DECISION ANALYSIS
Vol. 10, No. 1, March 2013, pp. 26–41
ISSN 1545-8490 (print) | ISSN 1545-8504 (online)

http://dx.doi.org/10.1287/deca.1120.0261

Approximating Joint Probability Distributions Given Partial Information

Luis V. Montiel, J. Eric Bickel
Graduate Program in Operations Research and Industrial Engineering, University of Texas at Austin, Austin, Texas 78712 {lvmontiel@utexas.edu, ebickel@mail.utexas.edu}

In this paper, we propose new methods to approximate probability distributions that are incompletely specified. We compare these methods to the use of maximum entropy and quantify the accuracy of all methods within the context of an illustrative example. We show that within the context of our example, the methods we propose are more accurate than existing methods.

Key words: maximum entropy; incomplete information; probabilistic dependence; analytic center; stochastic optimization; dynamic programming; sequential exploration; practice
History: Received on September 3, 2012. Accepted by former Editor-in-Chief L. Robin Keller on November 19, 2012, after 1 revision. Published online in *Articles in Advance* January 18, 2013.

1. Introduction

Decision analysts frequently encounter problems in which complete probabilistic information is unavailable. For example, Sarin (1978, 1979), Lowell (1994), Abbas (2006), and Bickel and Smith (2006) all described decisions where the joint probability distribution was underspecified. In these cases, we lack a unique probability distribution with which the decision can be analyzed.

Formally, suppose the value of an alternative depends upon the discrete probability distribution $\boldsymbol{\pi}$, which is a member of the set $\mathbb{T} = \{\boldsymbol{\pi}\colon \mathbf{A}\boldsymbol{\pi} = \mathbf{b}, \boldsymbol{\pi} \geq 0\}$. The linear system $\mathbf{A}\boldsymbol{\pi} = \mathbf{b}$ encodes any information that we do have, such as marginal and conditional assessments. We assume that the assessed information is consistent, such that $\mathbb{T}$ is not empty. We refer to $\mathbb{T}$ as the *truth set*, because any member of $\mathbb{T}$ is a probability distribution consistent with the assessed information and, as such, could be the "true" distribution.

This paper addresses how to evaluate a decision and formulate a recommendation when $\mathbb{T}$ is also not a singleton. In this case, there is an infinite (uncountable) number of possible distributions matching the assessed information. Although, we are primarily motivated by underspecification of probabilistic dependence, the methods and approximations we discuss could be applied in the case of incomplete marginal information, missing conditional probabilities, etc.

Decision analysts have long addressed the problem of underspecification by developing approximation methods that specify a unique probability distribution given partial information. For example, Jaynes (1957) developed the principle of maximum entropy (ME) and specifically recommended its use in decision analyses (Jaynes 1968). Under ME, one uses the distribution within $\mathbb{T}$ that is as close as possible to uniform, subject to honoring any information that one does have. Several authors have used ME to specify a joint distribution using marginal and conditional assessments (e.g., see Ireland and Kullback 1968, Lowell 1994, MacKenzie 1994, Abbas 2006, Bickel and Smith 2006). An equivalent procedure is to choose the distribution that minimizes the Kullback–Leibler (KL) divergence to a reference distribution that assumes independence (Kullback and Leibler 1951).

More recently, Keefer (2004) introduced the "underlying event" (UE) model, where it is assumed that all random variables are conditionally independent given another "hidden" or "underlying" random variable. In this case, the analyst simply needs to assess one conditional probability involving the underlying event, which determines the entire joint probability mass function (pmf).

Both Abbas (2006) and Bickel and Smith (2006) quantified the accuracy of ME and other methods such as UE within the confines of their illustrative

examples. ME was found to exceed the accuracy of UE and assumptions of independence.

Copulas (Sklar 1959) have also been used to specify a single joint distribution given lower-order component distributions such as marginals and pairwise correlation coefficients (Clemen and Reilly 1999). Montiel and Bickel (2012b) discussed copulas at length and contrasted their use with the simulation method presented in this paper. For this reason, we do not consider them further here.

Sarin (1978, 1979) proposed a sequential assessment procedure for binary random variables in which lower-order assessments were used to constrain the set of feasible higher-order distributions. For example, in a setting consisting of three binary random variables $(X1, X2, X3)$, Sarin (1979) derived bounds on the eight joint probabilities $p(X1, X2, X3)$ given a marginal assessment $p(X1)$ and the conditional distribution $p(X2 \mid X1)$. These bounds could then be provided to the expert, and may aid the assessment. Sarin's (1979) procedure does not specify a unique distribution. Rather, it is a structured process by which additional information is assessed and used to specify the full joint distribution. As such, it is fundamentally different from the approach taken in this paper, which addresses how to proceed given that further information is unavailable. Moskowitz and Sarin (1983) demonstrated that specification of marginal and conditional assessments in a binary setting significantly restricts the feasible region for joint probabilities. However, just how "tight" these bounds need to be depends upon the decision problem and the objective function. Montiel and Bickel (2012b) and Montiel (2012) showed that marginal and pairwise assessments might not significantly reduce the range in which a targeted level of performance may fall (e.g., the probability of achieving a positive net present value (NPV)).

Other approaches could be considered ad hoc. For example, Moore and Mudford (1999), Stabell (2000), and Delfiner (2003) all developed approaches where uncertainties are categorized as perfectly dependent or independent. In practice, it is common to simply ignore dependence because it complicates probability assessments (Korsan 1990, Bickel and Smith 2006).

The contribution of this paper is fourfold. First, we introduce two methods to approximate discrete joint probability distributions given partial information. As we show later, these methods might be better, in a sense to be made clear, than existing approximation methods such as ME. Second, we extend the accuracy studies conducted by Abbas (2006) and, in particular, Bickel and Smith (2006). Third, we illustrate the use of our approximations within the context of the illustrative example presented by Bickel and Smith (2006). Finally, we demonstrate the use of our simulation procedure, discussed herein and presented by Montiel and Bickel (2012a, b), to test the robustness of decision alternatives.

This paper is organized as follows: §2 summarizes our simulation procedure, which serves as a basis for our analysis and provides a benchmark for measuring the accuracy of the approximations; §3 summarizes existing methods to approximate probability distributions given partial information and introduces our two approximations; §4 introduces the sequential exploration problem that we use to motivate and explain our procedures; §5 quantifies and compares the accuracy of the approximation methods; §6 demonstrates how to use our simulation procedure to find robust decision alternatives; and §7 concludes the paper.

2. Generating a Collection of Feasible Joint Distributions

This section briefly explains the Monte Carlo simulation procedure that we use to sample from $\mathbb{T}$ and thereby generate millions of feasible pmfs. As demonstrated by Montiel and Bickel (2012a), this collection can be used to evaluate the decision under a set of distributions, all of which are consistent with the information we do have, while assuming nothing further. This paper uses this collection of distributions to quantify the accuracy of approximation methods that specify a single distribution within $\mathbb{T}$.

The joint distribution simulation approach (JDSIM) proposed by Montiel and Bickel (2012a, b) employs the hit-and-run sampler (HR) (Smith 1984) to uniformly sample $\mathbb{T}$. HR is the fastest known algorithm to sample the interior of an arbitrary polytope (Lovasz 1998). HR can be extended to sample nonuniformly (Bélisle et al. 1993). However, we are presently unsure how to specify a distribution over the set of all distributions in $\mathbb{T}$. Dyer et al. (1973) faced a similar problem

when specifying a utility function given partial preference information. They assumed that utility functions were uniformly distributed within their "truth set." We will also sample uniformly from $\mathbb{T}$ and regard our procedure as a form of sensitivity analysis. For example, we will be able to test optimal policies across the entire range of feasible distributions, with the caution that all distributions within $\mathbb{T}$ may not be equally likely.

