentenced to death than are black suspects *if we ignore victim's race*, but blacks are more likely to be sentenced to death *if we control for victim's race.*

ACT Scores

Average ACT composite scores went *down* between 1987 and 1988 among students who completed a core curriculum (at least four years of English and at least three years each of mathematics, natural sciences, and social sciences) and among students who did not complete a core curriculum, but the overall average went *up* -- see Table 4.

Table 4: ACT Composite Averages

		1987	1988	Change
Completed Core Curriculum	Yes	21.2	21.1	-0.1
	No	17.3	17.1	-0.2
	All	18.7	18.8	+0.1

Income Tax

Likewise, it is possible to reduce the income tax rate in every tax bracket while increasing total income tax revenues. As inflation pushes more and more persons into the higher tax brackets, revenues go up even as tax rates go down. (Congress has already figured out how this works.)

Jeff Witmer
Associate Professor of Mathematics
Oberlin College
Oberlin, OH 44074

* * * * *

References

Joel Cohen, "An Uncertainty Principle in Demography and the Unisex Issue," *American Statistician*, vol. 40 (Feb. 1986), 32-39.

Clifford Wagner, "Simpson's Paradox in Real Life," *American Statistician*, vol. 36 (Feb. 1982), 46-48.

The Berkeley Example, but without using the term "Simpson's Paradox," appears on pp. 12-15 of *Statistics* by David Friedman, Robert Pisani, and Roger Purves, Norton, NY, 1978.

DECISION TREE FOR ERICSSON METHOD

While many people love to gamble, they also grasp at any opportunity to change the odds in their favor. Thus, it is not surprising that many married couples are ready to spend money in an attempt to change the odds on the sex of a planned child.

A variety of methods have been advertised to increase your chances of having a boy or a girl. Some of these methods merely try to influence the time of conception during the period the woman is fertile (one method is based on late conception favoring a boy, another on early conception -- so there isn't even agreement on this factor). Other methods place the father's sperm in a centrifuge to separate them by weight. None of these methods is generally accepted as scientifically and medically sound.

One of the methods most widely advertised has been the Ericsson method (based on centrifuging the sperm), which a few years ago claimed to promise 80% success if the parents want a boy, but admitted no influence on the probabilities if the parents want a girl. Consequently, the procedure is used only if the couple desires a boy.

Suppose that the Ericsson method did work as advertised (or that in the near future we find a method with these results). Let us look at a group of fertile couples, each of which plans to have precisely two children. These couples conform to our cultural preferences and are especially anxious to have at least one boy, although they would like one child of each sex. (Many couples want the first child to be a boy, since they think the oldest child often has a higher IQ and a better chance for career success.) Our couple learns about the Ericsson method and also that it is only 80% effective in realizing a boy when that is desired. So to increase the odds that at least one child will be a boy, they decide to use the Ericsson procedure on the first child. Is this a sensible decision?

Decision Tree

In order to answer this question, we decide to construct a *decision tree* in the following steps:

(1) The first decision the couple must make is whether to use the Ericsson Method for the first child. To indicate a decision is required, we draw a square box, with the departing arrows labelled to show the options.

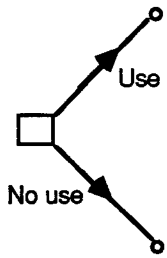


Fig. 1 Initial Decision.

(2) At each of the circles in Fig. 1, we know the resulting probabilities for a girl and a boy as the first child (that is, G_1 or B_1). If the Ericsson Method is used, the probability of B_1 is 8/10.

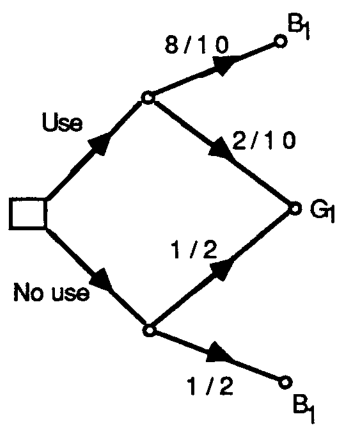


Fig. 2 Results of initial decision with probabilities shown.

