Entropy Methods for Joint Distributions in Decision Analysis

Ali E. Abbas, *Member, IEEE*

***Abstract*—A fundamental step in decision analysis is the elicitation of the decision maker's information about the uncertainties of the decision situation in the form of a joint probability distribution. This paper presents a method based on the maximum entropy principle to obtain a joint probability distribution using lower order joint probability assessments. The approach reduces the number of assessments significantly and also reduces the number of conditioning variables in these assessments. We discuss the order of the approximation provided by the maximum entropy distribution with each lower order assessment using a Monte Carlo simulation and discuss the implications of using the maximum entropy distribution in Bayesian inference. We present an application to a practical decision situation faced by a semiconductor testing company in the Silicon Valley.**

***Index Terms*—Decision analysis, eliciting probability distributions, maximum entropy, probabilistic dependence, project evaluation/selection.**

I. Introduction

ELICITING a representative probability distribution from a decision maker is a task that requires care if one is to avoid cognitive and motivational biases [1]–[3]. When eliciting a joint probability distribution, however, we are faced with the added difficulty of conditioning the probability assessments on several variables in order to capture the dependence relationships between them. In itself, this is not an easy task, but in addition, the number of assessments needed for a joint distribution of N variables, each discretized to k outcomes, can be on the order of k^N. For these reasons, eliciting a joint probability distribution may be a difficult task to perform in practice, and several methods have been proposed to incorporate dependence between the variables (see, for example [4]–[8]).

In this paper, we present a method that takes into account the difficulty of the elicitation process as a practical constraint on the side of the decision-maker. We use the discrete form maximum entropy principle to obtain a joint probability distribution using lower order assessments. This approach reduces the number of assessments significantly and facilitates the probability assessment process by conditioning on fewer variables. The maximum entropy formulation embeds several approaches for approximating joint distributions found in the literature [9]–[11], and provides the flexibility to model information that may be available in many different forms [12].

Manuscript received June 12, 2004; revised October 1, 2004, April 1, 2005, and June 12, 2005. Review of this manuscript was arranged by Department Editor A. E. Marucheck.

The author is with the School of Engineering, University of Illinois at Urbana–Champaign, Urbana, IL 61822 USA (e-mail: aliabbas@uiuc.edu).

Digital Object Identifier 10.1109/TEM.2005.861803

When used to construct joint probability distributions in decision analysis, the maximum entropy approach has several advantages: 1) it incorporates as much (or as little) information as there is available at the time of making the decision; 2) it makes no assumptions about a particular form of a joint distribution; 3) it applies to both numeric and nonnumeric variables (such as the categorical events Sunny/Rainy for weather); and 4) it does not limit itself to the use of only moments and correlation coefficients, which may be difficult to elicit in decision analysis practice.

We do observe, however, that there is a tradeoff between the accuracy of the maximum entropy joint distribution and the number of probability assessments that are incorporated into the formulation. We discuss this tradeoff in more detail in this paper and illustrate how a decision maker can assess the order of the approximation provided for a given decision problem. We also discuss a generalization of the maximum entropy principle known as the minimum cross-entropy principle.

Our focus in this paper is: 1) to familiarize the readers with the maximum entropy method for constructing representative probability distributions in decision analysis; 2) to emphasize the generality of the maximum entropy approach for constructing joint distributions using lower order assessments; and 3) to provide insights on the approximations obtained using each of the lower order assessments. This study enables a better understanding of both the number and order of lower order assessments needed for a given decision situation, before any probability assessments are made. We present an application to a real life decision situation faced by a semiconductor testing company in the San Francisco Bay area [13].

The remainder of this paper is structured as follows. Section II presents a motivating example for the maximum entropy approach when applied to a decision situation in practice. Section III reviews the definition of entropy, the maximum entropy principle, and the maximum entropy formulation for constructing a joint distribution given lower order assessments. Section IV presents a Monte Carlo simulation to investigate the order of the approximations provided. Section V presents the analysis of the decision situation, and Section VI investigates the order of the approximations when used for Bayesian inference.

II. The Executive's Problem

The decision-maker is the president of semiconductor operations for a multinational Fortune 500 company in the San Francisco Bay area. He has just attended a board meeting in which he heard news from the engineering department about the technical difficulties that may arise in the company's current line of

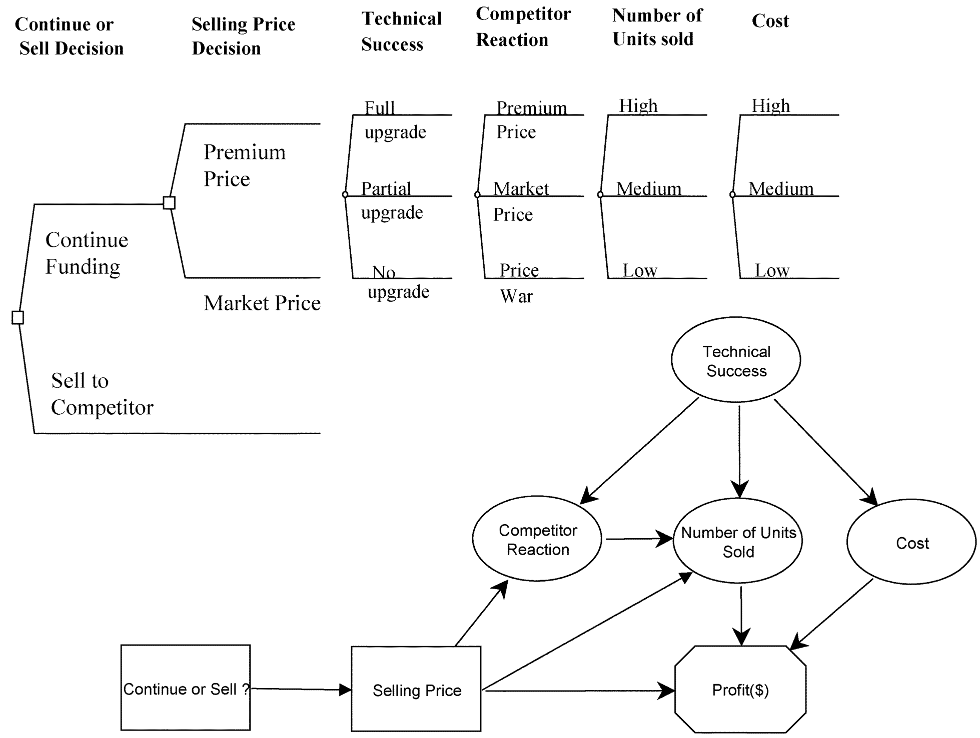

Fig. 1. Decision diagram and generic tree for the decision.

semiconductor test equipment: the product is currently operational but produces occasional intermittent failures during operation. The market for this line of semiconductor test equipment is a small niche market with two main players. The company's main competitor is also launching a similar product but lacks some technical know-how on design implementation. The competitor is interested in acquiring this product to expedite the time needed to launch his own product and to increase his market share.

The decision-maker realizes he has two decisions to make. The first decision is whether to continue funding this line of products till launch or to sell it to the competitor, who has submitted a bid of $25 million dollars. If the decision maker decides to continue funding the project, he will have to make a second decision about the selling price differentiation for this product: whether to sell at a premium price of $8 million or a market price of $5 million per machine.

The president consults the engineering and marketing departments to learn about their experience with similar types of Hi-Tech products. The directors of both departments emphasize that the following uncertainties are most important to the decision situation.

1) **Technical success**. This is determined by the ability of the company to include all of the desired technical upgrades, a partial modification upgrade, or no upgrades before launching the product.
2) **Cost**. These include the fixed overhead costs and variable costs needed to further support the product and implement the upgrades. The director of marketing also provided estimates about the Cost needed to market the product once it is out of the R&D phase.
3) **Number of units sold**. The director pointed out that the demand for the product would change significantly according to whether or not the product had the full technology upgrade.
4) **Competitor reaction**. The competitor may decide to sell their own product at the market price, sell at a premium price or enter into a price war and bring down their selling price. Judging by past experience and the size of the competitor, the data suggests that it is more likely the competitor will enter into a price war to force them out of the market or at least make their upgrade economically infeasible. The life cycle for this line of test equipment is short so it is unlikely that the competitor will start with a price war, and then change their prices later.

The decision diagram for this situation is shown in Fig. 1. The value node represents the profit in dollars coming from this product and is determined by

$$\text{Profit } (\$) = (\text{Number of Units sold}) * (\text{Selling Price} - \text{Cost}).$$

Since this line of products represents a small fraction of the overall projects for the multinational company, they are risk-neutral within this range of monetary prospects; they value deals at their expected value.

In some cases that arise in practice, conditional probability independence conditions may exist among the variables and reduce the number of assessments needed. In this particular decision, situation Cost is conditionally independent of both Competitor Reaction and Number of Units Sold given Technical Success. In other cases, however, probability independence conditions may not exist and, thus, we would need a number of assessments on the order of k^N. The large number of probability assessments needed makes eliciting a joint distribution, using the classical marginal-conditional approach, a difficult task to perform in practice.

In Section III, we provide a review of the maximum entropy method for constructing joint distributions and formulate this decision problem using the maximum entropy approach.