Figure 1 provides a graphical representation of the JDSIM algorithm in two dimensions. The sampler is initialized (Step 1) by generating an arbitrary point $x_i \in \mathbb{T}$ and setting the counter $i = 0$. Step 2 generates a set of directions $D \subseteq \mathbb{R}^n$ using an uncorrelated multivariate standard normal and standardizing the directions. Step 3 selects a uniformly distributed direction $d_i \in D$. Step 4 finds the line $L = \mathbb{T} \cap \{x \mid x = x_i + \lambda d_i, \lambda \text{ a real scalar}\}$ generated by extending the direction d_i in both directions until the boundary of $\mathbb{T}$ is reached. Step 5 selects a random point $x_{i+1} \in L$ uniformly distributed over the line. Finally, Step 6 evaluates the counter and stops if $i = N$ (where N is the desired number of samples); otherwise the counter is incremented by one, and the sampler returns to Step 2.

The interested reader should refer to Montiel and Bickel (2012a, b) for technical details regarding JDSIM and its performance characteristics.

3. Approximating the Joint Distribution

This section reviews two common approaches for addressing missing probabilistic information by specifying a single distribution that is consistent with the available information. Then, two new methods are proposed.

3.1. Maximum Entropy

Jaynes (1957) proposed using the distribution in $\mathbb{T}$ that has the highest entropy, which we denote $\boldsymbol{\pi}^{ME}$. Formally, $\boldsymbol{\pi}^{ME}$ is defined as

$$\begin{aligned} \boldsymbol{\pi}^{ME} = \underset{\boldsymbol{\pi}=\{\pi_1,\ldots,\pi_n\}}{\arg\max} \quad & -\sum_{i=1}^{n} \pi_i \ln \pi_i, \\ \text{s.t.} \quad & \mathbf{A}\boldsymbol{\pi} = \mathbf{b}, \\ & \boldsymbol{\pi} \geq 0. \end{aligned} \tag{1}$$

ME can be understood as the joint distribution in the interior of $\mathbb{T}$ that has the minimum self-weighted geometric mean. In the case of the unit simplex, this point is the uniform distribution.

3.2. Underlying Event

Keefer (2004) introduced an approximation known as the UE model, which assumes that the random variables are independent given another "hidden" or "underlying" random variable. This procedure produces a single, approximate pmf. However, this pmf may not be a member of $\mathbb{T}$ because it does not take into account all the conditional assessments (constraints). In fact, in the decision considered next, UE produces an approximation equal to an assumption of independence. For this reason, we do not consider it further, but will instead compare our results to the independent (IN) pmf, $\boldsymbol{\pi}^{IN}$.

3.3. Analytic Center

We augment these approximations by considering the analytic center (AC) of $\mathbb{T}$. The AC has been used extensively within the optimization community to initialize interior point algorithms (Bertsimas and Tsitsiklis 1997, Bertsekas 1999, Boyd and Vandenberghe 2004). The pmf $\boldsymbol{\pi}^{AC}$ is defined as

$$\begin{aligned} \boldsymbol{\pi}^{AC} = \underset{\boldsymbol{\pi}=\{\pi_1,\ldots,\pi_n\}}{\arg\max} \quad & \sum_{i=1}^{n} \ln \pi_i, \\ \text{s.t.} \quad & \mathbf{A}\boldsymbol{\pi} = \mathbf{b}, \\ & \boldsymbol{\pi} \geq 0. \end{aligned} \tag{2}$$

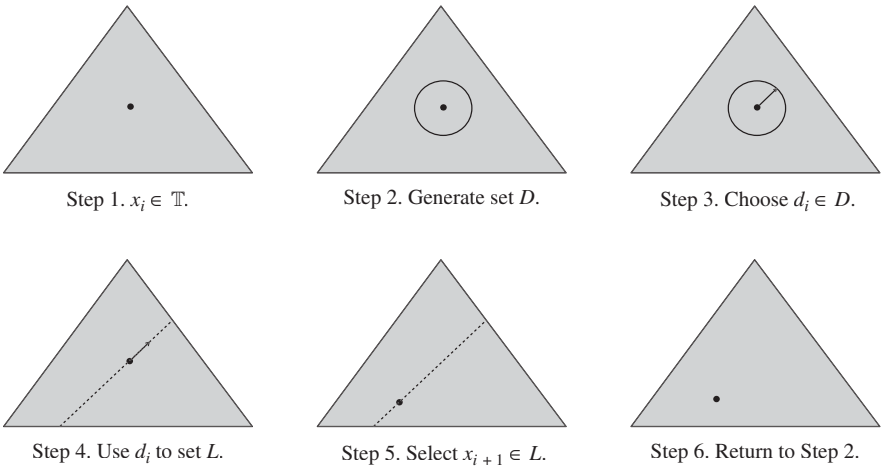
In our setting, the AC seeks to maximize the geometric mean, or the product of the joint probabilities.

Loosely speaking, the AC is the point within $\mathbb{T}$ that is farthest from all the constraints and is used within the optimization community because it possesses desirable "centrality" properties. In the unit simplex, the AC is the uniform distribution, as is the case for ME. However, these two "centers" might not coincide in the more general case.

3.4. Sample Average

Finally, we will consider another new approximation. Each point generated using JDSIM is a complete pmf. Nonetheless, one may want to evaluate the decision under a single pmf for reasons of computation or

Figure 1 Illustration of JDSIM in Two Dimensions



communication. To do so, we calculate the element-wise average of the m sampled pmfs, $\boldsymbol{\pi}^i \; \forall \, i = 1, \ldots, m$. We refer to this pmf as the sample average (SA) and denote it as $\boldsymbol{\pi}^{SA}$. Because $\boldsymbol{\pi}^{SA}$ is a convex combination of the points $\boldsymbol{\pi}^i$ within $\mathbb{T}$, which is convex, it is a feasible joint distribution. The pmf $\boldsymbol{\pi}^{SA}$ is defined as

$$\boldsymbol{\pi}^{SA} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{\pi}^i, \tag{3}$$

where the summation is taken element-wise.

3.5. Other Approximations

In the interest of space, other approximations such as the Chebyshev center and the maximum-volume-inscribed-ellipsoid center (Boyd and Vandenberghe 2004) are not considered here, but these are investigated in detail by Montiel (2012). Additionally, Dyer et al. (1973) suggested using the centroid of the corresponding polytope. In our case, this is infeasible because it requires specifying all of the polytope's vertices in advance. For high-dimensional polytopes, this is difficult, if not impossible, on a reasonable time scale. For example, consider a simple joint probability distribution comprised of eight binary random variables, whose marginal distributions are known. The polytope encoding these constraints could have up to 10^{13} vertices (McMullen 1970). Although this is an upper bound, the number of vertices likely to present in real-world problems is still enormous (Schmidt and Mattheiss 1977).

4. Illustrative Example

We now illustrate and compare our methods by using the example presented by Bickel and Smith (2006). Bickel and Smith (2006) presented an oil exploration decision and considered a field with six possible well locations. Each well could be wet (i.e., oil is present) or dry (i.e., oil is absent). Company experts believed that the wells were probabilistically dependent. The challenge was how to construct the joint pmf and determine the optimal exploration sequence.

For convenience, we summarize the problem presented by Bickel and Smith (2006) using their notation. Formally, let $w_i = 1$ if well i is wet, and let $w_i = 0$ if it is dry. Likewise, let the probability that well i is wet be $p_i \equiv P(w_i = 1)$. If well i is wet, the expected value of this success is s_i. However, if the well is dry, the expected value of this failure is f_i.

Let the vector $\mathbf{w} = (w_1, \ldots, w_n)$ represent the joint outcome for n wells. For example, $\mathbf{w} = (1, 0, 1, 1, 0, 0)$ represents the case where wells 1, 3, and 4 were wet and wells 2, 5, and 6 were dry. Let $\boldsymbol{\pi} \equiv \pi(\mathbf{w})$ be the joint pmf defined over the well outcomes $\mathbf{w}$.