(3) Now we turn to the second decision. If we are at B_1, we certainly do not use the Ericsson Method. In this case, the decision is obvious (we show the square decision box filled in). When we add the second decision, the tree now has the form:

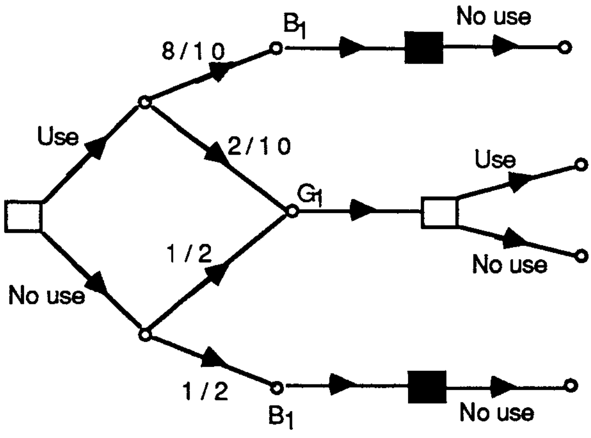


Fig. 3 Both decisions are in the tree.

(4) Next we show all the possible outcomes. There are four possible outcomes: B_1 B_2 (two boys), B_1 G_2, G_1 B_2, and G_1 G_2. The tree in Fig. 4 seems to indicate seven outcomes, but three identical pairs occur. The tree diagram is less messy and easier to interpret if we avoid lines crossing one another.

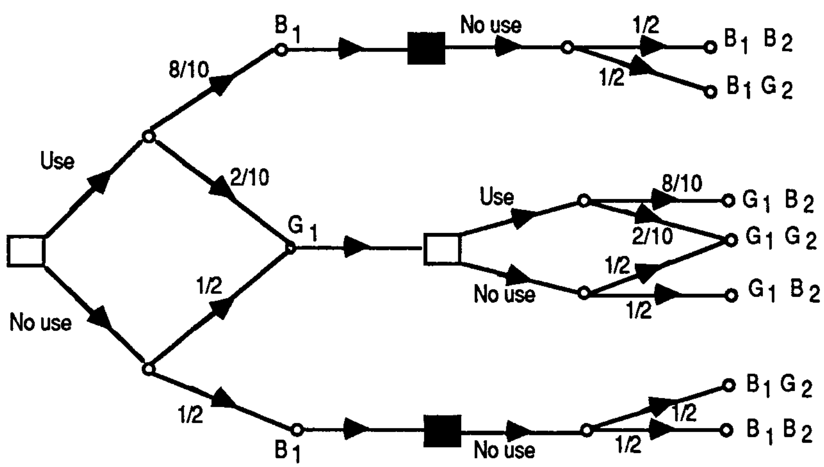


Fig. 4 Outcomes are included.

(5) Finally we have to pick *values* for each of the four possible outcomes. Usually in decision analysis, we pick

100 for the most favorable outcome (here B_1 G_2)
0 for the least favorable outcome (G_1 G_2)

The value of zero does not mean that two girls are of no value to the couple, but only that this is the least desirable outcome.

The choices of values for B_1 B_2 and G_1 B_2 is difficult and commonly subjective. In Fig. 5, we show the values 50 and 60, respectively, but any particular couple might select very different numbers.

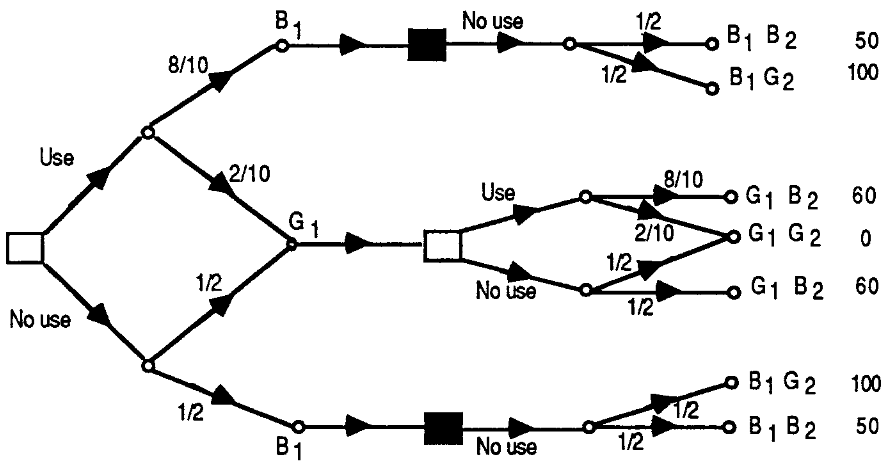


Fig. 5 Decision tree complete with values included.