III. Maximum Entropy Formulation

A. Interpretation of the Discrete Entropy Expression

In 1948, Shannon [14] introduced a measure of uncertainty about a discrete random variable X having a probability mass function $p(x)$ as

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log\left(p(x_i)\right). \tag{1}$$

He called this term entropy, which is also a measure of the amount of information needed to describe the outcome of a discrete random variable. Shannon's entropy expression marked the beginning of the information age and the dawn of a new science called "Information Theory." It has now become a unifying theory with profound intersections with Probability, Computer Science, Management Science, Economics, Bioinformatics, and many other fields. To get an intuitive feel for the entropy expression, let us consider the following example.

B. Example: The Entropy of a Discrete Random Variable

A random variable, X, can have four possible outcomes {0,1,2,3}, with probabilities (1/2,1/4,1/8,1/8), respectively. Let us calculate the entropy of X using base two for the logarithm in the entropy expression

$$\begin{aligned} H(X) = &-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \left(\frac{1}{4}\right)\log_2\left(\frac{1}{4}\right) \\ &-\left(\frac{1}{8}\right)\log_2\left(\frac{1}{8}\right) - \left(\frac{1}{8}\right)\log_2\left(\frac{1}{8}\right) \\ = &1\frac{3}{4}. \end{aligned} \tag{2}$$

One intuitive way to explain this number is to consider the minimum expected number of binary (Yes/No) questions needed to describe an outcome of X. The most efficient way to ask the questions in this example is to start by asking about the outcome with the highest probability of occurrence: i.e., we ask "Is $X = 0$?" If it is, then we have determined X in one question. If it is not, then we ask "Is $X = 1$?" Again, if it is correct, then we have determined X in two questions, if it is not, we ask "Is $X = 2$?" We do not need to ask "Is $X = 3$?" because if it is not 0, 1, or 2, then it must be 3. The expected number of binary questions needed to determine X is

$$\begin{aligned} E(X) &= 1\text{x}\frac{1}{2} + 2\text{x}\frac{1}{4} + 3\text{x}\frac{1}{8} + 3\text{x}\frac{1}{8} = 1\frac{3}{4} \\ &= H(X) \\ &= \text{Entropy of the random variable } X. \end{aligned} \tag{3}$$

This result provides an intuitive interpretation for the entropy of a discrete random variable as the expected number of questions needed to describe its outcome. While this equality does not always hold, an optimal question selection process always exists using coding theory such that the expected number of questions is bounded by the following well-known data compression inequality (explained in detail in [15, p. 87])

$$H(X) \leq \text{Expected number of questions} \leq H(X) + 1. \tag{4}$$

The entropy expression, thus, provides some notion of the amount of information needed to describe the outcome of a discrete random variable. The fact that binary questions were used in this example corresponds to the base two in the logarithm of the entropy expression. The interpretation of entropy as the expected number of questions also shows that the entropy of a discrete variable is nonnegative.

C. Interpretation of the Relative Entropy

Kullback and Leibler [16] generalized Shannon's definition of entropy and introduced a measure that is used frequently to determine the "closeness" between two probability distributions. The Kullback–Leibler (KL) measure of a distribution P from a reference distribution Q is

$$K(p:q) = \sum_{i=1}^{n} p(x_i) \log \frac{p(x_i)}{q(x_i)}. \tag{5}$$

The KL-measure is also known as the relative entropy or cross entropy. The KL-measure is nonnegative and is zero if and only if the two distributions are identical. However, it is not symmetric and does not follow the triangle inequality. As a result, it is not a distance measure but rather a measure of directed divergence of P from Q. The following example provides an intuitive interpretation for the KL-measure.

D. Example: The Relative Entropy

Consider another person wishing to determine the value of the outcome of X in the previous example. However, this person does not know the generating distribution P and believes it is generated by a different probability distribution $Q = (1/8, 1/8, 1/2, 1/4)$. The KL-measure of the distribution P from Q is equal to

$$\begin{aligned} K(p:q) = &\frac{1}{2}\log_2\left(\frac{\frac{1}{2}}{\frac{1}{8}}\right) + \frac{1}{4}\log_2\left(\frac{\frac{1}{4}}{\frac{1}{8}}\right) \\ &+ \frac{1}{8}\log_2\left(\frac{\frac{1}{8}}{\frac{1}{2}}\right) + \frac{1}{8}\log_2\left(\frac{\frac{1}{8}}{\frac{1}{4}}\right) \\ = &\frac{7}{8}. \end{aligned} \tag{6}$$

Now, let us calculate the expected number of binary questions needed for this second person to determine X. His question ordering based on the distribution Q would be "Is $X = 2$?," "Is $X = 3$?," "Is $X = 0$?" The expected number of binary questions needed to determine X is

$$E(X) = 1\text{x}\frac{1}{8} + 2\text{x}\frac{1}{8} + 3\text{x}\frac{1}{2} + 3\text{x}\frac{1}{4} = \frac{21}{8}. \tag{7}$$

Note that the difference in the expected number of questions needed to determine X for the two persons is equal to $(21/8) - 1(3/4) = 7/8 = K(p:q)$.

This example provides an intuitive interpretation for the KL-measure when used as a measure of "closeness" between a distribution P and another distribution, Q. We can think of the KL-measure as the increase in the number of questions that distribution Q would introduce if it were used to determine the outcome of a random variable that was generated by the distribution P. In coding theory, the relative entropy has an equivalent interpretation as the increase in expected length of the optimal code needed to describe a discrete random variable whose probability distribution is whose probability distribution is P and the coding is based on the distribution Q [15].

E. Maximum Entropy and Minimum Cross Entropy Principles

In 1957, Jaynes [17] built on the concept of entropy and proposed a method for assigning probabilities based on partial information. This is called the maximum entropy principle and is stated from his original paper as follows:

> *In making inferences on the basis of partial information, we must use that probability distribution which has maximum entropy subject to whatever is known. This is the only unbiased assignment we can make; to use any other would amount to arbitrary assumption of information, which by hypothesis we do not have.*

The maximum entropy principle is considered to be an extension of Laplace's principle of insufficient reason as it assigns equal probabilities to all outcomes when no further information is available except that the probabilities are nonnegative and sum to one. Furthermore, the maximum entropy principle allows us to incorporate additional information that may be available in many real-life decision problems. This additional information can have several forms such as moment constraints, probability constraints, and shape constraints. As a result, maximum entropy methods have found several applications for assigning univariate distributions in decision analysis (see, for example, [18] and [19]). In Appendix I, we review the solution to the maximum entropy formulation given moments and probability constraints.

The maximum entropy principle can also be applied to construct joint distributions using lower order assessments. When only the marginal distributions are available, the maximum entropy joint distribution assumes probabilistic independence between the variables (see, for example, [20]). This assignment, however, does not capture interdependencies that may occur in decision analysis practice. A natural extension is to consider the maximum entropy joint distribution given pairwise assessments between the variables. For example, the maximum entropy formulation of a joint distribution of four variables (not necessarily binary) given the knowledge of the pairwise joint assessments is

$$p^*_{i,j,k,l} = \underset{p_{i,j,k,l}}{\operatorname{argmax}} - \sum_{i,j,k,l} p_{i,j,k,l} \log(p_{i,j,k,l})$$

such that

$$\begin{aligned}
&\sum_{k,l} p_{i,j,k,l} = p_{ij..}, \sum_{j,l} p_{i,j,k,l} = p_{i.k.}, \sum_{j,k} p_{i,j,k,l} = p_{i..l},\\
&\sum_{i,l} p_{i,j,k,l} = p_{.jk.}, \sum_{i,k} p_{i,j,k,l} = p_{.j.l}, \sum_{i,j} p_{i,j,k,l} = p_{..kl.},\\
&\sum_{i,j,k,l} p_{i,j,k,l} = 1,\ p_{i,j,k,l} \geq 0 \quad \forall i,j,k,l,
\end{aligned} \tag{8}$$

where the subscripts refer to the variables in the order of assessment provided. For example, $p_{i,j,k,l}$ refers to the joint probability of the ith outcome (branch of the tree) of the first variable, the jth outcome of the second variable, the kth outcome of the third variable, and the lth outcome of the fourth variable. A dot "·" means that variable has been summed over. For example $p_{2...3}$ refers to the pairwise joint probability of the second branch of the first variable and the third branch of the forth variable.

Evaluating the Lagrange multiplier for (8), and equating its first partial derivative with respect to $p_{i,j,k,l}$ with zero gives

$$p^*_{i,j,k,l} = e^{-1-\lambda_o-\lambda_{ij...}-\lambda_{i.k.}-\lambda_{i...l}\lambda_{.jk.}-\lambda_{.j.l}-\lambda_{...kl}} \tag{9}$$

where λ's are the Lagrange multipliers for the constraints and the subscript notation is the same as that for the probability notation [for example, $\lambda_{ij...}$ corresponds to the constraint which has the pairwise assessment $p_{ij...}$ in (8)], and λ_0 is a normalizing constant.