Bickel and Smith's (2006) data are presented in Table 1. Expected values are in millions of dollars and represent NPVs for the period in which the well is drilled. The *intrinsic values* shown in Table 1 are the unconditional expected values: $p_i s_i + (1 - p_i) f_i$. In this example, the intrinsic values are all negative, implying that the company should not drill any of the wells if they are considered in isolation.

Table 1 Bickel and Smith (2006) Example Well Data

Well	Probability of success (p_i)	Expected values: Given success (s_i)	Given failure (f_i)	Intrinsic value
1	0.35	60	−35	−1.75
2	0.49	15	−20	−2.85
3	0.53	30	−35	−0.55
4	0.83	5	−40	−2.65
5	0.33	40	−20	−0.20
6	0.18	80	−20	−2.00

Table 3 Bickel and Smith (2006) Pairwise Joint Assessments of Well Success

	$p(w_j = 1, w_i = 1)$					
$i \backslash j$	1	2	3	4	5	6
1		0.2065	0.2205	0.2905	0.1365	0.1085
2			0.3185	0.4067	0.2695	0.1176
3				0.4399	0.2226	0.1643
4					0.2739	0.1494
5						0.0858
6						

The company believed that success or failure at one location may make the presence of hydrocarbons at another location more or less likely, respectively. However, this is not reflected in the data in Table 1. As detailed by Bickel and Smith (2006), assessing the full joint distribution would require 63 ($2^6 - 1$) probability assessments, including many that would be heavily conditioned. Although this task was deemed too difficult, the company was comfortable providing the pairwise conditional probabilities of success for well i given success of well j, $\forall\, i \neq j$, given in Table 2.

Well 4 was believed to be pairwise independent of all other wells. It will be seen as helpful to recast the information given in Tables 1 and 2 into pairwise joint probabilities, $p_{i,j} \equiv p(w_i = 1, w_j = 1) = p(w_j = 1 \mid w_i = 1)p(w_i = 1)$. These values are shown in Table 3.

The assessments in Tables 1 and 3, along with the requirement that the joint pmf sums to unity, define a polytope $\mathbf{A}\boldsymbol{\pi} = \mathbf{b}$ that contains all feasible joint pmfs. Again, we refer to this polytope as the truth set and define it as $\mathbb{T} = \{\boldsymbol{\pi}\colon \mathbf{A}\boldsymbol{\pi} = \mathbf{b}, \boldsymbol{\pi} \geq 0\}$. Any $\boldsymbol{\pi}$ that is a member of $\mathbb{T}$ is a feasible joint pmf consistent with the assessed information.

In this case, $\mathbf{A}$ is a 22-by-64 matrix of zeros and ones (Appendix A) that represents the 22 constraints (the requirement that the probabilities sum to one, six marginal constraints, and 15 pairwise constraints) and the 64 (2^6) elements of the joint distribution; $\boldsymbol{\pi}$ is the joint pmf, which in this case is a vector of 64 joint probabilities. The vector $\mathbf{b}$ encodes the 22 constraints described previously:

$$\mathbf{b} = [1, 0.35, 0.49, 0.53, 0.83, 0.33, 0.18, 0.2065, 0.2205, 0.2905, 0.1365, 0.1085, 0.3185, 0.4067, 0.2695, 0.1176, 0.4399, 0.2226, 0.1643, 0.2739, 0.1494, 0.0858].$$

The set $\mathbb{T}$ contains an infinite number of feasible pmfs, which is the challenge stated at the outset. In the following, we evaluate the four approximation methods that specify a single distribution within the truth set, using a collection of four million pmfs, each of which is consistent with the information in Tables 1 and 3. Four million pmfs are sufficient, in this case, to ensure that $\mathbb{T}$ is uniformly covered (Montiel and Bickel 2012a).

Once a *single* distribution has been specified by one of the procedures discussed in §3, we proceed

Table 2 Bickel and Smith (2006) Pairwise Conditional Assessments of Well Success

	Direct conditional assessments $p(w_j = 1 \mid w_i = 1)$							Implied moment correlation matrix (ρ_{ij})					
$i \backslash j$	1	2	3	4	5	6	Marginal p_i	1	2	3	4	5	6
1		0.59	0.63	0.83	0.39	0.31	0.35		0.147	0.147	0	0.094	0.248
2			0.65	0.83	0.55	0.24	0.49			0.236	0	0.459	0.153
3				0.83	0.42	0.31	0.53				0	0.203	0.359
4					0.33	0.18	0.83					0	0
5						0.26	0.33						0.146
6							0.18						

to determine the optimal exploration sequence. The structure of the decision is as follows: we decide which well to drill first (if any); based on whether that well is wet or dry, we decide which well to drill next, and so on, through the $n = 6$ stages. Following Bickel and Smith (2006), we find the optimal exploration sequence using dynamic programming (DP). In the interest of space, Bickel and Smith's (2006) formulation is not repeated here.

Figure 2 presents the optimal policies under the IN, ME, AC, and SA approximations. The pmfs associated with these optimal polices are given in Appendix B. In the case of IN, the optimal policy, shown in Figure 2(a), is to not drill, ensuring an NPV of \$0. Figure 2(b) shows the optimal policy under the ME approximation. This policy, which matches Bickel and Smith's (2006) Figure 2, begins by drilling well 3. If well 3 is wet, then well 6 is drilled; otherwise drilling stops. If wells 3 and 6 are wet, then well 1 is drilled, and so on. The dollar values in these figures are the continuation values (Bickel and Smith 2006, Equation (8)). So, for example, the value of the ME optimal policy is \$14.4 million. This contrasts with each well having had a negative value in isolation (i.e., ignoring the dependence structure). Under $\boldsymbol{\pi}^{ME}$ it is never optimal to drill well 4 because, as shown in Table 2, well 4 was specified as being pairwise independent of all the other wells, and, without additional information, ME will force higher-order assessments toward independence.

Figure 2(c) presents the AC optimal policy, whose value is \$14.6 million. This policy begins by drilling well 2 instead of ME's choice of well 3. If well 2 is wet, then well 5 is drilled, which is earlier than in the ME policy. Most strikingly, the AC optimal policy calls for drilling well 4, the pairwise independent well, if at least four of the previous wells were wet. Thus, the AC approximation has encoded the possibility of higher-order dependence that was assumed not to exist in the case of ME.

Finally, Figure 2(d) presents the SA optimal policy, with a value of \$14.71 million. Like the AC, the SA begins by drilling well 2. In fact, the only difference between the AC and SA optimal policies is that, under SA, we drill well 4 even if well 6 ends up being dry, after success at wells 2, 5, and 3.

The values of the optimal policies differing among ME (\$14.4 million), AC (\$14.6 million), and SA (\$14.71 million) is a result of differing optimal policies and that these expected values are being computed under differing probability distributions. Later in this paper we will isolate these two effects.

Given that many approaches exist for specifying a unique joint distribution within the truth set, it needs to be determined how "representative" or accurate they are, that is, how well the ME, AC, and SA policies perform across the set of distributions contained in $\mathbb{T}$.

5. Approximation Accuracy

We will consider two notions of accuracy. First, we will determine how "close" each of our approximations is to all other pmfs in the truth set, or how representative each approximation is of the entire set. Second, we will investigate how well the optimal policy associated with a particular approximation performs across the set of feasible distributions.

5.1. Accuracy Measures

We consider four measures of accuracy, defined in Equations (4a)–(4d), where $\boldsymbol{\pi}$ is a sampled pmf, π_i is the ith element of $\boldsymbol{\pi}$, n is the number of joint events in $\boldsymbol{\pi}$, and $\boldsymbol{\pi}^*$ is an approximate pmf, such as $\boldsymbol{\pi}^{ME}$, $\boldsymbol{\pi}^{AC}$, $\boldsymbol{\pi}^{IN}$, or $\boldsymbol{\pi}^{SA}$.