Solution or optimization

With the decision tree now complete, we can find the optimum first and second decisions. We look first at the second decision (we often work optimization problems back-to-front -- from the last decision steadily back toward the first),

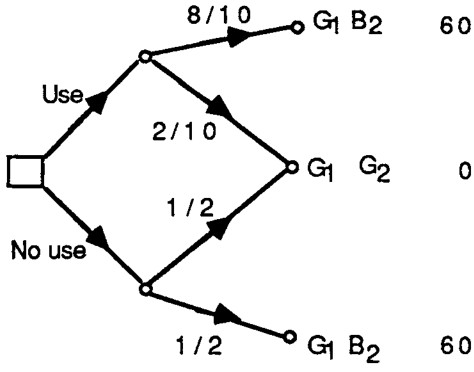


Fig. 6

The no-use option gives a value

$$\frac{1}{2}(0) + \frac{1}{2}(60) = 30$$

The Use option gives a value

$$\frac{8}{10}(60) + \frac{2}{10}(0) = 48$$

So we clearly should decide on Use, and if we do this, the

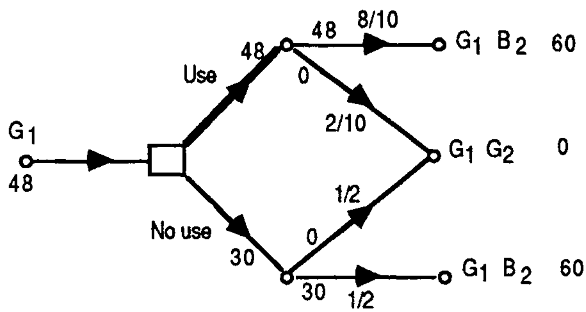


Fig. 7

value of the G_1 node is 48.

Then we can work back to the first decision. The value of the B_1 node is

$$\frac{1}{2}(50) + \frac{1}{2}(100) = 75$$

Thus, the value of the Use option in the first decision is

$$\frac{8}{10}(75) + \frac{2}{10}(48) = 69.6$$

The value of the No Use option in the first decision is

$$\frac{1}{2}(48) + \frac{1}{2}(75) = 61.5$$

and our best strategy is:

Use EM for first Child
If B_1, do not use EM for second
If G_1, use EM for second

To complete the study, we now need to investigate how the:

(1) Optimum strategy changes if our values in (5) change

(2) Optimum strategy changes if the probabilities with the EM change (what if the advertising literature is wrong)

(3) Gains expected from the use of the EM compare to the costs and risks (e.g., the expected value would be 52.5 if the couple had never heard of the EM. Is the increase to a maximum of 69.6 significant?)

A decision-tree problem

Another well-known problem is the Truel problem, used in an advanced management program. "Truel" is a word coined to describe a three-sided duel. Three competitors (A, B, and C) are at the apices of an equilateral triangle. Each has a pistol with two bullets.

They all know that they are very different in the quality of marksmanship:

- C hits whatever he aims at with certainty (a probability of 1)
- B has only 8/10 probability of hitting a target
- A hits only 6/10 of the time

To make the truel fair, A shoots one bullet first, then B (if he is still alive), then C, then A, B, and C their second bullets, in order.

There is one other condition: The only success lies in being the sole survivor. If two players survive, the value to each is the same as when all three survive. Thus, if all three are alive when C shoots C_2 (his second bullet), he does not choose one of the others to kill; he simply shoots in another direction.

(a) Which player would you like to be: A or B or C? Most people tend to pick B, some select C, and a few A. Actually the only rational choice is A.

(b) Construct the decision tree and determine each player's optimum strategy. Assign values to the outcomes and find the expected value for each player if everyone plays optimally.

This example is often cited as an analog of three companies in competition.

Detailed notes on the problem and its solution are available from the author of this monograph.

A PRODUCTION MANAGER'S DECISION PROBLEM

ABC Manufacturing Company fabricates a sensitive metal part that is assembled into the finished product elsewhere. Tolerances are agreed upon in the contract signed by ABC, but problems with varying quality of raw materials, machine stability, and inevitable human error can lead to flawed parts. Were it merely a matter of proper dimensions for the fabricated part, it would be a simple matter for ABC to check each part before shipment to the customer. But the part must also meet certain standards with respect to less easily measured features, such as quality of welds and uniformity of heat transmission from one end of the part to the other.

Under current conditions, 20 percent of all parts produced by ABC are unsatisfactory for one reason or another. Such a flawed part, if shipped to a customer, will be discovered only after it is used in the assembly. The entire assembly will have to be broken down, leaving an unhappy customer. Recognizing this, ABC's contract provides for a generous cash payment to the customer, not only to compensate for extra costs and related expenses, but also to indicate ABC's concern and goodwill. Moreover, at no additional charge (i.e., beyond the cost of the original part), ABC supplies the customer with a replacement part from a small inventory of parts that are sure to be satisfactory in all respects. The total cost to ABC of these consequences of shipping a defective part is estimated to be $400.

