AN ABSTRACT OF THE THESIS OF

LARRY ELWYN SHIRLAND for the DOCTOR OF PHILOSOPHY
(Name) (Degree)

in INDUSTRIAL ENGINEERING presented on March 17, 1972
(Major) (Date)

Title: MULTI-FACTOR INFORMATION THEORY MODELS AND THEIR INDUSTRIAL APPLICATIONS

Redacted for Privacy

Abstract approved: ______________________
Dr. M. S. Inoue

This dissertation treats information theory and its applications to the general area of decision making. Specifically, three areas are covered; (1) information theory applied to Bayesian analysis, (2) estimation using multi-factor information channel models, and (3) information theory applied to Markov chain analysis.

A major portion of this dissertation concerns the concept of conditional information which occurs as the result of the transmission of information for an experiment (Z) when the outcomes of another experiment (Y) are known. The gain in information is measured by computing the difference between the information transmitted when one set of values is known for an experiment, and another set obtained when certain experimental parameters are allowed to vary. When only one set of experimental results is available, the information gain is computed as the difference between the transmitted information

under the experimental conditions and the information transmitted assuming complete uncertainty. The latter is characterized by the condition which results when the events of the experiment are considered to be equally likely.

The information theory technique appears to be especially useful in the area of sampling. The cost of gathering information may be balanced against management's willingness to pay for the information in order to arrive at an optimal number of events to sample for a particular experiment.

Utilizing the concept of conditional information and information gain, estimates may be made by applying a multi-factor information channel analysis. In order to obtain the maximum amount of information from a sampling experiment, it may be desired to predict the strategies one should use. A case study is presented in which a research questionnaire was sent to prospective customers of several manufacturers of crushing and grinding equipment in an attempt to determine a particular company's standing with respect to its "image" and "progressiveness." The results of five specific questions were analyzed by the information theory approach in an attempt to predict the market shares for each of five companies. The information theory analysis showed that each of the five questions could be used independently as a market share predictor. This suggests that a person may subconsciously possess a pre-conceived opinion of a company which affects his answer to a specific question about that company.

A matrix method based on the work of Muroga (1959) is presented for solving multi-factor information channel problems. In order to solve a problem of this type it is easiest to first ignore the existence of non-positive solutions and solve the information maximization equations accordingly. If a non-positive solution occurs, one or more restrictions may then be imposed in order to force only positive values on the final solution. Non-positive solutions indicate that the maximum information gain occurs outside the realm of permissible values. The solution, then, involves maximizing the information gain while insuring that the probability of each event of the experiment is positive.

A multi-factor information theory analysis is applied to Markov chain problems in order to estimate at what point stochastic equilibrium occurs. This result is especially useful for computer simulations of Markov chains in which the equilibrium condition is of prime importance. By first employing the information theory analysis, the simulation may be started at or near stochastic equilibrium, thereby reducing the costs of unnecessary calculations during the transitional stages of the process. The information theory analysis shows that at least in some practical problems, stochastic equilibrium will not occur for a long period of time. In many applications, the transitional stages are of more interest than the steady-state conditions.

Current research points to several areas for further investigation. Models to allow for the heterogeneity among consumers, means to identify and quantify the important factors in a multi-factor information theory model, and learning models offer unique challenges for future research.

Also included in this dissertation is a computer program for solving any one-factor, two-factor or multi-factor information theory problem. A table of values for 1, 2, 3, or 4 levels of a one-factor model is also given.

Multi-Factor Information Theory Models and Their Industrial Applications

by

Larry Elwyn Shirland

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1972

APPROVED:

Redacted for Privacy

Associate Professor of Industrial Engineering

in charge of major

Redacted for Privacy

Head of Department of Industrial Engineering

Redacted for Privacy

Dean of Graduate School

Date thesis is presented March 17, 1972

Typed by Clover Redfern for Larry Elwyn Shirland

ACKNOWLEDGMENTS

I wish to express my sincere thanks to Dr. Michael S. Inoue for his advice and guidance in the preparation of this dissertation. His encouragement and criticisms are greatly appreciated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

MULTI-FACTOR INFORMATION THEORY MODELS AND THEIR INDUSTRIAL APPLICATIONS

I. INTRODUCTION

The Meaning of Information

The use of the term "information" in communication theory does not necessarily relate to what one does say but to what one could say. In other words, information measures one's freedom of choice in selecting a specific message out of all possible messages (Weaver, 1964). Also, it is not so much a property of an individual message, rather it is a property of the entire experimental situation which produces messages.

When evaluating alternatives using information theory, one has to consider not only the particular message sent but the set of all possible messages of which the chosen message is an element. The message source can be considered to consist of all the possible elements in a probabilistic experiment. These individual elements can be viewed as stimuli generating particular messages. For purposes of this discussion a discrete information source is assumed. Various experiments under consideration will consist of a finite number of elements or outcomes each having a definite probability of occurrence.

Justification for Using Logs to Base Two

Consider an experiment consisting of two equally likely outcomes such as in the tossing of a fair coin. On one toss of the coin there are two possibilities; "head" or "tail" each with probability one-half. This is the simplest experiment in which the outcome is uncertain. Before conducting the experiment we do not know for certain whether a head or tail will appear. This, the simplest case, has been chosen as a standard and is defined as transmitting one unit of information. If logarithms to base two are used, then the information transmitted by the two outcomes of the experiment would be:

$$H = \log_2 2 = 1$$

which does give a unit measure. Information of this type is defined as one "bit."

Now consider an experiment in which there is only one certain outcome. For example, suppose a card is drawn from a standard deck of playing cards from which all the black cards have been removed. We wish to determine the information transmitted by the result "a red card was drawn." There is only one outcome of this experiment; a red card occurs with certainty. The information transmitted, then, will be:

$$H = \log_2 1 = 0$$

Since the outcome of the experiment was a foregone conclusion, there is no information transmitted by the answer "a red card was drawn." Therefore, there is no uncertainty or equivocation concerning the outcome of an experiment and there has been no new information transmitted.

Given an experiment having two outcomes as a standard of measurement, one which consists of eight outcomes will transmit three times as much information. This result is not immediately obvious. However, consider tossing a fair coin three times in succession. The sample space of outcomes will consist of:

HHH
HHT
HTT
THH
TTH
THT
HTH
TTT

Since each toss is independent and each outcome is equally likely with probability one-eighth, the information transmitted by the combined experiment can be viewed as consisting of three separate elements or:

$$H(\text{experiment}) = H(\text{1st toss}) + (H(\text{2nd toss}) + H(\text{3rd toss})$$

so

$$H(\text{experiment}) = \log_2 2 + \log_2 2 + \log_2 2$$

and

$$H(\text{experiment}) = 1 + 1 + 1 = 3$$

Therefore, an experiment consisting of eight outcomes should transmit three bits of information. Indeed, $H = \log_2 8 = 3$ bits. A choice of logarithms to base two as a unit of measurement gives consistant results (Raisbeck, 1963). More rigorous proofs for the use of base two are given by Shannon (1949), Watanabe (1970) and others.

The information transmitted is usually expressed in the form of negative entropy.

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i$$

where H is the entropy in bits, p_i is the individual probability of each outcome of an experiment which must add to unity or $\sum_{i=1}^{n} p_i = 1$. For example, the entropy that results from tossing a coin once could have been determined from:

$$H = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}$$

$$H = -\frac{1}{2}(-1) - \frac{1}{2}(-1)$$

and

$$H = \frac{1}{2} + \frac{1}{2} = 1 \text{ bit}$$

Maximum Information and Information Gain

It may also be noted that the entropy of one bit is the maximum amount of information that can be transmitted by an experiment consisting of two possible outcomes. Figure 1 shows the change in entropy as the probability p of one outcome increases from 0 to 1.0 as the probability (1-p) of the other decreases from 1.0 to 0.

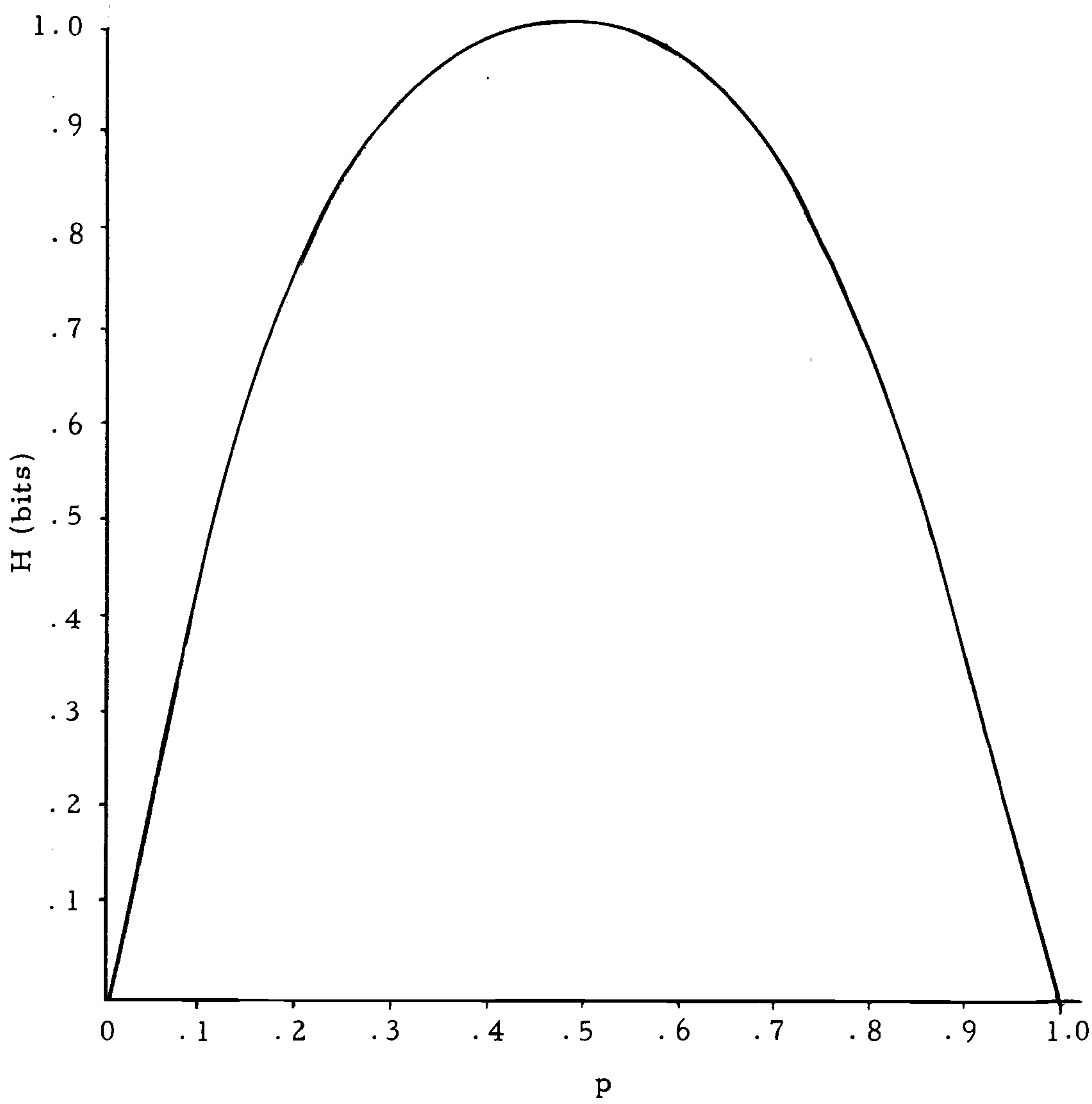

Figure 1. Entropy of a two outcome experiment.

It can be shown that for any experiment having n outcomes, information transmitted by the experiment will be a maximum when all outcomes are equally likely; this being the condition of greatest uncertainty or greatest equivocation. To illustrate, assume an experiment with a finite number of outcomes $X_1, X_2, X_3, \ldots, X_n$ having probabilities $p_1, p_2, p_3, \ldots, p_n$ respectively. The Lagrange Multiplier Technique may be employed to maximize the information subject to the restriction that $\sum_{i=1}^{n} p_i = 1$. In equation form:

$$[\max]H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i + \lambda(1 - \sum_{i=1}^{n} p_i) \qquad (1.0)$$

Since

$$\log_a b = \log_a e \log_e b$$

Then,

$$\log_2 p_i = (\log_2 e)(\log_e p_i)$$

Therefore, defining $K = \log_2 e = 1.4427$ Equation (1.0) may be written:

$$[\max]H(X) = -K \sum_{i=1}^{n} p_i \log_e p_i + \lambda(1 - \sum_{i=1}^{n} p_i)$$

Now, taking derivatives:

$$\frac{\partial[H(X)]}{\partial p_1} = -K\frac{p_1}{p_1} - K \log p_1 - \lambda = 0 \tag{1.1}$$

$$\frac{\partial[H(X)]}{\partial p_2} = -K\frac{p_2}{p_2} - K \log p_2 - \lambda = 0 \tag{1.2}$$

$$\vdots$$

$$\frac{\partial[H(X)]}{\partial p_n} = -K\frac{p_n}{p_n} - K \log p_n - \lambda = 0 \tag{1.3}$$

$$\frac{\partial[H(X)]}{\partial \lambda} = 1 - \sum_{i=1}^{n} p_i = 0 \tag{1.4}$$

Equations (1.1) through (1.3) may be solved to obtain:

$$p_i = e^{(-1-\lambda/K)}$$

Now, let

$$A = e^{(-1-\lambda/K)} \qquad i = 1, 2, \ldots, n$$

The individual probability values may now be calculated as:

$$p_1 = A$$

$$p_2 = A$$

$$\vdots$$

$$p_n = A$$

From Equation (1.4):

$$p_1 + p_2 + p_3 + \ldots + p_n = 1$$

Therefore:

$$A + A + A + \ldots + A = 1$$

or

$$nA = 1 \quad \text{and} \quad A = 1/n$$

so,

$$p_1 = p_2 = \ldots = p_n = 1/n$$

and information is maximized when all possible outcomes are equally likely. The equally likely concept also has an intrinsic intuitive appeal. If one knew absolutely nothing about the occurrence of various outcomes of an experiment, a logical choice would be to place equal weight on each. This condition represents one of maximum uncertainty. The point of maximum uncertainty will also be defined as zero information gain. Therefore, when we are able to place probabilities other than the equally likely conditions on outcomes in an experiment we will gain knowledge and experience an information gain. Information gain, then, will be defined as the difference between the information transmitted by the equally likely conditions and the information transmitted by a set of augmented probabilities based on our knowledge concerning the experiment. An equation for information gain, $G(X)$ may be written as:

$$G(X) = \log_2 n - \sum_{i=1}^{n} p_i \log_2 p_i$$

where n and p_i are as previously defined.

An Example

Shirland (1971) gave results of a survey to determine the airline ticket purchasing habits of college students. The following response frequencies were obtained upon asking students whether they would travel first class, tourist, or excursion given specific price levels for each.

First Class	10.2%
Tourist	30.8%
Excursion	59.0%

The information transmitted before the experiment is conducted, assuming complete ignorance is

$$H(X) = \log_2 3 = 1.585 \text{ bits}$$

The information gain then, would be calculated from:

$$G(X) = \log_2 3 - .102 \log_2 .102 - .308 \log_2 .308 - .590 \log_2 .590$$

$$G(X) = 1.585 - (.336 + .523 + .449)$$

or

$$G(X) = 1.585 - 1.308 = .244 \text{ bits}$$

which represented a 17.5% gain in information.

II. INFORMATION THEORY AND BAYESIAN DECISION ANALYSIS

Conditional Information

Consider an experiment X consisting of two sub-experiments Y and Z as shown in Figure 2.

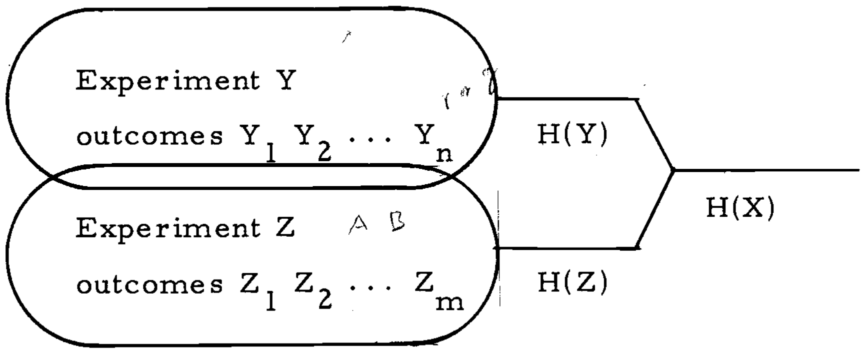

Figure 2. Information transmitted by two experiments.

If the outcomes of experiment Y and Z are independent, the information transmitted by the combined experiment X will be:

$$H(X) = H(Y) + H(Z)$$

which means that the information transmitted by the joint experiment is the sum of the individual subset experiments.

If the experiments Y and Z are not independent, then the information transmitted is given by:

$$H(X) \leq H(Y) + H(Z)$$

These results were proven by Kunisawa (1958), Shannon (1949) and others.

Considering the general case of two dependent experiments as shown in Figure 2, the joint probability of obtaining Y_i and Z_j may be expressed by $p(y_i, z_j)$ and the transmitted information by:

$$H(Y, Z) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(y_i, z_j) \log_2 p(y_i, z_j) \qquad (2.0)$$

The conditional probability distribution may be determined from the definition of conditional probability and the definition of Bayes Theorem.

$$p(z_j | y_i) = \frac{p(y_i z_j)}{p(y_i)} = \frac{p(z_j) p(y_i | z_j)}{\sum_{j=1}^{m} p(z_j) p(y_i | z_j)} \qquad (2.1)$$

Bayes formula gives the probability of an outcome Z_j given that outcome y_i of experiment Y has already been observed. In our information theory analysis of Bayesian decision problems, we will want to compute the uncertainty or equivocation after having learned of a result of experiment Y. The information transmitted by an outcome of this conditional analysis is:

$$H(Z_j | Y_i) = -\sum_{j=1}^{m} p(z_j | y_i) \log_2 p(z_j | y_i) \qquad (2.2)$$

and there will be exactly n terms of this type. In other words, we will have:

$$H(Z_j | Y_1) = - \sum_{j=1}^{m} p(z_j | y_1) \log_2 p(z_j | y_1)$$

$$H(Z_j | Y_2) = - \sum_{j=1}^{m} p(z_j | y_2) \log_2 p(z_j | y_2)$$

$$\vdots$$

$$H(Z_j | Y_n) = - \sum_{j=1}^{m} p(z_j | y_n) \log_2 p(z_j | y_n)$$

Since a particular $H(Z_j | Y_i)$ occurs only a fraction $p(y_i)$ of the time, the conditional information of the entire experiment may be expressed as a weighted sum of the information from the individual outcomes.

$$H(Z | Y) = - \sum_{i=1}^{n} \sum_{j=1}^{m} p(y_i) p(z_j | y_i) \log_2 p(z_j | y_i) \qquad (2.3)$$

The above equation may also be written as:

$$H(Z | Y) = - \sum_{i=1}^{n} \sum_{j=1}^{m} p(y_i, z_j) \log_2 p(z_j | y_i) \qquad (2.4)$$

since from the definition of conditional probability:

$$p(z_j | y_i) = \frac{p(y_i, z_j)}{p(y_i)} \qquad (2.5)$$

multiplying through by $p(y_i)$ gives:

$$p(y_i, z_j) = p(y_i)p(z_j | y_i) \tag{2.6}$$

which can be substituted into Equation (2.3) to obtain Equation (2.4).

Input/Output Information Tree

An input/output model may help to clarify calculations in an information theory problem. Figure 3 shows an input/output information tree which may be used to calculate the transmitted information $H(Z|Y)$ for a decision analysis problem. The tree is used as follows:

1. Information in, represented by the marginal probabilities $p(y_i)$ $i = 1, 2, \ldots, n$ are entered on dotted lines.
2. The information transmitted by various signals is calculated as $H(z_j | y_i) = p(z_j | y_i) \log_2 p(z_j | y_i)$.
3. The respective $H(z_j | y_i)$ are summed over j to determine $H(Z | y_i)$ for each $i = 1, 2, \ldots, n$.
4. The frequency of occurrence of each y_i is noted on a line terminating at the output node.
5. The information transmitted by the experiment is computed by summing $p(y_i)\, H(Z | y_i)\, f(y_i)$ over the various branches of the decision tree to get $H(Z | Y)$.

In this chapter, the comparative information gain between

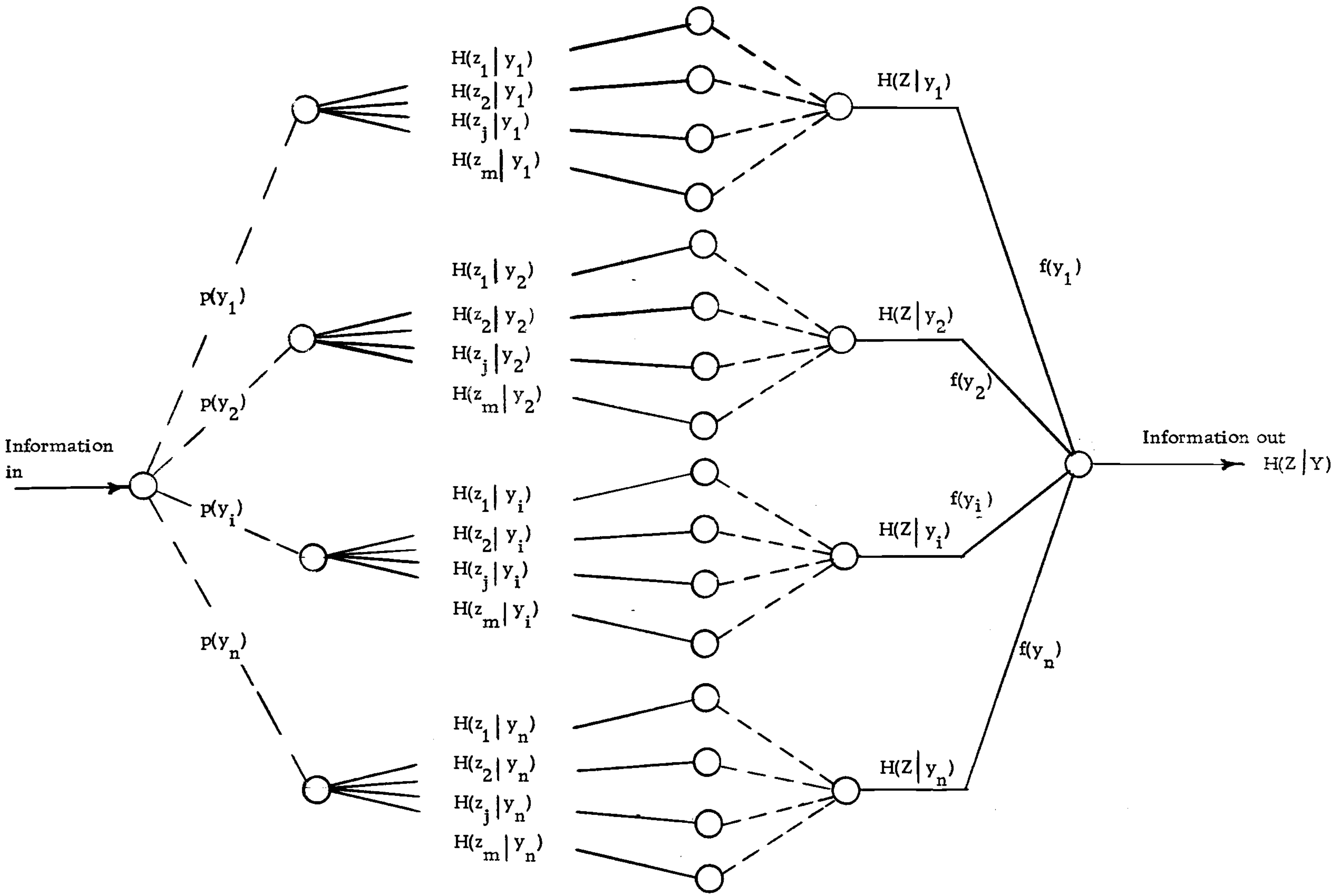

Figure 3. Input/output information tree.

experiments will be analyzed. In other words, the information gain due to additional information will be computed by considering experiments:

$$X \text{ composed of } Y \text{ and } Z$$

$$X' \text{ composed of } Y' \text{ and } Z'$$

The information gain will be computed from:

$$G(X) = H(X) = H(X') \tag{2.7}$$

Many decision making problems involving Bayesian analysis may also be solved using information theory. Instead of simply computing conditional probabilities as is done in Bayesian analysis, information theory allows one to view the entire experiment with respect to total information gain. The information theory calculations result in a sensitivity analysis answering questions such as: "Given slight changes in any or all probabilities in an experiment, what is being gained as far as information content is concerned?" It may even be possible to treat information theory in decision problems with respect to costs or profits. For example, a decision may be made concerning the cost value of information gain which can be balanced against the cost of collecting additional information. In this manner, an optimal point may be determined at which the cost of additional information is off-set by the amount that one is willing to pay for that information.

Two examples will serve to illustrate the decision analysis using information theory. The examples are drawn respectively from Halter (1971) and Riggs (1968).

Urn Problem

Formulation of the Problem

Two urns A and B are given. Urn A contains 4 red and 2 green balls and urn B contains 3 red and 7 green balls.

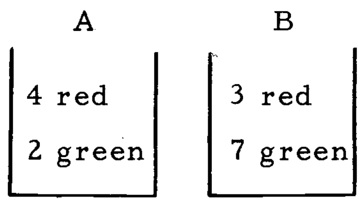

Figure 4. Distribution of balls for urn example.

Three sampling procedures are available:

a. Randomly choose an urn, draw one ball and examine it.

b. Randomly choose an urn, draw one ball, examine it, replace it and draw another.

c. Randomly choose an urn, draw two balls without replacement and examine them.

Decision Procedure

An outcome of the experiment is observed, but it is not known from which urn this particular outcome has been chosen. Which sampling procedure (a, b or c) more accurately predicts from which urn the sample has been selected?

Sampling Procedure (a)

Since we are choosing a ball from an urn in a random manner:

$$p(A) = p(B) = 1/2$$

The sample space for this experiment consists of two outcomes:

1. A green ball is drawn.

2. A red ball is drawn.

We would like, then, to determine the information transmitted by this experiment. Equation (2.3) may be used once we have calculated the respective probabilities. For this example, Equation (2.3) will be:

$$H(Z|Y) = p(g)H(g) + p(r)H(r)$$

which will be computed from:

$$\begin{aligned} H(Z|Y) = &-p(g)[p(A|g) \log p(A|g) + p(B|g) \log p(B|g)] \\ &-p(r)[p(A|r) \log p(A|r) + p(B|r) \log p(B|r)] \qquad (2.8) \end{aligned}$$

Definitions of the terms used above are as follows:

H(g) - Information transmitted by observing a green ball.

H(r) - Information transmitted by observing a red ball.

$p(g)$ - Probability of drawing a green ball.

$p(r)$ - Probability of drawing a red ball.

$p(A|g)$ - Probability that we were sampling from urn A given that we observed a green ball.

$p(B|g)$ - Probability that we were sampling from urn B given that we observed a green ball.

$p(A|r)$ - Probability that we were sampling from urn A given that a red ball was observed.

$p(B|r)$ - Probability that we were sampling from urn B given that a red ball was observed.

Bayes formula may be used to calculate $p(A|g)$ and $p(A|r)$. $p(B|g)$ may be calculated as: $1 - p(A|g)$ and $p(B|r) = 1 - p(A|r)$.

$$p(A|g) = \frac{p(A)p(g|A)}{p(A)p(g|A) + p(B)p(g|B)}$$

From the statement of the problem:

$$p(g|A) = 1/3$$

$$p(r|A) = 2/3$$

$$p(g|B) = 7/10$$

$$p(r|B) = 3/10$$

Therefore:

$$p(A|g) = \frac{(1/2)(1/3)}{(1/2)(1/3)+(1/2)(7/10)} = \frac{0.333}{1.033} = 0.321$$

$$p(B|g) = 1 - 0.321 = 0.679$$

$$p(A|r) = \frac{(1/2)(2/3)}{(1/2)(2/3)+(1/2)(3/10)} = \frac{0.667}{.967} = .690$$

$$p(B|r) = 1 - .690 = .310$$

Now, the probability of drawing a green ball $p(g)$ will be a weighted sum of all the ways in which a green ball can be drawn or:

$$p(g) = p(A)p(g|A) + p(B)p(g|B)$$

so

$$p(g) = (1/2)(1/3) + (1/2)(7/10) = .517$$

and

$$p(r) = 1 - p(g)$$

$$p(r) = 1 - .517 = .483$$

In tabular form, these results may be written: (z_j refers to urn A and urn B for $j = 1, 2$; y_i refers to a red or green ball for $i = 1, 2$).

Table 1. Bayesian probabilities for urn example--procedure (a).

z_j \ y_i	r	g
A	.690	.321
B	.310	.679
$p(y_i)$	.483	.517

The information transmitted can now be calculated from Equation (2.8).

$$H(Z|Y) = .517[.321 \log .321 + .679 \log .679]$$

$$+ .483[.690 \log .690 + .310 \log .310]$$

$$H(Z|Y) = .517\ (.9055) + .483\ (.8932)$$

$$H(Z|Y) = .8400 \text{ bits}$$

If we knew nothing concerning the outcomes; (1) we are drawing from urn A or (2) we are drawing from urn B, we would assign equal probabilities of .500 to both which would result in an information transmission of 1 bit. By using sampling procedure (a) we transmitted .8400 bit for a gain of .1600 bit. Sampling procedure (a), then, contributes 16% gain in information over complete uncertainty. If either sampling procedures (b) or (c) result in an information gain of more than 16% we may safely assume that we are even less uncertain about the urn from which we have been sampling. Therefore, the sampling procedure which results in the greatest amount of information is the best procedure to use.

Sampling Procedure (b)

Randomly choose an urn, draw one ball, examine it, replace it and draw another.

The sample space for this experiment consists of four outcomes:

1. (gg) green on both draws.

2. (gr) green first, then red.

3. (rg) red first, then green.

4. (rr) red on both draws.

The information transmitted by this experiment will be:

$$H(Z|Y) = p(gg)H(gg) + p(rr)H(rr) + 2p(gr)H(gr)$$

Since $p(rg) = p(gr)$, then $H(gr) = H(rg)$ which accounts for the factor of 2 in the above equation.

As in procedure (a) the respective conditional probabilities can be obtained from Bayes formula. The individual probabilities of each outcome can be computed as weighted sums of all the ways to achieve that outcome.

The results of the probability calculations are given in Table 2.

Table 2. Information gain analysis for urn example--procedure (b).

z_j \ y_i	gg	gr	rg	rr
A	.185	.514	.514	.832
B	.815	.486	.486	.168
H(i) bits	.6908	.9994	.9994	.6531
$p(y_i)$	.301	.216	.216	.268

The transmitted information will be:

$$H(Z|Y) = .301\ (.6908) + .216\ (2)(.9994) + .268\ (.6531)$$

$$H(Z|Y) = .8150 \text{ bits}$$

The information gain of procedure (b) over procedure (a), then, is:

$$G(X_{b-a}) = .8400 - .8150 = .0250 \text{ bit}$$

which represents a 3.08% increase in information gain over procedure (a).

We may conclude that drawing a second ball results in a gain in information of 3.08% or that we have learned 3.08% more about the experiment and are now in a better position to make predictions about from which urn we had been sampling given a particular outcome. Procedure (b) also contributes an 18.50% gain in information over complete uncertainty compared with 16% for procedure (a).

Sampling Procedure (c)

Randomly choose an urn, draw 2 balls and examine them.

The sample space for this experiment will be the same as in procedure (b). Since the first ball drawn will not be replaced, however, the probability distribution of outcomes will be hypergeometric rather than bionomial as in procedure (b).

The results of the calculations are given in Table 3.

Table 3. Information gain analysis for urn example--procedure (c).

z_j \ y_i	gg	gr	rg	rr
A	.125	.532	.532	.853
B	.875	.468	.468	.147
H(i) bits	.5436	.9970	.9970	.6023
p(i)	.267	.249	.249	.233

Transmitted information

$$H(Z|Y) = .267(.5436) + .249(2)(.9970) = .233(.6023)$$

$$H(Z|Y) = .7097 \text{ bits}$$

The information gain over procedure (a) is:

$$G(X_{c-a}) = .8400 - .7097 = .1393 \text{ bit}$$

which is a 15.5% gain.

$$G(X_{c-b}) = .8150 - .7097 = .1053 \text{ bit}$$

which is a 12.9% gain. The gain in information over complete uncertainty is 29.03%.

In answer to the question posed in the statement of the problem, our analysis shows that procedure (c), sampling two balls without replacement is the best technique to use since it results in our gaining the most information concerning the combined experiment. Figures 5, 6 and 7 show the calculations for the three sampling procedures in

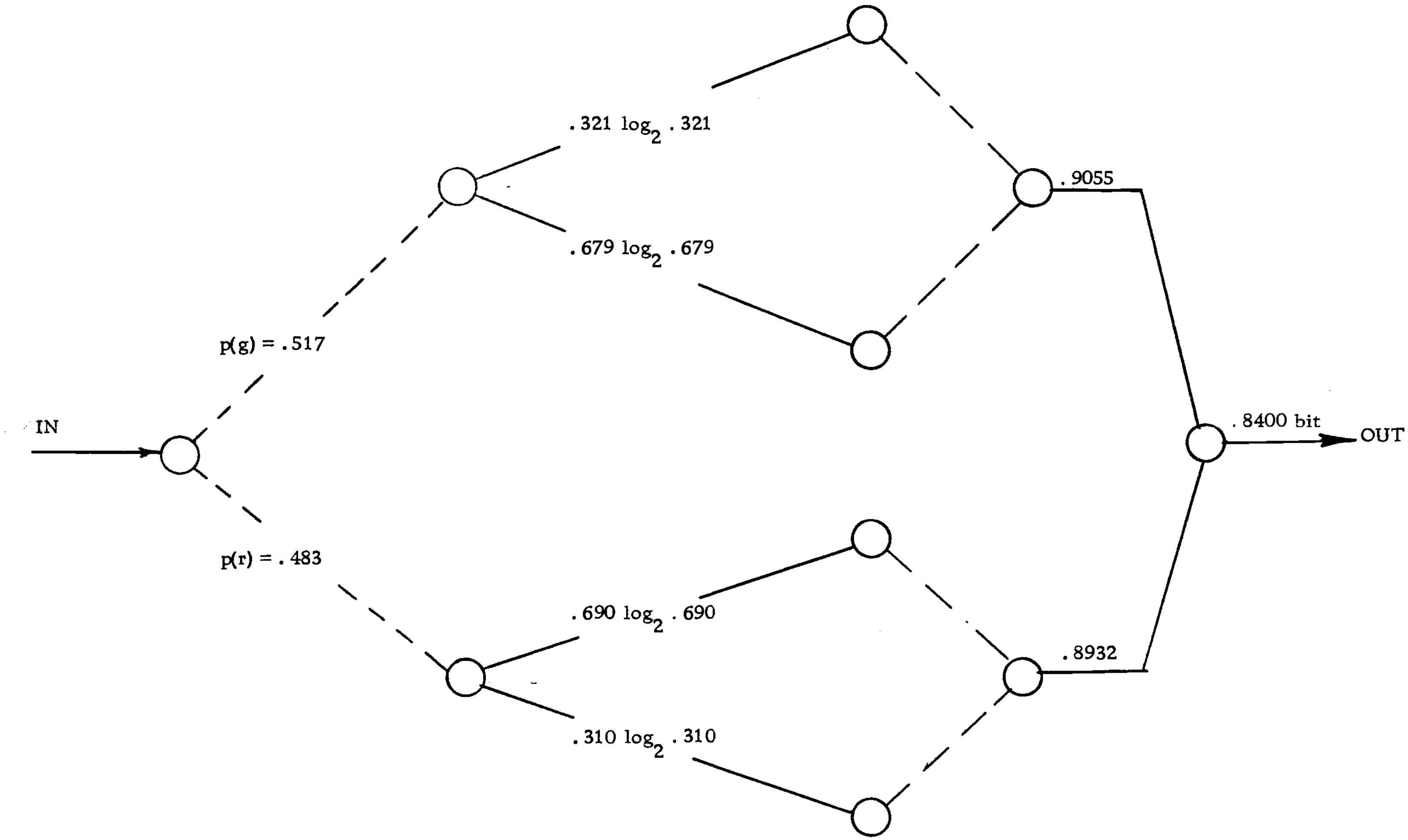

Figure 5. Information tree for sampling procedure (a).