The average absolute difference (**AAD**), or the scaled $\mathbf{L}^1$-norm, measures the average absolute difference between the elements of two pmfs. The maximum absolute difference ($\mathbf{L}^\infty$-norm) defines the difference between two pmfs to be equal to the maximum difference between any two elements. The Euclidean distance ($\mathbf{L}^2$-norm) is the straight line distance in n-dimensional space between the two pmfs. Finally, **KL** divergence measures the relative entropy between a pmf and a reference distribution, which we take to be one of the approximate pmfs. In this paper, we use the binary or base-2 logarithm:

$$\mathbf{AAD}_n(\boldsymbol{\pi}, \boldsymbol{\pi}^*) = \frac{1}{n}\mathbf{L}^1(\boldsymbol{\pi}, \boldsymbol{\pi}^*) = \frac{1}{n}\sum_{i=1}^{n} |\pi_i - \pi_i^*|, \quad (4a)$$

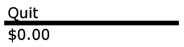
$$\mathbf{L}_n^\infty(\boldsymbol{\pi}, \boldsymbol{\pi}^*) = \max\{|\pi_1 - \pi_1^*|, \cdots, |\pi_n - \pi_n^*|\}, \quad (4b)$$

$$\mathbf{L}_n^2(\boldsymbol{\pi}, \boldsymbol{\pi}^*) = \left[\sum_{i=1}^{n} (\pi_i - \pi_i^*)^2\right]^{1/2}, \quad (4c)$$

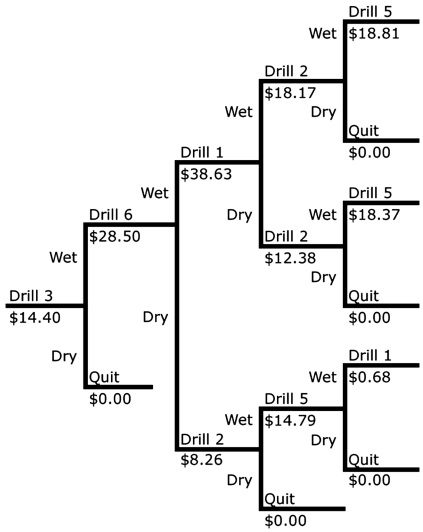
$$\mathbf{KL}_n(\boldsymbol{\pi} \| \boldsymbol{\pi}^*) = \sum_{i=1}^{n} \pi_i \log\left(\frac{\pi_i}{\pi_i^*}\right). \quad (4d)$$

Figure 2 Optimal Policy Under Four Approximation Methods

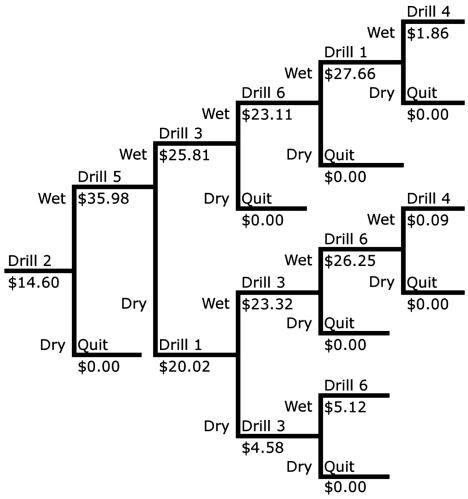
(a) IN optimal policy



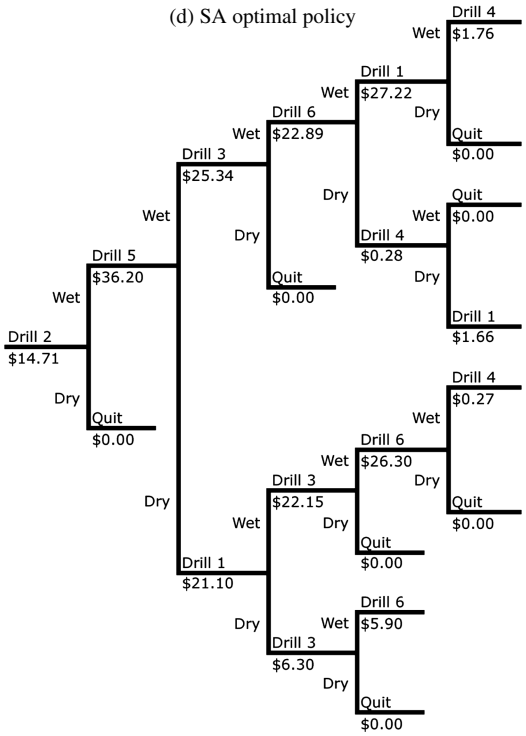
(b) ME optimal policy



(c) AC optimal policy



(d) SA optimal policy



Note. $ in millions.

5.2. Approximation Representativeness

The analysis in this section complements the study of accuracy presented by Bickel and Smith (2006), who used a simulation procedure proposed by Keefer (2004). Bickel and Smith (2006) generated a joint pmf at random, extracted its marginal and pairwise joint probabilities, executed an approximation procedure (ME, IN, and UE) given these marginal and pairwise joint probabilities, and then compared the simulated pmf to this approximation. Using the language in our current paper, Bickel and Smith (2006) generated a large number of random polytopes (truth sets), determined the approximate pmfs within these polytopes (again, ME, IN, and UE), and then compared these approximations to a *single* distribution in each polytope (i.e., the one they sampled). This procedure provides very little control over the structure of the polytope (marginal probabilities, correlations, etc.) and fails to measure the accuracy of an approximation within a single problem formulation. Likewise, the simulation procedure used in Abbas (2006), based on the method of uniform spacings (Devroye 1986), cannot be used to sample joint distributions within a given polytope. JDSIM overcomes these limitations by providing samples within a single polytope, matching known and easily controlled constraints.

Specifically, we compute our accuracy measures (Equations (4a)–(4d)) for each of the four million pmfs we sample, relative to the IN, ME, AC, and SA approximations. Histograms of our results are shown in Figure 3.