The manager could decide to make a careful inspection of each part before shipment in order to eliminate flawed parts and thus guarantee that any part delivered to a customer will be satisfactory. Such an inspection costs $100 per part.

Instead of such an inspection, there is an alternative test, relatively easy and inexpensive, that can be carried out on each part. This test costs only $20 per part. It is, however, far from a perfect diagnostic test. In fact, the test correctly diagnoses the part only 80 percent of the time; it is wrong 10 percent of the time; it is inconclusive and gives no verdict about the part's condition the remaining 10 percent of the time it is used.

What should the manager do?

The first task is to structure the problem by constructing a suitable *decision tree*. In Fig. 1, the left-most square #1 represents a *decision node* from which three paths emerge, one for each of the three possible actions to be taken: to deliver the manufactured part (without inspection or testing), to inspect it first, or to test it first.

Let us follow each of these path in turn. After delivery, chance determines whether the customer receives a flawed or satisfactory part. Note the circle #9 representing the *chance node* with the two possible consequences used as labels for the paths emanating from this node. At the terminal point of each path we write the total cost in dollars corresponding to that path (Fig. 2).

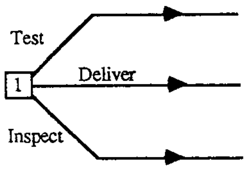


Fig. 1 First decision

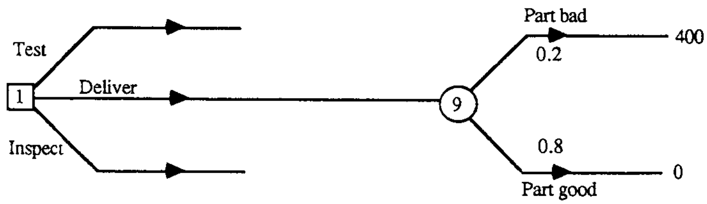


Fig. 2 Outcomes if decision is Deliver

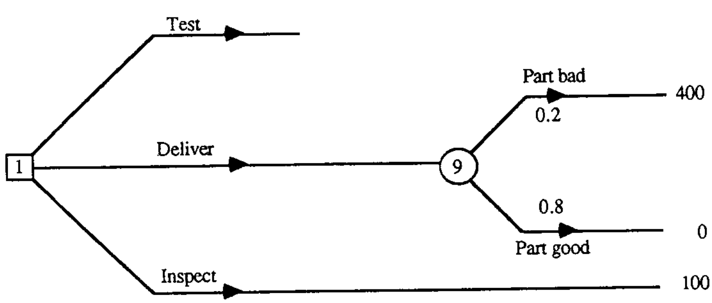


Fig. 3 Outcomes if decision is to Deliver or Inspect

If we choose the Inspect path at decision node #1, then we know all is well, and have only to write the cost of the inspection, $100, at the end of the path (Fig. 3).

Let us now follow the Test option at decision node #1. Chance then determines the outcome of the test, as is indicated by the three possible paths at chance node #2. If the test says the item is flawed, then we face another decision: to Deliver or to Inspect. These options are indicated by the two paths emerging at decision node #3 (Fig. 4). If we follow the Deliver option, then at chance node #6 we show the two possible outcomes: the item turns out to be flawed or it turns out to be satisfactory. The total cost of $420 is placed at the end of the topmost path in the decision tree since the test costs $20 and delivery of a flawed item costs an additional $400.

In this way, we continue and complete the decision tree (Fig. 5).

We have yet to determine the probabilities of the outcomes at each of the chance nodes in the tree. To do this, we construct the following joint frequency table:

Test Result

		Says Bad	Says Nothing	Says Good	Row Total
Nature of Product	Bad				
	Good				
	Column Total				100

We start by assuming 100 items (shown above in the lower right corner). Since the production process yields 20% flawed items, we can fill in the rightmost column

Test Result

		Says Bad	Says Nothing	Says Good	Row Total
Nature of Product	Bad				20
	Good				80
	Column Total				100