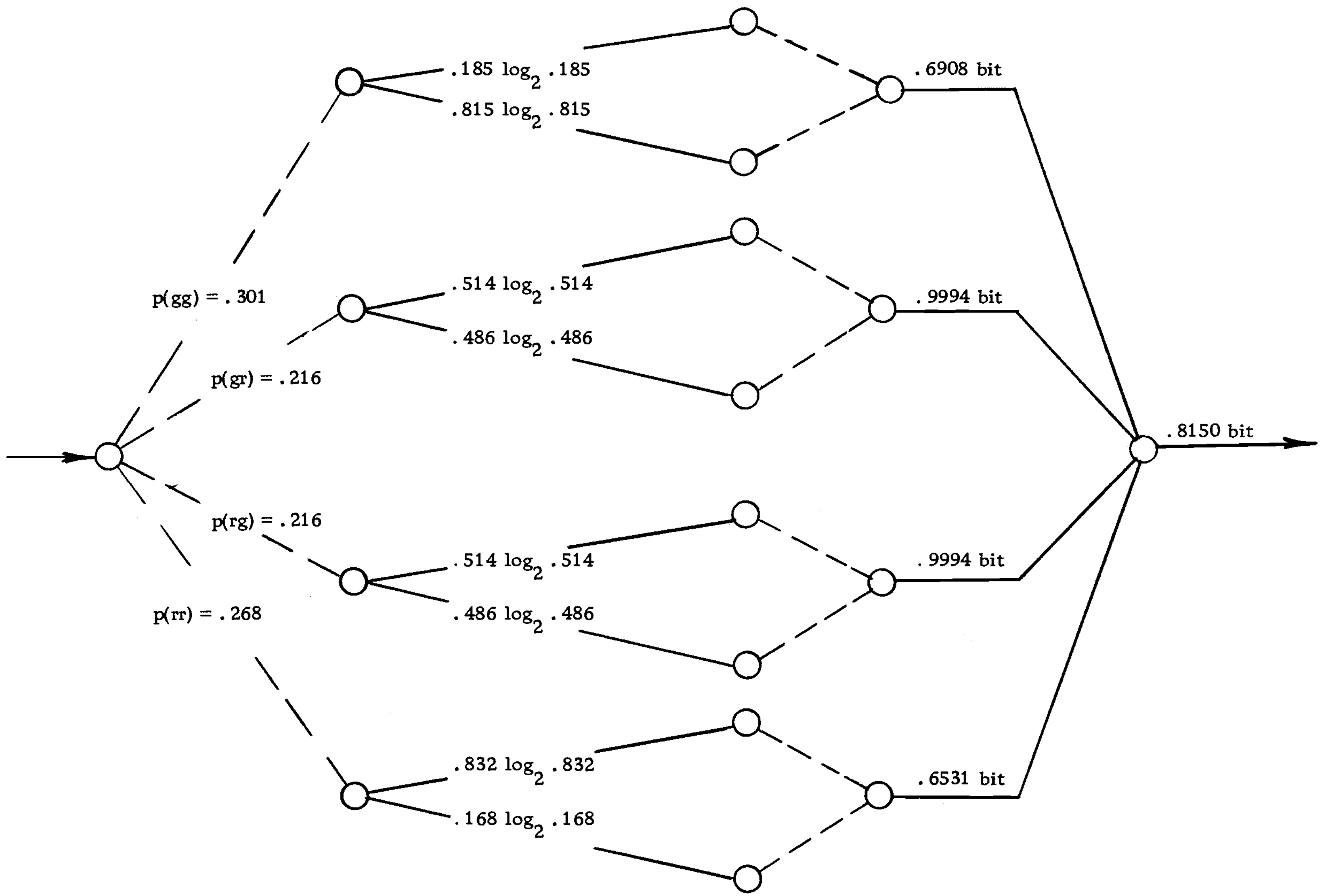

Figure 6. Information tree for sampling procedure (b).

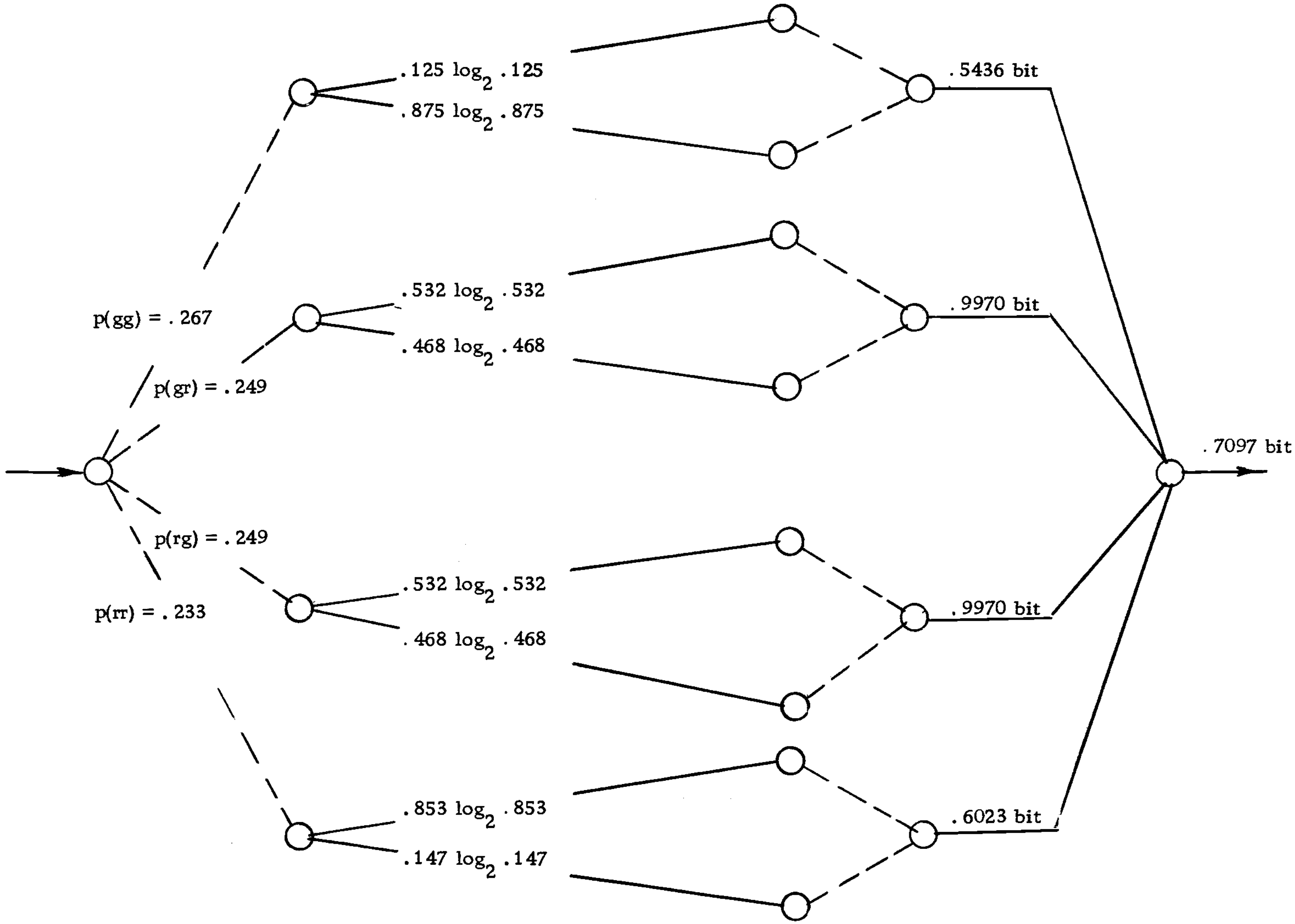

Figure 7. Information tree for sampling procedure (c).

the input/output information tree format.

Alternative Solution

In order to check the validity of the information theory analysis in the urn example, the following traditional solution method is presented using the concept of Maximum Likelihood.

Procedure (a): Referring to the table of probabilities obtained for this case we may determine the likelihood function for the experiment.

Table 4. Likelihood values for procedure (a) in the urn example.

z_j \ y_i	r	g
A	(.690)	.321
B	.310	(.679)
$p(y_i)$	.483	.517
$L(z_j \mid y_i)$	.690	.679
guess	A	B

The likelihood function, $L(z_j \mid y_i)$, indicates our guess as to which urn we have been sampling from given various outcomes. For example, if we observe a red ball, the probability that we have been sampling from urn A is (.690) and from B (.310). Therefore, if we observe a red ball we should guess that we have been sampling from urn A since it is more likely that we have been doing so. Similarly, if a green ball is observed, the probability that we have been sampling from urn B is (.679) and from A (.321). So, the likelihood function is (.690) and (.679) given a red and green ball respectively.

In order to determine which sampling method is best, we would like to determine the probability that we guess the right urn given a particular outcome.

From the likelihood function:

Outcome (r) results in a guess of (A)

Outcome (g) results in a guess of (B).

so:

P(Correct Guess) = P(red)P(Sampling from A given red)

+ P(green)P(Sampling from B given green)

P(Correct Guess) = (.483)(.690) + (.517)(.679)

and

P(Correct Guess) = .685

Procedure (b):

Table 5. Likelihood values for procedure (b) in the urn example.

$z_j \backslash y_i$	gg	gr	rg	rr
A	.185	.514	.514	.832
B	.815	.486	.486	.168
$p(y_i)$	.301	.216	.216	.268
$L(z_j \mid y_i)$	.815	.514	.514	.832
Guess	B	A	A	A

and

$$P(\text{Correct Guess}) = p(gg)p(B|gg) + 2p(gr)p(A|gr) + p(rr)p(A|rr)$$

$$P(\text{Correct Guess}) = (.301)(.815) + (.216)(2)(.514) + (.268)(.832)$$

$$P(\text{Correct Guess}) = .691$$

Procedure (c):

Table 6. Likelihood values for procedure (c) in the urn example.

z_j \ y_i	gg	gr	rg	rr
A	.125	(.532)	(.532)	(.853)
B	(.875)	.468	.468	.147
$p(y_i)$	.267	.249	.249	.233
$L(z_j \mid y_i)$	.875	.532	.532	.853
Guess	B	A	A	A

and

$$P(\text{Correct Guess}) = p(gg)p(B|gg) + 2p(gr)p(A|gr) + p(rr)p(A|rr)$$

$$P(\text{Correct Guess}) = (.267)(.875) + 2(.249)(.532) + (.233)(.853)$$

$$P(\text{Correct Guess}) = .697$$

To summarize:

Sampling procedure (a) = .685

Sampling procedure (b) = .691

Sampling procedure (c) = .697

The conclusion is that sampling procedure (c) is preferred since it results in a higher probability of guessing correctly given the various outcomes. However, it may be noted that the above probabilities are nearly equal. Therefore, the information theory analysis which results in the same conclusion appears to offer the investigator a ubiquitous look at the data in that he may view the entire process at once to obtain the relative gain in information between the various experiments. The information theory analysis also allows the investigator to set decision limits on the value of additional information. In other words, it may be determined before the experiment is conducted that an additional trial should result in at least say 10% gain in information to be significant or worthwhile. In this manner. a decision can be made at each stage of an experiment to determine whether another trial should be conducted. The example which follows illustrates an analysis of this type.

Concrete Mixing Problem

Problem Formulation

New types of concrete mixes are tested in a laboratory by conducting compression and other strength tests on one or more test cylinders. The probability that a trial batch will yield the specified strength is 0.90 if the mix is properly prepared and tested. Occasionally, about once every 20 times, the trial batch will be improperly handled or the ingredients inaccurately measured. The probability that a poorly prepared mix will yield the specified strength is 0.20.

We wish to determine the information transmitted and the information gain from sampling 1, 2, 3, , , n cylinders at random from a particular mix.

The following notation will be utilized:

z_1: The mix was properly prepared.

z_2: The mix was improperly prepared.

y_1: The mix is of the specified strength.

y_2: The mix is not of the specified strength.

Analysis

The transmitted information as a result of conducting an experiment with the above data may be computed from Equation (2.3). For this example, Equation (2.3) will be:

$$H(Z|Y) = p(y_1)[H(Z|y_1)] + p(y_2)[H(Z|y_2)]$$

or

$$H(Z|Y) = -p(y_1)[p(z_1|y_1) \log p(z_1|y_1) + p(z_2|y_1) \log p(z_2|y_1)]$$
$$-p(y_2)[p(z_1|y_2) \log p(z_1|y_2) + p(z_2|y_2) \log p(z_2|y_2)]$$

From the formulation of the problem, the following probabilities are available:

P(mix is of the specified strength given it was properly mixed)

$$p(y_1|z_1) = 0.90$$

P(mix is not of the specified strength given it was properly mixed)

$$p(y_2|z_1) = 0.10$$

P(mix is of the specified strength given it was improperly prepared)

$$p(y_1|z_2) = 0.20$$

P(mix is not of the specified strength given it was improperly prepared)

$$p(y_2|z_2) = 0.80$$

P(mix was properly prepared)

$$p(z_1) = 0.95$$

P(mix was improperly prepared)

$$p(z_2) = 0.05$$

The transmitted information will be calculated for sampling 1, 2, 3, 4, 5 and 6 cylinders respectively.

Case I: Determine the transmitted information by sampling one cylinder at random and observing whether it is of the specified strength.

Sample space of possible outcomes:

1. y_1: cylinder is of the specified strength.
2. y_2: cylinder is not of the specified strength.

Bayes formula may be used to determine the probabilities necessary for the equation of information transmission previously given. The results are as given in Table 7. In the ensuing analysis, the possible outcomes will be denoted by the number of y_1's in the outcome. For example, the two possible outcomes for Case I (sampling one cylinder) will be $1y_1$ and $0y_1$ respectively. The outcome $0y_1$ denotes the absence of y_1 or that y_2 is the outcome. The reason for this notation will be obvious when possible outcomes for sampling 2, 3, 4, 5 or 6 cylinders are enumerated.

Table 7. Concrete mixing problem--sampling one cylinder.

Outcome	$p(z_1 \mid y_i)$	$p(z_2 \mid y_i)$	$p(y_i)$	$H(Z \mid y_i)$
$1y_1$	0.989	0.011	0.856	0.0874
$0y_1$	0.702	0.298	0.135	0.8780

The transmitted information may be computed as:

$$H(Z \mid Y) = 0.865(0.0874) + 0.135(0.8780)$$

$$H(Z \mid Y) = 0.1932 \text{ bits}$$

<u>Case II:</u> Determine the transmitted information by sampling two cylinders at random and observing whether each is of the specified strength.

Sample space of outcomes:

1. The first cylinder is of the specified strength, the second is also.

2. The first cylinder is not of the specified strength, the second is not.

3. The first cylinder is of the specified strength, the second is not.

4. The first cylinder is not of the specified strength, the second is of the specified strength.

Since the probabilities of outcomes 3 and 4 are identical, we may compute one of them and note the existence of the other. The four possible outcomes are:

1. $y_1 y_1 = 2y_1$
2. $y_2 y_2 = 0y_1$
3. $y_1 y_2 = 1y_1$
4. $y_2 y_1 = 1y_1$

Table 8 shows the results needed in order to calculate the transmitted information.

Table 8. Concrete mixing problem--sampling two cylinders.

Outcome	$p(z_1 \mid y_i)$	$p(z_2 \mid y_i)$	$p(y_i)$	Frequency	$H(Z \mid y_i)$
1. $2y_1$	0.997	0.003	0.7720	1	0.0295
2. $0y_1$	0.229	0.771	0.0415	1	0.7763
3. $1y_1$	0.914	0.086	0.0935	2	0.4230

The transmitted information may be computed as:

$H(Z|Y) = 0.7720(0.0295) + 0.0415(0.7763) + 2(0.0935)(0.4230)$

$H(Z|Y) = 0.1340$ bits

Case III: Determine the transmitted information by sampling three cylinders at random and observing whether each is of the specified strength.

Using similar notation to that used in Case II the sample space of outcomes will be:

1. $y_1y_1y_1 = 3y_1$
2. $y_1y_1y_2 = 2y_1$
3. $y_1y_1y_1 = 1y_1$
4. $y_2y_2y_2 = 0y_1$

Table 9 shows the results needed in order to calculate the transmitted information.

Table 9. Concrete mixing problem--sampling three cylinders.

Outcome	$p(z_1 \mid y_i)$	$p(z_2 \mid y_i)$	$p(y_i)$	Frequency	$H(Z \mid y_i)$
1. $3y_1$	1.0	0.0	0.6540	1	0.000
2. $2y_1$	0.980	0.020	0.0786	3	0.1415
3. $1y_1$	0.572	0.428	0.01495	3	0.98499
4. $0y_1$	0.036	0.964	0.02655	1	0.2237

The transmitted information may be computed as:

$$H(Z|Y) = 0.6540(0.000) + 3(0.0786)(0.1415)$$
$$+ 3(0.01495)(0.98499) + 0.02655(0.2237)$$

$$H(Z|Y) = 0.0835 \text{ bits}$$

Case IV: Determine the transmitted information by sampling four cylinders at random and observing whether each is of the specified strength.

The sample space will be:

1. $y_1 y_1 y_1 y_1 = 4y_1$
2. $y_1 y_1 y_1 y_2 = 3y_1$
3. $y_1 y_1 y_2 y_2 = 2y_1$
4. $y_1 y_2 y_2 y_2 = 1y_1$
5. $y_2 y_2 y_2 y_2 = 0y_1$

Table 10 shows the results needed in order to calculate the transmitted information.

Table 10. Concrete mixing problem--sampling four cylinders.

Outcome	$p(z_1 \mid y_i)$	$p(z_2 \mid y_i)$	$p(y_i)$	Frequency	$H(Z \mid y_i)$
1. $4y_1$	1.000	0.000	0.622	1	0.00
2. $3y_1$	1.000	0.0000	0.069	4	0.00
3. $2y_1$	0.855	0.145	0.009	6	0.5972
4. $1y_1$	0.145	0.855	0.006	4	0.5972
5. $0y_1$	0.000	1.000	0.021	1	0.000

The transmitted information may be computed as:

$H(Z|Y) = 0.622(0.00) + 4(0.069)(0.00) + 6(0.009)(0.5972)$

$+ 4(0.006)(0.5972) + (0.021)(0.000)$

$H(Z|Y) = 0.0466$ bits

Case V: Determine the transmitted information by sampling five cylinders at random and observing whether each is of the specified strength.

The sample space will be:

1. $y_1y_1y_1y_1y_1 = 5y_1$
2. $y_1y_1y_1y_1y_2 = 4y_1$
3. $y_1y_1y_1y_2y_2 = 3y_1$
4. $y_1y_1y_2y_2y_2 = 2y_1$
5. $y_1y_2y_2y_2y_2 = 1y_1$
6. $y_2y_2y_2y_2y_2 = 0y_1$

Table 11 shows the results needed in order to calculate the transmitted information.

Table 11. Concrete mixing problem--sampling five cylinders.

Outcome	$p(z_1\|y_i)$	$p(z_2\|y_i)$	$p(y_i)$	Frequency	$H(Z\|y_i)$
1. $5y_1$	1.000	0.000	.5600	1	0.000
2. $4y_1$	1.000	0.000	.0620	5	0.000
3. $3y_1$	.973	.027	.0071	10.	0.1791
4. $2y_1$	.426	.574	.0018	10	0.9841
5. $1y_1$	.021	.979	.0042	5	0.1470
6. $0y_1$	.000	1.000	.0164	1	0.0000

The transmitted information is:

$$H(Z|Y) = .560(.000) + 5(.0620)(.0000) + 10(.0071)(.1791)$$
$$+ 10(.0018)(.9841) + 5(.0042)(.1470)$$

$$H(Z|Y) = .0290 \text{ bit}$$

Case VI: Determine the transmitted information by sampling six cylinders at random and observing whether each is of the specified strength.

The sample space will be:

1. $y_1y_1y_1y_1y_1y_1 = 6y_1$
2. $y_1y_1y_1y_1y_1y_2 = 5y_1$
3. $y_1y_1y_1y_1y_2y_2 = 4y_1$
4. $y_1y_1y_1y_2y_2y_2 = 3y_1$
5. $y_1y_1y_2y_2y_2y_2 = 2y_1$
6. $y_1y_2y_2y_2y_2y_2 = 1y_1$
7. $y_2y_2y_2y_2y_2y_2 = 0y_1$

Table 12 shows the results needed in order to calculate the transmitted information.

Table 12. Concrete mixing problem--sampling six cylinders.

Outcome	$p(z_1 \mid y_i)$	$p(z_2 \mid y_i)$	$p(y_i)$	Frequency	$H(Z \mid y_i)$
1. $6y_1$	1.00	0.00	0.503	1	0.00
2. $5y_1$	1.00	0.00	0.056	6.	0.00
3. $4y_1$	1.00	0.00	0.00621	15	0.00
4. $3y_1$	0.772	0.228	0.000896	20	0.7745
5. $2y_1$	0.085	0.915	0.000905	15	0.4196
6. $1y_1$	0.00	1.000	0.003365	6	0.00
7. $0y_1$	0.00	1.000	0.01310	1	0.00

The transmitted information may be computed as:

$$H(Z \mid Y) = 20(0.000896)(0.7745) + 15(0.000905)(0.4196)$$

$$H(Z \mid Y) = 0.0195 \text{ bits}$$

Figure 8 shows a plot of the transmitted information for the various sampling experiments given in the concrete mixing example.

Figures 9 through 14 summarize the six sampling procedures in terms of input/output information trees.

The concrete mixing example can be viewed as a decision process by calculating the information gained from taking successive samples. Suppose, for example, that due to the cost of sampling, it is felt that at least a 25% gain in information must result from sampling an additional cylinder over sampling just one cylinder. Since the transmitted information from sampling one cylinder is

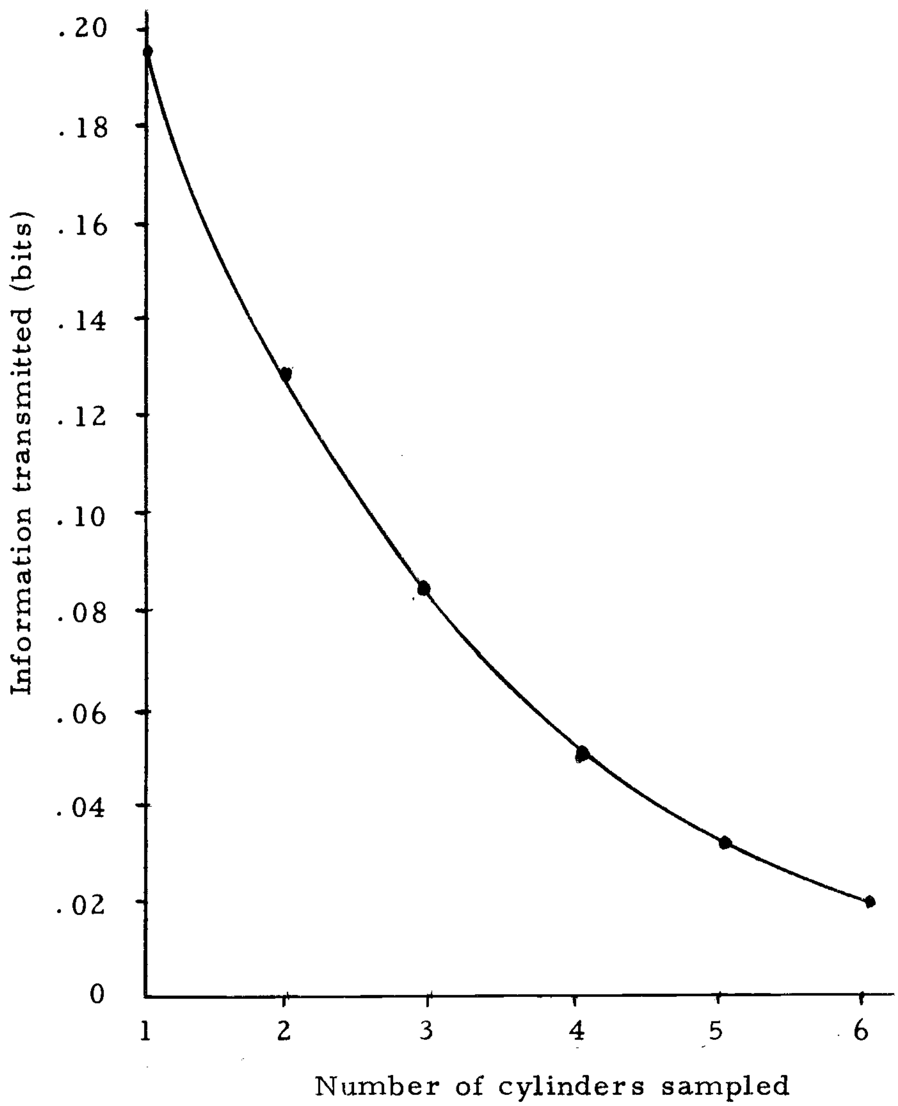

Figure 8. Transmitted information for sampling in a concrete mixing problem.

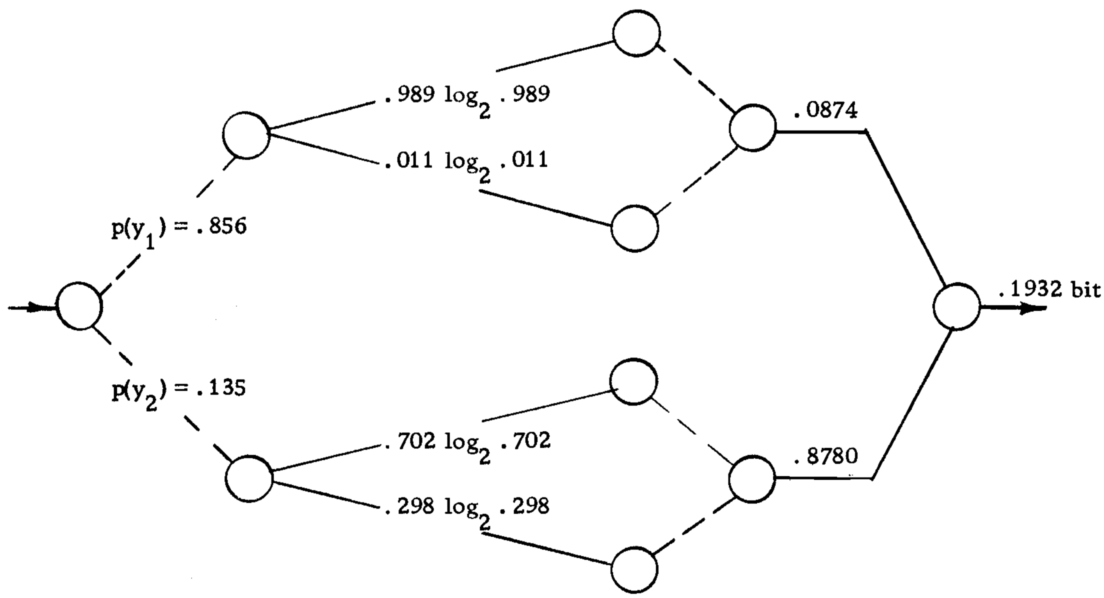

Figure 9. Information tree for sampling one cylinder.

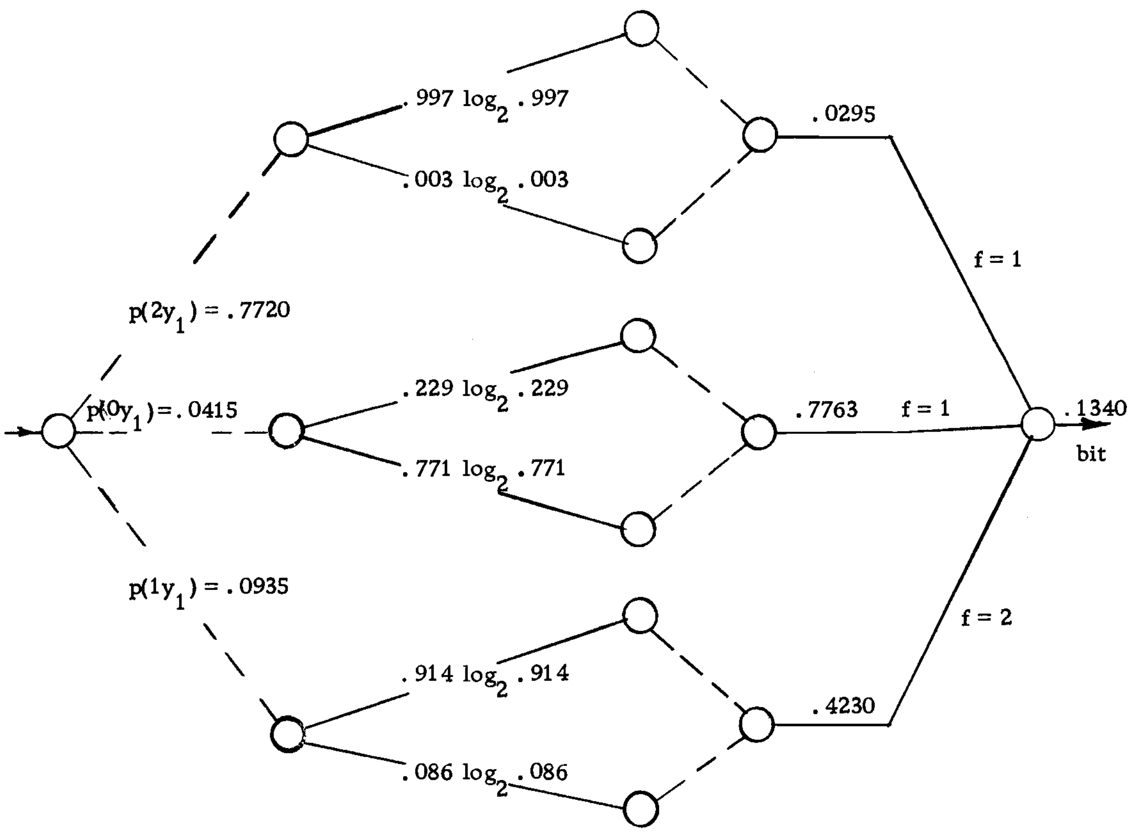

Figure 10. Information tree for sampling two cylinders.

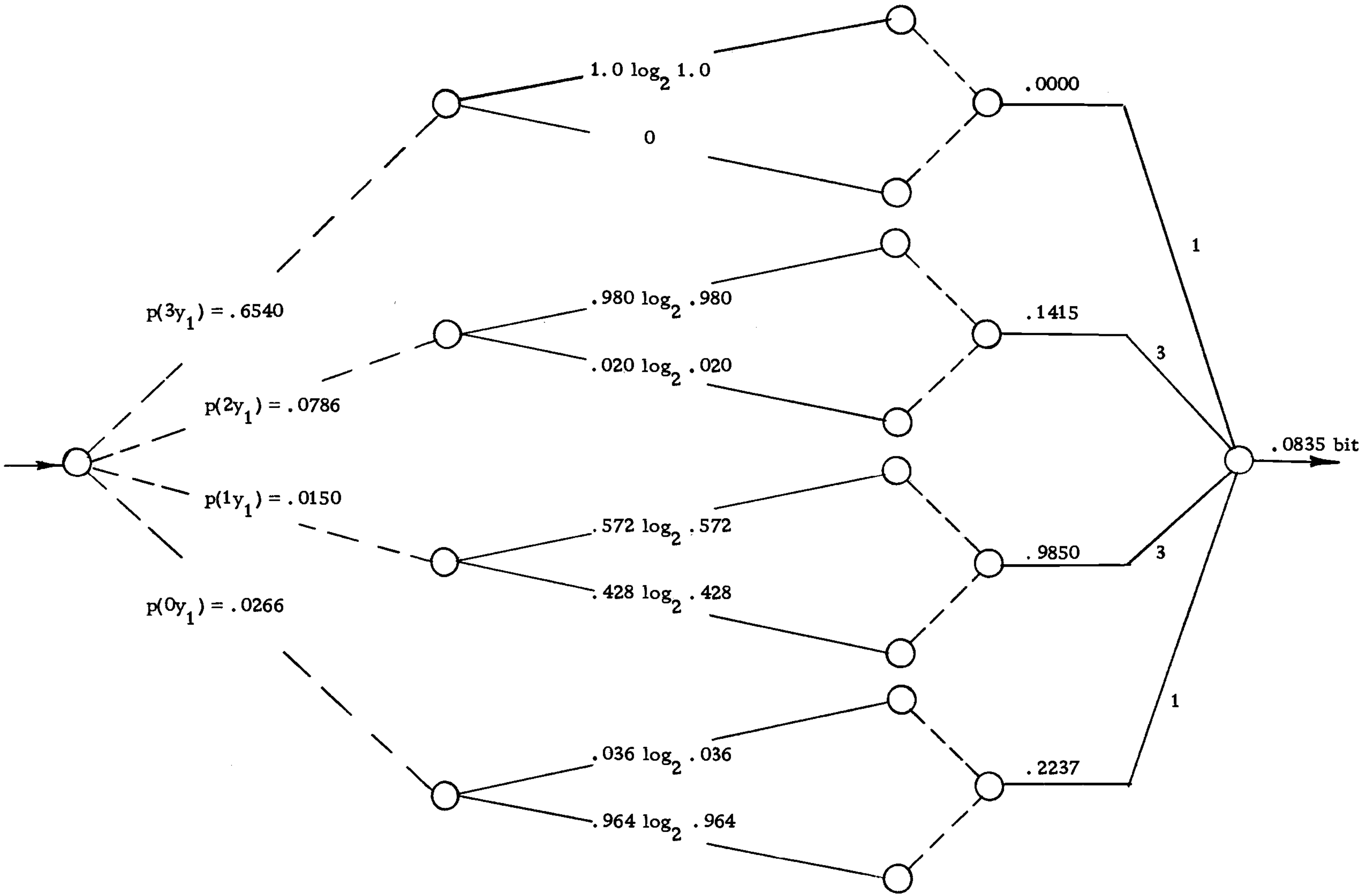

Figure 11. Information tree for sampling three cylinders.