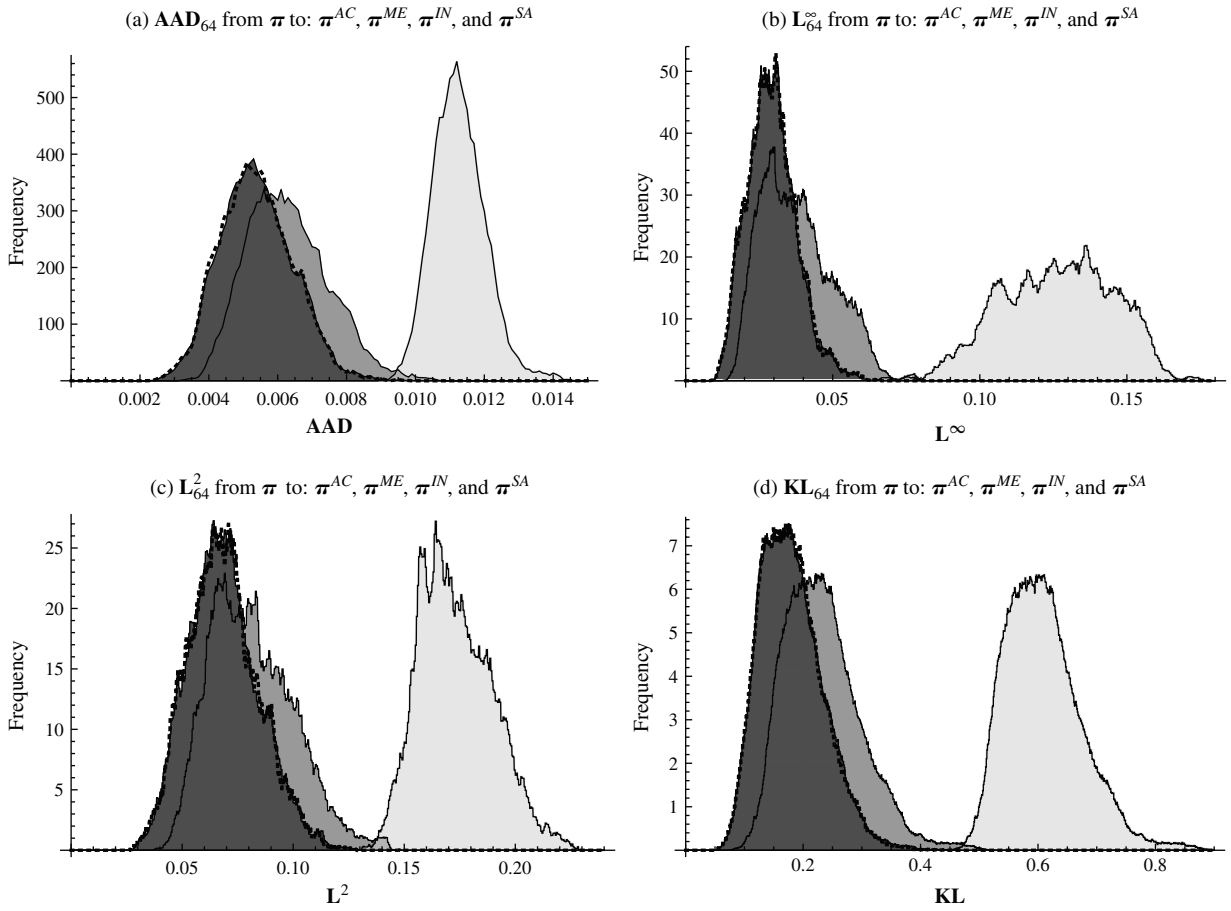
Each of the subfigures corresponds to one measure of accuracy (**AAD**, $\mathbf{L}^\infty$, $\mathbf{L}^2$, or **KL**) and presents the distribution of the distance (or divergence) for each of our sampled distributions to $\boldsymbol{\pi}^{IN}$ (light gray), $\boldsymbol{\pi}^{ME}$ (medium gray), $\boldsymbol{\pi}^{AC}$ (dark gray), and $\boldsymbol{\pi}^{SA}$ (dashed line). Table 4 provides the mean and standard deviation of our four error measures. In addition, we present the fraction of times a sampled distribution $\boldsymbol{\pi}$ was closer to one approximation than another.

Table 4 is analogous to Bickel and Smith's (2006) Table 6. However, direct comparison is difficult because, as discussed earlier, Bickel and Smith (2006) calculated the distance between a single pmf and an approximation in 5,000 different polytopes and averaged the results. In contrast, we calculate the average distance between each of the four million distributions and an approximation within the same polytope, which is consistent with the information in Tables 1 and 3. Although, as discussed next, the findings are broadly similar in the sense that ME is a better approximation than IN, we expand upon this finding by showing that the AC and SA are better than ME, using Bickel and Smith's (2006) accuracy measures.

Figure 3 and Table 4 yield several observations:

- IN, which in our case is equivalent to UE, provides a very poor approximation. This can be seen by noting that the IN frequency plots in Figure 3 are far to the right of the other approximations. Likewise, the mean **AAD** and $\mathbf{L}^\infty$-norm (Table 4) are 0.0113 and 0.1262, respectively. These values are between 1.79 and 3.34 times larger than the next best approximation, which is ME. This result is not surprising given that the IN approximation is not within the interior of $\mathbb{T}$.
- AC tends to be closer than either IN or ME to all other distributions in $\mathbb{T}$, on each of our four measures. For example, the mean **AAD** between the samples and AC is 0.0054, compared to 0.0063 for ME and 0.0113 for IN. Furthermore, considering **AAD** as an example, only 18.0% of the sampled pmfs were closer to ME than they were to AC, and none was closer to IN. Only 6.8% of our sampled pmfs were closer to ME than AC, as measured by **KL** divergence. In this sense, AC is more representative than ME of the set of possible distributions.
- AC is less extreme than either IN or ME, in that pmfs within $\mathbb{T}$ have a higher variation with respect to IN and ME, as measured by the standard deviation. For example, the standard deviation in the maximum absolute difference is 0.0083 for AC, 0.0119 for ME, and 0.0189 for IN.
- AC is very close to SA, as evidenced by their histograms in Figure 3 being almost indistinguishable (again, SA is the thin dashed line). The **AAD**, $\mathbf{L}^\infty$-norm, $\mathbf{L}^2$-norm, and **KL** divergence from AC to SA are 0.0011, 0.0082, 0.0171, and 0.0051, respectively. This is significant because determining the AC is less difficult than generating and averaging millions of pmfs. Therefore, these results appear to make a strong case for using the AC to determine the optimal exploration sequence under a single pmf.

Figure 3 Accuracy of Joint Distribution Approximations



Note. We show the distributions of the accuracy measures from the sampled pmfs to AC (dark gray), ME (medium gray), IN (light gray), and SA (dashed).

These results suggest that the AC may be a better approximation than the ME in the case of partial information. If $\mathbb{T}$ can be sampled, then SA also provides a good approximation. Although these findings are conditional upon a single decision situation, our conjecture is that similar results will hold in many other cases, because AC has desirable "centrality" properties.

5.3. Approximation Effect on Optimal Policies

Let $\nu^*(\boldsymbol{\pi})$ be the expected NPV of the optimal policy evaluated under approximation * (i.e., the value of the * optimal policy), evaluated under $\boldsymbol{\pi}$. Figure 4 displays the distribution of the value of the ME, AC, and SA optimal policies using our collection of four million pmfs. Table 5 reports the mean and standard deviation of these values under the (v) measure. Also shown in this table is the frequency that each approximation produced a lower expected NPV than another approximation.

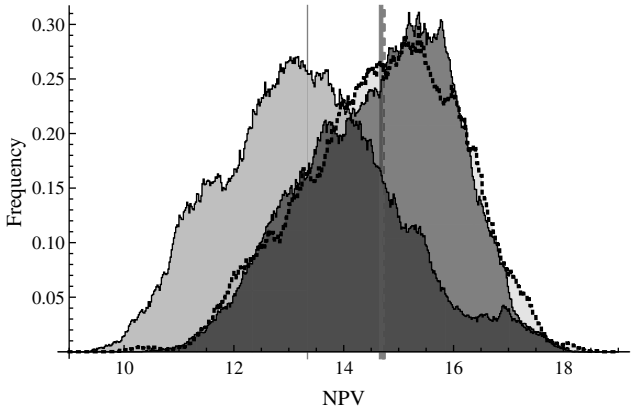
Recall that $\nu^{IN}(\boldsymbol{\pi}) = 0$ for any $\boldsymbol{\pi} \in \mathbb{T}$, because the IN optimal policy is not to drill. The SA policy displays the best performance, with an average NPV of \$14.71 million, followed by the AC policy at \$14.68 million, and the ME policy at \$13.34 million. Thus, the AC optimal policy is \$1.34 million better than the ME optimal policy used by Bickel and Smith (2006). In addition, the minimum and maximum NPVs achieved under the AC policy are greater than those of the ME policy. This suggests that the AC is more robust than ME. Likewise, the ME optimal policy underperformed either the AC or SA optimal policies in over 70% of the sampled distributions (Table 5).