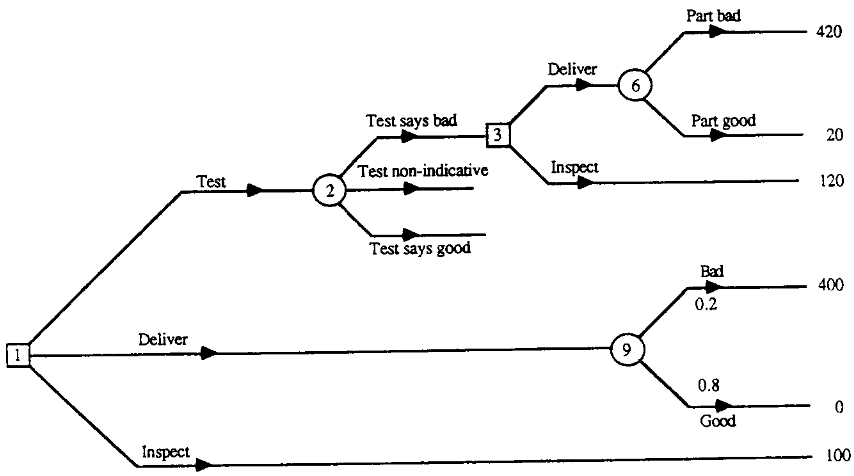


Fig. 4 Three possible results of the test

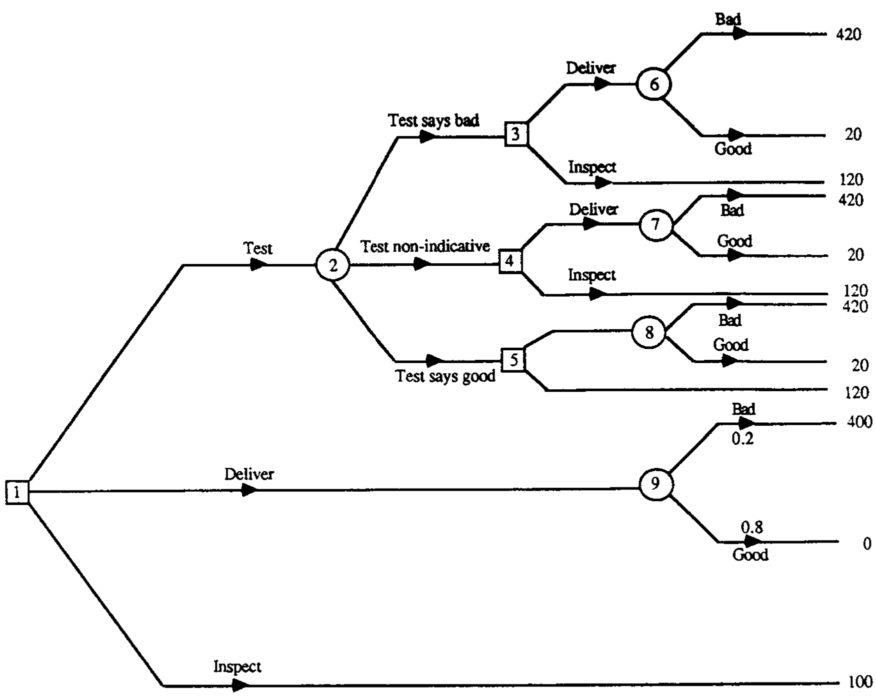


Fig. 5 Complete decision tree before probabilities evaluated

Of the 20 flawed items, the test correctly diagnosis 80% or 16, gives no indication for 10% or 2, and gives a false positive for 10% or 2.

Test Result

Nature of Product		Says Bad	Says Nothing	Says Good	Row Total
	Bad	16	2	2	20
	Good				80
	Column Total				100

We have assumed the percentages of the test are the same when the part is good, so we can complete the second row of numbers:

Test Result

Nature of Product		Says Bad	Says Nothing	Says Good	Row Total
	Bad	16	2	2	20
	Good	8	8	64	80
	Column Total				100

Finally, we can complete the table simply by adding the columns:

Test Result

Nature of Product		Says Bad	Says Nothing	Says Good	Row Total
	Bad	16	2	2	20
	Good	8	8	64	80
	Column Total	24	10	66	100

From the joint frequency table we are able to determine all the probabilities required for the decision tree, as demonstrated in the following listing:

Event	Probability
Test says product is flawed.	24/100 = .24
Test gives no verdict.	10/100 = .10
Test says product is not flawed.	66/100 = .66
Product is flawed, given that the test says product is flawed.	16/24 = 2/3
Product is flawed, given that the test gives no verdict.	2/10 = 1/5
Product is flawed, given that the test says product is okay.	2/66 = 1/33
Product is flawed.	20/100 = .20