Formulations similar to (8) can also be solved using iterative methods (see, for example, Ku and Kullback [21] who developed iterative solutions to approximate discrete maximum entropy distributions). Other solution approaches also take a look at the dual optimization problem [22]. With the advancement of high-speed computing in the last two decades, maximum entropy solutions can now be obtained numerically using many software packages. Some of these packages include the Solver in Excel, Matlab, Mathematica, as well as many other programs. In Appendix II, we provide, as an illustration, a Maltab code to determine the maximum entropy joint distribution using lower order assessments.

In a similar manner, the solution to the maximum entropy formulation for four variables using three-way joint assessments is

$$p^*_{i,j,k,l} = e^{-1-\lambda_o-\lambda_{ijk.}-\lambda_{j.kl}-\lambda_{ij.l}-\lambda_{.jkl}}. \tag{10}$$

We now have a whole range of maximum entropy joint distributions for the four-variable joint distribution, depending on the amount of information that is available (or is feasible to assess). These include 1) assigning equal probabilities to all outcomes when no further information is available; 2) assigning a joint distribution equal to the product of the marginal distributions when only the marginal distributions are available; 3) assigning the solution proposed by (9) when only the pairwise distributions are available; and 4) assigning the solution proposed by (10) when the three-way joint distributions are available. In addition, we may have other formulations that include only a fraction of the number of lower order assessments needed, or moments and pairwise correlation coefficients between the variables.

A natural question now is how we can assess whether some higher order marginals or some different types of marginals are needed for a given problem if the actual joint distribution is unknown. The answer to this question will be the focus of the next three sections. In particular, we will conduct a Monte Carlo simulation to test the accuracy of the approximations provided by the lower order probability assessments when applied to a particular decision situation. First, we discuss a generalization of the maximum entropy principle, known as the minimum cross-entropy principle below.

TABLE I
TOTAL VARIATION VERSUS TWICE MAXIMUM DEVIATION

Example: Total Variation is Greater than Twice Maximum Deviation

					Sum	
P	0.5988	0.1159	0.1877	0.0976	1.000	
Q	0.3502	0.3418	0.0341	0.2739	1.000	
Abs(P-Q)	**0.2487**	0.2259	0.1536	0.1763	0.804	*Total Variation*
					0.2487	*Max Deviation*

Example: Total Variation is Twice Maximum Deviation

					Sum	
P	0.0004	0.4560	0.1665	0.3770	1.000	
Q	0.3559	0.4306	0.0803	0.1332	1.000	
Abs(P-Q)	**0.3555**	0.0255	0.0863	0.2438	0.711	*Total Variation*
					0.3555	*Max Deviation*

The minimum cross-entropy principle incorporates into the formulation a prior probability distribution for the variable of interest. Minimum cross-entropy formulations minimize the KL-divergence measure of a distribution P from a prior distribution Q and satisfy the available information constraints. The maximum entropy principle is a special case of the minimum cross-entropy principle when the prior distribution Q is uniform.

Now, we review another common measure of "closeness" between two distributions, known as the total variation, and provide an interpretation for it below.

F. Total Variation

The total variation between a probability distribution P and a probability distribution Q with n outcomes is

$$\text{Total Variation} = \sum_{i=1}^{n} |p_i - q_i| \tag{11}$$

where p_i and q_i are the probabilities assigned to outcome i for distributions P and Q, respectively.

The total variation between two distributions is zero if and only if the two distributions are identical. Furthermore, half the total variation is the upper bound on the maximum deviation $\max_i |p_i - q_i|$ between any two discrete probabilities in the distributions P and Q [23]. For example, consider the two distributions P and Q shown in Table I. Each distribution has four possible outcomes. The maximum deviation between P and Q in the top table is 0.2487 and the total variation is 0.804. The total variation is higher than twice the maximum deviation. On the other hand, in the second table, we have a maximum deviation of 0.3554 and a total variation of 0.7109. The maximum deviation, thus, achieves its upper bound of half the total variation.

For the case of two-outcome and three-outcome distributions, the total variation is in fact always equal to twice the maximum deviation between them. However, as the number of variables increases, the total variation may exceed the maximum deviation by a factor greater than 2. We note here that Kullback [24] developed a lower bound on the KL-measure between two distributions in terms of their total variation. Several other authors have also compared distributions according to entropy measures. See, for example, [25]–[27] who provide several applications and unification of information indices.

Building on the previous discussion, we now discuss the performance of the maximum entropy approximation based on the total variation measure, the maximum deviation, and the relative entropy using a Monte Carlo simulation. As we discussed, this is equivalent to a uniform reference prior distribution Q in the minimum cross-entropy formulation. Other formulations can include different forms of the distribution Q. Next, we illustrate how a decision maker can determine the order of each approximation and the percent of error in dollar values for a given decision situation before conducting any of the probability assessments.

IV. MONTE CARLO SIMULATION

A. Simulation Steps When No Further Information is Available

In the analysis of our problem, each of the four variables is discretized to three possible outcomes. To determine the order of the approximation provided by each of the lower order assessments, we conducted a Monte Carlo simulation to simulate a discrete four-variable joint distribution, with each variable discretized to three outcomes. The simulation method uses a data set that is uniformly sampled from the space of all-possible $3 \times 3 \times 3 \times 3$ outcome joint distributions, and provides a general estimate of the order of the approximation provided. The simulation was run using the PC version of Matlab v 7 in the following steps.

Step 1) Generate a $3 \times 3 \times 3 \times 3$ outcome joint probability distribution uniformly from the space of possible 3^4 outcome joint probability distributions. To generate a joint distribution uniformly, we carried out the following procedure.

a) Generate $(3^4 - 1)$ independent samples, $x_1, x_2, \ldots, x_{3^4-1}$ from a uniform [0,1] distribution.

b) Sort the generated samples from lowest to highest to form an order statistic $u_1 \leq u_2 \leq \ldots \leq u_{3^4-1}$. This method generates an order statistic uniformly from the space of all possible values.

c) Take the difference between each two consecutive elements of the order statistic $\{u_1 - 0, u_2 - u_1, \ldots, u_{3^4-1} - u_{3^4-2}, 1 - u_{3^4-1}\}$.

The increments form a 3^4 outcome probability distribution that is uniformly sampled from the space of possible 3^4 outcome probability distributions. For more information on this sampling method and on the distributions of order statistics, we refer the readers to [28]. To further characterize the simulation data set given that there are categorical (nonnumeric) variables present, we assigned numerical values of 0, 1, and 2 to the three outcomes of each of the four variables and calculated the correlation coefficients. (Of course, the values of the correlation coefficients may change for different values assigned to the outcomes.) Table II shows the range and mean values of the pairwise correlation coefficients that were generated assuming values of 0, 1, and 2 for the outcomes.

From Table II, we see that the average value of the pairwise correlation coefficients generated

TABLE II
RANGE AND MEAN VALUES OF CORRELATION COEFFICIENTS

Pairwise Variables	Mean Value of Correlation	Min. Value	Max. Value
(Tech Success, Comp. Reac)	0.0000	-0.426	0.452
(Tech Success,# units sold)	0.0000	-0.428	0.452
(Tech Success, Cost)	-0.0002	-0.582	0.435
(Comp. Reac, # units sold)	-0.0003	-0.506	0.410
(Comp. Reac, Cost)	-0.0003	-0.442	0.480
(# units sold, Cost)	-0.0001	-0.419	0.440

TABLE III
SIMULATION RESULTS FOR MAXIMUM ENTROPY APPROXIMATION

	Maximum Deviation		Total Variation		Relative Entropy	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Equal Probability	0.0490	0.0140	0.7309	0.0555	0.5709	0.0906
Marginal Probability	0.0439	0.0123	0.6900	0.0571	0.5172	0.0896
Pairwise Probability	0.0269	0.0061	0.5245	0.0582	0.3277	0.0753
Three-way Probability	0.0085	0.0022	0.2121	0.0488	0.0760	0.0321

in the 5000 samples was zero, and that the range of the correlation coefficients was approximately $[-0.6, 0.5]$.

Step 2) Compute the lower order: 1) marginal distributions; 2) pairwise distributions; and 3) three-way joint distributions for the 3^4 outcome joint distributions that were generated in Step 1.

Step 3) Calculate the maximum entropy joint distribution given: 1) marginal distributions; 2) pairwise distributions; 3) three-way distributions; and 4) the equal probability distribution (no information is available except that probabilities are nonnegative and sum to one).

Step 4) Calculate the total variation, maximum deviation, and relative entropy measures between the probability distribution generated in Step 1 and the approximations obtained in Step 3.

Step 5) Repeat steps 1)–4).

The simulation results for the maximum deviation, total variation, and relative entropy are shown in Table III.

From Table III, we see the order of the approximation provided by the maximum entropy approach for the generated data set. We observe that there is a significant reduction in the mean values of the maximum deviation; total variation, and relative entropy as higher order probability assessments are incorporated into the formulation. For example, we see that ratio of the mean of the maximum deviations of the three way to the pairwise assessments is $(0.0085/0.0269) \approx 31\%$, this ratio for the total variation is $(0.2121/0.5245) \approx 40\%$, and for the relative entropy is $(0.076/0.3277) \approx 23.2\%$. Note also that the expected value of the maximum deviation obtained using three-way assessments is on the order of 0.0085, a small deviation in many problems. We do observe, however, that while these deviations appear to be small in magnitude, they are in fact comparable to the order of $1/n = 1/81 = 0.0123$.