Table 4 Probability Accuracy Results

	IN	ME	AC	SA
i. Average absolute difference				
Mean	0.0113	0.0063	0.0054	0.0054
Std. dev.	0.0008	0.0012	0.0011	0.0011
Fraction < SA (%)	0.0	19.2	45.5	—
Fraction < AC (%)	0.0	18.0	—	54.5
ii. Max absolute difference ($\mathbf{L}^\infty$-norm)				
Mean	0.1262	0.0378	0.0294	0.0294
Std. dev.	0.0189	0.0119	0.0083	0.0084
Fraction < SA (%)	0.0	26.8	54.2	—
Fraction < AC (%)	0.0	24.3	—	45.8
iii. Euclidean distance ($\mathbf{L}^2$-norm)				
Mean	0.1734	0.0815	0.0687	0.0686
Std. dev.	0.0166	0.0188	0.0159	0.0159
Fraction < SA (%)	0.0	24.3	45.2	—
Fraction < AC (%)	0.0	23.3	—	54.8
iv. KL divergence				
Mean	0.6095	0.2334	0.1802	0.1795
Std. dev.	0.0649	0.0649	0.0536	0.0528
Fraction < SA (%)	0.0	8.2	43.7	—
Fraction < AC (%)	0.0	6.8	—	56.3

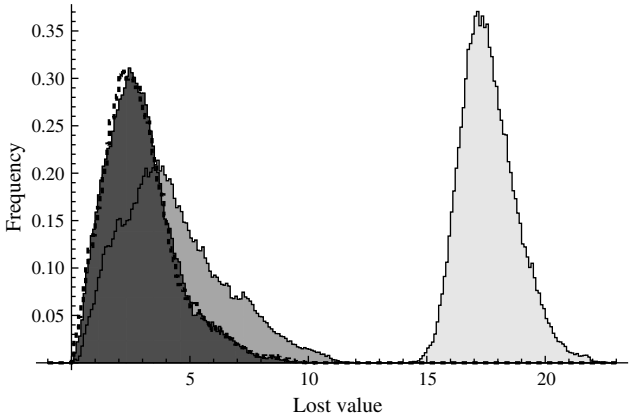
In addition to the distribution of the expected NPV, we also calculate what Bickel and Smith (2006) termed the "lost value" (LV), by using an approximation rather than the true (but unknown) distribution. The lost value represents the decrease in value that stems from using the approximate pmf in lieu of the true pmf. Let the NPV of the optimal policy generated

Figure 4 Distribution of the Optimal-Policy Value for ME (Medium Gray), AC (Dark Gray), and SA (Dashed), Evaluated Under Four Million Sampled Joint Distributions



Notes. The vertical lines indicate the average NPV for ME (thin solid line), AC (thick), and SA (dashed) policies. The IN optimal policy is to never drill.

Figure 5 Distribution of Lost Value for 160,000 Sampled Joint Distributions: IN (Light Gray), ME (Medium Gray), AC (Dark Gray), and SA (Dashed)



using $\boldsymbol{\pi} \in \mathbb{T}$ be $\nu(\boldsymbol{\pi})$. Then, the LV for approximation * is defined as:

$$\mathbf{LV}^*(\boldsymbol{\pi}) = \nu(\boldsymbol{\pi}) - \nu^*(\boldsymbol{\pi}) \quad \forall\, \boldsymbol{\pi} \in \mathbb{T}. \tag{5}$$

The calculation of $\mathbf{LV}^*(\boldsymbol{\pi})$ took an average of 20 seconds per distribution, which made it impractical to solve the complete sample of four million pmfs. Therefore, we selected a random subcollection of 160,000 samples to compute the **LV**.

Figure 5 shows the distribution of $\mathbf{LV}^*(\boldsymbol{\pi})$ for IN, ME, AC, and SA policies under 160,000 sampled joint pmfs. We present the mean and standard deviation of these results in Table 5, under measure (vi). We also compute the frequency that each approximate optimal policy outperformed either SA or AC.

Because $\nu^{IN}(\boldsymbol{\pi}) = 0$, the distribution of $\mathbf{LV}^{IN}(\boldsymbol{\pi})$ is simply the distribution of $\nu(\boldsymbol{\pi})$, which is the distribution of the optimal NPV evaluated under the possible true distributions. Thus, knowing the true joint distribution and following the corresponding optimal policy would earn an expected NPV between \$14.4 and \$22.3 million, with an average of \$17.6 million, which is the average **LV** by following IN optimal policy.

ME has an average **LV** of \$4.29 million that exceeds the corresponding values of \$2.95 and \$2.91 for AC and SA, respectively. Furthermore, ME outperformed AC in only 23.4% of the cases. SA outperformed AC approximately 57.7% of the time.

In sum, the results in Tables 4 and 5 demonstrate that within the context of the decision problem considered here, AC and SA are more representative of

Table 5 Optimal Policy Accuracy Results

	IN	ME	AC	SA
v. Value of optimal policy ($ millions)				
Mean	0.00	13.34	14.68	14.71
Std. dev.	0.00	1.55	1.33	1.39
Fraction < SA (%)	100.00	70.61	57.67	—
Fraction < AC (%)	100.00	70.55	—	42.33
vi. Optimal policy lost value ($ millions)				
Mean	17.62	4.29	2.95	2.91
Std. dev.	1.18	2.16	1.51	1.54
Fraction < SA (%)	0.00	29.30	42.30	—
Fraction < AC (%)	0.00	23.40	—	57.70

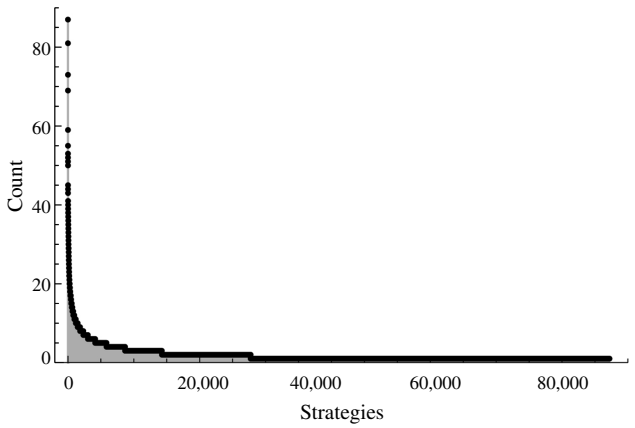
the set of all distributions that are consistent with the information provided in Tables 1 and 3 than either ME or IN. Furthermore, AC performs almost as well as SA. Therefore, if the time or resources are inadequate to use JDSIM, AC may provide a very good approximate distribution under which the decision can be analyzed.

6. Searching for Robust Policies

Having completed our study of approximation accuracy, we now demonstrate how to use the JDSIM procedure to solve Bickel and Smith's (2006) example in a new way. Rather than finding the optimal policy under a single approximation, we search for optimal policies that are "good" or "robust" under a large set of feasible distributions. By robust, we mean policies that are optimal across a range of distributions within our truth set.

As with the lost value calculation, we start by taking a subcollection of 160,000 sampled distributions from the full collection of four million and solve for the optimal policies. Excluding the 77,795 policies that are duplicates yields 82,205 policies that are optimal for at least one $\boldsymbol{\pi} \in \mathbb{T}$. This large number of optimal policies demonstrates the complexity of the decision situation and the degree to which the optimal policy can vary. Figure 6 shows the number of times that different optimal policies appear, where each dot represents a single optimal policy. Although many of the 160,000 optimal policies are possible, the vast majority appear less than 10 times; only nine optimal policies appear more than 50 times. This last group includes the policies that are optimal for the largest number of joint distributions sampled in the subcollection.

Figure 6 Most Common Optimal Policies



A complete analysis of the 82,205 policies is not feasible here. Therefore, we reduce our scope to the nine policies having the highest optimality count and present them in Figure 7. We also evaluate the three policies from the approximate distributions. For each of these optimal policies, we evaluate the expected NPV under all four million joint distributions in the original collection sampled from $\mathbb{T}$. Table 6 shows the mean, standard deviation, minimum, and maximum values for the expected NPVs generated by each policy given the collection sampled from $\mathbb{T}$. The first three columns (ME, AC, and SA) repeat the values reported in Table 5.

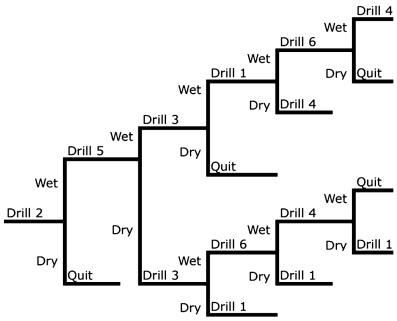
Of the policies shown in Table 6, the top four performers, in descending order, are SA, AC, S4, and ME. With a mean of $14.42 million, S4 outperforms ME ($13.34 million) and is close to SA ($14.71 million). Moreover, its minimum is the highest except for AC, and its maximum is the highest of them all.