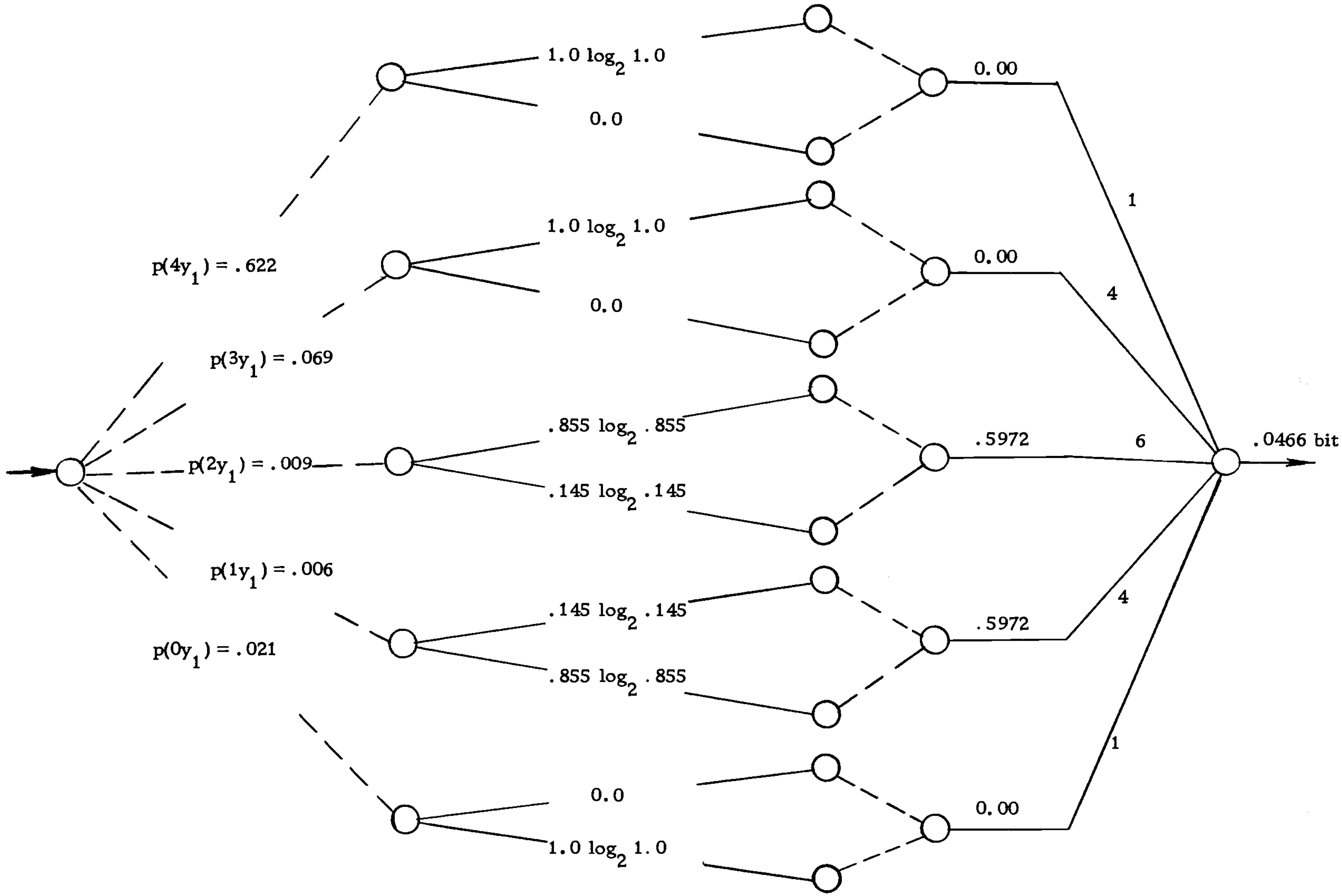

Figure 12. Information tree for sampling four cylinders.

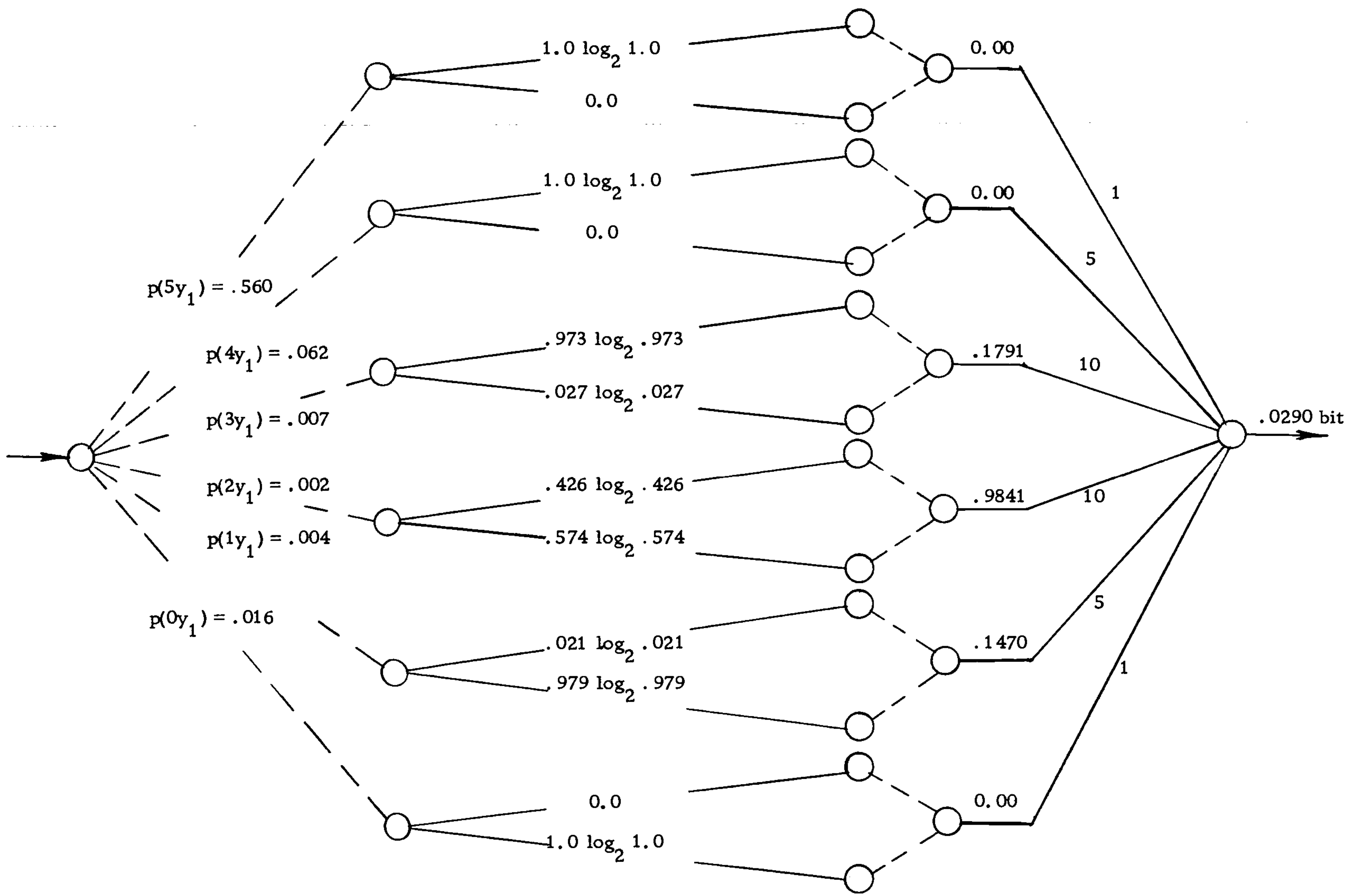

Figure 13. Information tree for sampling five cylinders.

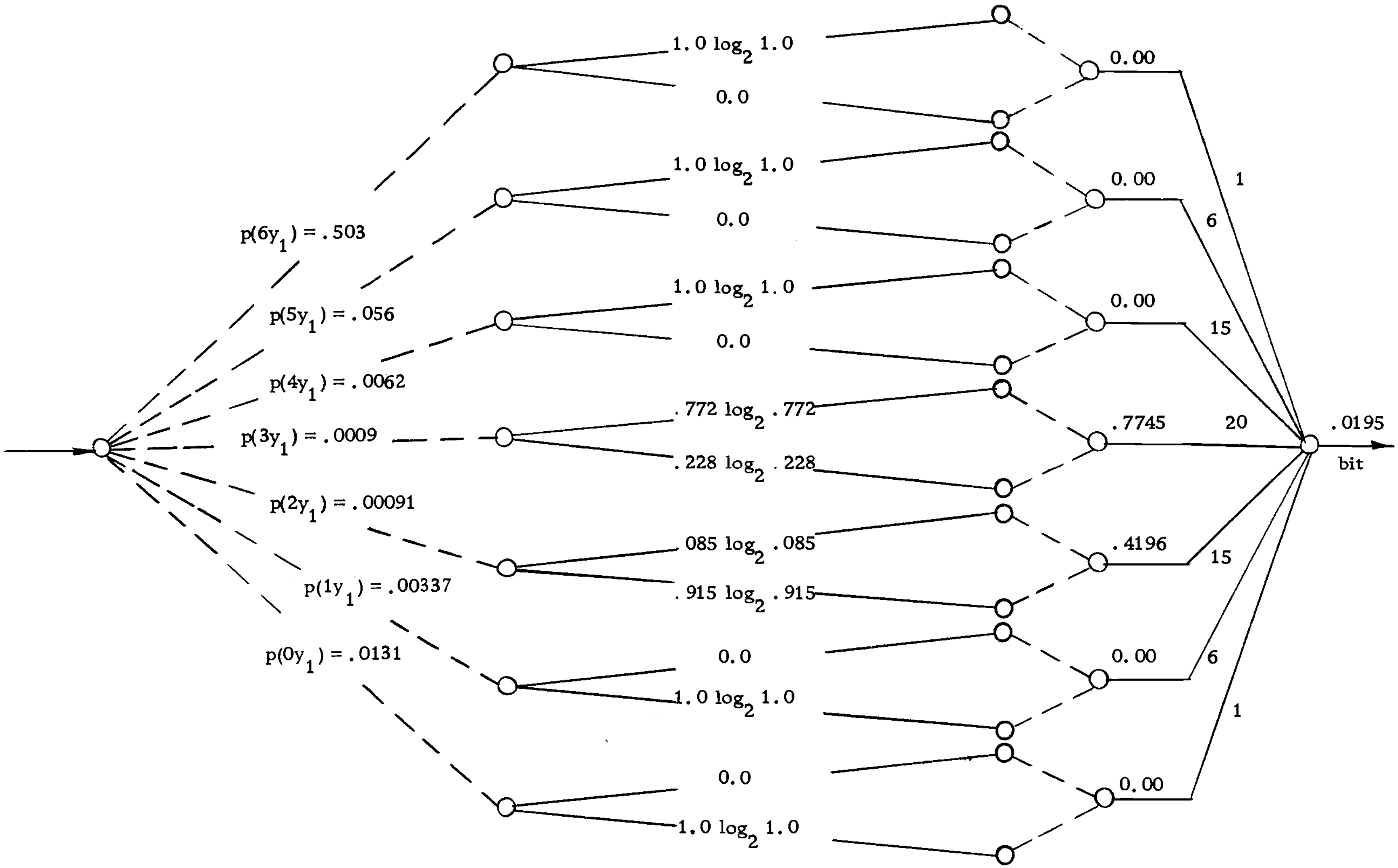

Figure 14. Information tree for sampling six cylinders.

0.1932 bits, we require that sampling 2, 3, 4, etc. cylinders will result in a gain of at least 0.0483 bits of information. Using the results previously obtained the following values may be determined.

Table 13. Summary of concrete mixing example results.

No. of Cylinders Sampled	Information Transmitted (bits)	Information Gain
1	0.1932	--
2	0.1340	0.0592
3	0.0835	0.0505
4	0.0466	0.0369
5	0.0290	0.0170
6	0.0195	0.0095

Based on the results of Table 13, we would sample 3 cylinders, since sampling a fourth cylinder only contributes 0.0466 bits of information which is less than the 25% requirement.

<u>Cost Considerations in Concrete Mixing Problem</u>

The concrete mixing example could also be viewed as a decision process that considers cost of additional information as a criterion. Suppose, for example, that the costs of testing cylinders (obtaining equipment, set up costs, direct labor, etc.) are as represented in Table 14.

Suppose also that management personnel have decided that they would be willing to pay for each additional test according to the schedule depicted in Figure 15.

Table 14. Testing costs for concrete mixing problem.

Number of Cylinders Sampled	Cost of Testing
1	$ 700
2	900
3	1320
4	1870
5	2570
6	3300

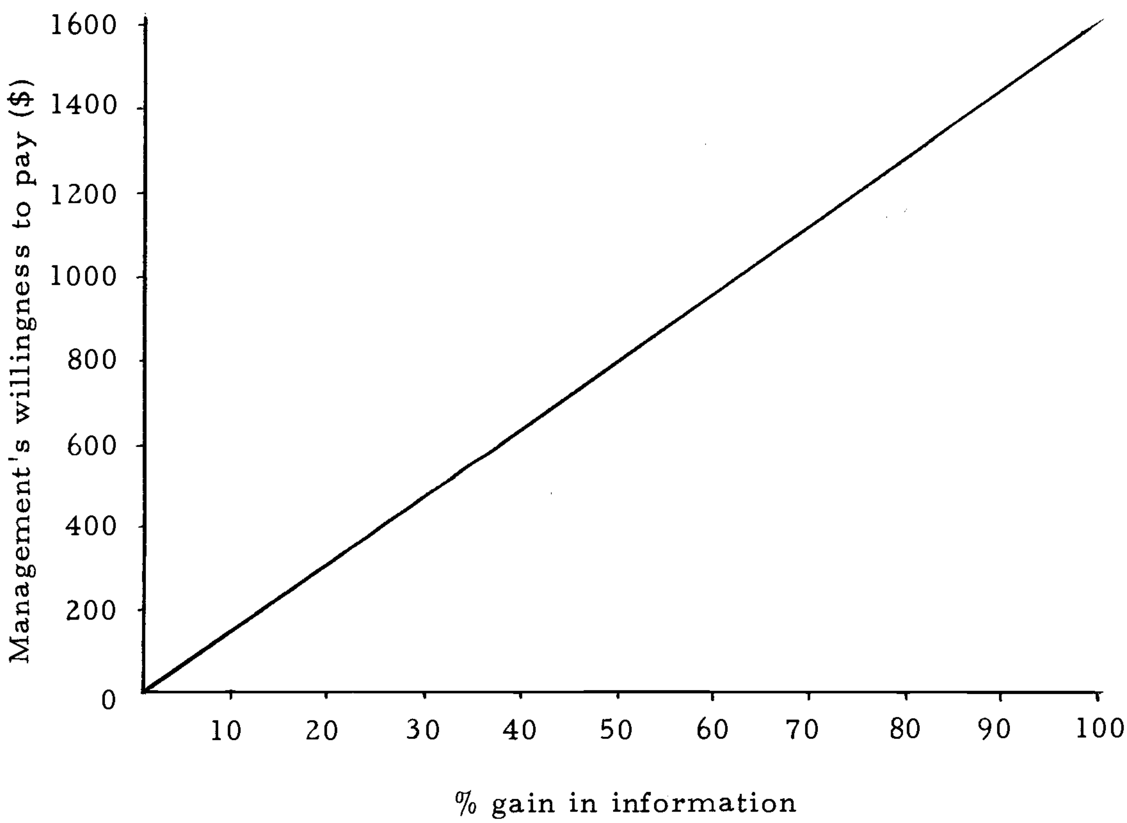

Figure 15. Management's willingness to pay for additional information in the concrete mixing example.

Using Table 14 and Figure 15 along with the results obtained for information gain, we can conclude that three cylinders should be sampled. The information gained from the three cylinder experiment is .0592 bit or 30% for which we are willing to pay $420. Sampling three cylinders results in a gain of .1097 bit or 56.6% for which we are willing to pay $792. The additional cost of sampling three cylinders is $620. Sampling four cylinders results in a gain of .1466 bit or 75.5% at a cost of $1060, so our decision would be to sample three cylinders if cost of information is a criterion in the decision making analysis. The relationship between information gain and cost is portrayed graphically in Figure 16.

Figure 16 shows that an additional sample should be tested for those points plotted for the cost of testing that are below the curve for management's willingness to pay. The last point which is below the curve for management's willingness to pay is the solution to the problem.

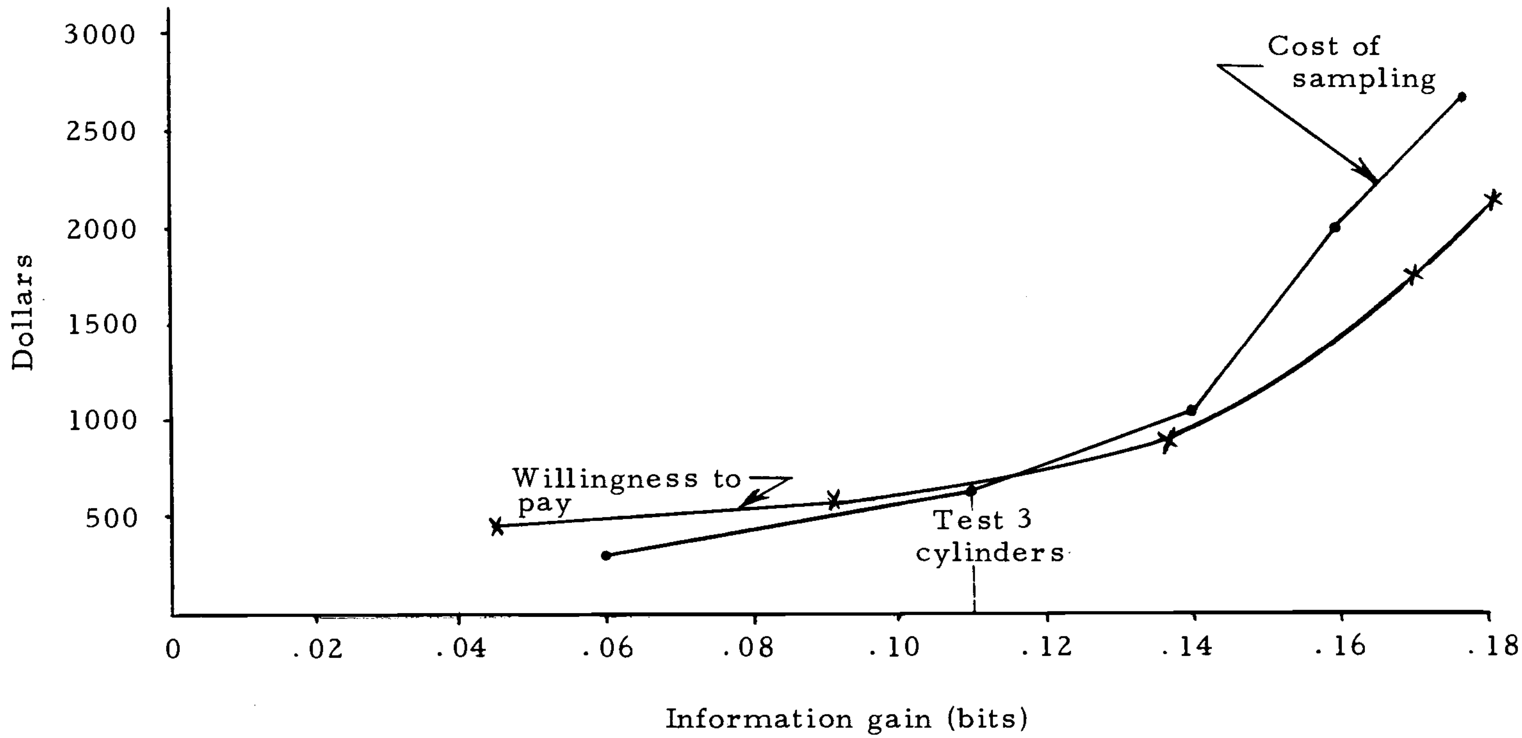

Figure 16. Decision making data--cost vs information gain for concrete mixing example.

III. ESTIMATION USING MULTI-FACTOR INFORMATION THEORY MODELS

Introduction and Theoretical Development

In this chapter, information theory will be applied in an attempt to predict values for $p(z_j)$ and $p(y_i)$ in a multi-factor conditional information analysis. In Chapter II the conditional information transmitted was determined from Equation (2.4) as:

$$H(Z|Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(y_i z_j) \log p(z_j|y_i) \qquad (3.0)$$

If Equation (2.5) for $p(z_j|y_i)$ is substituted into the above equation for the conditional information, the following relationship is obtained:

$$H(Z|Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(y_i z_j) \log\left[\frac{p(y_i z_j)}{p(y_i)}\right]$$

which may be written:

$$H(Z|Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(y_i z_j) \log p(y_i z_j) + \sum_{i}\sum_{j} p(y_i z_j) \log p(y_i) \qquad (3.1)$$

but $\sum_{i}\sum_{j} p(y_i z_j) \log p(y_i)$ may be written:

$$\sum_{i} \log p(y_i) \sum_{j} p(y_i z_j)$$

and employing Equation (2.5) again:

$$\sum_{j} p(y_i z_j) = \sum_{j} p(y_i) p(z_j | y_i) \qquad (3.2)$$

Since $p(y_i)$ is constant as far as the summation is concerned, Equation (3.2) may be written as:

$$\sum_{j} p(y_i z_j) = p(y_i) \sum_{j} p(z_j | y_i)$$

but $\sum_{j} p(z_j | y_i) = 1$ for all y_i $i = 1, 2, \ldots, n$ so Equation (3.1) becomes:

$$H(Z | Y) = -\sum_{i} \sum_{j} p(y_i z_j) \log p(y_i z_j) + \sum_{i} p(y_i) \log p(y_i)$$

and from Equation (2.0)

$$H(Y, Z) = -\sum_{i} \sum_{j} p(y_i z_j) \log p(y_i z_j)$$

and

$$H(Y) = \sum_{i} p(y_i) \log p(y_i)$$

Therefore

$$H(Z \mid Y) = H(Y, Z) - H(Y)$$

and the information transmitted by the joint experiment is:

$$H(Y, Z) = H(Y) + H(Z \mid Y) \tag{3.3}$$

In words, Equation (3.3) states that the information transmitted by the combined experiment is the information transmitted by Experiment Z when the results of Y are known. This result implies that the information transmitted by the combined experiment is greater than either of the sub-experiments or that:

$$H(Z, Y) \geq H(Y)$$

since $H(Z, Y)$, $H(Y)$ and $H(Z \mid Y)$ are all non-negative values.

We could also write:

$$H(Z) + H(Y) \geq H(Z, Y)$$

or

$$H(Z) + H(Y) \geq H(Y) + H(Z \mid Y)$$

which gives the result that:

$$H(Z) \geq H(Z \mid Y)$$

which shows that knowing the results of Experiment Y reduces the information transmitted by Experiment Z. Therefore, the amount of information gained by observing the results of Experiment Y may be defined as:

$$G(Y, Z) = H(Z) - H(Z|Y) \tag{3.4}$$

Equation (3.4) will be useful in deriving the prediction equations which follow.

Prediction Equations

Development

Given a conditional probability matrix:

$$\begin{array}{c} \\ z_1 \\ \\ z_j \\ \\ z_m \end{array} \begin{array}{c} \begin{array}{ccccc} y_1 & & y_i & & y_n \end{array} \\ \begin{bmatrix} p(z_1|y_1) & \cdots & p(z_1|y_i) & \cdots & p(z_1|y_n) \\ \vdots & & \vdots & & \vdots \\ p(z_j|y_1) & \cdots & p(z_j|y_i) & \cdots & p(z_j|y_n) \\ \vdots & & \vdots & & \vdots \\ p(z_m|y_1) & \cdots & p(z_m|y_i) & \cdots & p(z_m|y_n) \end{bmatrix} \end{array}$$

With no prior knowledge of $p(z_j)$ or $p(y_i)$ we would like to develop a method to predict these values. A logical approach is one which will determine $p(z_j)$ and $p(y_i)$ such that a maximum information gain is achieved. Equation (3.4) could be used then, to determine:

$$[\max]G(Y, Z) = H(Z) - H(Z|Y)$$

$$\text{subject to} \quad \sum_j p(z_j) = 1 \quad \text{and} \quad p(y_i) > 0 \quad i = 1, 2, \ldots, n$$

Using the Lagrange Multiplier technique, the maximization equation becomes:

$$[\max]G(Y, Z) = H(Z) - H(Z|Y) + \mu[\sum_j p(z_j) - 1] + \sum_j \lambda_i(p(y_i) + u_i^2) \tag{3.5}$$

The μ and λ_i terms are the Lagrange Multipliers and the u_i^2 terms are surplus variables introduced to form the necessary equality: $p(y_i) + u_i^2 = 0$. In most actual problems $p(y_i)$ will naturally turn out to be greater than zero. It is easiest then, to first ignore the terms $\lambda_i(p(y_i)+u_i^2)$ and solve for $p(z_j)$ and $p(y_i)$. If one or more of the $p(y_i)$ are negative, then we may insert the restriction on the $p(y_i)$ and employ the Kuhn-Tucker conditions to solve the resulting equations. The Kuhn-Tucker conditions will be:

1. $\dfrac{\partial[F(Z, \mu, \lambda_i, u)]}{\partial z_j} = 0$

2. $\mu(\sum_j p(z_j) - 1) = 0$

3. $\lambda_i(\sum_i p(y_i) + u_i^2) = 0$

4. $\lambda_i u_i = 0$

5. $p(z_j) > 0$

6. $\lambda_i \leq 0; \quad \mu \leq 0$

$j = 1, 2, \ldots, m$

$i = 1, 2, \ldots, n$

An alternative method of correcting for negative probability values is to first determine those values which are negative, set them equal to zero, and re-solve for the maximum information gain with these restrictions included.

An Example

Returning to the urn example of Chapter I, assume that we are interested in sampling one ball only. What strategies should we use in drawing from urns A or B that will result in obtaining the most information concerning the experiment?

The available data is:

$$p(g|A) = 1/3$$

$$p(r|A) = 2/3$$

$$p(g|B) = 7/10$$

$$p(r|B) = 3/10$$

In matrix form, $p(Y|Z)$, where $y_1 = g$, $y_2 = r$, $z_1 = A$, $z_2 = B$, becomes:

$$[p(Y|Z)] = \begin{matrix} & g & r \\ A & \multicolumn{2}{c}{} \end{matrix} \begin{matrix} \\ A \\ B \end{matrix}\begin{matrix} g & r \\ \left[\begin{matrix} 1/3 & 2/3 \\ 7/10 & 3/10 \end{matrix}\right] \end{matrix}$$

In this problem it will be easier to solve for $p(g)$ and $p(r)$

first, then to use the relationship that:

$$p(g) = p(A)p(g|A) + p(B)p(g|B)$$

and

$$p(r) = p(A)p(r|A) + p(B)p(r|B) \tag{3.6}$$

to obtain $p(A)$ and $p(B)$, (Appendix A gives a proof that the information gain may be computed as $H(Y) - H(Z|Y)$ or as $H(Z) - H(Y|Z)$).

As suggested in the theoretical development, the restriction that $p(A) > 0$ and $p(B) > 0$ will first be ignored. Then, if either of these probabilities turn out to be negative, the restriction will be imposed and the solution redetermined.

We wish to determine $p(g)$ and $p(r)$ such that:

$$[\max]G(Z, Y) = H(Y) - H(Y|Z) + \mu[p(y_1)+p(y_2)-1] \tag{3.7}$$

where

$$H(Y) = -p(g) \log_2 p(g) - p(r) \log_2 p(r) \tag{3.8}$$

and

$$H(Y|Z) = -p(A)[p(g|A) \log_2 p(g|A) + p(r|A) \log_2 p(r|A)]$$
$$-p(B)[p(g|B) \log_2 p(g|B) + p(r|B) \log_2 p(r|B)] \tag{3.9}$$

Equation (3.6) may be solved to obtain $p(A)$ and $p(B)$ needed for Equation (3.9).

$$p(A)(1/3) + p(B)(7/10) = p(g)$$

$$p(A)(2/3) + p(B)(3/10) = p(r)$$

Solving:

$$p(A) = \frac{-9p(g)+21p(r)}{11}$$

$$p(B) = \frac{20p(g)-10p(r)}{11}$$

and Equation (3.9) becomes:

$$H(Y|Z) = -[\frac{-9p(g)+21p(r)}{11}][1/3 \log_2 1/3 + 2/3 \log_2 2/3]$$

$$- [\frac{20p(g)-10p(r)}{11}][7/10 \log_2 7/10 + 3/10 \log_2 3/10]$$

Therefore,

$$H(Y|Z) = .848\, p(g) + .950\, p(r)$$

Equation (3.7) will now become:

$$[max]G(Z,Y) = -p(g) \log_2 p(g) - p(r) \log_2 p(r) - .848p(g)$$

$$- .950p(r) + \mu(p(g)+p(r)-1)$$

and

$$\frac{\partial[H(Z,Y)]}{\partial p(g)} = -\log_2 p(g) - 1 - .848 + \mu = 0 \qquad (3.10)$$

$$\frac{\partial[H(Z,Y)]}{\partial p(r)} = -\log_2 p(r) - 1 - .950 + \mu = 0 \qquad (3.11)$$

$$\frac{\partial[H(Z,Y)]}{\partial \mu} = p(g) + p(r) - 1 = 0 \qquad (3.12)$$

Solving Equations (3.10) and (3.11) for μ gives:[1]

$$-\log_2 p(r) - \log_2 p(g) = .102$$

or

$$\log_2 \left(\frac{p(g)}{p(r)}\right) = .102$$

and

$$p(g) = p(r)2^{(102)}$$

From Equation (3.12), $p(g) = 1 - p(r)$ so,

$$1 - p(r) = p(r)2^{(.102)}$$

and

$$p(r) = .482$$

$$p(g) = .518$$

Now the sampling strategies may be calculated as:

$$p(A) = \frac{-9(.518)+21(.482)}{11}$$

$$p(B) = 1 - p(A)$$

or,

$$p(A) = .532$$

$$p(B) = .468$$

Since $p(A) > 0$ and $p(B) > 0$, we are justified in neglecting that restriction from our model. The results suggest that in order to

---

[1] Appendix B gives a proof that derivatives of $\log_2 a$ may be differentiated the same as $\log_e a$ to obtain $1/a$.

obtain maximum information from our model we should sample from urns A and B with frequencies 0.532 and 0.468 respectively. Figure 17 shows a plot of the information gain for various values of p(A). Again, the maximum occurs for p(A) = .532 and p(B) = .468 as seen from the graph.

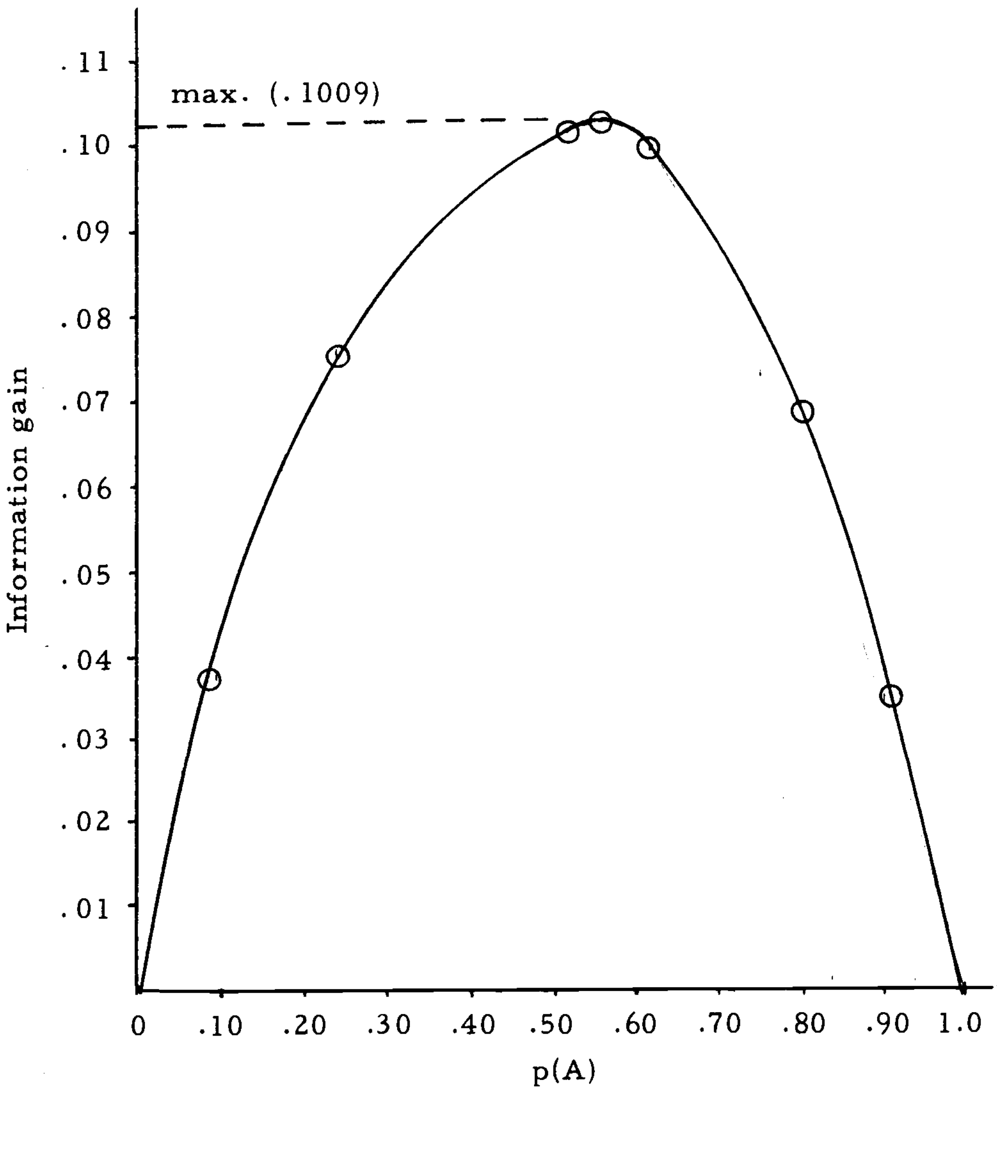

Figure 17. Plot of information gain for urn example.

Using the traditional method of maximum likelihood will result in a sampling strategy of 1.0 and 0.00 for A and B respectively. This means that a researcher should completely ignore one set of data. For example, in an R&D situation in which there are two possible tests for determining operating characteristics for a piece of equipment, the traditional method concludes that one will learn the most about the equipment by ignoring one of the tests completely. The information theory analysis gives a more logical and practical solution in which the information gained by conducting each test is determined. It could turn out, however, that one test is significantly better than the other. In this case the strategy would be to ignore that test. The information theory analysis, then, offers a semantic interpretation which follows from the logical assumption that a researcher would like to extract as much information as possible by utilizing all available data in the most efficient manner.