Now that the decision tree is complete, we can turn to the determination of the production manager's best strategy. What does "best" mean? Let us determine for each possible strategy, the corresponding expected (or average or mean) cost. The best strategy will be the one with the smallest expected cost. We move backward from right to left in the tree. For example, at chance node #6 we compute the expected cost as the weighted average of the two possible costs, each weighted by the probability with which it occurs:

$$420\ (2/3)\ +\ 20\ (1/3)\ =\ 287,\ \text{approx.}$$

This expected cost appears in the oval above chance node #6. Backing up to decision node #3, we compare the choice of Deliver with that of Inspect by comparing their expected costs, $287 and $120. The smaller value is preferred and we now know that if we ever arrive at decision node #3, the better decision is Inspect. The two short lines blocking the Deliver path indicate that this path is not to be followed. Continuing in this way, we obtain the various expected costs shown in ovals above chance nodes #7, #8, and #9. At both decision nodes #4 and #5, the better choice is seen to be the Deliver path. At chance node #2, the expected cost is then determined as follows:

$$120\ (.24)\ +\ 100\ (.10)\ +\ 32\ (.66)\ =\ 59.92,$$

rounded to 60 and placed in the oval at chance node #2. Moving backward to decision node #1, we compare Test (mean cost $60), Deliver (mean cost $80), and Inspect (mean cost $100). The preferred choice is Test. In this way, following the so-called "averaging out and folding back" technique, we identify the *manager's optimal strategy*:

> Test each item. If the test says the item is flawed, then inspect. But if the test gives no verdict or says the item is not flawed, then deliver the item (without inspection). The expected cost of using this optimal strategy is $60 per item.

The path corresponding to this optimal strategy is outlined by heavier lines in the decision tree of Fig. 6.

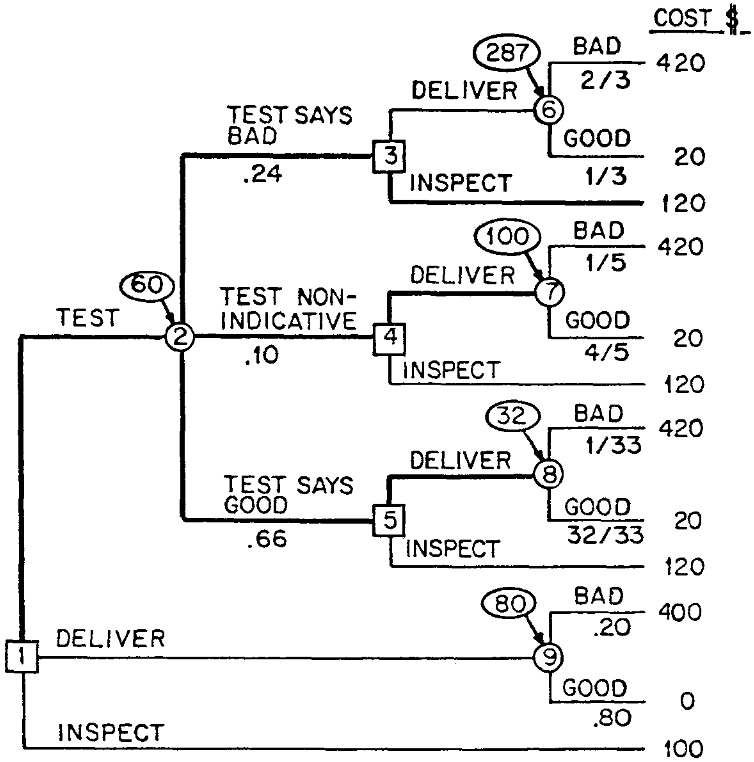


Fig. 6 Production Manager's Decision Tree

The following books on decision analysis are among many available for further study. They contain numerous types of applied problems, from medicine, business, and other fields, involving decision-making under uncertainty. All are introductory and intended for beginners. The required mathematical ideas and techniques from probability theory (to measure uncertain events) and from utility theory (to measure personal preferences and attitudes toward risk) are included in these texts.

Behn, Robert D. and Vaupel, James W., *Quick Analysis for Busy Decision Makers*, Basic Books, New York, 1982.

Holloway, Charles A., *Decision Making Under Uncertainty: Models and Choices*, Prentice-Hall, Englewood Cliffs, 1979.

Raiffa, Howard, *Decision Analysis (Introductory Lectures on Choices Under Uncertainty)*, Addison-Wesley, Reading, 1968.