We now test the relative robustness of the policies in Table 6. Specifically, we determine the fraction of times each policy outperforms the others across our set of four million distributions. These results are presented in Figure 8. Four of the policies appear to be the most robust: S4, SA, AC, and ME. However, S4 outperforms SA by nearly 11 percentage points. By selecting S4 we will be better off approximately 26% of the time (1.04 out of 4 million) than if we would have selected one of the other 11 policies.

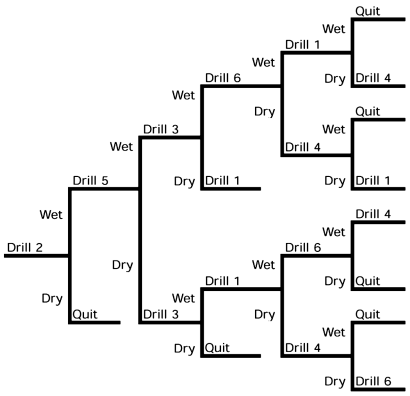
Before concluding, we address one further issue considered by Bickel and Smith (2006). They explained why it is optimal to start by drilling well 3 rather than one of the other five wells. To do so, they

Figure 7 Optimal Policies with Highest Frequencies

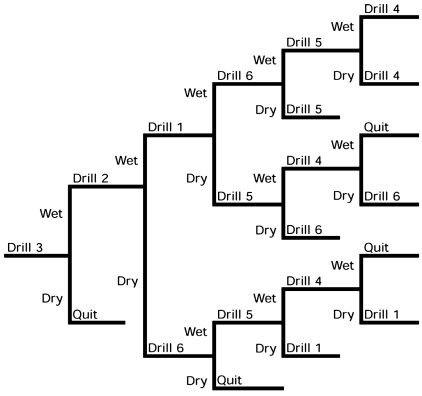
(a) Highest frequency (S1)



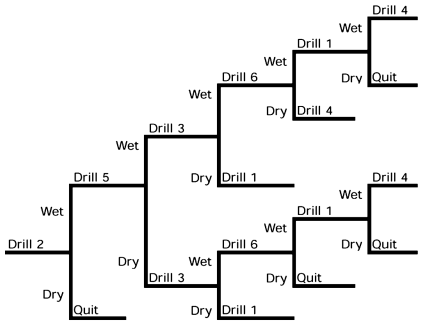
(b) Second highest frequency (S2)



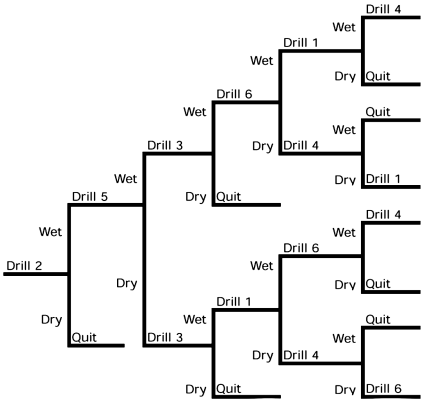
(c) Third highest frequency (S3)



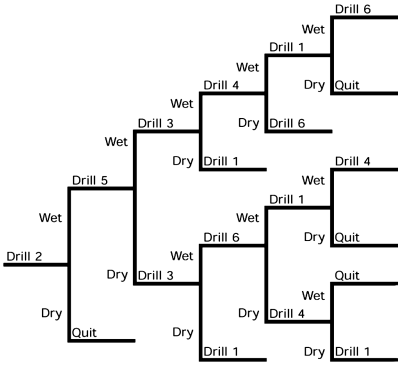
(d) Fourth highest frequency (S4)



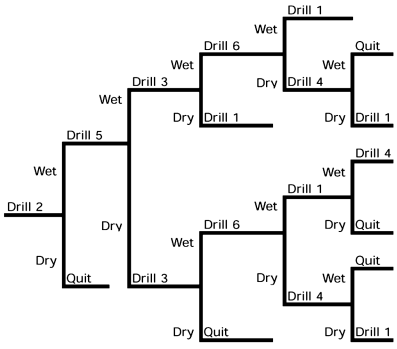
(e) Fifth highest frequency (S5)



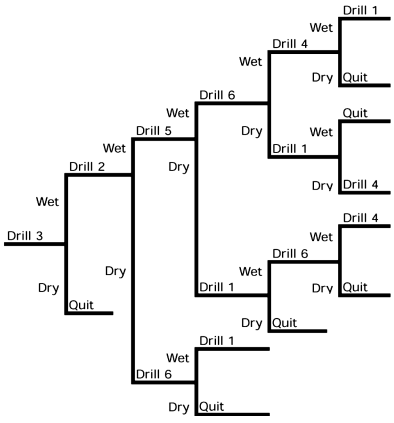
(f) Sixth highest frequency (S6)



(g) Seventh highest frequency (S7)



(h) Eighth highest frequency (S8)



(i) Ninth highest frequency (S9)

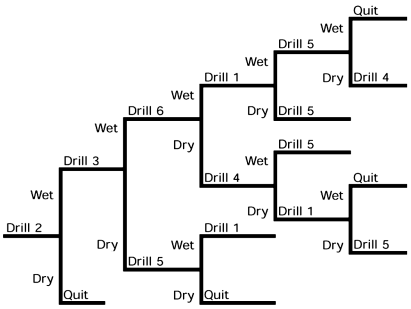


Table 6 Profit Statistics for Selected Optimal Policies ($ Millions)

	PMF approximations			High-frequency full optimal policies								
	ME	AC	SA	S1	S2	S3	S4	S5	S6	S7	S8	S9
Mean	13.34	14.68	14.71	12.84	13.27	13.05	14.42	13.46	12.99	13.39	13.04	13.26
Std. dev.	1.55	1.33	1.39	2.14	1.90	1.56	1.69	1.55	1.53	1.61	1.83	1.63
Min.	9.26	10.33	9.73	7.16	7.03	8.80	9.90	8.17	7.65	7.43	7.26	7.12
Max.	18.22	18.46	18.69	18.96	18.87	18.68	20.04	18.46	17.82	18.09	18.99	18.32

determined the expected NPV if one starts by drilling each of the six wells and then follows the optimal policy thereafter (see Bickel and Smith 2006, Table 5). They found that whereas the expected NPV if one starts with well 3 was $14.4 million, the expected NPV of starting with well 2 was $14.3 million. Thus, the optimal policies that begin with these two wells are close in value, being almost indistinguishable in this case. We now use JDSIM to provide additional insight into this finding.

Figure 9 plots the frequency with which each of our 160,000 optimal policies begins with a given well. Beginning with well 2 is seen to be optimal in 65.7% (105,131/160,000) of the cases, compared to approximately 33.4% (53,487/160,000) for well 3. Thus, although the values of the optimal policies beginning with well 2 or 3 are close, about twice as many optimal policies begin with well 2. Also, more than 1,000 policies begin by drilling well 4, which was never drilled under Bickel and Smith's (2006) ME optimal policy because it was pairwise independent of all the other wells.