B. Simulation Incorporating Specific Knowledge About the Decision Situation

In many cases, additional information may be available about the joint distribution under consideration. For example, we may have bounds on some of the joint probabilities or have knowledge of certain dependence structures between the variables. This information can be used to generate data sets that are specific to that particular decision situation, and can help the decision maker determine the order of the different approximations before any probability assessments are made. In this section, we present an example that incorporates knowledge of certain dependence structures that may be available to the decision maker in a practical decision situation before any probability assessments are made. For example, suppose the executive believes that

$$\begin{aligned}&\Pr(\text{Competitor Reaction}=\text{Price War}\\&\qquad|\text{Technical Success}=\text{Full Upgrade})\\&\quad>\Pr(\text{Competitor Reaction}=\text{Price War}\\&\qquad|\text{Technical Success}=\text{Partial Upgrade})\\&\quad>\Pr(\text{Competitor Reaction}=\text{Price War}\\&\qquad|\text{Technical Success}=\text{No Upgrade}).\end{aligned}$$

His rationale is that the competitor will be more likely to indulge into a price war if he feels that the executive's product is a superior product and that it has had great technical success. This information can be incorporated into the simulation data set as follows.

First, we generate the marginal probabilities for Technical Success. This is achieved by generating two samples, x_1, x_2 from a uniform [0,1] distribution, sorting them in ascending order to obtain u_1, u_2, and taking the increments $1-u_2, u_2-u_1, u_1-0$. The increments form the three marginal probabilities for the variable Technical Success.

Next, we incorporate the dependence structure for Competitor Reaction given Technical Success. To do that, we generate three samples from a uniform [0,1] distribution and sort them, to get $p_1 > p_2 > p_3$. The sorted probabilities represent the probability of Competitor Reaction="Price War" given Technical Success="Full Upgrade," "Partial Upgrade," and "No Upgrade," respectively.

We can also incorporate other dependence structures such as the Competitor Reaction being Premium Price given Technical

TABLE IV
RANGE AND MEAN VALUES OF CORRELATION COEFFICIENTS

Pairwise Variables	Mean Value of Correlation	Min. Value	Max. Value
(Tech Success, Comp. Reac)	0.3907	0.000	0.944
(Tech Success,# units sold)	0.1609	-0.664	0.883
(Tech Success, Cost)	0.3926	0.000	0.960
(Comp. Reac, # units sold)	0.4184	-0.397	0.874
(Comp. Reac, Cost)	0.1605	0.000	0.795
(# units sold, Cost)	0.0667	-0.407	0.590

TABLE V
SIMULATION RESULTS FOR MAXIMUM ENTROPY APPROXIMATION

	Maximum Deviation		Total Variation		Relative Entropy	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Equal Probability	0.2188	0.1180	1.2597	0.1490	1.9908	0.6637
Marginal Probability	0.1149	0.0533	0.7573	0.1953	0.7627	0.3923
Pairwise Probability	0.0064	0.0042	0.0837	0.0485	0.0222	0.0214
Three-way Probability	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000

Success. For example, suppose that the decision maker also believes that

$$\begin{aligned}&\Pr(\text{Competitor Reaction} = \text{Premium Price}\\&\quad|\text{Technical Success} = \text{Full Upgrade})\\&< \Pr(\text{Competitor Reaction} = \text{Premium Price}\\&\quad|\text{Technical Success} = \text{Partial Upgrade})\\&< \Pr(\text{Competitor Reaction} = \text{Premium Price}\\&\quad|\text{Technical Success} = \text{No Upgrade}).\end{aligned}$$

His rationale is that the competitor would be more likely to sell his product at a premium price if the executive's product did not include the upgrades. To incorporate this dependence structure, we now generate three samples, x_1, x_2, and x_3, from the uniform distributions $Uniform[0, 1-p_1]$, $Uniform[0, 1-p_2]$, and $Uniform[0, 1-p_3]$ and sort them to get $q_1 > q_2 > q_3$. The sorted probabilities represent the probability of Competitor Reaction="Premium Price" given Technical Success="No Upgrade," "Partial Upgrade," and "Full Upgrade," respectively. The conditional probabilities for Competitor Reaction="As Is" is determined by requiring the conditional probabilities to sum to one.

In a very similar manner, we can incorporate dependence between Technical Success and Cost using the knowledge that

$$\begin{aligned}&\Pr(\text{Cost} = \text{High}|\text{Technical Success} = \text{Full Upgrade})\\&> \Pr(\text{Cost} = \text{High}\\&\quad|\text{Technical Success} = \text{Partial Upgrade})\\&> \Pr(\text{Cost} = \text{High}|\text{Technical Success} = \text{No Upgrade})\end{aligned}$$

and that

$$\begin{aligned}&\Pr(\text{Cost} = \text{Low}|\text{Technical Success} = \text{Full Upgrade})\\&< \Pr(\text{Cost} = \text{Low}\\&\quad|\text{Technical Success} = \text{Partial Upgrade})\\&< \Pr(\text{Cost} = \text{Low}|\text{Technical Success} = \text{No Upgrade}).\end{aligned}$$

Furthermore, for a given level of Technical Success, we may have some beliefs that $\Pr(\#\text{units sold} = \text{High})$ is larger when the competitor's price is higher. For example

$$\begin{aligned}&\Pr(\text{Number of Units Sold} = \text{High}\\&\quad|\text{Technical Success} = \text{Full Upgrade},\\&\quad\text{Competitor Reaction} = \text{Premium})\\&> \Pr(\text{Number of Units Sold} = \text{High}\\&\quad|\text{Technical Success} = \text{Full Upgrade},\\&\quad\text{Competitor Reaction} = \text{As Is})\\&> \Pr(\text{Number of Units Sold} = \text{High}\\&\quad|\text{Technical Success} = \text{Full Upgrade},\\&\quad\text{Competitor Reaction} = \text{Price War}).\end{aligned}$$

This additional information can be incorporated using a procedure similar to the ones described above; we generate conditional probabilities and sort them in correspondence with the given information. Furthermore, to incorporate the three-way dependence between Technical Success, Competitor Reaction, and Number of Units Sold, we assumed that for a given level of Competitor Reaction that $\Pr(\text{Number of Units Sold} = \text{High})$ is larger when Technical Success is improved.

Recall also that, in this example, Cost was conditionally independent of both Competitor Reaction and Number of Units Sold given Technical Success. We incorporated this information into the simulation data set by generating a conditional probability distribution of Cost given Technical Success separately, and then multiplying it by the joint distribution of the remaining three variables. Note that this is performed prior to any numerical probability assessments. The mean and range of correlation coefficients generated assuming values of 0, 1, and 2 for the outcomes is shown in Table IV.

From Table IV, we see that incorporating dependence as described above has resulted in more representative values of correlation coefficients for this particular decision problem than the case of uniform sampling. For example, note that the range of correlation coefficients generated for Technical Success and Cost is nonnegative, which makes intuitive sense for this problem.

The simulation results for the maximum deviation, total variation, and relative entropy are shown in Table V.

The first observation we note from Table V is that the mean value for the maximum deviation of the joint distribution constructed using the marginal distribution is much smaller (approximately half) than that using the uniform probability assumption. This can be compared with the case of uniform sampling of Table III. Thus, the maximum entropy joint distribution constructed for the executive's problem improves significantly by including the marginal probabilities rather than simply assigning equal probabilities.

The second observation we note from Table V is that the maximum deviation using pairwise assessments 0.0064 is significantly less than that using only the marginal assessments 0.1149 (approximately 5% of the deviation with marginal assessments). These numbers illustrate the importance of incorporating dependence between the variables in the executive's problem, and also show how characterizing the properties of the existing set of test distributions for the given problem can provide a more representative assessment of the order of the approximation provided than merely uniformly sampling the space of possible test distributions. The value of the maximum deviation obtained for the pairwise assessments of Table V, suggests that the error in expected value calculations for the dollar amounts of each alternative will be small if the decision maker provides pairwise assessments in a decision situation which has the same dependence structure as the one considered in the simulation study.

Note that the maximum deviation, total variation, and relative entropy for the three-way assessments are all equal to zero in Table V. This should not come as a surprise since the data set did in fact incorporate the conditional independence of Cost on both Competitor Reaction and Number of Units Sold given Technical Success. Thus, the three-way assessments actually contain all the information needed to determine the joint distribution for this problem. The zero deviation results obtained for the three-way assessments confirm that the maximum entropy distribution yields the exact joint distribution if all the information about the joint distribution is incorporated into the constraints.

Figs. 2 and 3 show the cross-plots obtained from the simulation results. The clustering of points above the line indicates the improvement provided by a higher order assessment over a lower one.

C. Estimating the Error in Dollar Values

We now show how to place a dollar value on the accuracy of the maximum entropy approximations for a given decision alternative prior to making any probability assessments. Using the same data set that was generated for the Executive's Problem, we recorded the mean and standard deviation of the expected value of profit using the dollar values for "Price=Premium Price" alternative, for each of the maximum entropy approximations, for each of the generated distributions. The simulation results are shown in Table VI.