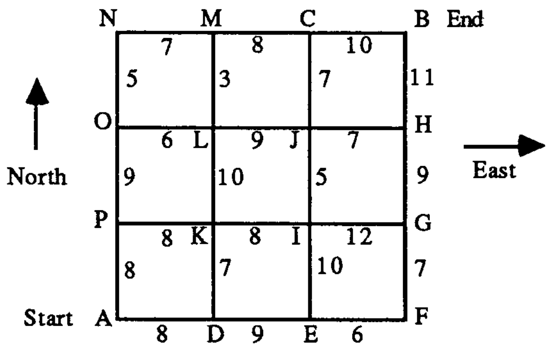
Weinstein, Milton C. and Fineberg, Harvey V., *Clinical Decision Analysis*, W. B. Saunders, Philadelphia, 1980.

DYNAMIC PROGRAMMING

Problem

In a central city, a serious accident occurs at location B. The nearest ambulance is located at point A. Since survival rates depend critically on the speed in receiving medical help, we want to find the minimum-time route from A to B.

Fortunately, the city has sensors in the pavement on each block and continually measures the traffic flow to estimate the time needed to travel each block. For example, travel from C to B takes 10 units of time.



How should we route the ambulance? What is the minimum-time path? (This routing scheme is used for ambulances in Tokyo.)

Approach

In many optimization problems like this, the critical step is finding the right approach. A problem may be horrendously difficult with the wrong approach, almost trivial with the right approach.

In this example, the wrong approach is simply to list all possible routes from A to B, find the total time for each, and finally look through the list of total times to find the smallest.

Actually in the simple example above, this is not too bad. There are only 20 possible routes (A D E F G H B is one, A D E I G H B another, and so on). If we had an example more complex than this, however, the number of routes rises sharply: a grid 20 blocks by 20 blocks has 137,846,528,820 different routes.

A much better approach involves a technique called *dynamic programming*. There are two ideas involved:

(1) We work from the end backward (back-to-front).

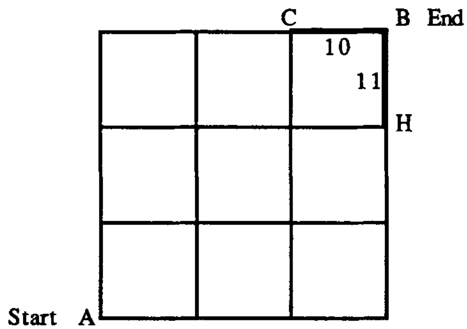
(2) Once we have found the best route from any point (say J) to B, we will always follow that route if we reach that point. This characteristic has the fancy name, *principle of optimality.*

Dynamic Programming Solution

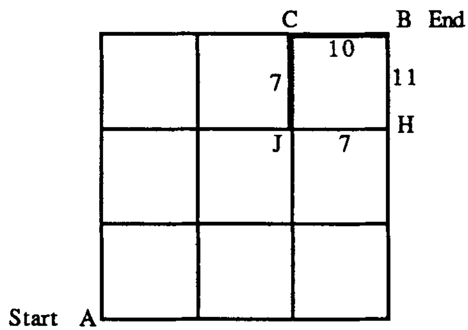
For the specific grid shown above, let's find the optimum route. Using dynamic programming, we proceed as follows:

(1) We start from B and work backward. From C, the best (and only) route is to the east and time is 10. We add the heavy line shown below with the time.

(2) From H, we can only go north and time is 11.



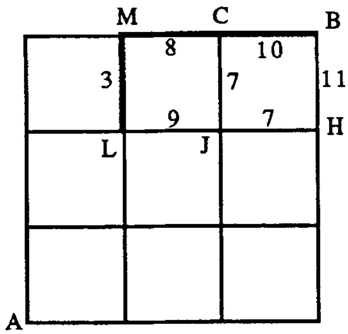
(3) From J, we have two choices: north for 7 to C, then 10 to B; or east for 7 to H, then 11 more to 18 to B. Obviously, if we are at J, we should go north.



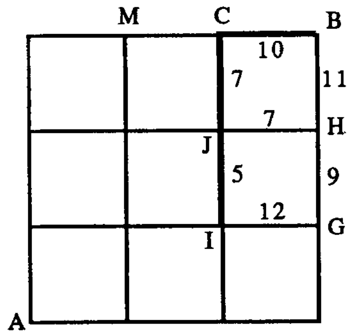
Optimum path from point J to B

(4) We turn to the next ring (M G L I K in that sequence). From M to B there is no choice, and time is 18; from G to B time is 20.

(5) We now find the optimum path from L. We can go north to M (3 + 18) or east to J (9 + 17), so clearly north is better. Similarly, we find that from I we should move north.