IV. MATRIX METHOD FOR SOLVING MULTI-FACTOR INFORMATION THEORY PROBLEMS

Introduction

Muroga (1958) suggests that a matrix method may be used to solve predictive equations in a multi-factor information channel analysis. Two types of problems are analyzed in this chapter; (1) problems that initially result in positive solutions and (2) problems that initially result in non-positive solutions.[2]

Theoretical Development

Given a conditional probability matrix of determinable values:

$$[p(Y|Z)] = \begin{array}{c} \\ z_1 \\ z_j \\ z_m \end{array} \begin{array}{c} \begin{array}{ccccc} y_1 & & y_i & & y_m \end{array} \\ \begin{bmatrix} p(y_1|z_1) & \cdots & p(y_i|z_1) & \cdots & p(y_m|z_1) \\ p(y_1|z_j) & \cdots & p(y_i|z_j) & \cdots & p(y_m|z_j) \\ p(y_1|z_m) & \cdots & p(y_i|z_m) & \cdots & p(y_m|z_m) \end{bmatrix} \end{array}$$

where

$$\sum_{i=1}^{m} p(y_i|z_j) = 1$$

---

[2] In this and subsequent chapters, [ ] will denote a matrix. [ ]' will denote a matrix transpose. All vectors are row vectors unless noted as a transpose.

The maximization equation, then, is:

$$[\max]G(Y, Z) = [H(Y)] - [H(Z|Y)] \tag{4.0}$$

Now, define

$$[H]' = [p(Y|Z)][B]' \tag{4.1}$$

where $[H]$ is a vector of information equations of the form:

$$h_1 = -\sum_i p(y_i|z_1) \log p(y_i|z_1)$$

$$\vdots$$

$$h_j = -\sum_i p(y_i|z_j) \log p(y_i|z_j)$$

$$\vdots$$

$$h_m = -\sum_i p(y_i|z_m) \log p(y_i|z_m)$$

and $[B]$ is a vector of constants b_i; $i = 1, 2, 3, \ldots, m$.

Solving Equation (4.1) for $[B]'$ gives:

$$[B]' = [p(Y|Z)]^{-1}[H]' \tag{4.2}$$

Writing Equation (3.4) in matrix form gives:

$$[p(Z)][p(Y|Z)] = [p(Y)] \tag{4.3}$$

Solving for $[p(Z)]$:

$$[p(Z)] = [p(Y)][p(Y|Z)]^{-1} \tag{4.4}$$

Combining Equations (3.0) and (3.2) into matrix form results in the following equation for $H(Y|Z)$.

$$[H(Y|Z)] = [H][p(Z)]' \tag{4.5}$$

Substituting Equation (4.1) and (4.4) into Equation (4.5) results in the following equation for $H(Y|Z)$.

$$[H(Y|Z)] = [p(Y|Z)]^{-1}[p(Y)][B][p(Y|Z)]$$

since $[p(Y)][B]$ is a one-by-one vector of constants $[H(Y|Z)]$ may be written as:

$$[H(Y|Z)] = [p(Y)][B][p(Y|Z)]^{-1}[p(Y|Z)]$$

but

$$[p(Y|Z)^{-1}[p(Y|Z)] = [I]$$

where $[I]$ is an identity matrix. Therefore,

$$[H(Y|Z)] = [p(Y)][B] = \sum_i b_i p(y_i) = \sum_j b_j p(z_j) \tag{4.6}$$

The above result occurs because we have assumed a square matrix of conditional probabilities initially.

Now, since we wish to maximize the information gain equation subject to $\sum_i p(y_i) = 1$, we may include this condition in the following equation:

$$[\max]G(Y, Z) = [H(Y)] - [H(Y|Z)] + \lambda \left(\sum_i p(y_i) - 1\right) \qquad (4.7)$$

where

$$[H(Y)] = -\sum_i p(y_i) \log_2 p(y_i) \qquad (4.8)$$

Substituting Equations (4.6) and (4.8) into Equation (4.7) gives the following equation to be maximized:

$$[\max]G(Y, Z) = \max\left[-\sum_i p(y_i) \log_2 p(y_i) - \sum_i b_i p(y_i) + \lambda \left(\sum_i p(y_i) - 1\right)\right] \qquad (4.9)$$

To maximize Equation (4.9) we take appropriate derivatives and solve.

$$\frac{\partial[G(Y, Z)]}{\partial y_i} = -\log_2 p(y_i) - 1 - b_i + \lambda = 0 \qquad (4.10)$$

$$\frac{\partial[G(Y, Z)]}{\partial \lambda} = \sum_i p(y_i) - 1 = 0 \qquad (4.11)$$

Solving Equation (4.10) in terms of $p(y_i)$ gives:

$$p(y_i) = 2^{\lambda - b_i + 1} = 2^{\lambda} \sum_i 2^{-b_i + 1} = 1$$

and

$$2^{\lambda} = \frac{1}{\sum_i 2^{-b_i + 1}}$$

Substituting this value for 2^{λ} into Equation (4.12) results in:

$$p(y_i) = \frac{2^{-b_i+1}}{\sum_i 2^{-b_i+1}} = \frac{2^{-b_i}2}{2\sum_i 2^{-b_i}}$$

or

$$p(y_i) = \frac{2^{-b_i}}{\sum_i 2^{-b_i}} \qquad (4.13)$$

Equation (4.13) may be solved in conjunction with Equation (4.2) to determine $p(y_i)$ for $i = 1, 2, 3, \ldots, m$ and $p(z_j)$ may be determined from Equation (4.4) (see also Appendix B).

Sampling Example

Let us return to the urn example presented in Chapter II in order to illustrate calculations using the matrix method.

The conditional probability matrix of data is:

$[p(Y|Z)] =$

		green y_1	red y_2
A	z_1	1/3	2/3
B	z_2	7/10	3/10

The inverse of $[p(Y|Z)]$ is:

$$[p(Y|Z)]^{-1} = \begin{bmatrix} -9/11 & 20/11 \\ 21/11 & -10/11 \end{bmatrix}$$

and

$$[H] = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}$$

or

$$h_1 = -1/3 \log 1/3 - 2/3 \log 2/3 = 0.917962$$

$$h_2 = -7/10 \log 7/10 - 3/10 \log 3/10 = 0.881291$$

The values for $[B] = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$ may be calculated using Equation (4.2).

$$[B] = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} -9/11 & 20/11 \\ 21/11 & -10/11 \end{bmatrix} \begin{bmatrix} .917962 \\ .881291 \end{bmatrix}$$

or

$$b_1 = .8550$$

$$b_2 = .9520$$

Therefore,

$$p(y_1) = \frac{2^{-.8550}}{2^{-.8550} + 2^{-.9520}}$$

and

$$p(y_1) = p(\text{green}) = .518$$

$$p(y_2) = p(\text{red}) = .482$$

which are the results obtained previously using the Lagrange Multiplier technique.

The probabilities p(A) and p(B) may now be calculated from Equation (4.3) as:

$$[p(A) \quad p(B)] = [.518 \quad .482] \begin{bmatrix} -9/11 & 20/11 \\ 21/11 & -10/11 \end{bmatrix}$$

and p(A) = .532; p(B) = .468 as previously calculated. The matrix method offers a quick and easy solution to many multi-factor information channel problems.

Example of Multi-Factor Information Theory Models with Non-Positive Solutions

As previously stated, there is no guarantee that non-positive solutions to the multi-factor information theory model will not occur. The following example illustrates a method for restricting the model to provide only positive solutions.

Consider a market research situation such as the one presented by Shirland (1971) in which questionnaires are sent to prospective customers posing the following four questions.

1. Given brands A, B, C, and D, which one would you purchase if price were of major importance to you?
2. Considering quality of product, which one would you purchase?
3. Considering reliability, which would you choose?

4. Considering the warranty of the product, which would you purchase?

For illustrative purposes, consider the following hypothetical data.

Table 15. Multi-factor negative probability example.

p(Y|Z) Brand	A y_1	B y_2	C y_3	D y_4
z_1 (price)	1/2	1/4	0	1/4
z_2 (quality)	0	1	0	0
z_3 (reliability)	1/4	1/4	1/4	1/4
z_4 (warranty)	0	0	1	0

The data of Table 15 imply that considering prices, 50% of those answering the questionnaire would purchase Brand A, 25% would purchase Brand B, none would buy Brand C and 25% would buy Brand D; etc.

Using the matrix method, the values for [H]' are:

$$[H]' = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix} = \begin{bmatrix} 1.500 \\ 0 \\ 2.00 \\ 0 \end{bmatrix}$$

and,

$$[B]' = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/4 & 0 & 1/4 \\ 0 & 1 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \end{bmatrix}$$

this yields,

$$[B]' = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} -2.00 \\ 0 \\ 0 \\ 10.00 \end{bmatrix}$$

Therefore,

$$p(y_i) = \frac{2^{-b_i}}{2^{+2.00} + 2^{0} + 2^{0} + 2^{-10.0}}$$

for $i = 1, 2, 3, 4.$

Putting in appropriate values for b_i in the equation for $p(y_i)$ gives the following results:

$$p(y_1) = .6666$$

$$p(y_2) = .1666$$

$$p(y_3) = .1666$$

$$p(y_4) = .0002$$

The values of $p(z_j)$ $j = 1, 2, 3, 4$ may now be calculated from the relationship that:

$$[.666 \quad .167 \quad .167 \quad 0] \begin{bmatrix} 1/2 & 1/4 & 0 & 1/4 \\ 0 & 1 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 \end{bmatrix}^{-1} = [p(z_1)p(z_2)p(z_3)p(z_4)]$$

Therefore,

$$p(z_1) = 2.6656$$

$$p(z_2) = 0.1665$$

$$p(z_3) = -2.6649$$

$$p(z_4) = .8329$$

The above probabilities for $p(z_j)$ satisfy the restriction that $\sum_j p(z_j) = 1$ but they violate the requirements that $p(z_j) \geq 0$ and $p(z_j) \leq 1$ for $j = 1, 2, 3, 4$ respectively. The result implies that the maximum information gain occurs outside the range of allowable values. Therefore, we must impose an additional restriction in order to limit our solution to positive probabilities only. Since $p(z_3) < 0$, a logical restriction is $p(z_3) = 0$. The Lagrange Multiplier technique may be used to re-solve the information gain equation. However, in order to maximize the information gain equation, we must first solve for the $p(z_j)$ in terms of the $p(y_i)$. In order to simplify that notation, $p(z_j)$ and $p(y_i)$ will be denoted by z_j and y_i. The necessary equations are:

$$
\begin{aligned}
2z_1 \quad\quad + z_3 \quad\quad &= 4y_1 \\
z_1 + 4z_2 + z_3 \quad\quad &= 4y_2 \\
z_3 + z_4 &= 4y_3 \\
z_1 \quad\quad + z_3 \quad\quad &= 4y_4
\end{aligned}
$$

Solving:

$$
\begin{aligned}
z_1 &= 4y_1 - 4y_4 \\
z_2 &= y_2 - y_4 \\
z_3 &= -4y_1 + 8y_4 \\
z_4 &= y_1 + y_3 - 2y_4
\end{aligned}
$$

The restriction that $p(z_3) = 0$ implies that:

$$-4y_1 + 8y_4 = 0$$

The restriction that $p(z_3) = 0$ could also be imposed using matrix notation. We need only to solve for $p(z_3)$ in terms of the $p(y_i)$ $i = 1, 2, 3, 4$.

We know that:

$$[p(Z)][p(Y|Z)] = [p(Y)]$$

Using the method of determinants to solve for the $p(z_i)$, the following result is obtained.

$$p(z_j) = \frac{\begin{array}{c|ccccc|} \underline{\text{row}} & & & & & \\ 1 & p(y_1|z_1) & \cdots & p(y_i|z_i) & \cdots & p(y_n|z_1) \\ \vdots & & & & & \\ j & p(y_1) & \cdots & p(y_i) & \cdots & p(y_n) \\ \vdots & & & & & \\ m & p(y_1|z_m) & \cdots & p(y_i|z_m) & \cdots & p(y_n|z_m) \end{array}}{\begin{array}{c|ccccc|} 1 & p(y_1|z_1) & \cdots & p(y_i|z_1) & \cdots & p(y_n|z_1) \\ \vdots & & & & & \\ j & p(y_1|z_j) & \cdots & p(y_i|z_j) & \cdots & p(y_n|z_j) \\ \vdots & & & & & \\ m & p(y_1|z_m) & \cdots & p(y_i|z_m) & \cdots & p(y_n|z_m) \end{array}}$$

The method of determinants may now be used to solve for $p(z_3)$ in our example.

$$p(z_3) = \frac{\begin{vmatrix} 1/2 & 1/4 & 0 & 1/4 \\ 0 & 1 & 0 & 0 \\ y_1 & y_2 & y_3 & y_4 \\ 0 & 0 & 1 & 0 \end{vmatrix}}{\begin{vmatrix} 1/2 & 1/4 & 0 & 1/4 \\ 0 & 1 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 \end{vmatrix}}$$

Solving:

$$p(z_3) = 8y_4 - 4y_1$$

which is the result previously obtained. By imposing the restriction that $p(z_3) = 0$, the information gain equation becomes:

$$[\max]G(Y, Z) = [H(Y)]-[H(Y|Z)]$$

subject to

$$\sum_i p(y_i) = 1$$

and

$$-4y_1 + 8y_4 = 0$$

Putting in appropriate values for $H(Y)$ and $H(Y|Z)$ results in the following equation.

$$[\max]G(Y, Z) = -\sum_i p(y_i) \log_2 p(y_i) - \sum_i \sum_j p(z_j)p(y_i|z_j) \log_2 p(y_i|z_j)$$

$$+ \lambda(\sum_i p(y_i)-1) + \mu(-4y_1+8y_4)$$

Plugging in the appropriate values of $p(z_j)$ in terms of the $p(y_i)$:

$$[\max]G(Y, Z) = -\sum_i p(y_i) \log_2 p(y_i) - 1.5(4y_1-4y_4) + \lambda(\sum_i p(y_i)-1)$$

$$+ \mu(8y_4-4y_1)$$

Taking derivatives:

$$\frac{\partial[G(Y, Z)]}{\partial y_1} = -1 - \log_2 y_1 - 6 + \lambda - 4\mu = 0 \qquad (4.14)$$

$$\frac{\partial[G(Y,Z)]}{\partial y_2} = -1 - \log_2 y_2 + \lambda = 0 \qquad (4.15)$$

$$\frac{\partial[G(Y,Z)]}{\partial y_3} = -1 - \log_2 y_3 + \lambda = 0 \qquad (4.16)$$

$$\frac{\partial[G(Y,Z)]}{\partial y_4} = -1 - \log_2 y_4 + 6 + \lambda + 8\mu = 0 \qquad (4.17)$$

$$\frac{\partial[G(Y,Z)]}{\partial \lambda} = y_1 + y_2 + y_3 + y_4 = 1 \qquad (4.18)$$

$$\frac{\partial[G(Y,Z)]}{\partial \mu} = 8y_4 - 4y_1 = 0 \qquad (4.19)$$

Solving Equations (4.14)-(4.17) for the $p(y_i)$ gives:

$$p(y_1) = 2^{(\lambda-4\mu-7)}$$
$$p(y_2) = 2^{(\lambda-1)}$$
$$p(y_3) = 2^{(\lambda-1)}$$
$$p(y_4) = 2^{(\lambda+8\mu+5)}$$

The above values for $p(y_i)$ may now be substituted into Equation (4.19) to obtain $\mu = -1.057$ (Appendix C gives the equations for computing μ). Now, Equation (4.18) may be solved for 2^{λ} which is:

$$2^{\lambda} = \frac{1}{2^{-1.772}+2^{0}+2^{0}+2^{-3.456}}$$

Therefore:

$$p(y_1) = .1274$$

$$p(y_2) = .4045$$

$$p(y_3) = .4045$$

$$p(y_4) = .0637$$

Now, solving for $p(z_j)$ $j = 1, 2, 3, 4$ gives:

$$[.1274 \quad .4045 \quad .4045 \quad .0637] \begin{bmatrix} 1/2 & 1/4 & 0 & 1/4 \\ 0 & 1 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 1 & 0 \end{bmatrix}^{-1} \quad [p(z_1)p(z_2)p(z_3)p(z_4)]$$

and

$$p(z_1) = .2548$$

$$p(z_2) = .3408$$

$$p(z_3) = 0$$

$$p(z_4) = .4044$$

These values for $p(z_j)$ satisfy all the restrictions. Each $p(y_i) \geq 0$ each $p(z_j) \geq 0$ and $\sum_i p(y_i) = \sum_j p(z_j) = 1$.

Based on the result obtained, we would predict that 12.74% would choose Brand A; 40.5% Brand B; 40.5% Brand C; and 6.37% Brand D. Also, the $p(z_j)$ $j = 1, 2, 3, 4$ indicate that the factor "Price" gives us 25.5% of the information concerning the various brands; that the factor "Quantity" gives 34.1% of the gain in information; that "Reliability" contributes zero percent and "Warranty" 40.4%. It is interesting to note that the various conditional $p(y_i | z_j)$ probabilities for Reliability are all equal to 1/4. Previously, we showed that Information Gain was measured in terms of a deviation from the equally likely condition. It is thus justified to set $p(z_3)$ equal to zero. In fact, we would expect this condition to be true indicating that the factor "Reliability" does not help us any in predicting which brand will be chosen (Appendix D gives a computer solution to this problem).

Capital Allocation Example

The symbiotic use of a one-factor and multi-factor information theory analysis is illustrated in the following example. A manufacturer produces two grades of a particular product; Grade A and Grade XA. Grade A controls 40% of the manufacturer's total market share while Grade XA controls 60%. Management feels that two factors alone determine the purchasing habits of customers, "price" and "advertising expense." The price of Grade A is \$5.00 and Grade XA is \$9.00. Currently, the company is spending \$100,000 per year on advertising with \$40,000 being allocated to Grade A and \$60,000 to Grade XA. It is felt that, perhaps, one or more other factors can be found which might influence purchasers to buy either Grade A or XA. Since funds are tight, the company would like to take \$20,000 of their annual advertising budget to search for new factors and to reallocate the remaining $80,000 in order to take advantage of their old belief that purchasers use only "price" and "advertising" when buying either Grade A or Grade XA.

If "price" alone is used as a factor to predict market share, the one-factor information channel equation will be (Shirland, 1971):

$$W^{-5} + W^{-9} = 1$$

and from the 2 level/one-factor table of Appendix E, the corresponding prediction of market share will to 50% for Grade A and 60% for XA. Taken alone, price is seen to be a poor predictor of market share since the respective shares are known to be 40% for A and 60% for XA. A multi-factor analysis may then be conducted using as the required conditional probabilities, $p(y_1|z_1) = .60$ and $p(y_2|z_1) = .40$. z_1 corresponds to the factor "price" and y_1 and y_2 correspond to the market shares of Grade A and Grade XA respectively. Using z_2 to represent advertising expense, the conditional probability matrix will be:

$$\begin{array}{r} \\ \text{price } z_1 \\ \text{advertising } z_2 \end{array} \begin{array}{c} \begin{array}{cc} y_1 & y_2 \end{array} \\ \begin{bmatrix} .60 & .40 \\ p(y_1|z_2) & p(y_2|z_2) \end{bmatrix} \end{array}$$

with

$$p(y_1) = .40$$

$$p(y_2) = .60$$

A multi-factor information theory analysis results in:

$$p(y_1 | z_2) = .20$$

$$p(y_2 | z_2) = .80$$

Letting .20 and .80 be predictions of market share when advertising alone is used as an estimator, we wish to determine the values of L_1 and L_2 such that:

$$W^{-L_1} + W^{-L_2} = 1$$

where L_1 and L_2 represent the advertising expense for Grades A and XA respectively. The tables of Appendix E may be used to determine L_1 and L_2. From the table for 2 levels/one-factor a ratio of 1:7.23 is obtained for the advertising expenditure. The conclusion, then, is to allocate $12,150 as the advertising budget for Grade A and $67,850 to Grade XA, while $20,000 additional is being spend in an attempt to define and utilize other factors that might possibly be important.

This example has pedagodic value in that it illustrates the multi-factor information theory analysis in reverse. However, in practical situations, it could prove dangerous to shift advertising resources to such a great extent as is recommended in this example. One reason is that it has been assumed that purchases rely solely on "price" and "advertising" expenditure when buying the product in question. If that were completely true, the analysis given here would be valid. However, in actual situations many subtle forces such as the media used or the method of advertising may be as dominant as the mere dollar volume of expenditures. Also, a drop of $20,000 in advertising may upset the market share balance in favor of competitive firms or the mere method of resource allocation by competing firms might severely affect any shifts of resources in the example. By the same token, it could turn out that the analysis is valid and that the recommended shift will not affect the market share, or it could even increase the respective shares.

V. A CASE ANALYSIS USING ONE-FACTOR AND MULTI-FACTOR INFORMATION THEORY MODELS

Introduction

Buzzel (1969) presents the following historical background for an attitude study of five companies involved in manufacturing crushing and grinding equipment. Buzzel's data will be used in conjunction with the information theory analysis presented in this dissertation. The theoretical development of a one-factor information channel analysis was discussed by Shirland (1971).

Case History

The Diamond Company

> In October 1966, Mr. Mathew West, Marketing Manager of the Diamond Company, was reviewing a series of studies of awareness of and attitudes toward his company and competing companies. He was concerned with assessing the validity of the measures and sampling procedures for two reasons: (1) two sets of studies had yielded apparently conflicting results; and (2) one set of studies indicated a deterioration in attitudes toward Diamond, and this seemed inconsistent with the fact that sales in 1966 were up substantially over previous years.
>
> Company Background. The Diamond Company manufactured crushing and grinding equipment for sale to the chemical industry in the United States. Diamond's equipment was sold by a 35-man sales force. Diamond and four other companies competed for the bulk of the business. Mr. West believed that their market shares were roughly:

Collins, Inc.	15%
Diamond Company	15
Ferguson, Inc.	25
Hammond Machine Co.	20
Maxwell Manufacturing	20
All Others	5
	100%

Crushing and grinding equipment was used in a variety of applications in the chemical industry. Corporate management, production management, process engineering, and/or research personnel might be involved in the purchase of a piece of equipment.

Although some companies have developed new types of equipment in recent years, Diamond had not made significant changes in its product line for several years. Nevertheless, 1966 sales were up sharply over previous years.

Advertising. The major purpose of the awareness and attitude studies was to assess the effects of the company's advertising program. For several years, Diamond had advertised relatively heavily in trade magazines. The main objective of the campaigns had been to encourage the feeling that Diamond stood for leadership and progressiveness in the industry. Diamond also maintained a group of 12 service engineers which was regarded as a part of the sales force. Since this was considered by the company to be a well-above-average service effort as measured by industry standards, Diamond was also interested in assessing its image for repair service.

Shares of total industry advertising expenditures in major media were estimated as given in Table 16.

The Studies. Mr. West has commissioned two research studies, one in 1964 and one in 1966. Each study used the same questionnaire, the same cover letter, the same universe (the 20,000 name circulation list of the major trade magazine in the industry), and the same systematic sampling procedure. The sample size had been increased in 1966 to 1,000 from 600 in 1964.

Both studies produced about the same proportion of usable questionnaires (149 in 1964 and 244 in 1966)--a response rate of 25%.

Analysis revealed that in both studies, the distribution of respondents' job classifications and geographical locations closely approximated that of the circulation list used as the basis for selecting the sample.

Table 16. Advertising expenditures for grinding and crushing equipment.

Company	1964-1966	1963
Collins, Inc.	14%	7%
Diamond Company	23	34
Ferguson, Inc.	11	24
Hammond Machine Co.	20	13
Maxwell Manufacturing	32	22
Total pages of advertising 1964-66 (all companies)	152	100%

The principal results of the 1964 and 1966 attitude and awareness studies are contained in Tables 17-21.

Magazine Studies. In addition to the 1964 and 1966 studies which were conducted for Diamond by an outside research agency, Mr. West became aware of a series of trend studies which had been conducted by a chemical industry magazine. Its circulation list had served as the basis for the sample in the 1964 and 1966 studies.

Analysis. Buzzel was interested in noting changes in attitudes of customers over a previous study. The information theory analysis will be used to make inferences about customer attitudes. The following five tables contain the relative frequencies of replies to five questions posed by the research questionnaire. In order to obtain values for the five levels of the one factor "Attitude Rating," the various categories; Excellent, Good, Fair, Poor, and Don't Know are rated as 1, 2, 3, 4, and 5 respectively. One composite rating is then computed and used to obtain predictions of market share for each company using each question singularly. Then a multi-factor analysis is used to test for independence between various questions.

The questionnaire was presented as follows:

> Listed below are five major factors which contribute to the overall impression you have of various companies. Under each of these are listed the names of five companies which manufacture crushing and grinding equipment for the chemical process industry. We wish to learn your opinion of these companies, and ask that you rate them on each characteristic by placing a check under the rating you select. Please bear in mind your opinion is important to us, even though it may represent an impression which is not based on first-hand knowledge or experience (Buzzel, 1959).

Question #1. "Please rate the companies below in regard to fairness of price and value represented by their crushing and grinding equipment."

Table 17. Fairness of price--Diamond Company case study.

Rating	Companies				
	Collins	Diamond	Ferguson	Hammond	Maxwell
Excellent	.074	.111	.131	.078	.107
Good	.303	.390	.421	.267	.340
Fair	.074	.094	.152	.107	.197
Poor	.008	.026	.016	.024	.028
Don't Know	.541	.377	.279	.525	.328

Note: Number of responses in the 1966 survey consisted of 244 usable questionnaires.

Question #2. "Please rate the companies below in regard to service."

Table 18. Service rating--Diamond Company case study.

Rating	Companies				
	Collins	Diamond	Ferguson	Hammond	Maxwell
Excellent	.053	.086	.115	.070	.143
Good	.266	.283	.390	.246	.311
Fair	.086	.160	.152	.115	.143
Poor	.020	.041	.045	.030	.030
Don't Know	.573	.430	.300	.537	.369

Question #3. "Quality - In addition to the matter of good materials and workmanship, consider performance measuring up to promised specifications in your ratings."

Table 19. Quality--Diamond Company case study.

Rating	Companies				
	Collins	Diamond	Ferguson	Hammond	Maxwell
Excellent	.139	.143	.139	.078	.156
Good	.254	.328	.369	.221	.324
Fair	.057	.135	.176	.115	.143
Poor	.016	.012	.020	.016	.041
Don't Know	.532	.382	.295	.570	.336

Question #4. "Progressiveness - consider product innovation and up-to-dateness of products in your ratings."

Table 20. Progressiveness--Diamond Company case study.

Rating	Companies				
	Collins	Diamond	Ferguson	Hammond	Maxwell
Excellent	.074	.111	.090	.065	.143
Good	.226	.262	.356	.258	.324
Fair	.119	.160	.185	.103	.119
Poor	.033	.053	.057	.012	.025
Don't Know	.550	.414	.312	.562	.389

Question #5. "Please rate the companies in regard to over-all leadership as a source for any type of crushing and grinding equipment."

Table 21. Leadership--Diamond Company case study.

Rating	Companies				
	Collins	Diamond	Ferguson	Hammond	Maxwell
Excellent	.074	.144	.209	.053	.168
Good	.217	.320	.394	.304	.357
Fair	.131	.143	.111	.127	.135
Poor	.020	.045	.033	.020	.037
Don't Know	.656	.348	.254	.496	.303

Analysis

Treating each question as a one-factor, five level information channel and using the ratings as previously stated, the following "attitude rating" and probability predictions are presented in Table 22.

The probability predictions of Table 22 may be arranged in a conditional probability matrix for a multi-factor information analysis. The $P(Y|Z)$ matrix is:

	Collins y_1	Diamond y_2	Ferguson y_3	Hammond y_4	Maxwell y_5
Question #1 z_1	.1652	.2174	.2396	.1648	.2131
Question #2 z_2	.1669	.1956	.2387	.1737	.2251
Question #3 z_3	.1734	.2096	.2322	.1561	.2288
Question #4 z_4	.1765	.2035	.2286	.1722	.2243
Question #5 z_5	.1212	.2212	.2585	.1672	.2320

Table 22. Information theory analysis of Diamond Company case study.

Company	Attitude Rating	Probability Prediction
Question #1		
Collins	3.639	16.52%
Diamond	3.080	21.74
Ferguson	2.888	23.96
Hammond	3.654	16.48
Maxwell	3.130	21.31
Question #2		
Collins	3.788	16.69
Diamond	3.446	19.56
Ferguson	3.031	23.87
Hammond	3.714	17.37
Maxwell	3.159	22.51
Question #3		
Collins	3.542	17.34
Diamond	3.162	20.96
Ferguson	2.960	23.22
Hammond	3.779	15.61
Maxwell	2.977	22.88
Question #4		
Collins	3.765	17.15
Diamond	3.397	20.35
Ferguson	3.145	22.86
Hammond	3.748	17.22
Maxwell	3.193	22.43
Question #5		
Collins	4.261	12.12
Diamond	3.133	22.12
Ferguson	2.728	25.85
Hammond	3.602	16.72
Maxwell	2.950	23.20

It may be observed from the matrix (p. 85) that every row is nearly the same as every other and therefore, the results may not be independent. In fact, a multi-factor analysis reveals that, indeed, the results are not independent. This suggests that any one of the questions may be used to predict the market shares for the various crushing and grinding companies. Since we have five independent predictions, an average of the values may be a better indicator of market share. The average predicted values, then, would be:

Collins, Inc.	15.20%
Diamond Company	19.71
Ferguson, Inc.	22.80
Hammond Machine Co.	15.80
Maxwell Manufacturing	21.49
All others	5.00
	100.00%

The above results agree very well with the rough estimates of Mr. West except those for the Hammond Company and for Diamond Company. A closer look at the original 244 usable questionnaires suggests a possible reason for the low estimate for the market share of the Hammond Company. Since it was estimated by Mr. West that the Hammond Company controlled 20% of the market for crushing and grinding equipment, one would expect that a representative sample of the population would have an opinion as to the various qualities of that company. However, upon close examination of the "Don't Know" answers it appears that a large number of the respondants are not aware of the Hammond Company. Table 23 presents the number answering "Don't Know" for the five questions.

Table 23. Answers of Don't Know--Diamond Company case study.

Question	Collins	Diamond	Ferguson	Hammond	Maxwell
1	132	92	68	128	80
2	140	105	73	131	90
3	130	93	72	139	82
4	134	101	76	137	95
5	160	85	62	121	74

Table 23 shows that fewer persons answered "Don't Know" about the Ferguson Company (25% Market Share) and for the questions, the Collins Company (15% Market Share) had the most "Don't Know" answers for three of the five questions. These results were to be expected since the Ferguson Company and Collins Company were first and last respectively in Market shares. On the other hand, the Hammond Company which was estimated to have the second largest (20%) market share had the most "Don't Know" answers to two of the questions and had second most of the other three.

It would appear, then, that the sample taken did not contain a representative number of persons familiar with Hammond Company's product. Suppose then that the Hammond Company's market share is set at the estimated value of 20% and the market shares re-computed using the one-factor information theory analysis. The various market shares would then be:

Collins, Inc.	14.90%
Diamond Company	18.80
Ferguson, Inc.	21.30
Hammond Machine Co.	20.00
Maxwell Manufacturing	20.00
All others	5.00
	100.00%

The above estimates using information theory agree almost identically with the "rough estimates" of Mr. West. The analysis also points out what might be a very important concept. The fact that a multi-factor information theory analysis showed that each question could be used independently to predict market share suggests that persons might not be able to consider a question such as "rate the quality" or "rate the progressiveness." Instead, it may be that a person has a preconceived opinion of a company in general which is prevalent no matter what question he is asked. In other words, a person may not be able to isolate his opinion of a company's "quality" or "progressiveness" as compared to its "acceptance." Instead, he rates each of these qualities in terms of a combination of all the factors that make up his opinion of a particular company.

VI. INFORMATION THEORY APPLIED TO MARKOV CHAIN ANALYSIS

Introduction

A first order Markov process is a process with the property that, given the value of a state of the process Y_i, the values $Y_j; j > i$, do not depend on the values of $Y_k; k < i$. In other words, the probability of any particular future behavior of the process, when its present state is positively known, does not depend on its past behavior. A discrete time Markov chain is a Markov process whose state space is a countable set (Karlan, 1969).

Some areas in which Markov chain analysis has been employed are:

1. One-dimensional random walks (Kemeney and Snell, 1965).
2. Queueing models (Pritsker, 1969).
3. Inventory analysis (Hadley and Whitin, 1963).
4. Genetic models (Karlan, 1969).
5. Simulation models (Inoue, 1971).
6. Marketing research applications (Green, 1964).
7. Quality control acceptance programs (Koyama, et al., 1970).

Usually, it is desired to determine the steady-state probabilities of a Markov chain after the process has reached stochastic equilibrium. The following four theorems are useful in determining these probabilities for any Markov chain.

Theorem 1. In a positive recurrent aperiodic class with states $j = 0, 1, 2, \ldots,$

$$\lim p_{jj}^{n} = \pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij}$$

$$\sum_{i=0}^{\infty} \pi_i = 1$$

and the π's are uniquely determined by the set of equations:

$$\pi_i \geq 0$$

$$\sum_{i=0}^{\infty} \pi_i = 1$$

$$\pi_j = \sum_{i=0}^{\infty} \pi_i p_{ij}$$

any set of $(\pi_i)_{i=0}^{\infty}$ satisfying the above equations is called a stationary probability distribution of the Markov chain (Karlan, 1969).

Theorem 2. A state i is recurrent if and only if:

$$\sum_{n=1}^{\infty} p_{ii}^{n} = \infty$$

Theorem 3. If state i communicates with state j and if i is recurrent then j is recurrent.

Theorem 4. The period of a state i is the greatest common divisor of all integers, $n \geq 1$, for which $p_{ii}^{n} > 0$.

Brand Switching Example

Consider three manufacturers of a product labeled Brand A, Brand B, and Brand C respectively. It has been found through a market research survey that there is a certain degree of brand loyalty apparent in this market system, but also, there is some brand switching by retail customers. It has been observed that of the persons purchasing Brand A in one period 10% will re-purchase Brand A during the next period, 10% will purchase Brand B and 80% Brand C. Of those purchasing Brand B in a particular period 70% will purchase A and 30% will repeat. For Brand C, 40% of those who purchase it in one period will switch to Brand A in the next period, 50% will switch to Brand B and 10% will repeat.

The individual manufacturers are interested in determining the answers to the following questions.

1. After n periods what will be the respective market shares for the three brands?
2. At stochastic equilibrium, what is the market share for each

manufacturer?

3. How many periods does it take for the Markov process to reach stochastic equilibrium?

Most Markov chain analyses involve determining answers to questions 1 and 2. However, information theory may be extremely useful in answering question 3. It may also be used in answering questions 1 and 2. Generally, it is very easy to determine the steady-state or stationary probability distribution by normal Markov chain analysis. If the transition matrix consists of a positive recurrent aperiodic class, the steady state probabilities are simply the left eigenvector $(\pi_1, \pi_2, \ldots, \pi_n)$ associated with the class. The eigenvalue associated with the class will be 1, and the right eigenvector $(\phi_1, \phi_2, \ldots, \phi_n)$ corresponding to the absorption probabilities will be of the form $(1, 0, 0, \ldots, 0)$. In many cases, stochastic equilibrium will not be reached until a large number of steps (periods in our example) have occurred. In these cases, it is not realistic to consider the steady-state probabilities since the conditions under which the initial transition probability matrix was obtained will likely change with time. Question 3, then, is very important in drawing conclusions concerning a Markov chain. At present, the easiest method of determining the answer to this question is through a trial and error procedure. However, recently there has been increasing interest in simulating Markov processes. Several computer programs are available which simulate each event with the intent of heuristically computing the steady-state probabilities along with other pertinent data concerning the process. For example GERT (Pritsker, 1970); MSIP (Inoue, 1970; Inoue et al., 1971; Ghaffari, 1970) are two such programs. Recent authors who have published their research concerning Markov process simulation include Hennie (1968), Howard (1960), Kemeny and Snell (1960), and Koyama et al. (1970).