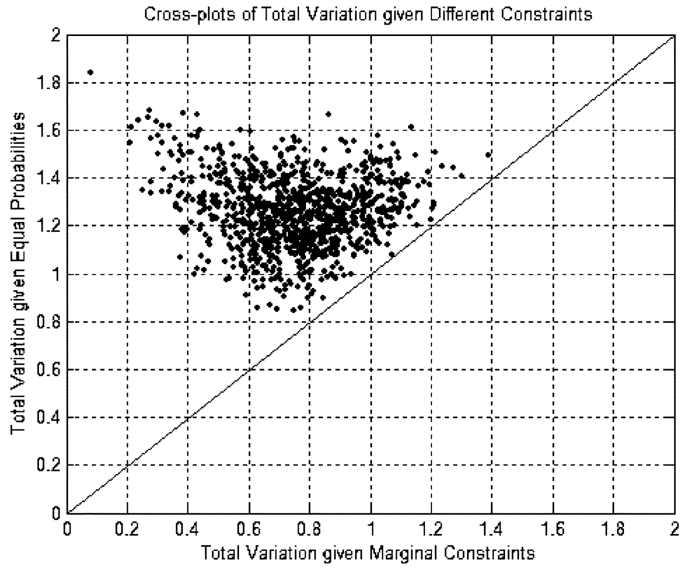

Fig. 2. Cross-plot of total variation between the equal probabilities approximation and the marginal constraints approximation.

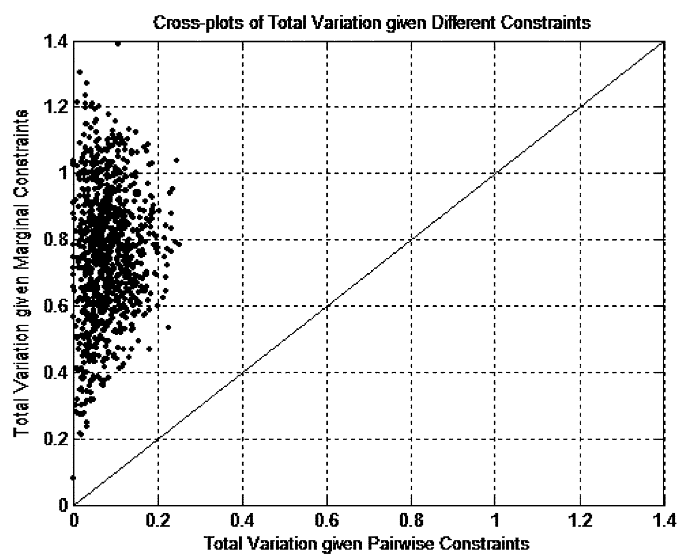

Fig. 3. Cross-plot of total variation between the marginal constraints approximation and the pairwise approximation.

TABLE VI
MEAN AND STANDARD DEVIATION FOR EXPECTED VALUE OF PROFIT FOR "PRICE=PREMIUM PRICE" USING MAXIMUM ENTROPY APPROXIMATIONS

	Equal Probability	Marginal Probability	Pairwise	Three-way
Mean	$41.6 M	$44.37 M	$43.38 M	$ 43.38 M
Std. Dev	$0	$4.09 M	$ 4.3 M	$4.2 M

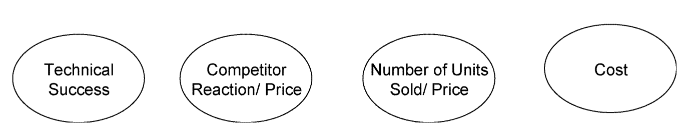

Fig. 4. Node ordering supplied by the decision-maker.

Note that the expected values of the "Price=Premium Price" alternative given the equal probability and marginal constraints assumptions are quite different than that obtained using the pairwise and three-way approximations. This can be explained by

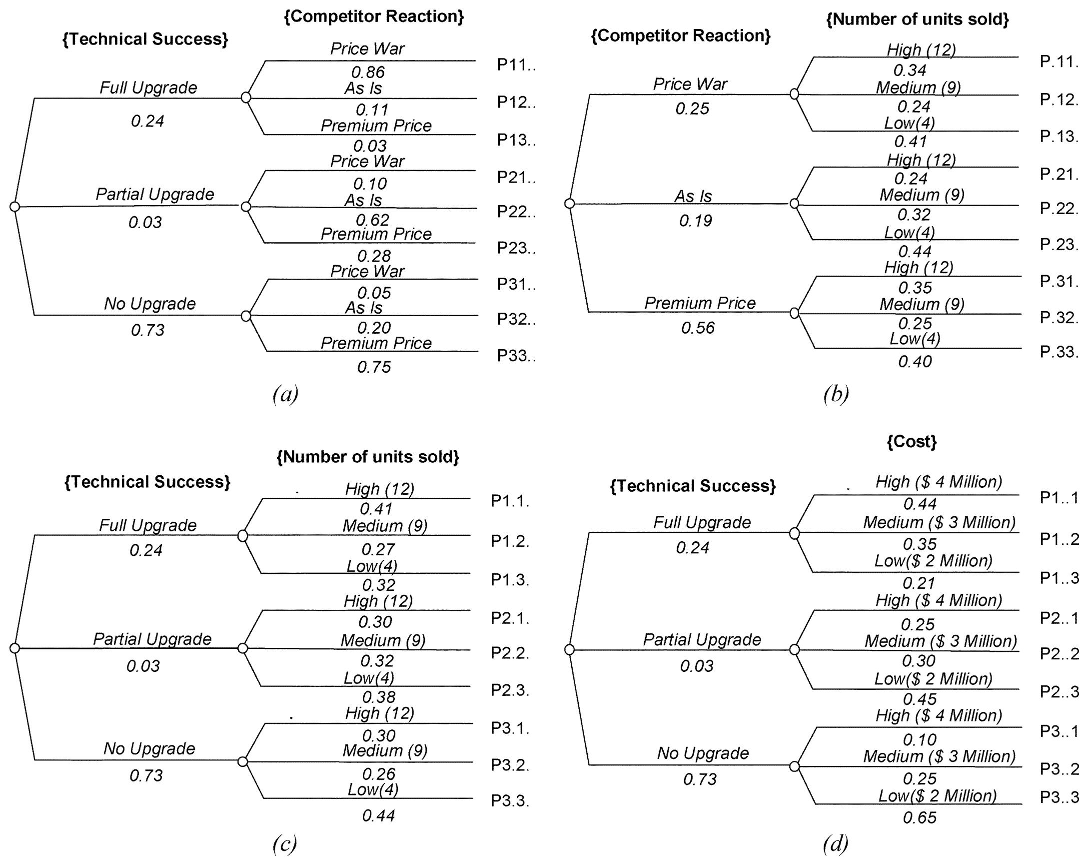

Fig. 5. Elicitation of pairwise joint distributions for the "Price=Premium Price" alternative in the executive's problem.

observing that the average maximum deviation for the pairwise approximation was very small (0.0064, with a standard deviation of 0.0042). With such small deviation values, the executive can determine whether it is worth proceeding from the pairwise approximation to the three-way assessments or to use only pairwise assessments and construct the joint distribution using the maximum entropy method. Table VI shows that, in our example, choosing the pairwise assessments would be satisfactory.

While Table VI corresponds to dollar values that were used in the actual analysis of this decision situation, we observe that the same methodology can be used to determine the errors provided by each approximation for any other decision situation, and before any probability assessments are made.

V. Back to the Executive's Problem

Now, we return to the executive's problem and show how he can use the previous analysis to think about constructing a joint distribution using lower order assessments. Recall that, in our example, the pricing decision influences the Competitor Reaction and the Number of Units Sold, so he will need to repeat the analysis to obtain a joint distribution for each pricing alternative. We will show the analysis for the "Price=Premium Price" alternative only but the same steps will apply for the "Price=Market Price" alternative.

Based on the previous discussion, the first step in the construction of the joint distribution is to determine the accuracy of each approximation provided. From Table V, we see that we can expect an average value for the Maximum deviation using pairwise assessments equal to 0.0064, and from Table VI, we see that the average value of the expected dollar values is very close to that using the three-way assessments for this decision situation. Using this information, the executive decides to use pairwise assessments for this problem.

As we have mentioned, the decision diagram of Fig. 1 implies that the Cost is conditionally independent of both the Number of Units Sold and the Competitor Reaction if we know the Technical Success. Therefore, he can assess Cost given Technical Success first and construct a joint distribution of the remaining three-variables (Technical Success, Competitor Reaction, and Number of Units Sold) separately.

Step 1) Ordering the nodes.

Now, the decision-maker orders the variables in the way he feels most comfortable reasoning with and assessing. This is an important step since any conditional probabilities elicited later will be conditioned in this order. The order provided by the decision maker in this example is shown in Fig. 4 from left to right.

Step 2) Eliciting the lower order joint distributions.

Now, the decision maker provides the pairwise assessments for his decision situation. Fig. 5(a)–(c) shows the assessments for the joint distribution of three variables (Technical Success, Number of Units Sold, and Competitor Reaction), and Fig. 5(d) shows the conditional assessment of Cost given Technical Success.