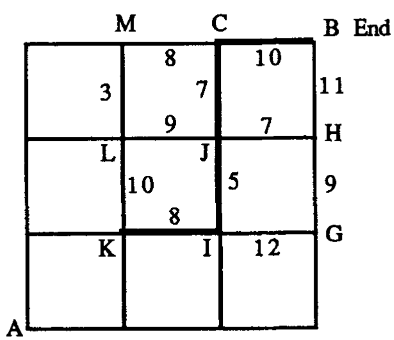


Optimum path from point L to B.



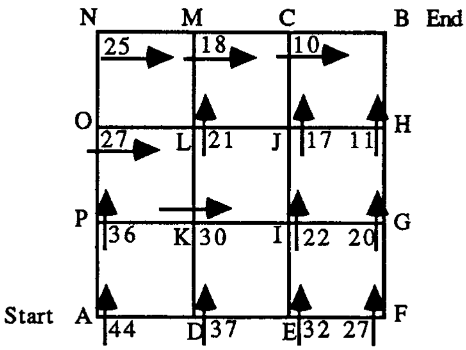
Optimum path from point I to B.

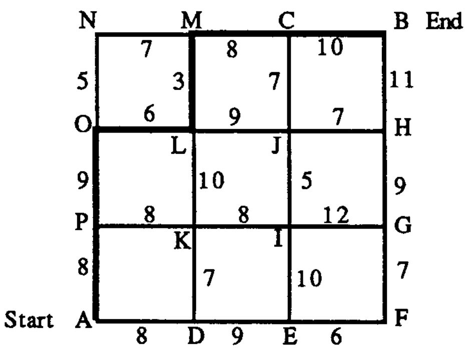
(6) We complete this ring by solving at K.



Optimum path from point K to B.

(7) We solve in sequence at N, F, O, E, P, D, and finally A.





The optimum path from A to B. The total travel requires 44 units of time.

Features of Solution

The solution takes appreciably more time to write than to do. There are 15 intersections where a decision must be made, but in six of these we have no choice (the three on the top and three on the right side). So there are really only nine decisions to be made: 3 x 3 for a three block by three block grid, so a grid 20 blocks on a side would require only 400 decisions.

Furthermore, each of these decisions is *binary*: a choice between two options, north or east. In each case, we choose the smaller of two numbers -- a trivial task. Thus, the solution of the original problem reduces to nine simple choices in our example.

This is the dominating feature of dynamic programming: the reduction of a problem involving a sequence of interdependent, complex decisions into a sequence of simple decisions.

Other Applications of Dynamic Programming

We can illustrate dynamic programming with a variety of other applications:

(1) With the same grid used above, we might seek the route requiring the maximum time (for the unethical taxi driver with a passenger obviously unfamiliar with the city).

(2) A power company wants to install a feeder cable from the generating station (A) to a new factory (B). The numbers on each block represent the cost of using that particular block.

(3) In a trip to the moon, the guidance/navigation system on the vehicle determines where the vehicle is. Dynamic programming allows determination of the strategy to reach an ultimate destination with a minimum consumption of fuel.

Origin of Dynamic Programming

While the two elements of dynamic programming (back-to-front approach and principle of optimality) were described in economics studies early in this century, Richard Bellman (then at the Rand Corporation and later at the University of Southern California) developed the modern theory in the 1950s. He applied the optimization approach to an enormous variety of problems, ranging from the guidance of space vehicles to the controlled release of drugs in medicine.

* * * * *

Expanded notes on dynamic programming (including additional examples, references, and material on Richard Bellman) are available from the author of this monograph.

ACKNOWLEDGMENT

The New Liberal Arts Program of the Alfred P. Sloan Foundation has involved a remarkable group of faculty from about 50 colleges and universities. Many of these individuals have contributed important ideas to the teaching of quantitative reasoning in addition to the credits given for specific examples. Every teacher repeatedly adopts innovative examples from discussion with friends in our profession, and often we forget the source; we apologize if we have overlooked such cases.

In this monograph, particular thanks go to Dr. Samuel Goldberg, Program Officer at the Sloan Foundation, who has guided and inspired NLA faculty and made notable contributions throughout this collection of examples. Any weaknesses in the monograph are clearly the responsibility of the author alone. In particular, the author is an engineer, not an applied mathematician, so that many of the examples are presented without careful mathematical development.

The author is also indebted to his colleague, Dr. Marian Visich, Jr., who has team-taught much of this material and has worked with me in every phase of the development.

Finally, Mrs. Carol Galdi has done all layout and formatting work, and my wife, Doris, has worked with me in every phase.