This dissertation presents an information theory analysis that may be used to predict the number of steps to reach the stochastic equilibrium conditions. This method should prove useful in making inferences about a particular Markov chain and to set limits on the number of iterations for a Markov chain simulation. Another advantage is that the information theory analysis will permit a researcher to begin his simulation at equilibrium, thus eliminating costly calculations while the process is in the transient states.

n-Step Transition Probabilities

It may be of interest to predict the probability distribution or market share for each brand after the system has been operating for 1, 2, 3, etc. up to n time periods. To calculate the market shares it is only necessary to multiply the one-step transition probability matrix by itself n times, where n is the number of steps desired, and then pre-multiply the resulting matrix by the initial market conditions at the start of the analysis. In our example, we know nothing concerning the respective market shares at the beginning of the analysis. We should choose, then, the condition of maximum uncertainty and use the equally likely case where

$P(A) = P(B) = P(C) = 1/3$.

From the statement of the problem, the one-step transition probability matrix is shown in Table 24.

Table 24. Transition probabilities for Markov chain analysis.

(Period i) From	(Period i + 1) to A	B	C
A	.1	.1	.8
B	.7	.3	0
C	.4	.5	.1

and

$$P(A_0) = 1/3$$

$$P(B_0) = 1/3$$

$$P(C_0) = 1/3$$

After n steps our market share prediction will be:

$$[P(A_0), P(B_0), P(C_0)] \begin{bmatrix} .1 & .1 & .8 \\ .7 & .3 & 0 \\ .4 & .5 & .1 \end{bmatrix}^n = [P(A_n), P(B_n), P(C_n)]$$

$$[P(A_2), P(B_2), P(C_2)] = [1/3, 1/3, 1/3] \begin{bmatrix} .40 & .44 & .16 \\ .28 & .16 & .56 \\ .43 & .24 & .33 \end{bmatrix}$$

or

$$P(A_2) = .37$$

$$P(B_2) = .28$$

$$P(C_2) = .35$$

Table 25 contains predictions for the market shares of Brands A, B, and C for the seven periods calculated.

Table 25. Markov chain prediction of market shares.

Period	Brand A	Brand B	Brand C
1	.4000	.3000	.3000
2	.3700	.2800	.3500
3	.3730	.2960	.3310
4	.3769	.2916	.3315
5	.3744	.2909	.3347
6	.3750	.2921	.3330
7	.3750	.2916	.3334

<u>Stochastic Equilibrium Probabilities</u>

Using Theorems 1 through 4 stated previously, we may calculate a set of π_j such that

$$\sum_{i=1}^{3} \pi_i p_{ij} = \pi_j , \qquad j = 1, 2, 3.$$

The calculations for our brand share example will be:

$$[\pi_A, \pi_B, \pi_C] \begin{bmatrix} .1 & .1 & .8 \\ .7 & .3 & .0 \\ .4 & .5 & .1 \end{bmatrix} = [\pi_A, \pi_B, \pi_C]$$

Solving the above system of equations results in:

$$\pi_A = .3750$$
$$\pi_B = .2917$$
$$\pi_C = .3333$$

Therefore, over a sufficiently long period of time the market share of Brands A, B, and C would be 37.50%, 29.17% and 33.33% respectively regardless of what the shares were initially. In this example, we assumed the equally likely condition of 1/3, 1/3, 1/3. However, the steady-state probabilities will be as calculated given any initial conditions.

Information Theory Analysis of n-Step Transition Probabilities

In the previous chapters of this dissertation, we have been using information theory to predict various probabilities for a conditional probability matrix. In the information theory analysis, we have predicted the probabilities of the various outcomes of Experiments Y and Z when we could observe the conditional probability of

Experiment Y given the results of Experiment Z. A Markov chain analysis may be viewed in the same manner. The states of the system at time i may be viewed as an Experiment Z and the states at time $i + 1$ as another Experiment called Y. The previously discussed multi-factor analysis may then be used to predict $p(y_i)$; $i = 1, 2, \ldots, n$, which corresponds to a prediction of the various market shares in the brand switching example. However, the information theory analysis is based on the results of conducting the experiment once. In other words, we may calculate the probabilities of $P(Y)$ and $P(Z)$ for the first step in the Markov chain analysis by using the matrix $[P(Y|Z)]$. For steps $2, 3, \ldots, n$ respectively, we would use the probabilities associated with the transition from step 2 to 3, 3 to 4, etc. up to $n - 1$, to n. More simply, we would use the probability matrix associated with $[P(Y|Z)]^2; [P(Y|Z)]^3; \ldots; [P(Y|Z)]^n$ respectively. It may be noted that at successive steps in the Markov chain analysis, as n gets large, the probabilities in each column approach a common value. In other words, when stochastic equilibrium has been reached $p_{11} = p_{21} = \cdots = p_{m1}$; $p_{1j} = p_{2j} = \cdots = p_{mj}$; and $p_{1m} = p_{2m} = \cdots = p_{im} = p_{mm}$. Stated simply, each row in the probability matrix will be identical. Therefore, the information gained by the matrix will be zero. Now, if we could use information theory to calculate the number of steps to reach the condition of zero information gain, we could use this value as a predictor of when stochastic equilibrium is reached.

The brand switching example will now be used to show that the multi-factor information theory analysis can be used to:

1. Predict the probability values in the transitional stages of a Markov process.
2. Predict the number of steps to reach stochastic equilibrium in a Markov process.

The multi-factor analysis of Chapter III was used to develop Table 26. This table contains the individual transition matrices that were used in the information theory analysis along with the pertinent results of the analysis. The following notation is used in Table 26.

Column 1. Step i: Gives the steps of the transitional probability matrix as:

$$\text{step } 1 = [p(Y|Z)]$$
$$\text{step } 2 = [p(Y|Z)]^2$$
$$\vdots$$
$$\text{step } i = [p(Y|Z)]^i$$
$$\vdots$$
$$\text{step } m = [p(Y|Z)]^m$$

Column 2. The states of the process as Brand A, Brand B and Brand C.

Column 3. The transitional probability matrix as:

$$[p(Y|Z)]^{\text{step } 1}$$
$$\vdots$$
$$[p(Y|Z)]^{\text{step } i}$$
$$\vdots$$
$$[p(Y|Z)]^{\text{step } m}$$

Column 4. $p(y_i)$: the prediction of the market shares of Brands A, B, and C respectively.

Column 5. $p(z_j)$: the calculation of the amount of importance or the relative amount of information gain contributed by each row in the transition probability matrix.

Column 6. H_i: calculated as:

$$\sum_{i=1}^{3} p(y_i|z_j) \log p(y_i|z_j) \quad \text{for} \quad j = 1, 2, 3$$

Column 7. $H(Y)$: the information transmitted by Experiment Y alone.

$$H(Y) = \sum_{i=1}^{3} p(y_i) \log_2 p(y_i)$$

Column 8. $H(Y|Z)$: the conditional information transmitted by Experiment Y when the results of Experiment Z are known.

Table 26. Summary of information theory data for brand switching example.

Step	State	Probability Matrix			$p(y_i)$	$p(z_j)$	H_i	H(Y)	H(Y|Z)
1	A	.1	.1	.8	.4362	.4396	.9219	1.5268	.899155
	B	.7	.3	.0	.2121	.5604	.8813		
	C	.4	.5	.1	.3517	.0000	1.3096		
2	A	.400	.440	.160	.3418	.5150	1.47294	1.44032	.141691
	B	.280	.160	.560	.3042	.4850	1.40568		
	C	.430	.240	.330	.3540	.0000	1.54552		
3	A	.412	.252	.336	.3744	.3207	1.55685	1.57928	1.569183
	B	.364	.356	.280	.3046	.4417	1.57539		
	C	.343	.280	.377	.3210	.2376	1.57429		
4	A	.352	.285	.363	.3713	.4667	1.57698	1.57788	1.575682
	B	.398	.283	.319	.2911	.2289	1.57038		
	C	.381	.307	.312	.3376	.3044	1.57768		
5	A	.380	.302	.318	.3728	.5018	1.57780	1.57810	1.577252
	B	.366	.284	.350	.2933	.4982	1.57670		
	C	.378	.286	.336	.3339	.0000	1.57581		
6	A	.377	.288	.335	.3762	.5004	1.57630	1.57732	1.577402
	B	.376	.297	.327	.2922	.4997	1.57819		
	C	.373	.292	.335	.3316	.0000	1.57782		
7	A	.373	.292	.335	.3748	.4292	1.57774	1.57749	1.577484
	B	.376	.290	.334	.2918	.1987	1.57694		
	C	.376	.293	.331	.3334	.3721	1.57748		

In order to prove the validity of the information theory calculations for predicting the various market shares for Brands A, B, and C, two statistical tests (the paired t-test and Wilcoxen Signed Rank Test) will be conducted using the results of the Markov chain analysis and the information theory analysis. Table 27, Table 28 and Table 29 summarize the data necessary for conducting a paired t-test for the predictions for Brands A, B, and C respectively.

Table 27. Market share predictions for brand A.

Step Number	Market Share by Markov Chain Analysis X_1	Market Share by Information Theory Analysis X_2	Difference X_1-X_2
1	.4000	.4362	-.0362
2	.3700	.3418	.0282
3	.3730	.3744	-.0014
4	.3769	.3716	.0056
5	.3744	.3728	.0016
6	.3750	.3762	-.0012
7	.3751	.3748	.0003

Table 28. Market share predictions for brand B.

Step Number	Market Share by Markov Chain Analysis X_1	Market Share by Information Theory Analysis X_2	Difference X_1-X_2
1	.3000	.2121	.0879
2	.2800	.3042	-.0242
3	.2960	.3046	-.0086
4	.2916	.2911	.0005
5	.2909	.2933	-.0024
6	.2921	.2922	-.0001
7	.2917	.2918	-.0001

Table 29. Market share predictions for brand C.

Step Number	Market Share by Markov Chain Analysis X_1	Market Share by Information Theory Analysis X_2	Difference X_1-X_2
1	.3000	.3517	-.0517
2	.3500	.3540	-.0040
3	.3310	.3210	.0100
4	.3315	.3376	-.0061
5	.3347	.3339	.0008
6	.3330	.3316	.0014
7	.3333	.3334	-.0001

Figure 18 shows a plot of the prediction of the market shares of Brands A, B, and C using the two methods; Markov chain analysis and information theory analysis.

Wine (1964) gives the following statistic for conducting a paired t-test:

$$t = \frac{\bar{d} - \delta}{\frac{s_d}{\sqrt{n}}} \quad \text{with} \quad n-1 \text{ degrees of freedom,}$$

where:

$\bar{d}$ = mean of the difference between pairs of observations.

s_d = standard deviation of the paired differences.

$$s_d = \frac{\sum_{i=1}^{n} d_i^2 - (\sum_i d)^2/n}{n-1}$$

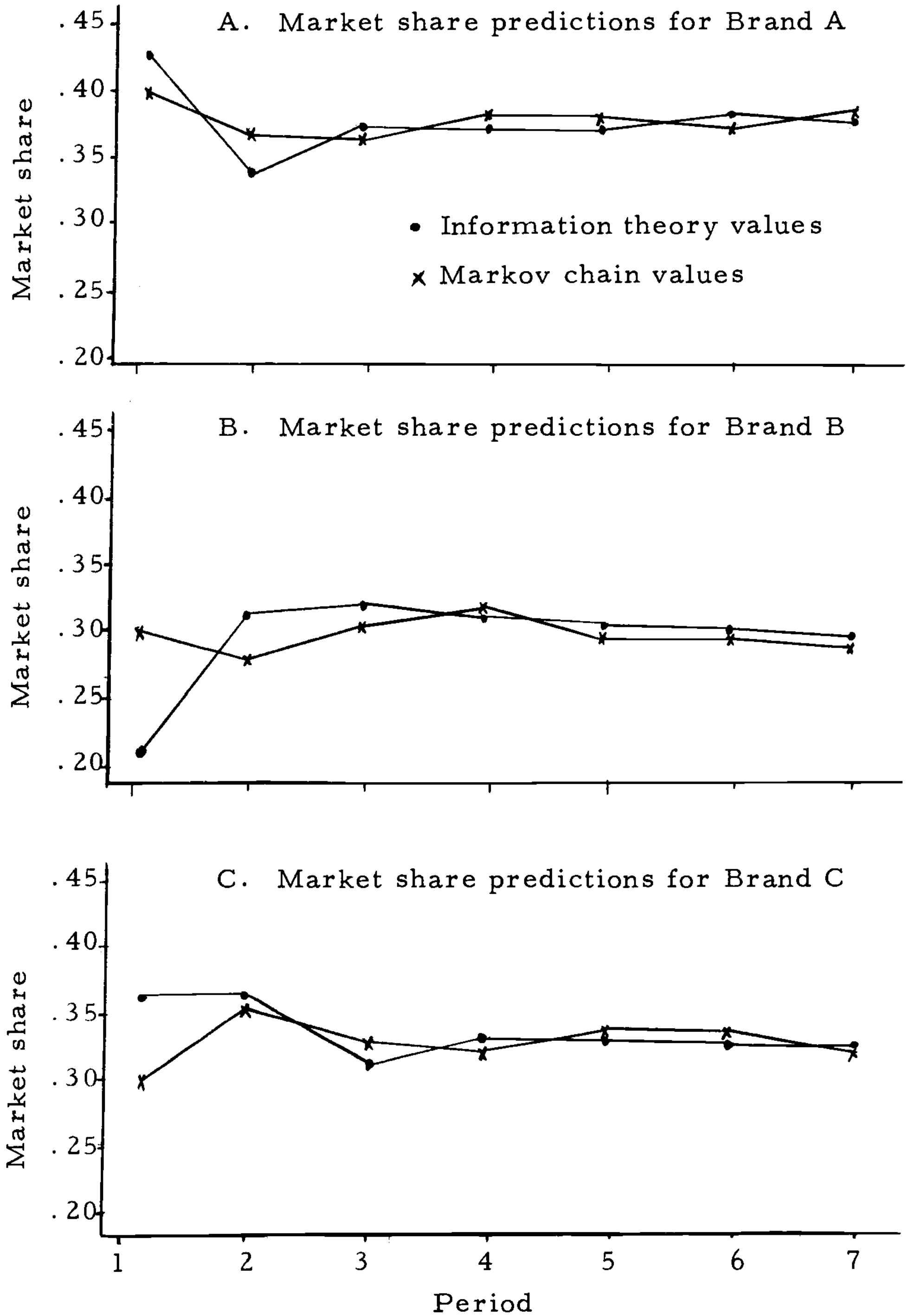

Figure 18. Plot of market share predictions for Brands A, B, and C.

n = number of observations.

δ = hypothesized population mean.

Table 30 summarizes the t-test data for the three Brands A, B, and C. In this example, we wish to test whether the two prediction methods are identical by setting $\delta = 0$. In Table 30,

$$\sum_{i=1}^{7} d_i = \text{the sum of the paired differences.}$$

$$\sum_{i} d_i^2 = \text{the sum of squares of the paired differences.}$$

t_{crit} = the t value at the .05 significance level with 6 degrees of freedom.

Table 30. Summary of t-test data.

Variable	d_i	d_i^2	$\bar{d}$	s_d	t	$t_{crit(.05)}$
A	-.0231	.00214309	-.0033	.0185601	-0.470417	2.447
B	.0530	.00839204	.00757	.0364937	0.548816	2.447
C	.0487	.00282871	.00696	.0188600	.976375	2.447

Since each calculated t value is less than t_{crit} at the .05 significance level, we cannot reject the hypothesis that the two methods of predicting market share are the same. We may conclude that in the absence of additional information the information theory analysis gives predictions not significantly different from the Markov chain analysis.

In order to use the paired t-test we must assume that the two populations from which data are drawn are normally distributed. As a further test of our data a distribution free test as suggested by Wilcoxon will be conducted. The Wilcoxon signed rank test uses both the order and magnitude properties of the paired t-test but the normality requirement is not necessary for the Wilcoxon test (Wine, 1964).

In order to conduct a Wilcoxon signed rank test, the paired differences are first arranged according to increasing order of the absolute value of the differences; that is, for Brand A:

Table 31. Wilcoxon test data for Brand A.

Order	.0003	-.0012	-.0014	.0016	.0056	.0282	-.0362
Rank	1	2	3	4	5	6	7

Next, assign ranks from 1 to n for each difference as was done above. Finally, find the sum T of the positive or negative ranks, whichever is smaller, and compare with the appropriate critical value from Table 34. If T is smaller than the critical value, reject the null hypothesis that the two methods are the same; otherwise fail to reject the null hypothesis.

For brand A: $T_+ = 1 + 4 + 5 + 6 = 16$

$T_- = 2 + 3 + 7 = 12$

Since $T_- = 12$ is smaller we use it and fail to reject the null hypothesis since for seven pairs of values, we reject if $T < 2$ at the .05 significance level.

Similarly, for Brand B the ordered values and their ranks are:

Table 32. Wilcoxon test data for Brand B.

Order	-.0001	-.0001	.0005	-.0024	-.0086	-.0242	.0879
Rank	1.5	1.5	3	4	5	6	7

Since there are two values of .0001 we assign as rank the average of their respective ranks or 1 + 2 divided by 2 = 1.5.

Now, $T_+ = 10$

$T_- = 18$

Since $T_+ = 10 > 2$, we cannot reject the hypothesis that the two methods are the same.

Table 33. Wilcoxon test data for Brand C.

Order	-.0001	.0008	.0014	-.0040	-.0061	.0100	-.0517
Rank	1	2	3	4	5	6	7

For Brand C: $T_+ = 11$

$T_- = 17$

Again, since $T_+ = 11 > 2$, we cannot reject the hypothesis that the two methods are the same.

Table 34. Critical values of T for Wilcoxon's signed rank two sided test.

Pairs	Probability		
n	.05	.02	.01
6	0	-	-
7	2	0	-
8	4	2	0
9	6	3	2

From the results of the two statistical tests conducted on the data in the brand switching example we can conclude that the two methods of predicting the marketing shares for the various brands are not significantly different.

Information Theory Used to Predict the Number of Steps to Reach Stochastic Equilibrium

It was stated previously that the gain in information $G(Y, Z)$ will be zero when stochastic equilibrium is reached since each row of the probability matrix will be identical. To show that this statement is true consider the following:

Given the stochastic equilibrium probability matrix:

$$
\begin{array}{llc}
\text{Column} & & \begin{matrix} 1 & \;\;2 & \;\;\; j & \;\;\; m \end{matrix} \\
\text{Row} & 1 & \left[\begin{matrix} P_1 & P_2 \cdots P_j \cdots P_m \\ \vdots & \\ P_1 & P_2 \cdots P_j \cdots P_m \\ \vdots & \\ P_1 & P_2 \cdots P_j \cdots P_m \end{matrix}\right] \\
 & i & \\
 & m &
\end{array}
$$

Where:

$$P_1 = p(y_1|z_1) = p(y_1|z_2) = \ldots = p(y_1|z_m)$$

$$P_2 = p(y_2|z_1) = p(y_2|z_2) = \ldots = p(y_2|z_m)$$

$$\vdots$$

$$P_m = p(y_m|z_1) = p(y_m|z_2) = \ldots = p(y_m|z_m)$$

Since each row in the probability matrix is the same, one row alone contributes 100% of the information concerning the matrix or $p(z_1) = 1$ and $p(z_2) = \ldots = p(z_m) = 0$.

Then $p(y_1) = P_1$; $p(y_2) = P_2$; $\ldots$; $p(y_m) = P_m$.

The information gain equation:

$$G(Y, Z) = \sum_i p(y_i) \log_2 p(y_i) - \sum_j \sum_i z_j p(y_i|z_j) \log_2 p(y_i|z_j)$$

Plugging in appropriate values for the z_j:

$$G(Y, Z) = \sum_i P_i \log_2 P_i - 1 \sum_i P_i \log_2 P_i = 0$$

Therefore, at stochastic equilibrium, the information gain will be zero. If it were possible, then, to predict the number of iterations needed to reach the point of zero information gain, we could have a predictor of the number of iterations to reach equilibrium. The brand switching example will be used to show that the information gain equation follows an exponential type distribution. Linear regression will then be employed using as data the information gain values for the first few iterations in order to obtain an estimate of the number of steps to reach a point of zero information gain.

The gain in information for the first seven iterations of the brand switching example are presented in Table 35.

Table 35. Information gain for brand switching example.

Step	Information Gain (bits) H(Y) - H(Y/Z)	Log_e (Gain)
1	.627673	- 0.465736
2	.141691	- 1.954107
3	.010098	- 4.595418
4	.002200	- 6.119298
5	.000849	- 7.071451
6	.000084	- 9.350000
7	.000004	-10.000000

Table 35 shows that after six iterations the gain in information is zero to four decimal places. In fact, from the Markov chain analysis for iteration six the respective probabilities for Brands A, B, and C were .375, .292, and .333 which are the same as the computed stationary probability distribution using Theorems 1 through 4 listed previously. Figure 19 shows a plot of the information gain for the seven iterations for the brand switching example.

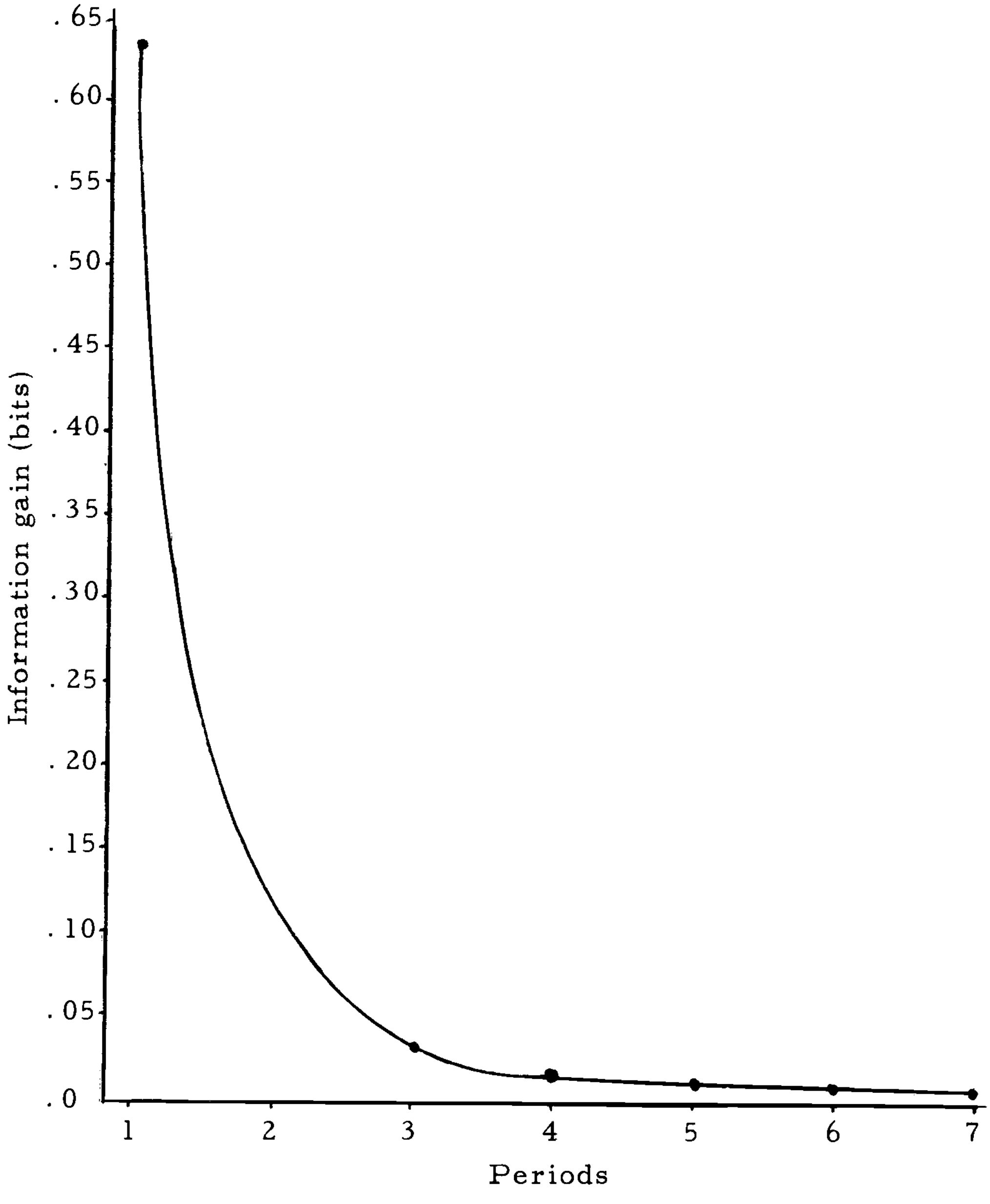

Figure 19. Plot of information gain for brand switching example.

From Figure 19 it may be seen that the gain in information follows an exponential type distribution. A non-linear regression analysis could be used directly to determine the equation of this curve. However, we wish to make our predictions based on as few data points as possible so it will be expedient to make a log (base e) transformation of the data. Table 35 gives the logarithms to base e for the seven data points. Figure 20 shows a plot of the transformed data points.

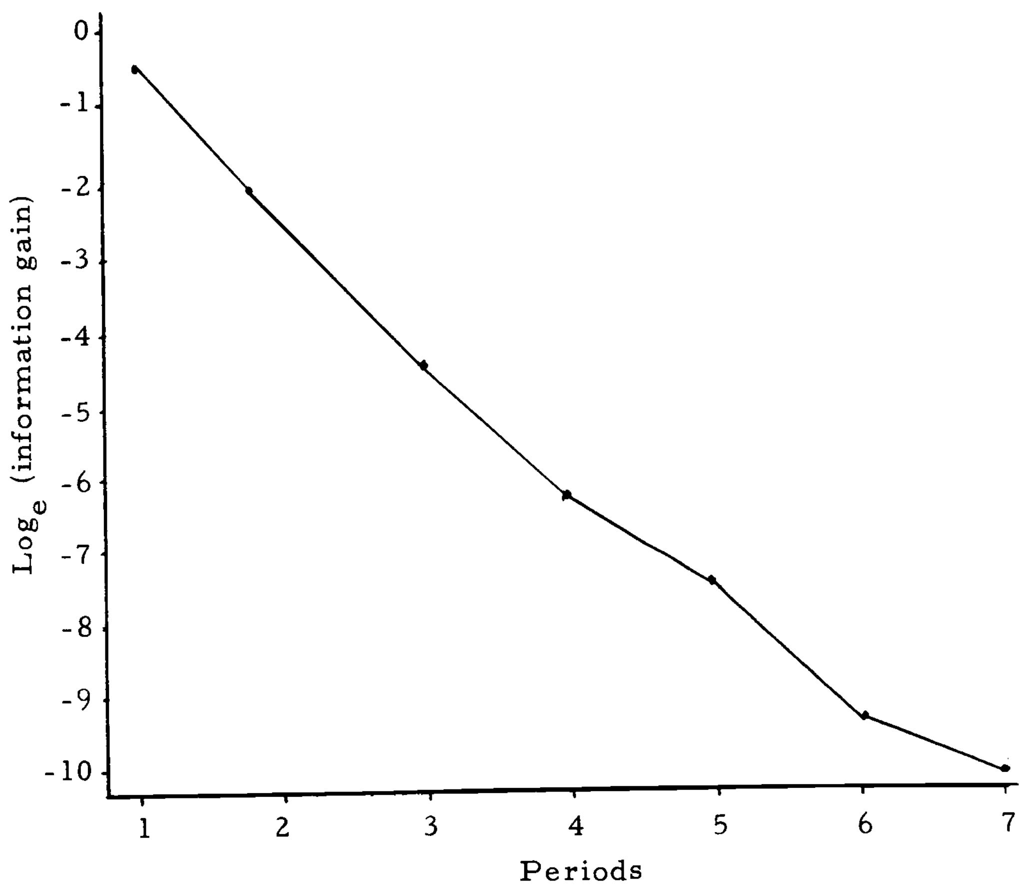

Figure 20. Plot of transformed data for the information gain in a brand switching example.

The plot of the transformed data does not follow exactly a straight line. However, we must decide on the number of iterations to use for the predictive equation. If all seven data points were used, a quadratic or cubic term as well as a linear term might be desirable. For this particular example we will employ simple linear regression using only the first two data points to make our prediction. Then, successively three, four and five points will be used to note any change in the prediction. Table 36 summarizes the linear regression data and gives the predictions for the number of steps to reach stochastic equilibrium. For cases II, III, and IV the regression equation is based on values of X of -1, 0, 1, 2 etc. instead of X = 1, 2, 3, 4 to facilitate ease of computations; X being the (i-2) iteration.

Table 36. Predictions of the number of steps for a Markov chain to reach stochastic equilibrium.

Prediction Based on Steps	Prediction Equation	Predicted Number of Trials to Reach Stochastic Equilibrium
1, 2	Y = -1.48837X + 1.022635	7.4 ≈ 8
1, 2, 3	Y = -2.33842X - 2.530577	5.2 ≈ 6
1, 2, 3, 4	Y = -2.21039X - 2.146494	5.6 ≈ 6
1, 2, 3, 4, 5	Y = -2.21039X - 1.830819	5.7 ≈ 6

From Table 36 it may be noted that using the first two iterations of the Markov process, the prediction is eight steps to reach stochastic equilibrium while using three, four, or five iterations results in a prediction of six. This is an important result in several respects. For computer programs that rely on matrix multiplication to compute steady-state probabilities, the usual method is to set an arbitrarily large limit on the number of iterations to insure that stochastic equilibrium has been reached. Instead, the information theory analysis could be used to predict an acceptable limit for the number of iterations, thus reducing unnecessary costs of intermediate steps. In Monte Carlo simulations, only one event is sampled at a time. Therefore, it takes an even greater number of calculations to determine the steady-state probabilities. The information theory approach could be even more useful in these simulations since the simulation could be started at a point at or near stochastic equilibrium, thereby saving the costs of many calculations.

VII. IMPLICATIONS AND EXTENSIONS

Implications of Multi-Factor Information Theory Model

This dissertation has presented three areas in which multi-factor information theory models may be applied; (1) Bayesian decision analysis, (2) estimation and (3) Markov chain analysis. In Bayesian decision analysis, it was shown that information theory could be used to obtain the same results as with traditional methods. An advantage of the information theory approach is that it offers a measure of sensitivity in a particular problem. For example, in the urn problem it was found that sampling procedures (a), (b), and (c) resulted in values of .685, .691, and .697 respectively for the probabilities of guessing correctly the urn from which one had been sampling. The information theory approach showed that sampling procedure (a) resulted in a 16% gain in information over complete uncertainty; procedure (b) showed a gain of 18.5% over complete uncertainty and a 3.08% gain over procedure (a); procedure (c) was concluded to be the best since it contributed a 29.03% gain over complete uncertainty and a 12.9% gain over procedure (b). Information theory, then, may be used as a decision tool in determining before an experiment has been conducted whether enough will be learned to make the experiment worthwhile. In this way the feasibility of collecting additional information can be examined.

Information gain may also be used to determine an optimal point for management's willingness to pay for additional information by equating it to the cost of collecting that information. This concept should be especially useful in the area of Quality Control and Sampling. The main advantage over traditional methods is that the information theory approach answers the question: "Exactly how much is learned from the collection of additional information and is it worthwhile costwise to gather additional information?" Of secondary importance is the fact that the information theory approach bases its interpretation of the problem situation on all the outcomes of an experiment and not on just one or two specific outcomes.

When utilizing information theory as an estimating technique, the results obtained showed that sampling strategies that produce a maximum information gain are identical to those obtained by traditional methods. However, the information theory analysis is easier to apply, and offers a novel interpretation and a clearer understanding of the information processing mechanism.

A multi-factor information channel analysis is very useful in many areas of marketing research. For example, the results of questionnaires may be analyzed in an attempt to determine market share. The basic hypothesis is that consumers are influenced by a variety of factors when they make purchases; price, advertising, color, location, packaging, quality of product or company image, etc. The information theory analysis can help to identify those factors which are being considered. Once defined, these factors can be altered in an attempt to increase market share or to improve the overall image of a particular company. In the case analysis presented in Chapter V, five questions were analyzed in an attempt to determine if they could be used to predict market share. The solution showed that any of the five questions could be used independently as an estimator of market share. This result was significant because it suggested that persons were not able to separate their opinions of a particular company into levels of "fairness of price," "service value," "quality," "progressiveness" and "leadership." Instead, the analysis led one to believe that, perhaps, it was a combination of these factors that influenced a customer's opinion of a particular company but he was simply not able to differentiate between different factors. A manufacturer should then be cautious in his use of information supplied by questionnaires of this type. For example, if a manufacturer were interested in determining customer attitude concerning the quality of his product, he should remember that answers may well contain feelings about many other aspects of his company as well and not on quality alone.

The information theory analysis may serve to illuminate problem areas of negative or positive feelings about a company. If upon analyzing the results of questionnaires, several factors are seen to be independent as far as estimating market share is concerned while one is not, that factor may distinguish positive or negative reactions about the company. For example, in the case study presented in this dissertation, suppose that four of the questions could be used independently to predict market share but the question concerning quality could not. This might indicate that customers are heavily influenced by this factor. If the analysis showed that using quality as a factor resulted in a low estimate of market share, perhaps the company could improve its position through advertising or by improving its image as far as quality was concerned. If the factor, quality, results in a high market share estimate, it may indicate that the company's quality is considered exceptional. Again, a decision could be made to further capitalize on this fact through advertising or other means.

In any practical problem concerning Markov processes there are two phases, a transitional phase characterized by constantly fluctuating probability values and a stochastic equilibrium or steady-state phase corresponding to long run aspects of the process. It may be of practical interest to determine exactly how many steps it takes to reach stochastic equilibrium. For example, in an inventory problem one might like to develop a stocking policy for merchandise. The usual approach is to calculate the steady-state probabilities and base a decision on those values. However, it may take an unusually large number of periods for the particular process to reach the equilibrium phase. In this case, the stochastic equilibrium condition is of no interest since estimates generated to obtain values for the study will vary with time and are therefore only reliable over a short planning horizon. It would be very desirable, then, to calculate the number of steps to reach equilibrium. This dissertation presented a procedure using the concept of information gain which could be used to predict the number of iterations of the process needed to reach stochastic equilibrium.

Montgomery (1969) states that the Markov chain model of customer behavior is subject to certain limitations, one of which is the need for the development of statistical methods to render a model more empirically viable. In recommendations for future research, Montgomery states that there is a need for the development of multi-dimensional Markov, diffusion and learning models. The information theory model presented in this dissertation is one answer to both of these recommendations.

Extensions for Future Research

In its present state, the Markov model of consumer behavior is subject to limitations which identify areas for future research. First, the model is not able to account for heterogeneity among consumers in terms of their transition probabilities. Another generalization would be to allow customers to vary their opinions or reactions with time. Second, developments in additional methods to test and validate the assumptions are needed.

In the area of human behavior, it would be extremely desirable to show that in any given situation, consumer behavior can be analyzed as a communication channel. If this were the case, any situation could be analyzed by identifying those factors upon which consumers rely when making purchases. The challenge, then, involves first, studying a wide variety of specific problems in an attempt to identify the specific factors involved. Secondly and of more importance, is the need for a method of systematically searching for the pertinent identifiable factors that explain consumer behavior. Of course, as in most mathematical techniques, data collection is an important problem area. If reliable data were available the criticism of most mathematical techniques would be alleviated.