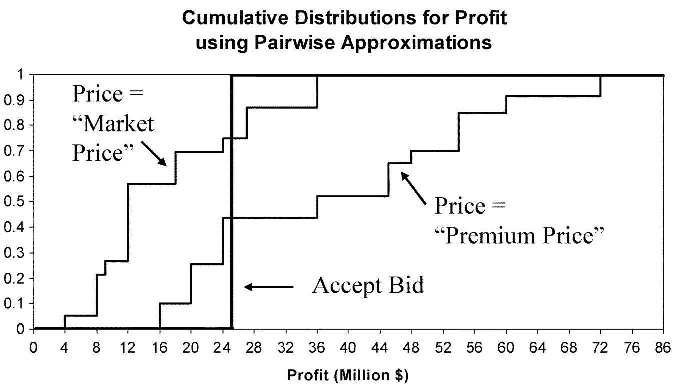

Fig. 6. Cumulative distributions for profit for the two alternatives and the bid.

TABLE VII
MEAN AND STANDARD DEVIATION FOR EXPECTED PROFIT FOR EACH DECISION ALTERNATIVE USING THE PAIRWISE ASSESSMENTS

	Price = "Premium Price"	Price = "Market Price"	"Accept Bid"
Mean	$38.37 M	$17.1 M	$25 M
Std. Dev	$18	$9.72 M	$ 0 M

Once again, the subscripts for the probabilities at the end of each tree in Fig. 5 refer to the variables in the order provided by the decision maker in Step 1. For example, the notation $P2\ldots3$ refers to the sum of all joint probabilities with the first variable equal to the second branch (Technical Success="Partial Upgrade") and the forth variable equal to the third branch (Cost="Low"). Note that the probability assessments have the same dependence structures that were used to generate the data set for the simulation study.

Step 3) Constructing the joint distribution.

Given the marginal and pairwise distributions, we can now solve formulation (8) to obtain a maximum entropy joint distribution, or any other minimum cross-entropy formulation.

Fig. 6 shows the maximum entropy cumulative distribution for profit for each of the decision alternatives using the pairwise approximation. The curves show stochastic dominance for the Price="Premium Price" alternative over Price="Market Price" alternative.

Table VII compares the mean and standard deviations for profit calculated for the different alternatives using the pairwise assessments. They can be compared with the $25 million bid submitted by the competitor.

From the previous analysis, the best alternative was to continue funding the project and to choose the "Price=Premium Price" alternative as the selling price to the customers.

A. Comparison With the Test Distribution

Fig. 7 shows a comparison between the $3 \times 3 \times 3 \times 3$ variable joint distribution that was used to construct the pairwise assessments of Fig. 5, and the maximum entropy joint distribution using pairwise assessments. The numbers on the horizontal axis represent the 81 discrete probabilities, where the number 1 represents the probability p_{1111}, and the number 81 represents the probability p_{3333}. The total variation between the two distributions is 0.02, which is a fairly small difference, as we discussed earlier. The relative entropy is 0.0024, and the maximum deviation is 0.0016.

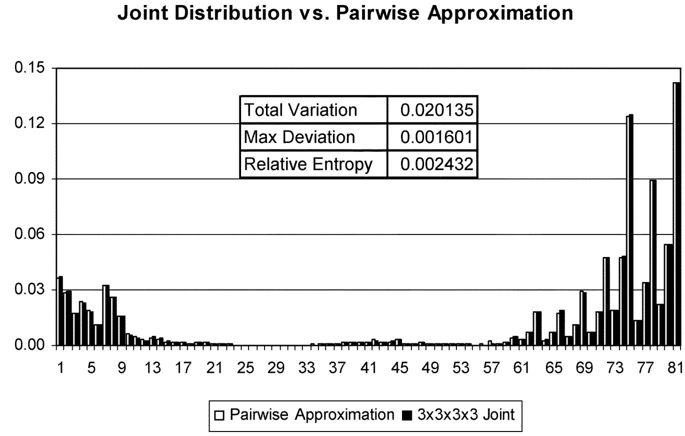

Fig. 7. Comparison of the joint distribution with its pairwise approximation.

TABLE VIII
NUMERICAL VALUES FOR THE JOINT DISTRIBUTION USED IN THE EXAMPLE AND THE MAXIMUM ENTROPY PAIRWISE APPROXIMATION

	FullJoint	Pairwise		FullJoint	Pairwise		FullJoint	Pairwise
P_{1111}	0.0369	0.0360	P_{2111}	0.0002	0.0001	P_{3111}	0.0001	0.0003
P_{1112}	0.0294	0.0287	P_{2112}	0.0002	0.0001	P_{3112}	0.0002	0.0008
P_{1113}	0.0176	0.0172	P_{2113}	0.0004	0.0002	P_{3113}	0.0005	0.0021
P_{1121}	0.0231	0.0239	P_{2121}	0.0001	0.0001	P_{3121}	0.0007	0.0005
P_{1122}	0.0184	0.0190	P_{2122}	0.0002	0.0002	P_{3122}	0.0018	0.0014
P_{1123}	0.0110	0.0114	P_{2123}	0.0002	0.0003	P_{3123}	0.0047	0.0036
P_{1131}	0.0323	0.0324	P_{2131}	0.0003	0.0005	P_{3131}	0.0028	0.0028
P_{1132}	0.0257	0.0258	P_{2132}	0.0004	0.0005	P_{3132}	0.0071	0.0069
P_{1133}	0.0154	0.0155	P_{2133}	0.0006	0.0008	P_{3133}	0.0185	0.0181
P_{1211}	0.0053	0.0062	P_{2211}	0.0011	0.0011	P_{3211}	0.0029	0.0027
P_{1212}	0.0043	0.0049	P_{2212}	0.0013	0.0013	P_{3212}	0.0073	0.0068
P_{1213}	0.0026	0.0030	P_{2213}	0.0019	0.0020	P_{3213}	0.0190	0.0177
P_{1221}	0.0048	0.0040	P_{2221}	0.0015	0.0016	P_{3221}	0.0044	0.0045
P_{1222}	0.0038	0.0032	P_{2222}	0.0018	0.0019	P_{3222}	0.0109	0.0113
P_{1223}	0.0023	0.0019	P_{2223}	0.0027	0.0028	P_{3223}	0.0284	0.0295
P_{1231}	0.0018	0.0017	P_{2231}	0.0017	0.0016	P_{3231}	0.0073	0.0073
P_{1232}	0.0014	0.0014	P_{2232}	0.0021	0.0020	P_{3232}	0.0182	0.0183
P_{1233}	0.0009	0.0008	P_{2233}	0.0031	0.0029	P_{3233}	0.0474	0.0477
P_{1311}	0.0019	0.0020	P_{2311}	0.0008	0.0009	P_{3311}	0.0191	0.0191
P_{1312}	0.0015	0.0016	P_{2312}	0.0009	0.0010	P_{3312}	0.0479	0.0477
P_{1313}	0.0009	0.0009	P_{2313}	0.0014	0.0016	P_{3313}	0.1244	0.1241
P_{1321}	0.0005	0.0005	P_{2321}	0.0006	0.0005	P_{3321}	0.0137	0.0137
P_{1322}	0.0004	0.0004	P_{2322}	0.0007	0.0006	P_{3322}	0.0342	0.0342
P_{1323}	0.0003	0.0003	P_{2323}	0.0010	0.0009	P_{3323}	0.0889	0.0890
P_{1331}	0.0003	0.0002	P_{2331}	0.0006	0.0005	P_{3331}	0.0219	0.0219
P_{1332}	0.0002	0.0002	P_{2332}	0.0007	0.0007	P_{3332}	0.0547	0.0548
P_{1333}	0.0001	0.0001	P_{2333}	0.0010	0.0010	P_{3333}	0.1422	0.1424

Table VIII shows the numerical values for the 81 discrete probabilities and the maximum entropy solution using pairwise constraints.

The previous example illustrates a decision situation where the pairwise approximation was sufficient to approximate the joint distribution. In the next section, we discuss the accuracy of the lower order approximations when some of the outcomes of the uncertainties are revealed.

TABLE IX
SIMULATION STATISTICS FOR "NUMBER OF UNITS SOLD" AFTER BAYESIAN INFERENCE

	Maximum Deviation		Total Variation		Relative Entropy	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Equal Probability	0.5276	0.1027	1.0551	0.2053	1.3638	0.7532
Marginal Probability	0.2844	0.1382	0.5687	0.2764	0.6177	0.5716
Pairwise Probability	0.0327	0.0423	0.0654	0.0847	0.0181	0.0497
Three-way Probability	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

VI. BAYESIAN INFERENCE WITH THE MAXIMUM ENTROPY DISTRIBUTION

Nine months later, the executive heard from the engineering department that they had a "Full Upgrade" in Technical Success. He also heard from the marketing department that their competitor would start a "Price War" because their product was still inferior. Furthermore, the weakening of the dollar against foreign currencies incurred "High Costs" in manufacturing. In a final attempt to gain market share, and with concerns about the growing technical success of the product, the competitor offered a new bid of $38 million.