Another area of interest for future research is in the area of learning models. Traditionally, information theory was applied to the human operator in order to determine learning rates. Similar models could be studied using consumer behavior as a communication mechanism. A fundamental aspect of learning theory is the idea that the occurrence of a response will increase the probability that a consumer will repurchase the product at a later date. In fact, most available learning models explicitly state that purchase event feedback affects the future response probability. Kuehn (1962) postulated a linear learning model for consumer brand choice. Information theory suggests a logarithmic function. Several specific examples exist that suggest that information theory may be very applicable in the area of consumer learning. Haines (1964) developed a learning model using aggregate market measures which were found to be asymptotic. He then used regression analysis which resulted in a modified linear learning model. Future research could show that a model based on information theory is more realistic in certain situations than either Kuehn's or Haines' models.

Montgomery (1969) states that there is a need to test the empirical viability of stochastic models for many product classes and to compare various models. The information theory model presented in this dissertation should be compared to other models for many sets of data.

Additional work is needed in the area of evaluating the costs of gathering information versus gain in information. Usually, cost consideration problems are analyzed in an attempt to optimize manufacturing a product at the point of lowest total cost. However, in many cases, especially in R&D situations, minimum cost might not be of interest. For example, suppose a specified amount of capital is available for quality and reliability studies of a product. Several tests can be conducted each costing a specified amount. The questions of interest are; "Which tests should be conducted?" and "How much testing should be done?" Instead of minimum total cost, the problem might concern itself with allocating capital in a manner that will provide the most information about the product. One example of this type of analysis was described in this dissertation. More research is needed, however, to develop improved methods which may be tested in actual problem situations.

More research is needed in the area of identifying random effects or noise that severely affect an information theory model. Also, the actions of competing alternatives should be studied in an attempt to determine the effect a change in one will have on the information theory model.

In models similar to the capital allocation model of page 76, calculations become increasingly cumbersome as more factors are introduced into the model. More research is needed to develop methods to handle these problems.

Shannon's early research into stochastic processes prompted Kunisawa (1958) to suggest that a department of a company might be analyzed by studying the flow of memos and communications between members of a work group. It may be possible to treat one person in the group as a transmitter sending information using the paperwork channel to a receiver represented by the person receiving the information. A similar analysis could be made using the flow of machine parts in a factory. Various departments such as the assembly area, inspection area, manufacturing area, etc. would act as transmitters and receivers and the volume of movement of various parts as the information channel. If information theory could be shown to be applicable to either of these situations, a new and powerful application will have been found which will extend the usefulness of information theory models.

VIII. CONCLUSIONS

This dissertation has attempted to present the extreme versatility of information theory. Claude Shannon (1949) who pioneered information theory research, avoided the criticisms of other researchers of his time by specifying that his theory only applied to the technical problems of communication and not to the semantic or pragmatic problems with which social scientists were concerned. Technical problems as investigated by Shannon involve the accuracy of transmitting symbols of communication. Semantic problems, on the other hand, deal with how precisely transmitted symbols convey the desired meaning. Problems involving motivation and understanding fall into this category. Pragmatic problems are concerned with how successful the meaning of a message transmitted to the receiver leads to the desired result.

A mathematical treatment of consumer motivation has been presented in this dissertation. The premise is that consumers are deluged by positive and negative signals and messages which motivate them to purchase or not to purchase a particular product. These signals are transmitted to consumers by means of specific factors such as price, advertising, packaging, function, appeal, etc. Some factors may act as positive motivators such as a low price, snob appeal or other desirable feature and negative motivators such as a high price, poor quality, etc. Acting collectively, these factors determine consumer behavior. Stated categorically, any consumer behavior problem could be analyzed mathematically by identifying and measuring the specific factor or factors that influence consumers when purchasing a particular product. Practically speaking, however, it would be an impossible task to identify every factor involved. Instead, it may be possible to identify a sufficient number of factors which can be used to predict behavior. All the undefined factors would then be considered as noise. In many cases, one factor will be sufficient, in others two or more will be required. Each situation would have to be individually analyzed in order to determine the particular information theory model that applied.

Markov chains have been extensively used as consumer learning models, game theory models, queueing models, etc. The mathematical treatment of consumer motivation presented in this dissertation can be used to explain Markov chain learning models. The Markov chain model is predicated on the assumption that once a person has purchased a particular product, there is a distinct probability that he will repurchase or switch products during the next purchase period. Steady-state probabilities are then calculated to determine the long run purchasing habits of consumers. Employing the multi-factor analysis suggests that the fact that a person had or had not purchased a particular product would act as a factor in conjunction with all others to either positively or negatively influence the repurchase of the product during the next purchase period. A multi-factor information theory model, as well as a Markov chain analysis can be used, then, to predict purchase habits of consumers. At steady-state the Markov chain analysis and the multi-factor information theory model result in identical solutions. During transitional phases, simulation of the Markov process and the information theory models are seen to be closely correlated.

Another important application of information theory models is in the prediction of the point at which steady-state conditions of a Markov chain analysis will be reached. Using the results of the first few steps in a Markov chain analysis, the information theory models give a fairly accurate prediction of the number of steps to reach stochastic equilibrium. An analysis of this type could prove useful in simulation situations. For those simulations in which stochastic equilibrium is of prime importance, the information theory model could be used to predict the point at which equilibrium is reached. The simulation could then be begun at this point saving many costly calculations.

The information theory models presented in this dissertation can also be useful in economic decision situations in which it is desired to obtain a maximum amount of information per dollar and not simply to operate at minimum total cost. For example, in the area of quality and reliability control, several tests may be available which can be used to determine specific characteristics for a product. The information theory model could be used to determine the number of tests to conduct that will result in allocation of capital such that a maximum amount of information about that product results.

Again, the power of the information theory models presented in this dissertation appears to be in their potential for explaining how consumers behave, how Markov chain analyses are used as learning models, and how decisions can be made. For these reasons, information theory should prove to be another useful addition to the fields of Operations Research and decision analysis.

BIBLIOGRAPHY

Abramson, Norman. 1963. Information theory and coding. New York, McGraw-Hill. 201 p.

Ash, Robert. 1965. Information theory. New York, Interscience Publishers. 339 p.

Bell, David Arthur. 1962. Information theory and its engineering applications. 3rd ed. London, Pitman Publishing Corporation. 196 p.

Bierman, Harold, Jr., Charles P. Bonini and Warren H. Hausman. 1969. Quantitative analysis for business decisions. Homewood, Richard D. Irwin. 358 p.

Bierman, Harold, Jr., Lawrence E. Fouraker and Robert K. Jaedicke. 1961. Quantitative analysis for business decisions. Homewood, Richard D. Irwin. 358 p.

Brillouin, Leon. 1964. Scientific uncertainty and information. New York, Academic Press. 164 p.

Brown, Douglas B. and Martin R. Warshaw. 1965. Media selection by linear programming. Journal of Marketing Research 2:83-88.

Bush, Robert R. and Frederick Mosteller. 1951. A mathematical model for simple learning. Psychological Review 58:313-328.

Buzzel, Robert D. and Charles C. Slater. 1962. Decision theory and marketing management. Journal of Marketing 26:7-16.

Buzzel, Robert D., Donald F. Cox and Rex V. Brown. 1969. Marketing research and information systems. New York, McGraw-Hill. 788 p.

Cox, Donald P. Editor. 1967. Risk taking and information handling in consumer behavior. Boston, Division of Research, Graduate School of Business Administration Harvard University. 667 p.

Darden, William R. 1968. An operational approach to product pricing. Journal of Marketing 32:29-33.

Davies, R.H. and B.P. Welford. 1969. A simple way to analyze multi-way tables using information theory. Journal of the Market Research Society 11:147-151.

Duncan, Acheson J. 1965. Quality control and industrial statistics. Homewood, Richard D. Irwin. 992 p.

Farley, John V. 1964. Why does brand loyalty vary over products? Journal of Marketing Research 1:9-14.

Field, J.G. 1970. Whether psychology. Journal of the Market Research Society 12:12-24.

Flagle, Charles D., William H. Huggins and Robert H. Roy. 1960. Operations research and systems engineering. Baltimore, Md. The Johns Hopkins Press. 889 p.

Frank, Ronald E., Alfred A. Kuen and William F. Massy. 1962. Quantitative techniques in marketing analysis. Homewood, Richard D. Irwin. 556 p.

Goldman, Stanford. 1953. Information theory. Englewood Cliffs, Prentice-Hall. 385 p.

Green, Paul E. 1963. Bayesian decision theory in pricing strategy. Journal of Marketing 27:5-14.

Green, Paul E. 1964. Bayesian classification procedures in analyzing customer characteristics. Journal of Marketing Research 1:44-50.

Green, Paul E., Patrick J. Robinson and Peter T. Fitzroy. 1967. Experiments on the value of information in simulated marketing environments. Boston, Allyn and Bacon, Inc. 194 p.

Gue, Ronald L. and Michael E. Thomas. 1968. Mathematical methods in operations research. New York, The MacMillan Company. 385 p.

Hadley, G. and T.M. Whitin. 1963. Analysis of inventory systems. Englewood Cliffs, N.J., Prentice-Hall. 452 p.

Hake, Harold W. and Ray Hyman. 1953. Perception of the statistical structure of a random series of binary symbols. Journal of Experimental Psychology 45:64-74.

Haines, G. 1964. A theory of market behavior after innovation. Management Science 10:634-658.

Halter, Albert N. and Gerald W. Dean. 1971. Decisions under uncertainty. Cincinnati, South-Western Publishing Co. 266 p.

Hennie, Frederick C. 1968. Finite-state models for logical machines. New York, John Wiley and Sons. 466 p.

Hermann, Robert O. and Leland L. Beik. 1968. Shoppers movements outside their local retail areas. Journal of Marketing 32:45-51.

Howard, Ronald A. 1960. Dynamic programming and markov processes. Cambridge, Mass., M.I.T. Press. 136 p.

Inoue, Michael S. 1971. Markov simulation by iterative process: A pedagogic tool in man-machine symbiosis. In: Fourth Annual Simulation Symposium Record of Proceedings, ed. by Dennis K. Caldwell and Barbara Dawson. New York, Gordon and Beach Science Publishers. p. 295-341.

Irwin, Francis and W.A.S. Smith. 1957. Value, cost and information as determiners of decision. Journal of Applied Psychology 54:229-232.

Karlan, Samuel. 1969. A first course in stochastic processes. New York, Academic Press. 502 p.

Kemeny, John G. and J. Laurie Snell. 1960. Finite markov chains. Princeton, N.J., D. Van Nostrand. 216 p.

Kotz, Samuel. 1966. Recent results in information theory. London, Methuen Co. 96 p.

Koyama, Takeshi, Yoshitsugu Ohmae, Ryonosuke Suga, Taro Yamamoto, Tsuneo Yokoh, and William R. Rabst. Mil-std-105 D and the Japanese modified standard. Journal of Quality Technology. Vol. 2 No. 2 (April 1970).

Kuehn, A.A. 1962. Consumer brand choice--a learning process? Journal of Advertising Research 2:10-17.

Kullback, Solomon. 1959. Information theory and statistics. New York, John Wiley and Sons, Inc. 395 p.

Kunisawa, K. 1959. Introduction to information theory for operations research. Japanese Union of Scientists and Engineers (JUSE). in Japanese. 150 p.

Lipstein, Benjamin. 1968. The dynamics of brand loyalty and brand switching. In: Analytical viewpoints in marketing management, ed. by Keith K. Cox, Englewood Cliffs, Prentice-Hall. p. 177-191.

Luce, R. Duncan. 1960. Mathematical psychology. Glencoe, Illinois, The Free Press. 294 p.

Luce, R. Duncan. 1959. Individual choice behavior. New York John Wiley and Sons, Inc. 153 p.

Machol, Robert E. 1960. Information and decision processes. New York, McGraw-Hill. 185 p.

Machol, Robert E. and Paul Gray. 1962. Recent developments in information and decision processes. New York, The MacMillan Company. 197 p.

McMillan, Calude and Richard F. Gonzalez. 1968. Systems analysis: a computer approach to decision models. Homewood, Richard D. Irwin. 520 p.

Moles, Abraham. 1968. Information theory and esthetic perception, tr. by Joel E. Cohen. Urbana, University of Illinois Press. 255 p.

Montgomery, David B. and Glen L. Urban. 1969. Management science in marketing. Englewood Cliffs, Prentice-Hall. 376 p.

Montgomery, David B. and Frederick Webster, Jr. 1968. Application of operations reserach to personal selling strategy. Journal of Marketing 32:50-57.

Muroga, Saburo. 1953. On the capacity of a discrete channel. Journal of the Physical Society of Japan 8:484-494.

Parfitt, J.M. and B.J.K. Collins. 1968. Use of consumer panels for brand-share prediction. Journal of Marketing Research 32:1-8.

Pritsker, A. Alan B. and Phillip J. Kiviat. 1969. Simulation with GASP II. Englewood Cliffs, Prentice-Hall. 332 p.

Quastler, Henry. 1955. Information theory in psychology; problems and methods. Proceedings of a conference on the estimation of information flow, Monticello, Illinois. July 5-9, 1954, and related papers. Glencoe, Illinois, Free Press. 436 p.

Raisbeck, Gordon. 1963. Information theory. Cambridge, Mass., Massachusetts Institute of Technology Press. 105 p.

Riggs, James L. 1968. Economic Decision Models. New York, McGraw-Hill. 401 p.

Roberts, Harry V. 1963. Bayesian statistics in marketing. Journal of Marketing 27:1-4.

Rothstein, Jerome. 1958. Communication, organization and science. Indian Hills, Colorado, Falcon's Wing Press. 110 p.

Shannon, Claude E. and Warren Weaver. 1949. The mathematical theory ofcommunication. Urbana, University of Illinois Press. 125 p.

Shirland, L. E.. 1971. An application of one-factor and two-factor information channel models to marketing research problems. Master's Thesis, Corvallis, Oregon State University. 97 numb. leaves.

Styen, George and Harry Smith, Jr. 1964. Markov chains applied to marketing. Journal of Marketing Research 1:50-55.

Theil, Henri. 1967. Economics and information theory. Amsterdam, North-Holland Publishing Company. 488 p.

Udell, Jon G. and Evan E. Anderson. 1968. The product warranty as an element of competitive strategy. Journal of Marketing 32:1-8.

Wagner, Harvey M. 1963. Practical slants on operations research. Harvard Business Review 41:61-71.

Wasson, Chester R. 1965. Research analysis for marketing decision. New York, Appleton-Century-Crofts. 278 p.

Watanabe, Satosi. 1969. Knowing and guessing. New York, John Wiley and Sons. 592 p.

Webster, Frederick E., Jr. 1968. On the applicability of communication theory to industrial markets. Journal of Marketing Research 5:425-428.

Weiss, Doyale L. 1968. Determinants of market share. Journal of Marketing Research 5:290-295.

Wine, Lowell R. 1964. Statistics for scientists and engineers. Englewood Cliffs, Prentice-Hall. 671 p.

Wolff, Ronald. 1966. Lecture notes in queueing. Berkeley, University of California. 67 numb. leaves.

Woodward, P. M. 1953. Probability and information theory, with applications to radar. London, Pergamon Press Ltd. 123 p.

APPENDICES

APPENDIX A

Information Gain Equation Proof

Proof that

$$G(Y, Z) = H(Y) - (Y|Z) = H(Z) - H(Z|Y)$$

where

$$H(Y) = -\sum_{i} p(y_i) \log p(y_i)$$

and

$$H(Y|Z) = -\sum_{i} \sum_{j} p(z_j)[p(y_i|z_j) \log p(y_i|z_j)]$$

We know from Bayes theorem that:

$$p(y_i|z_j) = \frac{p(y_i)p(z_j|y_i)}{p(z_j)}$$

Substituting the value of $p(y_i|z_j)$ as determined from Bayes theorem into the information gain equation, the following result is obtained:

$$G(Y, Z) = -\sum_{i} p(y_i) \log p(y_i) + \sum_{i} \sum_{j} p(y_i)p(z_j|y_i) \log \frac{p(y_i)p(z_j|y_i)}{p(z_j)}$$

Expanding the terms containing logs:

$$G(Y, Z) = -\sum_i p(y_i) \log p(y_i) + \sum_i \sum_j p(y_i)p(z_j|y_i) \log p(y_i)$$

$$+ \sum_i \sum_j p(y_i)p(z_j|y_i) \log p(z_j|y_i)$$

$$- \sum_i \sum_j p(y_i)p(z_j|y_i) \log p(z_j)$$

By definition,

$$H(Z|Y) = -\sum_i \sum_j p(y_i)p(z_j|y_i) \log p(z_j|y_i)$$

From above, $\sum_i \sum_j p(y_i)p(z_j|y_i) \log p(z_j|y_i)$ may be written as:

$$\sum_i p(y_i) \log p(y_i) \sum_j p(z_j|y_i)$$

but

$$\sum_j p(z_j|y_i) = 1$$

by definition, so

$$\sum_i \sum_j p(y_i)p(z_j|y_i) \log p(z_j|y_i) = \sum_i p(y_i) \log p(y_i)$$

which cancels with the like term in the information gain equation.

Now $-\sum_i \sum_j p(y_i)p(z_j|y_i) \log p(z_j)$ can be written

$$-\sum_j \log p(z_j) \sum_i p(y_i)p(z_j|y_i)$$

but

$$p(z_j) = \sum_i p(y_i)p(z_j|y_i)$$

Therefore:

$$G(Y, Z) = -\sum_j p(z_j) \log p(z_j) + \sum_i \sum_j p(y_i)p(z_j|y_i) \log p(z_j|y_i)$$

and by definition

$$H(Z) = -\sum_j p(z_j) \log p(z_j)$$

so:

$$G(Y, Z) = H(Z) - H(Z|Y) = H(Y) - H(Y|Z)$$

which completes the proof.

APPENDIX B

Justification for Equation 3. 10

The following is a proof that derivatives of $x \log_2 x$ in a multi-channel information theory model may be differentiated as $1 + \log_2 x$ instead of first transforming $\log_2 x$ to $\frac{\log_e x}{\log_e 2}$ and then differentiating.

Consider the objective function (Equation 3. 7):

$$\max [G(Y, Z)] = \max - \sum_{i=1}^{m} p(y_i) \log_2 p(y_i) - \sum_{i=1}^{m} b_i p(y_i)$$

$$+ \lambda \left(\sum_{i=1}^{m} p(y_i) - 1 \right)$$

In order to differentiate $\sum_i p(y_i) \log_2 p(y_i)$, $\log_2 p(y_i)$ should be transformed to $\frac{\log_e p(y_i)}{\log_e 2}$; since $\log_a b = \frac{\log_c b}{\log_c a}$. For notational purposes $1/\log_e 2$ will be defined to be C. The information gain equation may now be written as:

$$\max [G(Y, Z)] = \max - C \sum_i p(y_i) \log_e p(y_i) - \sum_i b_i p(y_i)$$

$$+ \lambda \left(\sum_i p(y_i) - 1 \right)$$

Differentiating:

$$\frac{\partial[G(Y, Z)]}{\partial(y_i)} = -C - C \log_e p(y_i) - b_i + \lambda = 0 \qquad \text{(B. 1)}$$

$$\frac{\partial[G(Y, Z)]}{\partial(\lambda)} = \sum_i p(y_i) - 1 = 0 \qquad \text{(B. 2)}$$

Dividing Equation (B. 2) by C gives:

$$-1 - \log_e p(y_i) - \frac{b_i - \lambda}{C} = 0$$

or

$$p(y_i) = e^{\left(\frac{-b_i + \lambda}{C} - 1\right)} \qquad \text{(B. 3)}$$

Summing over i from $i = 1, 2, \ldots, m$ and noting Equation (B. 2):

$$1 = \sum_i e^{\left(\frac{-b_i + \lambda}{C} - 1\right)}$$

which may be written as:

$$1 = e^{\lambda/C} \sum_i e^{\left(\frac{-b_i}{C} - 1\right)}$$

or

$$e^{(\lambda/C)} = \frac{1}{\sum_i e^{\left(\frac{-b_i}{C} - 1\right)}} \qquad \text{(B. 4)}$$

Plugging the value of $e^{\lambda/C}$ from Equation (B. 4) into Equation (B. 3) gives:

$$p(y_i) = \frac{e^{\left(\frac{-b_i}{C} - 1\right)}}{\sum_i e^{\left(\frac{-b_i}{C} - 1\right)}}$$

Factoring out an e^{-1} from both numerator and denominator gives:

$$p(y_i) = \frac{e^{\frac{-b_i}{C}}}{\sum_i e^{\frac{-b_i}{C}}} \qquad \text{(B. 5)}$$

Now, we wish to solve the following identity for x.

$$e^{\frac{-b_i}{C}} = 2^x$$

Taking the natural logarithm of both sides:

$$-\frac{b_i}{C} = x \log_e 2$$

But

$$\log_e 2 = 1/C$$

so,

$$-\frac{b_i}{C} = \frac{x}{C}$$

or

$$x = -b_i$$

Therefore,

$$e^{\left(\frac{-b_i}{C}\right)} = 2^{-b_i}$$

So Equation (B. 5) may be written:

$$p(y_i) = \frac{2^{-b_i}}{\sum_{i=1}^{m} 2^{-b_i}} \tag{B. 6}$$

The above result, Equation (B. 6) completes the desired proof of the validity of Equation (3. 11) of Chapter III.

APPENDIX C

Justification for Using $\sum_{i=1}^{m} a_i 2^{(-b_i+a_i\mu)}$ to Calculate the Lagrange Multiplier Constant μ in a Multi-Factor Information Theory Problem with Negative Probabilities

When a multi-channel information theory problem results in non-positive probabilities, an additional restriction must be imposed for each negative value. The necessary restrictions are of the form:

$$\sum_{i=1}^{m} a_i p(y_i) = 0$$

where a_i; $i = 1, 2, \ldots, m$; are constants determined by the particular problem data.

The information gain equation which must be maximized will now become:

$$\max [G(Y, Z)] = \max - \sum_i p(y_i) \log_2 p(y_i) - \sum_i b_i p(y_i)$$

subject to:

$$\sum_i p(y_i) = 1$$

and

$$\sum_i a_i p(y_i) = 0$$

Using the Lagrange Multiplier technique with multipliers λ and μ results in the following maximization equation.

$$\max\,[G(Y,Z)] = \max\; - C\sum_i p(y_i)\log_e p(y_i) - \sum_i b_i p(y_i)$$

$$+ \lambda\left(\sum_i p(y_i) - 1\right) + \mu\left(\sum_i a_i p(y_i)\right) \qquad (C.0)$$

In Equation (C. 0), it may be noted that the $\log_2 p(y_i)$ term has been transformed to $\log_e p(y_i)$ by the use of the relationship that $\log_a b = (\log_c b/\log_c a)$.

Taking appropriate derivatives:

$$\frac{\partial[G(Y,Z)]}{\partial(y_i)} = -C - C\log_e p(y_i) - b_i + \lambda + a_i\mu = 0 \qquad (C.1)$$

$$\frac{\partial[G(Y,Z)]}{\partial(\lambda)} = \sum_i p(y_i) - 1 = 0 \qquad (C.2)$$

$$\frac{\partial[G(Y,Z)]}{\partial(\mu)} = \sum_i a_i p(y_i) = 0 \qquad (C.3)$$

Solving Equation (C. 1) for $p(y_i)$ gives:

$$p(y_i) = e^{\left(\frac{-b_i+\lambda+\mu a_i}{C} - 1\right)} \qquad (C.4)$$

Inserting the result of Equation (C. 4) into Equation (C. 3) results in the following equation:

$$\sum_i a_i e^{\left(\frac{-b_i+\lambda+\mu a_i}{C}-1\right)} = 0 \qquad \text{(C. 5)}$$

Now, $e^{\left(\frac{-\lambda}{C}-1\right)}$ may be factored out of Equation (C. 5) to get:

$$\sum_i a_i e^{\left(\frac{-b_i+\mu a_i}{C}\right)} = 0 \qquad \text{(C. 6)}$$

If we let $(-b_i+\mu a_i) = K$, Equation (C. 6) becomes:

$$\sum_i a_i e^{K/C} = 0 \qquad \text{(C. 7)}$$

But, in Appendix B, it was shown that $e^{K/C} = 2^K$. Therefore, Equation (C. 7) may be written as

$$\sum_i a_i 2^K = 0$$

or

$$\sum_i a_i 2^{(-b_i+a_i\mu)} = 0 \qquad \text{(C. 8)}$$

which is the desired result. Equation (C. 8) is employed in the computer program of Appendix D. For purposes of hand calculations, it is probably easier to use logarithms to base e. In that case use:

$$\sum_i a_i e^{\left(\frac{-b_i + a_i \mu}{C}\right)} = 0$$

where $C = 1/\log_e 2$.

When μ has been obtained, λ, or more appropriately $e^{\lambda/C}$ may be computed from Equation (C. 2) as:

$$\sum_i e^{\left(\frac{-b_i + \lambda + \mu a_i}{C} - 1\right)} = 1$$

Factoring out $e^{\lambda/C}$ gives:

$$e^{\lambda/C} \sum_i e^{\left(\frac{-b_i + \mu a_i}{C} - 1\right)} = 1$$

or

$$e^{\lambda/C} = \frac{1}{\sum_i e^{\left(\frac{-b_i + \mu a_i}{C} - 1\right)}}$$

Now $p(y_i)$ may be calculated by plugging in the appropriate value for $e^{\lambda/C}$ into Equation (C.4).

$$p(y_i) = \frac{e^{\left(\frac{-b_i+\mu a_i}{C} - 1\right)}}{\sum_i e^{\left(\frac{-b_i+\mu a_i}{C} - 1\right)}}$$

factoring out e^{-1} gives:

$$p(y_i) = \frac{e^{\left(\frac{-b_i+\mu a_i}{C}\right)}}{\sum_i e^{\left(\frac{-b_i+\mu a_i}{C}\right)}}$$

Let $(-b_i+\mu a_i) = K$ then,

$$p(y_i) = \frac{e^{K/C}}{\sum_i e^{K/C}}$$

But as shown in Appendix B; $e^{K/C} = 2^K$, so

$$p(y_i) = \frac{2^K}{\sum_i 2^K}$$

or,

$$p(y_i) = \frac{2^{(-b_i+\mu a_i)}}{\sum_i 2^{(-b_i+\mu a_i)}} \qquad (C.9)$$

APPENDIX D

Computer Program for Solving Multi-Factor Information Channel Problems

The following computer program calculates all values for one-factor, two-factor and multi-factor information theory problems as described in this dissertation. The user must enter a value of 1, 2 or 3 depending upon whether he wishes a one, two, or multi-factor analysis respectively. The one-factor analysis portion of the program calculates the respective p(i) for i = 1, 2, . . . , n for the various levels of the factor, conducts a Chi-square test if one is desired and determines whether to accept or reject the hypothesis that the actual and theoretical values are identically distributed. An initial value to start the approximation must be entered by the user along with the values of the various levels of the factor being considered.

In the two-factor analysis, the program calculates the importance values Y(1) and Y(2) for the two factors. The probability values for each row must be entered by the user along with the marginal frequencies for each level.

In a multi-factor analysis, the user must enter data as follows:

1. The number of rows in the conditional probability matrix.
2. The specific values of the conditional probability matrix by rows as; p(1, 1), p(1, 2), . . . , p(1, m); p(2, 1), p(2, 2), . . . , p(2,m); . . . ; p(m, 1), p(m, 2), . . . , p(m, m).

After the data have been entered, the computer will then calculate

1. The inverse of the conditional probability matrix if one exists.
2. The conditional information values $h(i)$, $i = 1, \ldots, n$.
3. The necessary constants, $b(i)$; $i = 1, \ldots, n$ for calculating $p(y_i)$.
4. The $p(y_i)$.
5. The $p(z_j)$.
6. A check is conducted to insure that all restrictions are satisfied, i.e., $\sum_i p(y_i) = 1$, $\sum_j p(z_j) = 1$, $p(y_i) \geq 0$, $p(z_j) \geq 0$, for each i and j.
7. If one or more of these restrictions are violated, the program will impose additional restrictions and recalculate the values of $b(i)$, $p(y_i)$ and $p(z_j)$.

The program is written in conversational form and is designed to be used with the CDC-3300; OS-3 time sharing system with Fortran IV compilor at Oregon State University. A unique feature of the system is the availability of the statement TTYIN which allows data to be entered directly from a remote unit without the necessity of READ statements. For example, if a program were written in which it were desired to enter an array of values for the variable X, the following statements would be sufficient.

```
      DO 1 I=1,N
    1 X(I)= TTYIN (4HX= )
```