After receiving this new information, the only remaining uncertainty is Number of Units Sold. From Fig. 5, we can determine the marginal distribution of Number of Units Sold prior to receiving this information as $\Pr(\text{Number of Units Sold} = \text{High}) = .33$, $\Pr(\text{Number of Units Sold} = \text{Medium}) = .26$, $\Pr(\text{Number of Units Sold} = \text{Low}) = .41$. As we shall see, however, using this prior distribution may lead to errors in decision making. The executive is now interested in the posterior distribution for the Number of Units Sold given this new information, and wonders about the order of the approximation provided by the maximum entropy distribution when used in Bayesian inference. To answer this question, we conducted a simulation study to determine the order of the approximations provided. We generated joint distributions using the simulation data set of the Executive's problem and, for each distribution, determined the conditional probabilities of $\text{Number of Units Sold} = x_i$ given Technical Success=Full upgrade, Competitor Reaction=Price War, and Cost=High for all values of $x_i = \text{High}$, Medium, and Low.

We compared these values to the values obtained using each of the maximum entropy approximations. Table IX shows the total variation, maximum deviation, and relative entropy for the updated distribution of Number of Units Sold obtained using the test distributions and the pairwise maximum entropy approximations.

From Table IX, we note that the maximum deviation using the marginal distributions is on the order of 0.28. This implies that the joint distributions constructed using the independence assumption may lead to errors in the probabilities that are significant when additional information is revealed. Furthermore, we note that there is significant reduction in maximum deviation when we incorporate the pairwise constraints. The mean value of the maximum deviation using pairwise assessments is 0.0327 (about 90% reduction from the maximum deviation of 0.287 obtained using only the marginal assessments). Table VIII shows, again, that the three-way distributions are sufficient to determine the full joint distribution, so the maximum deviation, total variation, and relative entropy are all equal to zero.

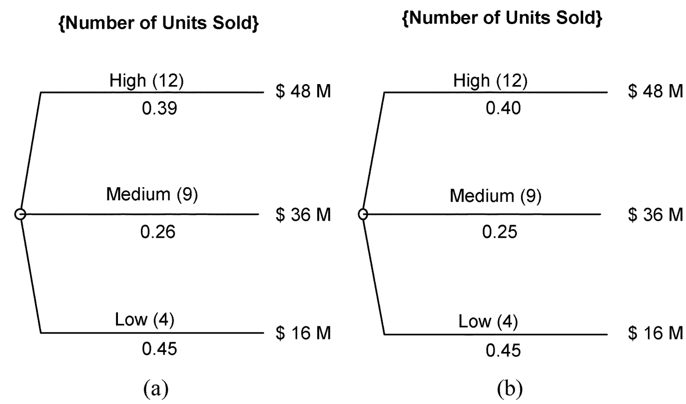

Fig. 8. Posterior distributions using (a) pairwise approximation and (b) actual distribution, after outcomes of technical success, competitor reaction, and cost have been revealed.

Fig. 8(a) shows the updated distribution for Number of Units Sold using pairwise assessments obtained after revelation of the new information. The expected value of the "Price=Premium Price" alternative is now 35.3 M. Fig. 8(b) shows the actual distribution for Number of Units Sold for the $3\times3\times3\times3$ joint distribution after conditioning on the new information. The latter distribution gives an expected value of $35.4 M. The two distributions have a maximum deviation of 0.0096, a total variation of 0.0192, and relative entropy of 0.0003. These results are within the ranges of the results of Table IX, by which the executive decided to perform the updating using the maximum entropy pairwise approximation. The optimal decision was to accept the new bid and sell the line of testing equipment to the competitor.

We observe here that using the prior distribution for Number of Units Sold to assess the new situation would have yielded an expected value of $31.8 M, which would lead the semiconductor company to undervalue the deal by a significant amount. For example, if the competitor had offered them a bid of $33 M, they would have accepted it when, in fact, their expected value for the deal is now $35.3 M. This situation illustrates that when we do have knowledge of a certain dependence structure between the variables, using only marginal assessments may produce significant errors in the decision making process, as they would not update with new information.

Using pairwise assessments helps incorporate some form of dependence between the variables. In this example, pairwise assessments led to a maximum entropy joint distribution that provided satisfactory results when used in the analysis of the problem and in Bayesian inference. In general, however, this may not be the case, and the performance of the pairwise assessments may not be as satisfactory when the number of variables increases. In these situations, which can be determined by simulation, it would be prudent to reassess the probability distribution for a particular variable in light of the new information. Note, however, that this new assessment is easier, since it involves fewer probability assessments, as the outcomes of several variables are revealed.

VII. Conclusion

In many situations that arise in practice, decision makers often ignore the dependence relations among the variables of the decision situation in their analysis. As we have seen, this independence assumption leads to errors that can lead to suboptimal decision alternatives. On the other hand, the difficulty of the joint probability assessment may be a large barrier to incorporating dependence in the analysis because of the need to condition the assessments on several variables.

The primary purpose of this paper was to illustrate the use of the maximum entropy method for constructing joint distributions in decision analysis practice, and to illustrate the order of the approximations that may result when incorporating lower order dependence relations before any probability assessments are made. We have seen how simulations can provide insights into the accuracy of the approximations provided, in terms of maximum deviations and errors in dollar values. Furthermore, we have shown how to incorporate qualitative knowledge about the nature of the dependencies into the simulation data set.

In the executive's problem, the error we encountered using the pairwise assessments was sufficiently small to justify their use in the analysis of the problem. In other situations, this may not be the case, particularly when the number of variables increases. However, the methodology that we have used in this example can be applied in many other decision situations to determine the order of the approximations provided by the lower order marginals before any probability assessments are made. Using simulation analysis, we can then determine the tradeoff between the number of assessments needed and the accuracy of the results.

The second purpose of this paper was to familiarize the readers with the maximum entropy approach and its ability to incorporate many types of partial information. Maximum entropy applications have found great popularity in source separation, natural language processing, space communications, and many other fields. The applications in this paper emphasized the use of the maximum entropy formulation in solving discrete decision analysis problems using lower order joint probability assessments. Other applications include aggregating expert forecasts using lower order dependence relations between the experts, and maximum entropy and minimum cross entropy formulations for continuous variables.

The convenience of the approximations provided by the maximum entropy and minimum cross-entropy formulations, and the generality of their applications provide an incentive for their use in constructing joint probability distributions in decision analysis practice. The calculations presented in this paper used numerical software packages that were run on a personal computer, but many maximum entropy formulations can even be derived graphically or using spreadsheets with simple sensitivity analysis. See, for example, [29] that also includes formulations for maximum entropy distributions with upper and lower bounds on the probabilities. We also look forward to seeing enhancements in existing decision analysis software packages that construct maximum entropy joint distributions using lower order assessments, or other types of information that is available. These enhancements will provide the convenience for the decision analyst to work on one common platform throughout the analysis, and to incorporate many types of partial information about the variables of the decision situation using the maximum entropy joint distribution.

Appendix I
Review of Maximum Entropy Solution

Consider the following maximum entropy formulation for a distribution with N outcomes and the following constraints on its probabilities:

$$p^*(x_i) = \underset{p(x_i)}{\operatorname{argmax}} - \sum_{i=1}^{n} p(x_i) \log(p(x_i))$$

such that

$$\sum_i h_j(x_i)p(x_i) = \mu_j,\ j = 1, \ldots, m$$
$$\sum_i p(x_i) = 1,\ p(x_i) \geq 0. \tag{12}$$

These constraints can have several forms. For example, if no further information is available except the probability distribution is nonnegative and sums to one, then $h_j(x_i) = 0$, $j = 1, \ldots, m$. On the other hand, if we know the first moment of the distribution, then $h_1(x_i) = x_i$ and $h_j(x_i) = 0, j = 2, \ldots, m$ . . ., etc.

Using the Method of Lagrange multipliers, we take the Lagrange operator for (12) as

$$L = -\sum_{i=1}^{n} p(x_i) \log(p(x_i)) - \sum_{j=1}^{m} \lambda_j \left(\sum_i h_j(x_i)p(x_i) - \mu_j \right) - \lambda_0 \left(\sum_{i=1}^{n} p(x_i) - 1 \right). \tag{13}$$

Now, we take the first partial derivative of (13) with respect to $p(x_i)$, and equate it to zero to get

$$-\log(p(x_i)) - 1 - \sum_{j=1}^{m} \lambda_j h_j(x_i) - \lambda_0 = 0. \tag{14}$$

Rearranging (14) gives

$$p^*(x_i) = e^{-1-\lambda_0-\lambda_1 h_1(x_i)-\lambda_2 h_2(x_i)-\ldots-\lambda_m h_m(x_i)}. \tag{15}$$

When $h_j(x_i) = 0$, $j = 1, \ldots, m$, the maximum entropy probability density is uniform. If only the first moment is available, the maximum entropy solution is an exponential distribution on the positive domain. If only the first and second moments are available, the maximum entropy solution is a Gaussian distribution on the real domain [30].