During execution of the program, the computer will ask for data to be entered by printing, X = . The operator will then enter his value for X_1 and the computer will again print, X = . The process is repeated until all N values of X have been entered.

```
      PROGRAM SHIRLAND
      DIMENSION P(15,15),E(15,15),H(15),D(15,15),B(15)
      DIMENSION Y(15),Z(15),F(15),X(15),A(15),P1(15),OBS(15)
      DIMENSION TOB(15)
      DIMENSION P2(2,50),X1(50)
 37   CONTINUE
      WRITE (61,508)
 508  FORMAT(' TYPE 1 IF THIS IS A ONE-FACTOR',
     1' ANALYSIS',/,' TYPE 2 IF TWO-FACTOR OR',
     2' 3 IF MULTI-FACTOR'//)
      NTYPE=TTYIN(4H?=  )
      IF(NTYPE .EQ. 1)GO TO 56
      IF(NTYPE .EQ. 2)GO TO 36
      IF(NTYPE .EQ. 3)GO TO 34
      IF(NTYPE .EQ. 0)GO TO 37
      IF(NTYPE .GT. 3)GO TO 37
C
C
C     THIS PART OF THE PROGRAM CALCULATES THE ROOT OF AN
C     EQUATION OF THE FORM 1/X + 1/X**A1 + 1/X**A2 ETC = 1.0
C     FOR ANY COMBINATION OF UP TO TEN VALUES FOR X, USING THE
C     NEWTON RHAPSON METHOD.
C
C
 510  FORMAT(1H ,F10.5)
56    CONTINUE
      DO 38 I=1,10
38    X(I)=0.0
      WRITE(61,108)
 108  FORMAT(//,' ENTER AN INITIAL VALUE TO START THE',
     I' APPROXIMATION',/,' USUALLY A NUMBER BETWEEN',
     2'  1 AND 2',//)
      R=TTYIN(4HINIT,4H=   )
      WRITE(61,100)
 100  FORMAT(//,' ENTER THE POWERS TO WHICH THE VALUES',
     1' ARE TO BE RAISED'/' ENTER A ZERO FOR THE LAST ',
     2' VALUE IN ORDER TO STOP ENTERING DATA',//)
      DO 39 I=1,10
      A(I)=TTYIN (4H A= )
      IF(A(I) .EQ.0.0)GO TO 40
39    CONTINUE
40    DO 41 I=1,10
      IF(A(I).EQ.0.0)42,41
42    X(I)=0.0
      GO TO 43
 41   X(I)=1./R**A(I)
43    FUNCT=X(1)+X(2)+X(3)+X(4)+X(5)+X(6)+X(7)+X(8)
     1+X(9)+X(10) -1.0
      IF(FUNCT.GT.  .00009)44,45
45    CONTINUE
      IF(FUNCT .LT. -.00009)44,46
 44   DERIV=-A(1)/R**(A(1)+1.)-A(2)/R**(A(2)+1.)
     1 -A(3)/R**(A(3)+1.)-A(4)/R**(A(4)+1.)-A(5)/R**(A(5)+1.)
     2 -A(6)/R**(A(6)+1.)-A(7)/R**(A(7)+1.)
     3 -A(8)/R**(A(8)+1.)-A(9)/R**(A(9)+1.)
     4 -A(10)/R**(A(10)+1.)
      R=R-FUNCT/DERIV
      GO TO 40
46    CONTINUE
      WRITE(61,510)R
C
C
C     THIS PART OF THE PROGRAM CALCULATES THE
C     PERCENTAGES OF THE DIFFERENT LEVELS IN THE ROOT
C     EQUATION.
      WRITE(61,104)
 104  FORMAT(//,' ENTER THE NUMBER OF VARIABLES IN THE',
     1' ROOT EQUATION.',//)
      NUMW=TTYIN(4HNUM=)
      C=ALOG(R)
      E=2.71828
      DO 47 I=1,NUMW
      XPON=C*A(I)
      P1(I)=100./(E**XPON)
47    WRITE(61,511)I,P1(I)
511   FORMAT(' P',I2,' = ',F10.2,' %')
C
C
C     THIS PART OF THE PROGRAM EXECUTES A CHI-SQUARE
C     GOODNESS OF FIT TEST.
      WRITE(61,205)
 205  FORMAT(//,' DO YOU WISH TO RUN A CHI-SQUARE TEST?',/,
     1' ENTER  1  FOR YES OR  0  FOR NO'/)
      NCHI=TTYIN(4H?=  )
      IF(NCHI .EQ. 0)GO TO 57
48    WRITE(61,106)
 106  FORMAT(/,' ENTER THE TOTAL NUMBER OF OBSERVATIONS',/)
      NOBS=TTYIN(4HNOBS,4H=   )
      WRITE (61,105)
 105  FORMAT(/,' ENTER THE OBSERVED VALUES TO BE TESTED',/)
      CHISQRE=0.0
      DO 49 I=1,NUMW
      OBS(I)=TTYIN(4HOBS=)
      TOB(I)=NOBS*P1(I)/100.
49    CHISQRE=(OBS(I)-TOB(I))**2/TOB(I) + CHISQRE
      WRITE(61,109)
 109  FORMAT(/,' ARE THE DATA JUST ENTERED CORRECT?',
     1' ENTER  I FOR YES OR  0  FOR NO.',/)
      ICHEK=TTYIN(4HCHEC,4HK=  )
      IF(ICHEK .EQ. 0)GO TO 48
      NDF=NUMW-1
      IF(NDF .EQ. 1)CRIT=3.84146
      IF(NDF .EQ. 2)CRIT=5.99147
      IF(NDF .EQ. 3)CRIT=7.81473
      IF(NDF .EQ. 4)CRIT=9.48773
      IF(NDF .EQ. 5)CRIT=11.0705
      IF(NDF .EQ. 6)CRIT=12.5916
      IF(NDF .EQ. 7)CRIT=14.0671
      IF(NDF .EQ. 8)CRIT=15.5073
      IF(NDF .EQ. 9)CRIT=16.9190
      IF(CRIT .GT. CHISQRE)GO TO 53
      WRITE(61,107)CHISQRE,NDF
107   FORMAT(1H1,' SINCE',F10.4,' IS GREATER THAN THE',
     1' CRITICAL VALUE OF',F10.4,/' WITH',I3,' DEGREES OF',
     2' FREEDOM, REJECT THE HYPOTHESIS OF INDEPENDENCE',/
     3' AT THE  .05  SIGNIFICANCE LEVEL.'//)
      GO TO 54
53    WRITE(61,110)CHISQRE,CRIT,NDF
110   FORMAT(1H1,' SINCE',F10.4,' IS LESS THAN THE CRITICAL',
     1' VALUE OF',F10.4,/' WITH',I3,' DEGREES OF FREEDOM,',
     2' ACCEPT THE HYPOTHESIS',/' OF INDEPENDENCE',
     3' AT THE  .05  SIGNIFICANCE LEVEL.'//)
 57   CONTINUE
54    WRITE(61,512)
 512  FORMAT(' ENTER A  1  IF MORE RUNS ARE DESIRED',
     1/' ENTER A  0  IF NO EXTRA RUNS ARE DESIRED'/)
      NRUNS=TTYIN(4HRUNS,4H=     )
      IF(NRUNS .EQ. 1)37,58
C
C
C
C     THIS PART OF THE PROGRAM EXECUTES A TWO FACTOR INFORMATION
 THEORY
C     ANALYSIS.
C
C
 36   CONTINUE
      WRITE(61,200)
 200  FORMAT(' THIS PROGRAM CALCULATES THE VALUES OF'
     1' THE VARIABLES',/,' Y1 AND Y2 IN A TWO FACTOR',
     2' INFORMATION CHANNEL PROBLEM.'///)
      WRITE(61,201)
```

```
201  FORMAT(' ENTER THE NUMBER OF TERMS IN EACH ',
    1' ROW OF THE CHANNEL MATRIX.'///)
     M=TTYIN (4HNOEQ,4HS=  )
     DO 60 K=1,2
     DO 60 L=1,25
60   P2(K,L)=0.0
     DO 61 K=1,25
61   X1(K)=0.0
     WRITE(61,202)
202  FORMAT(' ENTER THE VALUES FOR THE P(I,J)',
    1' IN THE ORDER',/' P(11),P(12) . . . P(1M)',
    2' P(21),P(22), . . ., P(2M).'///)
62   DO 63 I=1,2
     DO 63 J=1,M
63   P2(I,J)=TTYIN (4HP=  )
     WRITE(61,203)
203  FORMAT(' ENTER THE OBSERVED MARGINAL PROBABILITIES',
    1' AS',/,' X(1),X(2) . . . X(M).'///)
     DO 64 J=1,M
64   X1(J)=TTYIN (4HX=  )
     WRITE(61,300)
300  FORMAT(' ARE THE DATA ENTERED CORRECT?',
    1' ENTER A  1 FOR YES  OR  0 FOR NO.')
     NDATA =TTYIN(4H? = )
     IF (NDATA .EQ. 0)GO TO 62
     Y1=0.0
     Y2=1.0-Y1
65   CALL GUESS (M,P2,Y1,Y2,X1,DIFF1)
     CALL PROB (0.1,Y1,Y2)
     CALL GUESS(M,P2,Y1,Y2,X1,DIFF2)
     IF(DIFF2 .LT. DIFF1)65,66
66   Y1=Y1-0.1
     Y2=1. - Y1
67   CALL GUESS(M,P2,Y1,Y2,X1,DIFF1)
     CALL PROB(0.01,Y1,Y2)
     CALL GUESS(M,P2,Y1,Y2,X1,DIFF2)
     IF(DIFF2 .LT. DIFF1)67,68
68   Y1=Y1 - 0.01
     Y2=1.0-Y1
69   CALL GUESS(M,P2,Y1,Y2,X1,DIFF1)
     CALL PROB(0.001,Y1,Y2)
     CALL GUESS(M,P2,Y1,Y2,X1,DIFF2)
     IF(DIFF2 .LT. DIFF1)69,59
59   Y1=Y1-0.001
     Y2=1-Y1
     WRITE(61,204)Y1,Y2
204  FORMAT(' Y1 = ',F6.4,' Y2 = ',F6.4)
     WRITE(61,400)
400  FORMAT(' IF MORE RUNS ARE DESIRED TYPE',
    1' A  1  IF NOT TYPE A  0')
     NRUNS=TTYIN(4HNRUN,4HS=  )
     IF(NRUNS .EQ. 1)GO TO 37
     GO TO 58
C
C
C    THIS PROGRAM CALCULATES ALL THE NECESSARY VALUES FOR
C    A MULTI-FACTOR INFORMATION CHANNEL PROBLEM.
C    ARRANGE YOUR DATA SO THAT THE CONDITIONAL P(Y/Z)
C    ROW PROBABILITIES SUM TO ONE.
34   CONTINUE
C
C
C    THE INVERSE OF A MATRIX
C
     WRITE(61,607)
607  FORMAT(' THE NUMBER OF EQUATIONS ARE',/)
     NOEQS=TTYIN(4HNOEQ,4HS=  )
     WRITE(61,608)
608  FORMAT(//,' THE PROBABILITIES P(Y/Z) ARE',/)
     DO 1 I=1,NOEQS
     DO 1 J=1,NOEQS
     P(I,J)=TTYIN(4HP=  )
1    D(I,J)=P(I,J)
     WRITE(61,609)
609  FORMAT(//,' ARE THE DATA CORRECT'/,
    1' ENTER A  1 FOR YES OR  0 FOR NO',//)
     NCHECK=TTYIN(4H?=  )
     IF(NCHECK .EQ. 0)GO TO 34
     DO 2 I=1,NOEQS
     DO 2 J=1,NOEQS
2    E(I,J)=0.0
     DO 6 M=1,NOEQS
6    E(M,M)=1.0
     DO 13 MPIVRO=1,NOEQS
     NPIVCO=MPIVRO
     T=P(MPIVRO,NPIVCO)
     DO 15 N=1,NOEQS
     E(MPIVRO,N)=E(MPIVRO,N)/T
15   P(MPIVRO,N)=P(MPIVRO,N)/T
     M=1
10   CONTINUE
     IF(MPIVRO .EQ. M) GO TO 8
     CM=-P(M,NPIVCO)
     DO 11 N=1,NOEQS
     TM=P(MPIVRO,N)*CM
     TA=E(MPIVRO,N)*CM
     E(M,N)=E(M,N)+TA
11   P(M,N)=P(M,N)+TM
8    M=M+1
     IF(M .LE. NOEQS) GO TO 10
13   CONTINUE
     I=1
     K=1
71   CONTINUE
     CIDN=0.0
     DO 72 J=1,NOEQS
72   CIDN=E(J,K)*D(I,J)+CIDN
     IF(CIDN .EQ. 0)GO TO 73
     IF(CIDN .EQ. 1.)73,74
73   K=K+1
     IF(K .EQ. NOEQS+1)I=I+1
     IF(K .EQ. NOEQS+1)K=1
     IF(I .EQ. NOEQS+1)GO TO 75
     GO TO 71
74   WRITE(61,205)
205  FORMAT(' INVERSE OF THE PROBABILITY MATRIX',
    1' DOES NOT EXIST',/)
     GO TO 58
75   WRITE(61,101)
101  FORMAT(/,' THE INVERSE OF THE P(Y/Z) MATRIX IS;')
     WRITE(61,102)((M,N,E(M,N),N=1,NOEQS),M=1,NOEQS)
102  FORMAT(1H ,I2,I2,2X,F10.4)
C
C
C
C    CALCULATIONS FOR H= SUM P(Y/Z) LOG P(Y/Z)
     DO 3 K=1,NOEQS
3    H(K)=0.0
     WRITE(61,509)
509  FORMAT(/,' THE CONDITIONAL INFORMATION '
    1'SUM P(Y/Z) LOG P(Y/Z) VALUES ARE;'/)
     DO 4 I=1,NOEQS
     DO 4 J=1,NOEQS
     G=D(I,J)
     IF(G .EQ. 0.0)20,21
20   TLOG=0.0
     GO TO 4
21   TLOG=-1.4426950409*ALOG(G)
4    H(I)=H(Y)+D(I,J)*TLOG
     DO 50 L=1,NOEQS
     WRITE(61,500)L,H(L)
```

```
50    CONTINUE
C
C
C
C     THE NECESSARY  B  CONSTANTS FOR DETERMINING PROB (Y)
C
      WRITE(61,502)
502   FORMAT(///,' THE NECESSARY CONSTANTS'
     1' FOR CALCULATING PROB(Y) ARE;'/)
500   FORMAT(' H',I2,' = ',F12.5)
      DO 7 K=1,NOEQS
7     B(K)=0.0
      DO 9 I=1,NOEQS
      DO 9 J=1,NOEQS
9     B(I)=B(I)+E(I,J)*H(J)
      DO 51 N=1,NOEQS
      WRITE(61,501)N,B(N)
51    CONTINUE
C
C
C
C     CALCULATIONS FOR PROB (Y)
C
      WRITE(61,504)
504   FORMAT(///,' THE PROBABILITIES P(Y) ARE;'/)
501   FORMAT(' B',I2,' = ',F12.5)
      YDEN=0.0
      DO 12 K=1,NOEQS
12    YDEN=2.00**(-B(K))+YDEN
      DO 14 L=1,NOEQS
      Y(L)=2.00**(-B(L))/YDEN
      WRITE(61,503)L,Y(L)
14    CONTINUE
C
C
C
C     CALCULATIONS FOR PROB (Z) AND CHECK FOR NEGATIVE Z'S
C     CALCULATION OF LAGRANGE MULTIPLIER  U  AND
C     RECALCULATION OF P(Y) AND P(Z) PLUS THE CORRECTED
C     B  CONSTANTS
C
      WRITE(61,505)
505   FORMAT(///,' THE PROBABILITIES P(Z) ARE;'/)
503   FORMAT(' Y',I2,' = ',F6.4)
      DO 16 K=1, NOEQS
16    Z(K)=0.0
      DO 17 I=1,NOEQS
      DO 17 J=1,NOEQS
17    Z(I)=Z(I)+Y(J)*E(J,I)
      DO 52 L=1,NOEQS
      WRITE(61,506)L,Z(L)
52    CONTINUE
506   FORMAT(' Z',I2,' = ',F8.4)
      DO 35 K=1,NOEQS
      IF(Z(K) .LT. 0.0)GO TO 5
35    CONTINUE
      GO TO 32
5     H(K)=0.0
      NZERO=K
      DO 19 L=1,NOEQS
      B(L)=0.0
19    F(L)=E(L,NZERO)
      DO 26 M=1,NOEQS
      DO 26 N=1,NOEQS
      B(M)=B(M)+E(M,N)*H(N)
26    CONTINUE
      U=0.0
      CHECK1=U
      CALL TRIAL(F,B,NOEQS,U,EQN1)
      IF(EQN1 .LT. 0.0)GO TO 22
23    U=U-1.0
      CHECK2=U
      CALL TRIAL(F,B,NOEQS,U,EQN2)
      IF(EQN2 .GT. 0.0)GO TO 23
24    U=(CHECK1+CHECK2)/2.0
      CHECK3=U
      CALL TRIAL(F,B,NOEQS,U,EQN3)
      IF(EQN3 .GT.  .00009)CHECK1=CHECK3
      IF(EQN3 .LT. -.00009)CHECK2=CHECK3
      IF(EQN3 .GT.  .00009)GO TO 24
      IF(EQN3 .LT. -.00009)GO TO 24
      GO TO 33
22    U=U+1.0
      CHECK2 =U
      CALL TRIAL(F,B,NOEQS,U,EQN2)
      IF(EQN2 .LT. 0.0)GO TO 22
25    U=(CHECK1+CHECK2)/2.0
      CHECK3=U
      CALL TRIAL(F,B,NOEQS,U,EQN3)
      IF(EQN3 .GT.  .00009)CHECK2=CHECK3
      IF(EQN3 .LT. -.00009)CHECK1=CHECK3
      IF(EQN3 .GT.  .00009)GO TO 25
      IF(EQN3 .LT. -.00009)GO TO 25
33    CONTINUE
      WRITE(61,601)
601   FORMAT(///,' SINCE SOME P(Z) ARE NEGATIVE',
     1' NEW VALUES MUST BE CALCULATED',/,' WITH THE ',
     2' RESTRICTION THAT P(Z) EQUAL ZERO FOR SOME Z.',
     3///,' THE CORRECTED CONSTANTS ARE;'/)
      DO 27 M=1,NOEQS
      WRITE(61,600)M,B(M)
600   FORMAT(' B',I2,' = ',F10.4)
27    CONTINUE
      WRITE(61,507)U
507   FORMAT(///,' THE LAGRANGE MULTIPLIER CONSTANT',
     1' IS',F10.4,///)
      WRITE(61,603)
603   FORMAT(' THE CORRECTED P(Y) VALUES ARE;'/)
      YDEN2=0.0
      DO 28 L=1,NOEQS
28    YDEN2=YDEN2+2.00**(F(L)*U-B(L))
      DO 29 K=1,NOEQS
      Y(K)=2.00**(F(K)*U-B(K))/YDEN2
      WRITE(61,602)K,Y(K)
29    CONTINUE
602   FORMAT(' Y',I2,' =',F10.4)
      WRITE(61,605)
605   FORMAT(///,' THE CORRECTED P(Z) VALUES ARE;'/)
      DO 30 M=1,NOEQS
30    Z(M)=0.0
      DO 31 J=1,NOEQS
      DO 31 K=1,NOEQS
31    Z(J)=Z(J)+Y(K)*E(K,J)
      DO 32 J=1,NOEQS
      WRITE(61,604)J,Z(J)
32    CONTINUE
604   FORMAT(' Z',I2,' = ',F10.4)
      WRITE(61,606)
606   FORMAT(//,' IF MORE RUNS ARE DESIRED, TYPE',
     1' A  1;  IF NOT TYPE A 0 '/)
      NRUN=TTYIN(4HNRUN,4H=    )
      IF(NRUN .EQ. 1)GO TO 37
58    CONTINUE
      END
C
C
C
```

```
      SUBROUTINE TRIAL(F,B,NOEQS,U,EQN)
      DIMENSION F(15),B(15)
      EQN=0.0
      DO 1 I=1,NOEQS
    1 EQN=F(I)*2.00**(F(I)*U)*2.00**
     1(-B(I))+EQN
      RETURN
      END
C
C
      SUBROUTINE GUESS(M,P2,Y1,Y2,X1,DIFF)
      DIMENSION P2(2,50),X1(50)
      DIFF=0.0
      DO 1 K=1,M
      DIF=P2(1,K)*Y1+P2(2,K)*Y2-X1(K)
      D=ABS(DIF)
    1 DIFF= D  + DIFF
      RETURN
      END
C
C
C
      SUBROUTINE PROB(Z,A,B)
      A=A+Z
      B=1.0-A
      RETURN
      END

]
```

Solved Examples

Following are the computer printouts for six examples illustrating methods discussed in this dissertation. These six examples are briefly discussed below.

Example #1. A multi-factor analysis illustrating a non-positive initial solution for $p(z_j)$, $(j = 1, 2, 3, \ldots,)$. This example was given in the text on page 68.

Example #2. A multi-factor analysis illustrating an example with a positive initial solution for $p(z_j)$.

Example #3. A multi-factor analysis with non-positive initial solution. This example was given by Kunisawa (1958).

Example #4. A multi-factor analysis. An example given incorrectly by Kunisawa (1958) on page 107. Kunisawa presents a two-factor two-level example based on the results of a survey questionnaire. The following conditional probability matrix is given.

	Brands	y_1 A	y_2 B
z_1	Practicality	.60	.40
z_2	Quality	.55	.45

Kunnisawa gives for results:

$p(y_1) = p(\text{Brand A}) = 57\%$

$p(y_2) = p(\text{Brand B}) = 43\%$

$p(z_1) = p(\text{practicality}) = 60\%$

$p(z_2) = p(\text{quality}) = 40\%$

The error in Kunisawa's calculations occurs in his inversion of the $p(Y|Z)$ matrix. He gives:

$$\begin{bmatrix} 9 & -11 \\ -8 & 12 \end{bmatrix}$$

as the required inverse which is incorrect since:

$$\begin{bmatrix} 9 & -11 \\ -8 & 12 \end{bmatrix} \begin{bmatrix} .6 & .4 \\ .55 & .45 \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

The above relationship, $p(Y|Z)^{-1}p(Y|Z) = I$ is a requirement for a proper inverse to exist. The correct inverse is:

$$\begin{bmatrix} 9 & -8 \\ -11 & 12 \end{bmatrix}$$

which gives the following results for the probability predictions.

$p(y_1) = p(\text{Brand A}) = .5713$

$p(y_2) = p(\text{Brand B}) = .4249$

$$p(z_1) = p(\text{practicality}) = .5013$$

$$p(z_2) = p(\text{quality}) = .4987$$

Example #5. A multi-factor analysis. An example given incorrectly by Kunisawa (1958) on page 90. Kunisawa presents a two-factor two-level example based on a race prediction problem. A cyclist's performance is predicted using a racing form and it is desired to determine if the racing form is actually helpful. The following conditional probability matrix is given.

	Actual Results	
Racing Form Prediction	y_1 win	y_2 lose
Predicts Winner z_1	.6	.4
Predicts Loser z_2	.2	.8

Kunisawa gives the following results:

$$p(\text{win}) = p(y_1) = .3$$

$$p(\text{lose}) = p(y_2) = .7$$

Kunisawa gives the following for the matrix inverse.

$$\begin{bmatrix} 2 & -.5 \\ -1.0 & 1.5 \end{bmatrix}$$

but,

$$\begin{bmatrix} 2 & -.5 \\ -1 & 1.5 \end{bmatrix} \begin{bmatrix} .6 & .4 \\ .2 & .8 \end{bmatrix} \neq \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The correct inverse is:

$$\begin{bmatrix} 2 & -1 \\ -.5 & 1.5 \end{bmatrix}$$

which gives results of:

$$p(\text{win}) = p(y_1) = .3938$$

$$p(\text{lose}) = p(y_2) = .6062$$

Example #6. A multi-factor analysis. An example incorrectly given by Kunisawa (1958) on page 87. Kunisawa presents an example based on weather forcasting using a barometer. The following conditional probability matrix is given:

	Fair weather y_1	Rainy weather y_2
Low Barometer reading z_1	.25	.75
High Barometer reading z_2	.667	.333

Kunisawa predicts:

$$p(\text{fair weather}) = p(y_1) = .34315$$

$$p(\text{rainy weather}) = p(y_2) = .65685$$

$$p(\text{low reading}) = p(z_1) = .7764$$

$$p(\text{high reading}) = p(z_2) = .2236$$

The correct results are:

$$p(\text{fair weather}) = p(y_1) = .4558$$

$$p(\text{rainy weather}) = p(y_2) = .5442$$

$$p(\text{low reading}) = p(z_1) = .5065$$

$$p(\text{high reading}) = p(z_2) = .4935$$

EXAMPLE 1

```
RUN
TYPE 1 IF THIS IS A ONE-FACTOR ANALYSIS
TYPE 2 IF TWO-FACTOR OR 3 IF MULTI-FACTOR


?=  3
THE NUMBER OF EQUATIONS ARE

NOEQS=  4


THE PROBABILITIES P(Y/Z) ARE

P=  .5 P=  .25 P=  .00 P=  .25
P=  .0 P=  1.0 P=  .00 P=  .00
P=  .25 P=  .25 P=  .25 P=  .25
P=  .0 P=  .00 P=  2.\\1. P=  .00


ARE THE DATA CORRECT
ENTER A  1 FOR YES OR  0 FOR NO


?=  1

THE INVERSE OF THE P(Y/Z) MATRIX IS;
 1 1        4.0000
 1 2             0
 1 3       -4.0000
 1 4        1.0000
 2 1             0
 2 2        1.0000
 2 3             0
 2 4             0
 3 1             0
 3 2             0
 3 3             0
 3 4        1.0000
 4 1       -4.0000
 4 2       -1.0000
 4 3        8.0000
 4 4       -2.0000

THE CONDITIONAL INFORMATION SUM P(Y/Z) LOG P(Y/Z) VALUES ARE;

H 1 =        1.50000
H 2 =              0
H 3 =        2.00000
H 4 =              0
```

```
THE NECESSARY CONSTANTS FOR CALCULATING PROB(Y) ARE;

B 1 =      -2.00000
B 2 =             0
B 3 =             0
B 4 =      10.00000


THE PROBABILITIES P(Y) ARE;

Y 1 =  .6666
Y 2 =  .1666
Y 3 =  .1666
Y 4 =  .0002


THE PROBABILITIES P(Z) ARE;

Z 1 =   2.6656
Z 2 =    .1665
Z 3 =  -2.6649
Z 4 =    .8329

SINCE SOME P(Z) ARE NEGATIVE NEW VALUES MUST BE CALCULATED
WITH THE  RESTRICTION THAT P(Z) EQUAL ZERO FOR SOME Z.


THE CORRECTED CONSTANTS ARE;

B 1 =       6.0000
B 2 =            0
B 3 =            0
B 4 =      -6.0000

THE CORRECTED P(Y) VALUES ARE;

Y 1 =      .1274
Y 2 =      .4045
Y 3 =      .4045
Y 4 =      .0637


THE CORRECTED P(Z) VALUES ARE;

Z 1 =      .2548
Z 2 =      .3408
Z 3 =      .0000
Z 4 =      .4044
```

```
RUN
TYPE 1 IF THIS IS A ONE-FACTOR ANALYSIS                EXAMPLE 2
TYPE 2 IF TWO-FACTOR OR 3 IF MULTI-FACTOR

?=  3
THE NUMBER OF EQUATIONS ARE

NOEQS=  4

THE PROBABILITIES P(Y/Z) ARE

P=  .50 P=  .25 P=  .25
P=  .00 P=  1.0 P=  .00
P=  22
P=  44 P=  44
P=  44 P=  44 P=  44
P=  44
P=  44
P=  44
P=  44

ARE THE DATA CORRECT
ENTER A  1 FOR YES OR  0 FOR NO

?=  0
THE NUMBER OF EQUATIONS ARE

NOEQS=  4

THE PROBABILITIES P(Y/Z) ARE

P=  .50 P=  .25 P=  .00 P=  .25
P=  .00 P=  1.0 P=  .00 P=  .00
P=  .25 P=  .00 P=  .25 P=  .50
P=  .00 P=  .00 P=  1.0 P=  .00

ARE THE DATA CORRECT
ENTER A  1 FOR YES OR  0 FOR NO

?=  1

THE INVERSE OF THE P(Y/Z) MATRIX IS:
 1 1       2.6667
 1 2      -0.6667
 1 3      -1.3333
 1 4        .3333
 2 1            0
 2 2       1.0000
 2 3            0
 2 4            0
 3 1            0
 3 2            0
 3 3            0
 3 4       1.0000
 4 1      -1.3333
 4 2        .3333
 4 3       2.6667
 4 4      -0.6667

THE CONDITIONAL INFORMATION SUM P(Y/Z) LOG P(Y/Z) VALUES ARE:

H 1 =       1.50000
H 2 =             0
H 3 =       1.50000
H 4 =             0
```

```
THE NECESSARY CONSTANTS FOR CALCULATING PROB(Y) ARE:

B 1 =       2.00000
B 2 =             0
B 3 =             0
B 4 =       2.00000

THE PROBABILITIES P(Y) ARE:

Y 1 =  .1000
Y 2 =  .4000
Y 3 =  .4000
Y 4 =  .1000

THE PROBABILITIES P(Z) ARE:

Z 1 =    .1333
Z 2 =    .3667
Z 3 =    .1333
Z 4 =    .3667

IF MORE RUNS ARE DESIRED, TYPE A  1;  IF NOT TYPE A 0

NRUN=   1
TYPE 1 IF THIS IS A ONE-FACTOR ANALYSIS
TYPE 2 IF TWO-FACTOR OR 3 IF MULTI-FACTOR

?=  3                                                   EXAMPLE 3
THE NUMBER OF EQUATIONS ARE

NOEQS=  4

THE PROBABILITIES P(Y/Z) ARE

P=  .50 P=  .00 P=  .00 P=  .50
P=  .50 P=  .50 P=  .00 P=  .00
P=  .00 P=  .00 P=  1.0 P=  .00
P=  .25 P=  .25 P=  .25 P=  .25

ARE THE DATA CORRECT
ENTER A  1 FOR YES OR  0 FOR NO

?=  1

THE INVERSE OF THE P(Y/Z) MATRIX IS:
 1 1       2.0000
 1 2       2.0000
 1 3       1.0000
 1 4      -4.0000
 2 1      -2.0000
 2 2            0
 2 3      -1.0000
 2 4       4.0000
 3 1            0
 3 2            0
 3 3       1.0000
 3 4            0
 4 1            0
 4 2      -2.0000
 4 3      -1.0000
 4 4       4.0000
```

```
THE CONDITIONAL INFORMATION SUM P(Y/Z) LOG P(Y/Z) VALUES ARE;

H 1 =        1.00000
H 2 =        1.00000
H 3 =              0
H 4 =        2.00000


THE NECESSARY CONSTANTS FOR CALCULATING PROB(Y) ARE;

B 1 =      -4.00000
B 2 =       6.00000
B 3 =             0
B 4 =       6.00000


THE PROBABILITIES P(Y) ARE;

Y 1 =  .9394
Y 2 =  .0009
Y 3 =  .0587
Y 4 =  .0009


THE PROBABILITIES P(Z) ARE;

Z 1 =    1.8771
Z 2 =    1.8771
Z 3 =     .9963
Z 4 =  -3.7505


SINCE SOME P(Z) ARE NEGATIVE NEW VALUES MUST BE CALCULATED
WITH THE  RESTRICTION THAT P(Z) EQUAL ZERO FOR SOME Z.

THE CORRECTED CONSTANTS ARE;

B 1 =       4.0000
B 2 =      -2.0000
B 3 =            0
B 4 =      -2.0000


THE LAGRANGE MULTIPLIER CONSTANT IS    -0.8750


THE CORRECTED P(Y) VALUES ARE;

Y 1 =        .2929
Y 2 =        .1464
Y 3 =        .4142
Y 4 =        .1464


THE CORRECTED P(Z) VALUES ARE;

Z 1 =        .2929
Z 2 =        .2929
Z 3 =        .4142
Z 4 =      -0.0000
```

```
IF MORE RUNS ARE DESIRED, TYPE A  1;  IF NOT TYPE A 0

NRUN=   1
TYPE 1 IF THIS IS A ONE-FACTOR ANALYSIS
TYPE 2 IF TWO-FACTOR OR 3 IF MULTI-FACTOR
```

EXAMPLE 4

```
?=  3
THE NUMBER OF EQUATIONS ARE

NOEQS=  2


THE PROBABILITIES P(Y/Z) ARE

P=  .60 P=  .40
P=  .55 P=  .45


ARE THE DATA CORRECT
ENTER A  1 FOR YES OR  0 FOR NO

?=  1

THE INVERSE OF THE P(Y/Z) MATRIX IS;
 1 1       9.0000
 1 2      -8.0000
 2 1     -11.0000
 2 2      12.0000

THE CONDITIONAL INFORMATION SUM P(Y/Z) LOG P(Y/Z) VALUES ARE;

H 1 =        .97095
H 2 =        .99277


THE NECESSARY CONSTANTS FOR CALCULATING PROB(Y) ARE;

B 1 =        .79636
B 2 =       1.23284


THE PROBABILITIES P(Y) ARE;

Y 1 =  .5751
Y 2 =  .4249


THE PROBABILITIES P(Z) ARE;

Z 1 =    .5013
Z 2 =    .4987


IF MORE RUNS ARE DESIRED, TYPE A  1;  IF NOT TYPE A 0

NRUN=   1
TYPE 1 IF THIS IS A ONE-FACTOR ANALYSIS
TYPE 2 IF TWO-FACTOR OR 3 IF MULTI-FACTOR
```

EXAMPLE 5

```
?=  3
THE NUMBER OF EQUATIONS ARE

NOEQS=  2
```

```
THE PROBABILITIES P(Y/Z) ARE

P=  .60 P=  .40
P=  .20 P=  .80


ARE THE DATA CORRECT
ENTER A  1 FOR YES OR  0 FOR NO


?=  1

THE INVERSE OF THE P(Y/Z) MATRIX IS;
 1 1       2.0000
 1 2      -1.0000
 2 1      -0.5000
 2 2       1.5000

THE CONDITIONAL INFORMATION SUM P(Y/Z) LOG P(Y/Z) VALUES ARE;

H 1 =         .97095
H 2 =         .72193


THE NECESSARY CONSTANTS FOR CALCULATING PROB(Y) ARE;

B 1 =        1.21997
B 2 =         .59742


THE PROBABILITIES P(Y) ARE;

Y 1 =  .3938
Y 2 =  .6062


THE PROBABILITIES P(Z) ARE;

Z 1 =     .4844
Z 2 =     .5156

IF MORE RUNS ARE DESIRED, TYPE A  1;  IF NOT TYPE A 0

NRUN=    1
TYPE 1 IF THIS IS A ONE-FACTOR ANALYSIS
TYPE 2 IF TWO-FACTOR OR 3 IF MULTI-FACTOR
```

EXAMPLE 6

```
?=  3
THE NUMBER OF EQUATIONS ARE

NOEQS=  2


THE PROBABILITIES P(Y/Z) ARE

P=  .250 P=  .750
P=  .667 P=  .333


ARE THE DATA CORRECT
ENTER A  1 FOR YES OR  0 FOR NO


?=  1

THE INVERSE OF THE P(Y/Z) MATRIX IS;
 1 1      -0.7986
 1 2       1.7986
 2 1       1.5995
 2 2      -0.5995

THE CONDITIONAL INFORMATION SUM P(Y/Z) LOG P(Y/Z) VALUES ARE;

H 1 =        .81128
H 2 =        .91796


THE NECESSARY CONSTANTS FOR CALCULATING PROB(Y) ARE;

B 1 =       1.00316
B 2 =        .74732


THE PROBABILITIES P(Y) ARE;

Y 1 =  .4558
Y 2 =  .5442


THE PROBABILITIES P(Z) ARE;

Z 1 =     .5065
Z 2 =     .4935
```

APPENDIX E

Table of Values for a One-Factor Information Theory Model

For a theoretical development of a one-factor information channel analysis see Shirland (1971). Following is a table for two, three and four values in a one-factor model. The table gives the root of the channel equation along with the probability predictions for the various levels. The information channel equation is:

$$W^{-L_1} + W^{-L_2} + \ldots + W^{-L_n} = 1$$

where,

$$n = 2, 3, 4$$

and, $L_1, L_2, \ldots, L_n$ are the values of the various levels of the particular factor in question.

In explanation of the use of the tables, consider the following example. Suppose that you have a four level one-factor information channel problem. For example, given four brands of a product with prices of \$2, \$3, \$6, and \$8 respectively, it is desired to predict the demands for the various products using only "Price" as a factor. The information channel equation, then, is:

$$W^{-2} + W^{-3} + W^{-6} + W^{-8} = 1$$

Turn to the portion of the table with the heading "Four Levels/One Factor" and search for the 2, 3, 6 heading in the extreme left column. Then read across to the column heading 8 and read:

$$p_1 = .4874$$

$$p_2 = .3403$$

$$p_3 = .1158$$

$$p_4 = .0565$$

$$W = 1.4323$$

The p_i $(i = 1, 2, 3, 4)$ are the predictions for the various demands and W is the root of the information channel equation given the particular levels specified.

For a three level problem look under the heading "Three Level/One-Factor" and read the first two levels from the left hand column and the third level from its appropriate column. A two level problem is done similarly.

This table does not contain all the combinations of possible values for a 2, 3, or 4 level one-factor information channel problem. The table gives only whole number combinations. For combinations containing fractions of whole numbers use the computer program of Appendix D which is capable of performing one-factor information channel calculations for any combination of values for as many as ten levels for a factor.

TWO LEVEL

	2	3	4	5	6	7	8	9	10
1									
P1	.6180	.6823	.7245	.7549	.7781	.7965	.8117	.8243	.8351
P2	.3820	.3177	.2755	.2451	.2219	.2035	.1883	.1758	.1650
W	1.6180	1.4656	1.3803	1.3247	1.2852	1.2554	1.2321	1.2131	1.1975
2									
P1	0	.5699	.6180	.6541	.6823	.7053	.7245	.7408	.7549
P2	0	.4302	.3820	.3460	.3177	.2947	.2755	.2592	.2451
W	0	1.3247	1.2720	1.2365	1.2106	1.1907	1.1749	1.1619	1.1510
3									
P1	0	0	.5497	.5877	.6180	.6431	.6642	.6823	.6981
P2	0	0	.4503	.4123	.3820	.3569	.3358	.3177	.3019
W	0	0	1.2207	1.1939	1.1740	1.1586	1.1461	1.1359	1.1272
4									
P1	0	0	0	.5386	.5698	.5959	.6181	.6372	.6540
P2	0	0	0	.4614	.4302	.4041	.3820	.3628	.3460
W	0	0	0	1.1673	1.1510	1.1382	1.1278	1.1192	1.1120
5									
P1	0	0	0	0	.5316	.5581	.5808	.6006	.6180
P2	0	0	0	0	.4684	.4420	.4192	.3994	.3820
W	0	0	0	0	1.1347	1.1237	1.1148	1.1074	1.1010
6									
P1	0	0	0	0	0	.5267	.5497	.5699	.5877
P2	0	0	0	0	0	.4733	.4503	.4302	.4123
W	0	0	0	0	0	1.1128	1.1049	1.0983	1.0926
7									
P1	0	0	0	0	0	0	.5231	.5435	.5615
P2	0	0	0	0	0	0	.4769	.4566	.4385
W	0	0	0	0	0	0	1.0970	1.0910	1.0859
8									
P1	0	0	0	0	0	0	0	.5204	.5386
P2	0	0	0	0	0	0	0	.4796	.4614
W	0	0	0	0	0	0	0	1.0851	1.0804
9									
P1	0	0	0	0	0	0	0	0	.5183
P2	0	0	0	0	0	0	0	0	.4818
W	0	0	0	0	0	0	0	0	1.0758

THREE LEVEL

	3	4	5	6	7	8	9	10
1 2								
P1	.5437	.5698	.5865	.5975	.6047	.6095	.6126	.6146
P2	.2956	.3247	.3440	.3570	.3657	.3715	.3753	.3777
P3	.1607	.1054	.0694	.0455	.0296	.0190	.0121	.0077
W	1.8393	1.7549	1.7049	1.6736	1.6536	1.6407	1.6324	1.6271
1 3								
P1	0	.6180	.6369	.6500	.6593	.6660	.6707	.6742
P2	0	.2361	.2583	.2746	.2866	.2954	.3018	.3064
P3	0	.1459	.1048	.0754	.0541	.0387	.0275	.0194
W	0	1.6180	1.5701	1.5385	1.5168	1.5016	1.4909	1.4833
1 4								
P1	0	0	.6680	.6823	.6929	.7007	.7067	.7111
P2	0	0	.1991	.2168	.2305	.2411	.2494	.2558
P3	0	0	.1330	.1009	.0767	.0581	.0439	.0331
W	0	0	1.4971	1.4656	1.4433	1.4271	1.4151	1.4062
1 5								
P1	0	0	0	.7044	.7158	.7245	.7312	.7364
P2	0	0	0	.1734	.1879	.1996	.2090	.2166
P3	0	0	0	.1222	.0963	.0759	.0598	.0469
W	0	0	0	1.4196	1.3970	1.3803	1.3676	1.3579
1 6								
P1	0	0	0	0	.7325	.7418	.7491	.7549
P2	0	0	0	0	.1544	.1666	.1767	.1851
P3	0	0	0	0	.1131	.0917	.0743	.0601
W	0	0	0	0	1.3653	1.3481	1.3349	1.3247
1 7								
P1	0	0	0	0	0	.7549	.7627	.7689
P2	0	0	0	0	0	.1397	.1501	.1589
P3	0	0	0	0	0	.1055	.0873	.0722
W	0	0	0	0	0	1.3247	1.3112	1.3006
1 8								
P1	0	0	0	0	0	0	.7733	.7799
P2	0	0	0	0	0	0	.1279	.1369
P3	0	0	0	0	0	0	.0989	.0832
W	0	0	0	0	0	0	1.2932	1.2822
1 9								
P1	0	0	0	0	0	0	0	.7887
P2	0	0	0	0	0	0	0	.1181
P3	0	0	0	0	0	0	0	.0932
W	0	0	0	0	0	0	0	1.2679
2 3								
P1	0	.4656	.4896	.5076	.5213	.5320	.5403	.5468
P2	0	.3177	.3426	.3616	.3764	.3880	.3971	.4043
P3	0	.2168	.1678	.1308	.1023	.0801	.0626	.0489
W	0	1.4656	1.4291	1.4036	1.3850	1.3711	1.3605	1.3523
2 4								
P1	0	0	.5249	.5437	.5583	.5698	.5791	.5865
P2	0	0	.2755	.2956	.3117	.3247	.3353	.3440
P3	0	0	.1996	.1607	.1300	.1054	.0856	.0694
W	0	0	1.3802	1.3562	1.3383	1.3247	1.3141	1.3057
2 5								
P1	0	0	0	.5698	.5850	.5972	.6071	.6151
P2	0	0	0	.2451	.2618	.2756	.2871	.2968
P3	0	0	0	.1850	.1532	.1272	.1058	.0881
W	0	0	0	1.3247	1.3074	1.2940	1.2835	1.2750

THREE LEVEL

	3	4	5	6	7	8	9	10
2 6								
P1	0	0	0	0	.6054	.6180	.6283	.6369
P2	0	0	0	0	.2219	.2361	.2481	.2583
P3	0	0	0	0	.1727	.1459	.1236	.1048
W	0	0	0	0	1.2852	1.2720	1.2615	1.2531
2 7								
P1	0	0	0	0	0	.6345	.6452	.6540
P2	0	0	0	0	0	.2035	.2157	.2263
P3	0	0	0	0	0	.1621	.1392	.1197
W	0	0	0	0	0	1.2554	1.2450	1.2365
2 8								
P1	0	0	0	0	0	0	.6588	.6680
P2	0	0	0	0	0	0	.1883	.1991
P3	0	0	0	0	0	0	.1529	.1330
W	0	0	0	0	0	0	1.2321	1.2235
2 9								
P1	0	0	0	0	0	0	0	.6795
P2	0	0	0	0	0	0	0	.1757
P3	0	0	0	0	0	0	0	.1448
W	0	0	0	0	0	0	0	1.2131
3 4								
P1	0	0	.4302	.4509	.4675	.4810	.4922	.5014
P2	0	0	.3247	.3458	.3629	.3769	.3886	.3984
P3	0	0	.2451	.2033	.1696	.1421	.1192	.1002
W	0	0	1.3247	1.3041	1.2885	1.2763	1.2666	1.2587
3 5								
P1	0	0	0	.4784	.4955	.5095	.5211	.5308
P2	0	0	0	.2927	.3103	.3250	.3374	.3480
P3	0	0	0	.2289	.1943	.1656	.1415	.1211
W	0	0	0	1.2786	1.2637	1.2521	1.2427	1.2350
3 6								
P1	0	0	0	0	.5174	.5317	.5437	.5538
P2	0	0	0	0	.2677	.2827	.2956	.3067
P3	0	0	0	0	.2149	.1856	.1607	.1395
W	0	0	0	0	1.2456	1.2344	1.2252	1.2177
3 7								
P1	0	0	0	0	0	.5497	.5620	.5723
P2	0	0	0	0	0	.2475	.2606	.2720
P3	0	0	0	0	0	.2028	.1775	.1557
W	0	0	0	0	0	1.2207	1.2118	1.2044
3 8								
P1	0	0	0	0	0	0	.5771	.5877
P2	0	0	0	0	0	0	.2308	.2423
P3	0	0	0	0	0	0	.1922	.1700
W	0	0	0	0	0	0	1.2011	1.1939
3 9								
P1	0	0	0	0	0	0	0	.6006
P2	0	0	0	0	0	0	0	.2166
P3	0	0	0	0	0	0	0	.1828
W	0	0	0	0	0	0	0	1.1852
4 5								
P1	0	0	0	.4098	.4278	.4428	.4554	.4663
P2	0	0	0	.3279	.3460	.3612	.3742	.3853
P3	0	0	0	.2623	.2263	.1961	.1704	.1484
W	0	0	0	1.2499	1.2365	1.2259	1.2173	1.2102

THREE LEVEL

	3	4	5	6	7	8	9	10
4 6								
P1	0	0	0	0	.4503	.4656	.4786	.4896
P2	0	0	0	0	.3022	.3177	.3310	.3426
P3	0	0	0	0	.2475	.2168	.1905	.1678
W	0	0	0	0	1.2207	1.2106	1.2023	1.1955
4 7								
P1	0	0	0	0	0	.4843	.4975	.5088
P2	0	0	0	0	0	.2812	.2947	.3065
P3	0	0	0	0	0	.2346	.2078	.1847
W	0	0	0	0	0	1.1987	1.1907	1.1840
4 8								
P1	0	0	0	0	0	0	.5134	.5249
P2	0	0	0	0	0	0	.2636	.2755
P3	0	0	0	0	0	0	.2231	.1996
W	0	0	0	0	0	0	1.1814	1.1749
4 9								
P1	0	0	0	0	0	0	0	.5386
P2	0	0	0	0	0	0	0	.2485
P3	0	0	0	0	0	0	0	.2129
W	0	0	0	0	0	0	0	1.1673
5 6								
P1	0	0	0	0	.3965	.4123	.4259	.4376
P2	0	0	0	0	.3296	.3454	.3590	.3709
P3	0	0	0	0	.2739	.2423	.2151	.1915
W	0	0	0	0	1.2032	1.1939	1.1862	1.1797
5 7								
P1	0	0	0	0	0	.4314	.4451	.4570
P2	0	0	0	0	0	.3082	.3220	.3341
P3	0	0	0	0	0	.2605	.2329	.2089
W	0	0	0	0	0	1.1831	1.1757	1.1695
5 8								
P1	0	0	0	0	0	0	.4614	.4735
P2	0	0	0	0	0	0	.2901	.3023
P3	0	0	0	0	0	0	.2485	.2242
W	0	0	0	0	0	0	1.1673	1.1613
5 9								
P1	0	0	0	0	0	0	0	.4877
P2	0	0	0	0	0	0	0	.2745
P3	0	0	0	0	0	0	0	.2378
W	0	0	0	0	0	0	0	1.1545
6 7								
P1	0	0	0	0	0	.3872	.4012	.4135
P2	0	0	0	0	0	.3306	.3446	.3569
P3	0	0	0	0	0	.2822	.2542	.2295
W	0	0	0	0	0	1.1713	1.1644	1.1586
6 8								
P1	0	0	0	0	0	0	.4177	.4302
P2	0	0	0	0	0	0	.3123	.3247
P3	0	0	0	0	0	0	.2700	.2451
W	0	0	0	0	0	0	1.1566	1.1510
6 9								
P1	0	0	0	0	0	0	0	.4446
P2	0	0	0	0	0	0	0	.2964
P3	0	0	0	0	0	0	0	.2590
W	0	0	0	0	0	0	0	1.1447

THREE LEVEL

	3	4	5	6	7	8	9	10
7 8								
P1	0	0	0	0	0	0	.3803	.3929
P2	0	0	0	0	0	0	.3312	.3438
P3	0	0	0	0	0	0	.2885	.2633
W	0	0	0	0	0	0	1.1481	1.1428
7 9								
P1	0	0	0	0	0	0	0	.4075
P2	0	0	0	0	0	0	0	.3153
P3	0	0	0	0	0	0	0	.2773
W	0	0	0	0	0	0	0	1.1368
8 9								
P1	0	0	0	0	0	0	0	.3749
P2	0	0	0	0	0	0	0	.3317
P3	0	0	0	0	0	0	0	.2934
W	0	0	0	0	0	0	0	1.1305

FOUR LEVEL

	4	5	6	7	8	9	10
1 2 3							
P1	.5188	.5295	.5357	.5392	.5412	.5423	.5429
P2	.2691	.2804	.2870	.2908	.2929	.2941	.2948
P3	.1396	.1485	.1537	.1568	.1585	.1595	.1601
P4	.0724	.0416	.0236	.0133	.0074	.0041	.0022
W	1.9276	1.8885	1.8668	1.8545	1.8477	1.8439	1.8418
1 2 4							
P1	0	.5518	.5591	.5635	.5662	.5677	.5686
P2	0	.3044	.3126	.3176	.3205	.3223	.3233
P3	0	.0927	.0977	.1009	.1027	.1039	.1045
P4	0	.0511	.0305	.0180	.0106	.0061	.0035
W	0	1.8124	1.7885	1.7745	1.7663	1.7615	1.7587
1 2 5							
P1	0	0	.5735	.5786	.5818	.5837	.5848
P2	0	0	.3289	.3348	.3385	.3407	.3420
P3	0	0	.0620	.0649	.0666	.0678	.0684
P4	0	0	.0356	.0217	.0131	.0079	.0047
W	0	0	1.7437	1.7282	1.7189	1.7132	1.7098
1 2 6							
P1	0	0	0	.5882	.5918	.5940	.5954
P2	0	0	0	.3460	.3502	.3529	.3545
P3	0	0	0	.0414	.0430	.0439	.0446
P4	0	0	0	.0244	.0150	.0092	.0056
W	0	0	0	1.7001	1.6898	1.6835	1.6795
1 2 7							
P1	0	0	0	0	.5983	.6007	.6023
P2	0	0	0	0	.3579	.3609	.3627
P3	0	0	0	0	.0274	.0282	.0287
P4	0	0	0	0	.0164	.0102	.0063
W	0	0	0	0	1.6714	1.6646	1.6604
1 2 8							
P1	0	0	0	0	0	.6051	.6068
P2	0	0	0	0	0	.3661	.3682
P3	0	0	0	0	0	.0180	.0184
P4	0	0	0	0	0	.0109	.0068
W	0	0	0	0	0	1.6527	1.6481
1 2 9							
P1	0	0	0	0	0	0	.6096
P2	0	0	0	0	0	0	.3717
P3	0	0	0	0	0	0	.0116
P4	0	0	0	0	0	0	.0071
W	0	0	0	0	0	0	1.6403
1 3 4							
P1	0	.5934	.6023	.6080	.6117	.6140	.6155
P2	0	.2090	.2185	.2247	.2288	.2315	.2332
P3	0	.1240	.1316	.1366	.1400	.1421	.1435
P4	0	.0736	.0477	.0307	.0196	.0124	.0078
W	0	1.6851	1.6603	1.6448	1.6349	1.6286	1.6247
1 3 5							
P1	0	0	.6180	.6245	.6288	.6316	.6334
P2	0	0	.2361	.2435	.2486	.2519	.2542
P3	0	0	.0902	.0950	.0983	.1005	.1020
P4	0	0	.0557	.0370	.0244	.0160	.0104
W	0	0	1.6180	1.6013	1.5904	1.5833	1.5787

FOUR LEVEL

	4	5	6	7	8	9	10
1 3 6							
P1	0	0	0	.6355	.6403	.6435	.6457
P2	0	0	0	.2567	.2625	.2665	.2692
P3	0	0	0	.0659	.0689	.0710	.0725
P4	0	0	0	.0419	.0283	.0189	.0126
W	0	0	0	1.5735	1.5617	1.5539	1.5487
1 3 7							
P1	0	0	0	0	.6483	.6518	.6543
P2	0	0	0	0	.2724	.2769	.2801
P3	0	0	0	0	.0481	.0500	.0513
P4	0	0	0	0	.0312	.0212	.0144
W	0	0	0	0	1.5426	1.5342	1.5284
1 3 8							
P1	0	0	0	0	0	.6576	.6603
P2	0	0	0	0	0	.2844	.2879
P3	0	0	0	0	0	.0350	.0361
P4	0	0	0	0	0	.0230	.0157
W	0	0	0	0	0	1.5206	1.5145
1 3 9							
P1	0	0	0	0	0	0	.6645
P2	0	0	0	0	0	0	.2934
P3	0	0	0	0	0	0	.0253
P4	0	0	0	0	0	0	.0168
W	0	0	0	0	0	0	1.5049
1 4 5							
P1	0	0	.6445	.6518	.6569	.6604	.6628
P2	0	0	.1726	.1805	.1862	.1902	.1930
P3	0	0	.1112	.1177	.1223	.1256	.1279
P4	0	0	.0717	.0500	.0347	.0239	.0164
W	0	0	1.5515	1.5342	1.5224	1.5143	1.5086
1 4 6							
P1	0	0	0	.6637	.6693	.6732	.6759
P2	0	0	0	.1941	.2006	.2054	.2088
P3	0	0	0	.0855	.0899	.0931	.0954
P4	0	0	0	.0567	.0403	.0284	.0199
W	0	0	0	1.5067	1.4942	1.4855	1.4794
1 4 7							
P1	0	0	0	0	.6781	.6823	.6854
P2	0	0	0	0	.2114	.2168	.2207
P3	0	0	0	0	.0659	.0689	.0711
P4	0	0	0	0	.0447	.0321	.0229
W	0	0	0	0	1.4748	1.4656	1.4590
1 4 8							
P1	0	0	0	0	0	.6890	.6923
P2	0	0	0	0	0	.2253	.2297
P3	0	0	0	0	0	.0508	.0528
P4	0	0	0	0	0	.0350	.0253
W	0	0	0	0	0	1.4514	1.4445
1 4 9							
P1	0	0	0	0	0	0	.6973
P2	0	0	0	0	0	0	.2365
P3	0	0	0	0	0	0	.0390
P4	0	0	0	0	0	0	.0272
W	0	0	0	0	0	0	1.4340

FOUR LEVEL

	4	5	6	7	8	9	10
1 5 6							
P1	0	0	0	.6823	.6884	.6929	.6961
P2	0	0	0	.1479	.1546	.1597	.1634
P3	0	0	0	.1009	.1065	.1107	.1138
P4	0	0	0	.0689	.0505	.0368	.0267
W	0	0	0	1.4656	1.4526	1.4432	1.4366
1 5 7							
P1	0	0	0	0	.6978	.7026	.7061
P2	0	0	0	0	.1654	.1712	.1755
P3	0	0	0	0	.0806	.0845	.0875
P4	0	0	0	0	.0562	.0417	.0308
W	0	0	0	0	1.4331	1.4233	1.4162
1 5 8							
P1	0	0	0	0	0	.7098	.7136
P2	0	0	0	0	0	.1801	.1850
P3	0	0	0	0	0	.0644	.0672
P4	0	0	0	0	0	.0457	.0342
W	0	0	0	0	0	1.4089	1.4014
1 5 9							
P1	0	0	0	0	0	0	.7192
P2	0	0	0	0	0	0	.1924
P3	0	0	0	0	0	0	.0515
P4	0	0	0	0	0	0	.0370
W	0	0	0	0	0	0	1.3905
1 6 7							
P1	0	0	0	0	.7117	.7169	.7208
P2	0	0	0	0	.1309	.1358	.1402
P3	0	0	0	0	.0925	.0973	.1011
P4	0	0	0	0	.0658	.0500	.0379
W	0	0	0	0	1.4051	1.3949	1.3873
1 6 8							
P1	0	0	0	0	0	.7245	.7287
P2	0	0	0	0	0	.1446	.1497
P3	0	0	0	0	0	.0759	.0795
P4	0	0	0	0	0	.0550	.0422
W	0	0	0	0	0	1.3803	1.3724
1 6 9							
P1	0	0	0	0	0	0	.7347
P2	0	0	0	0	0	0	.1572
P3	0	0	0	0	0	0	.0624
P4	0	0	0	0	0	0	.0458
W	0	0	0	0	0	0	1.3611
1 7 8							
P1	0	0	0	0	0	.7354	.7398
P2	0	0	0	0	0	.1163	.1213
P3	0	0	0	0	0	.0855	.0897
P4	0	0	0	0	0	.0629	.0491
W	0	0	0	0	0	1.3599	1.3517
1 7 9							
P1	0	0	0	0	0	0	.7461
P2	0	0	0	0	0	0	.1287
P3	0	0	0	0	0	0	.0717
P4	0	0	0	0	0	0	.0535
W	0	0	0	0	0	0	1.3403

FOUR LEVEL

	4	5	6	7	8	9	10
1 8 9							
P1	0	0	0	0	0	0	.7549
P2	0	0	0	0	0	0	.1054
P3	0	0	0	0	0	0	.0796
P4	0	0	0	0	0	0	.0601
W	0	0	0	0	0	0	1.3247
2 3 4							
P1	0	.4249	.4369	.4454	.4514	.4557	.4587
P2	0	.2769	.2888	.2973	.3033	.3076	.3106
P3	0	.1805	.1909	.1984	.2038	.2076	.2104
P4	0	.1177	.0834	.0590	.0415	.0291	.0203
W	0	1.5342	1.5129	1.4984	1.4884	1.4814	1.4765
2 3 5							
P1	0	0	.4562	.4656	.4723	.4772	.4808
P2	0	0	.3082	.3177	.3246	.3297	.3333
P3	0	0	.1406	.1479	.1533	.1573	.1603
P4	0	0	.0950	.0689	.0498	.0358	.0257
W	0	0	1.4805	1.4656	1.4550	1.4476	1.4422
2 3 6							
P1	0	0	0	.4801	.4874	.4929	.4969
P2	0	0	0	.3326	.3403	.3460	.3502
P3	0	0	0	.1106	.1158	.1197	.1227
P4	0	0	0	.0767	.0565	.0414	.0303
W	0	0	0	1.4433	1.4323	1.4244	1.4187
2 3 7							
P1	0	0	0	0	.4986	.5045	.5089
P2	0	0	0	0	.3521	.3583	.3630
P3	0	0	0	0	.0875	.0912	.0940
P4	0	0	0	0	.0618	.0460	.0341
W	0	0	0	0	1.4162	1.4079	1.4018
2 3 8							
P1	0	0	0	0	0	.5133	.5180
P2	0	0	0	0	0	.3677	.3728
P3	0	0	0	0	0	.0694	.0720
P4	0	0	0	0	0	.0497	.0373
W	0	0	0	0	0	1.3958	1.3895
2 3 9							
P1	0	0	0	0	0	0	.5249
P2	0	0	0	0	0	0	.3803
P3	0	0	0	0	0	0	.0550
P4	0	0	0	0	0	0	.0398
W	0	0	0	0	0	0	1.3803
2 4 5							
P1	0	0	.4855	.4957	.5032	.5088	.5130
P2	0	0	.2357	.2457	.2532	.2588	.2631
P3	0	0	.1643	.1730	.1796	.1846	.1885
P4	0	0	.1145	.0857	.0641	.0478	.0355
W	0	0	1.4351	1.4204	1.4098	1.4020	1.3962
2 4 6							
P1	0	0	0	.5107	.5188	.5249	.5295
P2	0	0	0	.2608	.2691	.2755	.2804
P3	0	0	0	.1332	.1396	.1446	.1485
P4	0	0	0	.0952	.0724	.0550	.0416
W	0	0	0	1.3993	1.3884	1.3803	1.3742

FOUR LEVEL

	4	5	6	7	8	9	10
2 4 7							
P1	0	0	0	0	.5305	.5371	.5421
P2	0	0	0	0	.2815	.2885	.2939
P3	0	0	0	0	.1088	.1135	.1173
P4	0	0	0	0	.0792	.0610	.0468
W	0	0	0	0	1.3729	1.3645	1.3582
2 4 8							
P1	0	0	0	0	0	.5464	.5518
P2	0	0	0	0	0	.2986	.3044
P3	0	0	0	0	0	.0891	.0927
P4	0	0	0	0	0	.0659	.0511
W	0	0	0	0	0	1.3528	1.3463
2 4 9							
P1	0	0	0	0	0	0	.5593
P2	0	0	0	0	0	0	.3128
P3	0	0	0	0	0	0	.0732
P4	0	0	0	0	0	0	.0547
W	0	0	0	0	0	0	1.3372
2 5 6							
P1	0	0	0	.5323	.5409	.5475	.5526
P2	0	0	0	.2068	.2152	.2218	.2270
P3	0	0	0	.1509	.1583	.1641	.1688
P4	0	0	0	.1101	.0856	.0665	.0515
W	0	0	0	1.3766	1.3596	1.3514	1.3452
2 5 7							
P1	0	0	0	0	.5531	.5601	.5656
P2	0	0	0	0	.2275	.2348	.2405
P3	0	0	0	0	.1258	.1315	.1360
P4	0	0	0	0	.0936	.0736	.0579
W	0	0	0	0	1.3446	1.3362	1.3297
2 5 8							
P1	0	0	0	0	0	.5698	.5756
P2	0	0	0	0	0	.2451	.2514
P3	0	0	0	0	0	.1054	.1098
P4	0	0	0	0	0	.0796	.0632
W	0	0	0	0	0	1.3247	1.3180
2 5 9							
P1	0	0	0	0	0	0	.5836
P2	0	0	0	0	0	0	.2602
P3	0	0	0	0	0	0	.0886
P4	0	0	0	0	0	0	.0677
W	0	0	0	0	0	0	1.3090
2 6 7							
P1	0	0	0	0	.5698	.5772	.5831
P2	0	0	0	0	.1850	.1923	.1982
P3	0	0	0	0	.1397	.1461	.1514
P4	0	0	0	0	.1054	.0843	.0674
W	0	0	0	0	1.3247	1.3162	1.3096
2 6 8							
P1	0	0	0	0	0	.5873	.5934
P2	0	0	0	0	0	.2026	.2090
P3	0	0	0	0	0	.1190	.1240
P4	0	0	0	0	0	.0912	.0736
W	0	0	0	0	0	1.3049	1.2981

FOUR LEVEL

	4	5	6	7	8	9	10
2 6 9							
P1	0	0	0	0	0	0	.6017
P2	0	0	0	0	0	0	.2178
P3	0	0	0	0	0	0	.1017
P4	0	0	0	0	0	0	.0789
W	0	0	0	0	0	0	1.2892
2 7 8							
P1	0	0	0	0	0	.6008	.6072
P2	0	0	0	0	0	.1680	.1744
P3	0	0	0	0	0	.1302	.1359
P4	0	0	0	0	0	.1010	.0825
W	0	0	0	0	0	1.2902	1.2834
2 7 9							
P1	0	0	0	0	0	0	.6157
P2	0	0	0	0	0	0	.1831
P3	0	0	0	0	0	0	.1127
P4	0	0	0	0	0	0	.0885
W	0	0	0	0	0	0	1.2745
2 8 9							
P1	0	0	0	0	0	0	.6268
P2	0	0	0	0	0	0	.1543
P3	0	0	0	0	0	0	.1222
P4	0	0	0	0	0	0	.0967
W	0	0	0	0	0	0	1.2631
3 4 5							
P1	0	0	.3803	.3916	.4004	.4071	.4124
P2	0	0	.2755	.2865	.2951	.3017	.3069
P3	0	0	.1996	.2096	.2175	.2236	.2285
P4	0	0	.1446	.1122	.0871	.0675	.0522
W	0	0	1.3803	1.3668	1.3568	1.3492	1.3435
3 4 6							
P1	0	0	0	.4078	.4171	.4244	.4302
P2	0	0	0	.3025	.3117	.3190	.3247
P3	0	0	0	.1663	.1740	.1801	.1850
P4	0	0	0	.1234	.0972	.0765	.0601
W	0	0	0	1.3484	1.3384	1.3306	1.3247
3 4 7							
P1	0	0	0	0	.4302	.4379	.4440
P2	0	0	0	0	.3247	.3325	.3388
P3	0	0	0	0	.1397	.1456	.1504
P4	0	0	0	0	.1054	.0840	.0668
W	0	0	0	0	1.3247	1.3169	1.3108
3 4 8							
P1	0	0	0	0	0	.4485	.4550
P2	0	0	0	0	0	.3433	.3500
P3	0	0	0	0	0	.1179	.1225
P4	0	0	0	0	0	.0902	.0725
W	0	0	0	0	0	1.3064	1.3001
3 4 9							
P1	0	0	0	0	0	0	.4639
P2	0	0	0	0	0	0	.3591
P3	0	0	0	0	0	0	.0998
P4	0	0	0	0	0	0	.0773
W	0	0	0	0	0	0	1.2918

FOUR LEVEL

	4	5	6	7	8	9	10
3 5 6							
P1	0	0	0	.4302	.4400	.4477	.4539
P2	0	0	0	.2451	.2545	.2620	.2681
P3	0	0	0	.1850	.1936	.2005	.2061
P4	0	0	0	.1397	.1120	.0898	.0719
W	0	0	0	1.3247	1.3148	1.3071	1.3012
3 5 7							
P1	0	0	0	0	.4533	.4615	.4681
P2	0	0	0	0	.2675	.2756	.2822
P3	0	0	0	0	.1579	.1646	.1701
P4	0	0	0	0	.1213	.0983	.0796
W	0	0	0	0	1.3017	1.2940	1.2879
3 5 8							
P1	0	0	0	0	0	.4725	.4794
P2	0	0	0	0	0	.2866	.2936
P3	0	0	0	0	0	.1354	.1408
P4	0	0	0	0	0	.1055	.0862
W	0	0	0	0	0	1.2839	1.2777
3 5 9							
P1	0	0	0	0	0	0	.4885
P2	0	0	0	0	0	0	.3030
P3	0	0	0	0	0	0	.1166
P4	0	0	0	0	0	0	.0918
W	0	0	0	0	0	0	1.2697
3 6 7							
P1	0	0	0	0	.4711	.4796	.4865
P2	0	0	0	0	.2219	.2300	.2367
P3	0	0	0	0	.1727	.1801	.1862
P4	0	0	0	0	.1343	.1103	.0906
W	0	0	0	0	1.2852	1.2775	1.2714
3 6 8							
P1	0	0	0	0	0	.4909	.4981
P2	0	0	0	0	0	.2409	.2481
P3	0	0	0	0	0	.1499	.1559
P4	0	0	0	0	0	.1183	.0979
W	0	0	0	0	0	1.2677	1.2615
3 6 9							
P1	0	0	0	0	0	0	.5075
P2	0	0	0	0	0	0	.2576
P3	0	0	0	0	0	0	.1307
P4	0	0	0	0	0	0	.1043
W	0	0	0	0	0	0	1.2537
3 7 8							
P1	0	0	0	0	0	.5054	.5129
P2	0	0	0	0	0	.2035	.2106
P3	0	0	0	0	0	.1621	.1686
P4	0	0	0	0	0	.1291	.1080
W	0	0	0	0	0	1.2554	1.2493
3 7 9							
P1	0	0	0	0	0	0	.5225
P2	0	0	0	0	0	0	.2199
P3	0	0	0	0	0	0	.1427
P4	0	0	0	0	0	0	.1149
W	0	0	0	0	0	0	1.2415

FOUR LEVEL

	4	5	6	7	8	9	10
3 8 9							
P1	0	0	0	0	0	0	.5347
P2	0	0	0	0	0	0	.1883
P3	0	0	0	0	0	0	.1529
P4	0	0	0	0	0	0	.1241
W	0	0	0	0	0	0	1.2321
4 5 6							
P1	0	0	0	.3540	.3643	.3727	.3795
P2	0	0	0	.2730	.2830	.2912	.2979
P3	0	0	0	.2106	.2199	.2275	.2338
P4	0	0	0	.1624	.1327	.1085	.0887
W	0	0	0	1.2965	1.2871	1.2798	1.2740
4 5 7							
P1	0	0	0	0	.3781	.3869	.3941
P2	0	0	0	0	.2965	.3052	.3123
P3	0	0	0	0	.1823	.1898	.1961
P4	0	0	0	0	.1430	.1181	.0975
W	0	0	0	0	1.2752	1.2679	1.2621
4 5 8							
P1	0	0	0	0	0	.3985	.4060
P2	0	0	0	0	0	.3166	.3241
P3	0	0	0	0	0	.1588	.1648
P4	0	0	0	0	0	.1262	.1050
W	0	0	0	0	0	1.2586	1.2527
4 5 9							
P1	0	0	0	0	0	0	.4158
P2	0	0	0	0	0	0	.3339
P3	0	0	0	0	0	0	.1388
P4	0	0	0	0	0	0	.1115
W	0	0	0	0	0	0	1.2453
4 6 7							
P1	0	0	0	0	.3961	.4052	.4128
P2	0	0	0	0	.2493	.2580	.2652
P3	0	0	0	0	.1978	.2058	.2126
P4	0	0	0	0	.1569	.1310	.1095
W	0	0	0	0	1.2605	1.2534	1.2476
4 6 8							
P1	0	0	0	0	0	.4170	.4249
P2	0	0	0	0	0	.2693	.2770
P3	0	0	0	0	0	.1739	.1805
P4	0	0	0	0	0	.1398	.1177
W	0	0	0	0	0	1.2444	1.2386
4 6 9							
P1	0	0	0	0	0	0	.4349
P2	0	0	0	0	0	0	.2868
P3	0	0	0	0	0	0	.1536
P4	0	0	0	0	0	0	.1247
W	0	0	0	0	0	0	1.2314
4 7 8							
P1	0	0	0	0	0	.4320	.4401
P2	0	0	0	0	0	.2302	.2378
P3	0	0	0	0	0	.1866	.1937
P4	0	0	0	0	0	.1513	.1285
W	0	0	0	0	0	1.2335	1.2278

FOUR LEVEL

	4	5	6	7	8	9	10
4 7 9							
P1	0	0	0	0	0	0	.4503
P2	0	0	0	0	0	0	.2475
P3	0	0	0	0	0	0	.1661
P4	0	0	0	0	0	0	.1361
W	0	0	0	0	0	0	1.2207
4 8 9							
P1	0	0	0	0	0	0	.4630
P2	0	0	0	0	0	0	.2144
P3	0	0	0	0	0	0	.1768
P4	0	0	0	0	0	0	.1459
W	0	0	0	0	0	0	1.2123
5 6 7							
P1	0	0	0	0	.3365	.3460	.3538
P2	0	0	0	0	.2707	.2798	.2874
P3	0	0	0	0	.2177	.2263	.2335
P4	0	0	0	0	.1751	.1480	.1252
W	0	0	0	0	1.2433	1.2365	1.2309
5 6 8							
P1	0	0	0	0	0	.3579	.3661
P2	0	0	0	0	0	.2915	.2995
P3	0	0	0	0	0	.1932	.2004
P4	0	0	0	0	0	.1574	.1340
W	0	0	0	0	0	1.2281	1.2226
5 6 9							
P1	0	0	0	0	0	0	.3764
P2	0	0	0	0	0	0	.3096
P3	0	0	0	0	0	0	.1723
P4	0	0	0	0	0	0	.1417
W	0	0	0	0	0	0	1.2158
5 7 8							
P1	0	0	0	0	0	.3729	.3814
P2	0	0	0	0	0	.2513	.2593
P3	0	0	0	0	0	.2063	.2139
P4	0	0	0	0	0	.1694	.1454
W	0	0	0	0	0	1.2181	1.2126
5 7 9							
P1	0	0	0	0	0	0	.3919
P2	0	0	0	0	0	0	.2694
P3	0	0	0	0	0	0	.1852
P4	0	0	0	0	0	0	.1536
W	0	0	0	0	0	0	1.2061
5 8 9							
P1	0	0	0	0	0	0	.4047
P2	0	0	0	0	0	0	.2352
P3	0	0	0	0	0	0	.1963
P4	0	0	0	0	0	0	.1638
W	0	0	0	0	0	0	1.1983
6 7 8							
P1	0	0	0	0	0	.3242	.3327
P2	0	0	0	0	0	.2687	.2770
P3	0	0	0	0	0	.2227	.2306
P4	0	0	0	0	0	.1846	.1598
W	0	0	0	0	0	1.2065	1.2013

FOUR LEVEL

	4	5	6	7	8	9	10
6 7 9							
P1	0	0	0	0	0	0	.3433
P2	0	0	0	0	0	0	.2873
P3	0	0	0	0	0	0	.2011
P4	0	0	0	0	0	0	.1683
W	0	0	0	0	0	0	1.1951
6 8 9							
P1	0	0	0	0	0	0	.3561
P2	0	0	0	0	0	0	.2524
P3	0	0	0	0	0	0	.2125
P4	0	0	0	0	0	0	.1789
W	0	0	0	0	0	0	1.1878
7 8 9							
P1	0	0	0	0	0	0	.3149
P2	0	0	0	0	0	0	.2669
P3	0	0	0	0	0	0	.2263
P4	0	0	0	0	0	0	.1919
W	0	0	0	0	0	0	1.1795

	APPLICATIONS	COMPUTER APPLICATIONS	BAYESIAN ANALYSIS	REGRESSION ANALYSIS	ANALYSIS OF VARIANCE	INFORMATION THEORY	PRESENT CONTRIBUTIONS OF INFORMATION THEORY	AREAS FOR FUTURE RESEARCH
RANKING	MINIMUM COST ANALYSIS (CHURCHMAN, 1954)	*SORT (OS3)	X	X	X	ONE-FACTOR INFOR. GAIN	ILLUSTRATES LEARNING ALLOWS COST/INFORMATION TRADEOFF	LEARNING MODELS PRACTICAL APPLIC. (QUALITY CONTROL, ETC.)
	DECISIONS AMONG COMPET. ALTERNATIVES (BARISH, 1962)		X			INFOR. GAIN		
HYPOTHESIS TESTING	HUMAN COMMUNICATIONS (CHERRY, 1952)	*SIPS (OS3)				INFORMATION TRANSMITTED	ESTABLISHES MEANS TO IDENTIFY "FACTORS" WHICH INFLUENCE DECISION MAKER'S ACTIONS	LARGE SCALE PRACTICAL APPLIC.
	ADVERTISING & MARKETING (MACHOL, 1960) (BUZZEL, 1969)		X	X	X	ONE-FACTOR MULTI-FACTOR		
ESTIMATION	ECONOMETRICS (THIEL,1967)	*MULREG (GE)		X	X	X	ILLUSTRATES LEARNING AIDS IN EXPLAINING HUMAN BEHAVIOR	LEARNING MODELS EXPECTED VALUE HUMAN FACTORS
	ESTHETIC PERCEPTION (MOLES, 1968)					X		
	OPERATIONS RESEARCH (MUROGA, 1953) (KUNISAWA, 1958)		X			MULTI-FACTOR		
	INFORMATION THEORY (SHIRLAND, 1971)	*LES1 (OS3)				ONE-FACTOR		
SIMULATION (MARKOV CHAINS)	CODING (SHANNON, 1949)	MSIP (OS3) GERT				X	OFFERS EXPLANATION OF BEHAVIOR IN MARKOV CHAINS AND QUEUEING	GAME THEORY LARGE SCALE STOCHASTIC MODELS
	MANAGEMENT INFORMATION SYSTEMS (KUNISAWA, 1959		X			MULTI-FACTOR		
	BRAND SWITCHING (GREEN, 1967)		X			MULTI-FACTOR	IDENTIFIES NUMBER OF STEPS TO REACH STEADY-STATE IN MARKOV CHAINS	
	STOCHASTIC EQUILIBRIUM (KARLAN, 1969)			X		ONE-FACTOR MULTI-FACTOR	ALLOWS SIMULATIONS TO BEGIN AT STEADY-STATE	
	GENERAL INFORMATION MODELLING					ONE-FACTOR MULTI-FACTOR	ALLOWS FOR LARGE SCALE MODELLING	GENERALIZED MODELLING LANG.