TABLE X
GIVEN $2 \times 2 \times 2$ JOINT DISTRIBUTION

P_{111}	P_{112}	P_{121}	P_{122}	P_{211}	P_{212}	P_{221}	P_{222}
0.0153	0.1835	0.0734	0.1729	0.0209	0.2808	0.185	0.0682

TABLE XI
INDEPENDENT PAIRWISE ASSESSMENTS FOR GIVEN $2 \times 2 \times 2$ DISTRIBUTION

$P_{.12}$	$P_{.22}$	$P_{2.2}$	$P_{11.}$	$P_{12.}$	$P_{21.}$	$P_{22.}$
0.46433	0.24109	0.34898	0.19881	0.24628	0.30169	0.25321

TABLE XII
MAXIMUM DEVIATION, TOTAL VARIATION, AND RELATIVE ENTROPY

Maximum Deviation	Total Variation	Relative Entropy
0.010	0.081	0.005

APPENDIX II
MATLAB CODE FOR MAXIMUM ENTROPY SOLUTION

Suppose we wish to find the maximum entropy distribution for the following $2 \times 2 \times 2$ joint distribution using its pairwise constraints (Table X).

Note that while there are a total of 12 pairwise assessments that can be made, there are only seven independent pairwise assessments. Thus, even the full pairwise assessments are redundant. The independent pairwise assessments are shown in Table XI.

We can solve this problem in Matlab using the function "fmincon," which minimizes a convex function (maximizes a concave function) subject to linear equality or inequality constraints. The pairwise constraints can be expressed as $A\mathrm{x} = b$, where A is a matrix that selects the pairwise probabilities from the maximum entropy distribution, b is a vector that enters the pairwise probability assessments (of Table XII) into Matlab, "fmincon" is the function we use to solve the problem, $P0$ is the initial guess of the maximum entropy solution, and lb and ub are lower and upper bounds on the probability values.

Matlab code for the maximum entropy $2 \times 2 \times 2$ joint distribution given pairwise assessments:

$$
\begin{aligned}
&> lb = zeros(8,1) + .000\,001;\\
&> ub = ones(8,1) - .000\,001;\\
&> P0 = \frac{1}{8} * ones(8,1);\\
&> A = [\begin{matrix}0 & 1 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1\\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1\end{matrix}];\\
&> b = [0.4643;\quad 0.2411;\quad 0.3490;\quad 0.1988;\\
&\qquad 0.2463;\quad 0.3017;\quad 0.2532];\\
&> P = \mathrm{fmincon}\left(inline\left('sum\left(x.*\log(x)\right)'\right),\right.\\
&\qquad \left. P0, [\,], [\,], A, b, lb, ub, [\,]\right)
\end{aligned}
$$

Solution

$$
\begin{aligned}
P = [&0.0051 \quad 0.1937 \quad 0.0835 \quad 0.1627\\
&0.0311 \quad 0.2706 \quad 0.1749 \quad 0.0783].
\end{aligned}
$$

$$
\begin{aligned}
&> lb = zeros(8,1) + .000\,001;\\
&> ub = ones(8,1) - .000\,001;\\
&> P0 = \frac{1}{8} * ones(8,1);\\
&> A = [\begin{matrix}0 & 1 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1\\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1\end{matrix}];\\
&> b = [0.4643;\quad 0.2411;\quad 0.3490;\quad 0.1988;\\
&\qquad 0.2463;\quad 0.3017;\quad 0.2532];\\
&> P = \mathrm{fmincon}\left(inline\left('sum\left(x.*\log(x)\right)'\right),\right.\\
&\qquad \left. P0, [\,], [\,], A, b, lb, ub, [\,]\right)
\end{aligned}
$$

$Solution:$

$$
\begin{aligned}
P = [&0.0051 \quad 0.1937 \quad 0.0835 \quad 0.1627\\
&0.0311 \quad 0.2706 \quad 0.1749 \quad 0.0783]
\end{aligned}
$$

Table XII shows the maximum deviation, total variation, and relative entropy of the $2 \times 2 \times 2$ joint distribution and the maximum entropy pairwise approximation.

ACKNOWLEDGMENT

The author thanks S. Taing for her help with some of the earlier simulation work and for edits of several figures in the paper. The author also thanks the editor and two anonymous reviewers for their many helpful suggestions on the exposition and clarity of this work.

REFERENCES

[1] D. Kahneman and A. Tversky, "Subjective probability: A judgment of representativeness," *Cogn. Psychol.*, vol. 3, pp. 430–454, 1972.

[2] A. Tversky and D. Kahneman, "Judgment under uncertainty: Heuristics and biases," *Science*, vol. 185, pp. 1124–1131, 1974.

[3] C. S. Spetzler and A. S. von Holstein, "Probability encoding in decision analysis," *Manage. Sci.*, vol. 22, pp. 340–358, 1975.

[4] M. W. Merkhofer, "Quantifying judgmental uncertainty: Methodology, experiences, and insights," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-17, pp. 741–752, 1987.

[5] W. Yi and V. M. Bier, "An application of copulas to accident precursor analysis," *Manage. Sci.*, vol. 44, pp. S257–S270, 1998.

[6] R. T. Clemen and T. Reilly, "Correlations and copulas for decision and risk analysis," *Manage. Sci.*, vol. 45, no. 2, pp. 208–224, 1999.

[7] R. T. Clemen, G. W. Fischer, and R. L. Winkler, "Assessing dependence: Some experimental results," *Manage. Sci.*, vol. 46, no. 8, pp. 1100–1115, 2000.

[8] D. Keefer, "The underlying event model for approximating probabilistic dependence among binary events," *IEEE Trans. Eng. Manage.*, vol. 51, no. 2, pp. 173–182, May 2004.

[9] C. K. Chow and C. N. Liu, "Approximating discrete probability distributions with dependence trees," *IEEE Trans. Inf. Theory*, vol. 14, no. 3, pp. 462–467, May 1968.

[10] R. M. Cooke, "Markov and entropy properties of tree and vine-dependent variables," in *Proc. ASA Section on Bayesian Statistical Science*, 1997, pp. 166–175. Amer. Statist. Assoc., Alexandria, VA.
[11] G. R. MacKenzie, "Approximately maximum-entropy multivariate distributions with specified marginals and pairwise correlations," Ph.D. dissertation, Univ. Oregon, Eugene, OR, 1994.
[12] E. T. Jaynes, "Prior probabilities," *IEEE Trans. Syst., Sci. Cybern.*, vol. SSC-4, no. 3, pp. 227–240, 1968.
[13] A. E. Abbas, "Entropy Methods in Decision Analysis," Ph.D. dissertation, Stanford Univ., Stanford, CA, 2003.
[14] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379 and 623–656–423, 1948.
[15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[16] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 2, pp. 79–86, 1951.
[17] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, no. 4, pp. 620–630, 1957.
[18] D. Freund and U. Saxena, "An algorithm for a class of discrete maximum entropy problems," *Oper. Res.*, vol. 32, no. 1, 1984.
[19] J. Smith, "Moment methods in decision analysis," *Manage. Sci.*, vol. 39, no. 3, 1993.
[20] J. N. Kapur, *Measures of Information and Their Application*. New York: Wiley, 1994.
[21] H. H. Ku and S. Kullback, "Approximating discrete probability distributions," *IEEE Trans. Inf. Theory*, vol. IT–15, no. 4, pp. 444–447, Jul. 1969.
[22] R. T. Rockafeller, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.
[23] L. Devroye and G. Lugosi, *Combinatorial Methods in Density Estimation*. New York: Springer-Verlag, 2000.
[24] S. Kullback, "A lower bound for discrimination information in terms of variation," *IEEE Trans. Inf. Theory*, vol. IT–13, pp. 126–127, Jan. 1967.
[25] D. V. Gokhale and S. Kullback, *The Information in Contingency Tables*. New York: Marcel Dekker, 1978.
[26] D. V. Gokhale, "On entropy-based goodness-of-fit tests," *Comput. Statist. Data Anal.*, vol. 1, pp. 157–165, 1983.
[27] E. S. Soofi and J. J. Retzner, "Information indices: Unification and application," *J. Econ.*, vol. 107, pp. 17–40, 2002.
[28] H. A. David, *Order Statistics*. New York: Wiley, 1981.
[29] A. E. Abbas, "Maximum entropy distributions between upper and lower bounds," in *Proc. AIP Conf. 803*, K. Knuth, A. Abbas, R. Morris, and P. Castle, Eds., San Jose, CA, Aug. 2005, pp. 25–42.
[30] M. Tribus, *Rational Descriptions, Decisions and Designs*. Elmsford, NY: Pergamon, 1969.

Ali E. Abbas (M'95) received the M.S. degree in electrical engineering in 1998, the M.S. degree in engineering economic systems and operations research in 2001, the Ph.D. degree in management science in 2003, and the Ph.D. degree (minor) in electrical engineering from Stanford University, Stanford, CA.

He is currently an Assistant Professor at the School of Engineering, University of Illinois at Urbana–Champaign, Urbana. He previously worked as a Lecturer in the Department of Management Science and Engineering, Stanford University. He has also worked in Schlumberger Oilfield Services from 1991 to 1997, where he held several positions in wireline logging, operations management, and international training. He has also worked on several consulting projects for mergers and acquisitions in California, and co-taught several executive seminars on decision analysis at Strategic Decisions Group, Menlo Park. His current research interests include decision making with incomplete information and preferences, dynamic programming, information theory, and applications of operations research tools in bioinformatics.

Dr. Abbas is also a member of the Institute for Operations Research and the Management Sciences (INFORMS).