7. Discussion and Conclusion

In this paper, we have challenged long-standing methods such as maximum entropy by demonstrating, within a single decision problem, that it may not produce an approximation that is the most representative of the set of all feasible distributions. Other approximations, such as the analytic center, may provide better performance. This finding is limited, however, by our assumption of uniform sampling. Nonetheless, it highlights the assumptions underlying the use of maximum entropy. Bickel and Smith (2006) motivated their use of maximum entropy by noting that higher-order assessments were either too difficult or would expend excessive time or resources, not that experts had "no idea" regarding higher-order assessments. In this case, maximum entropy may not be the most reasonable approximation. An analyst's lacking the time or resources to perform an assessment differs from carefully considering the assessment and having "no idea." In the maximum-entropy framework, "no idea" has a precise mathematical meaning: maximum entropy will try to force any unspecified relationships toward an assessment of independence, irrespective

Figure 8 Fraction of Time Each Policy is the Best Alternative in a Sample of Four Million Distributions

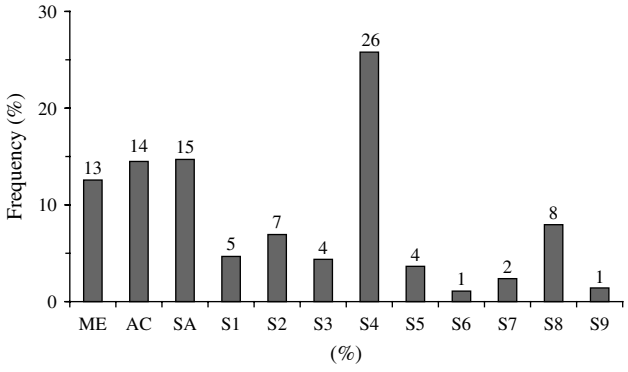
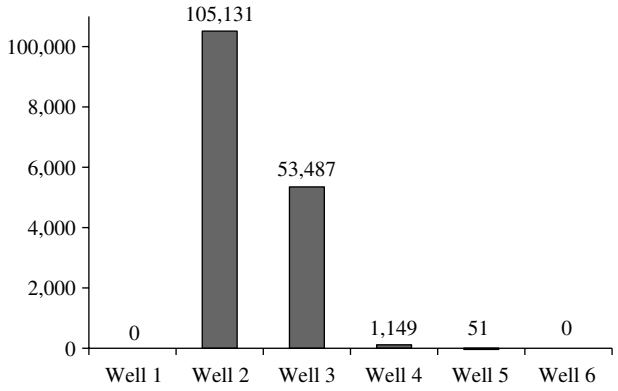


Figure 9 Number of Times Each Optimal Policy Begins with the Listed Well



of the analyst's intent. In the oil and gas problem we considered, well 4 was pairwise independent of all other wells. Maximum entropy interpreted the lack of any higher-order assessments as a statement of independence and thus assumed that well 4 was independent of all other wells. The analyst might allow for the possibility that higher-order dependence exists, even though the time and/or resources to assess it might have been unavailable.

Continuing this line of thought, rather than assuming a single distribution, be it maximum entropy, analytic center, or some other approximation, the analyst might want to test a large collection of feasible distributions. This testing will provide insight into the robustness of different decision alternatives. The simulation procedure presented in this paper offers the possibility of testing alternatives to a greater degree than has been done in the past. This may improve the confidence decision makers have in their choices. Likewise, this simulation procedure could be used iteratively to determine whether further assessments are worthwhile. For example, Figure 4 shows that the range of expected NPVs obtained by following the analytic-center optimal policy is between approximately \$11 and \$18 million. This uncertainty range is driven by the fact that higher-order assessments (e.g., all three-way joint assessments) have not been specified. At this point, one could ask whether performing these assessments is worthwhile, or which assessment should be gathered next.

Finally, in this paper we have analyzed the accuracy of the independent, maximum-entropy, and analytic-center approximations within a single probabilistic and decision context. In future work, our simulation procedure could also be used to test the accuracy of these approximations across a wide range of polytopes. This research would seek to understand the accuracy of the various approximations as the number of random variables is increased, as the number of outcomes per random variable is increased, and across different dependence structures. This would help analysts understand the accuracy of the analytic-center approximation, for example, in situations that might arise in practice.

Acknowledgments
The authors thank the editor-in-chief, the associate editor, and two referees for helpful comments and suggestions regarding a previous draft of this paper. This work was supported by the National Science Foundation under J. E. Bickel's CAREER [Grant No. SES-0954371].

Appendix A. Elements of the Truth Set

We define the truth set as $\mathbb{T} \equiv \{\boldsymbol{\pi} \mid \mathbf{A}\boldsymbol{\pi} = \mathbf{b}, \boldsymbol{\pi} \geq 0\}$. The first row of $\mathbf{A}$ constrains the probabilities to sum to one. Rows 2–7 select the joint elements of $\boldsymbol{\pi}$ necessary to generate the marginal event probabilities for all six wells being wet. Rows 8–22 select the joint elements of $\boldsymbol{\pi}$ necessary to generate the pairwise event probabilities of wells i and j, $i \neq j$, being wet. $\mathbf{A}$ is then defined as follows:

$$
\mathbf{A} = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0
\end{bmatrix}.
$$

Appendix B. Approximate Joint Distributions

	Independent	Maximum entropy	Analytic center	Sample average
Event 1 (all wells wet)	0.0045	0.0383	0.0380	0.0396
Event 2	0.0204	0.0372	0.0320	0.0252
Event 3	0.0091	0.0209	0.0259	0.0251
Event 4	0.0414	0.0271	0.0273	0.0295
Event 5	0.0009	0.0078	0.0028	0.0028
Event 6	0.0042	0.0076	0.0051	0.0051
Event 7	0.0019	0.0043	0.0032	0.0030
Event 8	0.0085	0.0056	0.0065	0.0061
Event 9	0.0040	0.0019	0.0012	0.0011
Event 10	0.0181	0.0205	0.0236	0.0265
Event 11	0.0081	0.0017	0.0012	0.0012
Event 12	0.0368	0.0238	0.0264	0.0279
Event 13	0.0008	0.0004	0.0008	0.0009
Event 14	0.0037	0.0042	0.0050	0.0053
Event 15	0.0017	0.0003	0.0009	0.0009
Event 16	0.0075	0.0049	0.0067	0.0066
Event 17	0.0047	0.0044	0.0057	0.0065
Event 18	0.0212	0.0052	0.0061	0.0067
Event 19	0.0095	0.0195	0.0199	0.0186
Event 20	0.0431	0.0305	0.0324	0.0361
Event 21	0.0010	0.0009	0.0020	0.0021
Event 22	0.0044	0.0011	0.0032	0.0033
Event 23	0.0019	0.0040	0.0031	0.0031
Event 24	0.0088	0.0062	0.0073	0.0077
Event 25	0.0041	0.0004	0.0010	0.0010
Event 26	0.0188	0.0054	0.0061	0.0062
Event 27	0.0084	0.0030	0.0012	0.0011
Event 28	0.0383	0.0508	0.0427	0.0383
Event 29	0.0008	0.0001	0.0008	0.0008
Event 30	0.0039	0.0011	0.0032	0.0035
Event 31	0.0017	0.0006	0.0009	0.0009
Event 32	0.0078	0.0104	0.0080	0.0076
Event 33	0.0083	0.0207	0.0191	0.0168
Event 34	0.0379	0.0619	0.0774	0.0857
Event 35	0.0169	0.0117	0.0139	0.0157
Event 36	0.0770	0.0466	0.0389	0.0368
Event 37	0.0017	0.0042	0.0032	0.0031
Event 38	0.0078	0.0127	0.0088	0.0076
Event 39	0.0035	0.0024	0.0035	0.0036
Event 40	0.0158	0.0095	0.0129	0.0130
Event 41	0.0074	0.0013	0.0011	0.0011
Event 42	0.0336	0.0419	0.0421	0.0390
Event 43	0.0150	0.0011	0.0011	0.0011
Event 44	0.0683	0.0501	0.0376	0.0346
Event 45	0.0015	0.0003	0.0009	0.0009
Event 46	0.0069	0.0086	0.0084	0.0090
Event 47	0.0031	0.0002	0.0009	0.0009
Event 48	0.0140	0.0103	0.0136	0.0147
Event 49	0.0087	0.0037	0.0051	0.0051
Event 50	0.0395	0.0134	0.0073	0.0064
Event 51	0.0176	0.0171	0.0130	0.0134
Event 52	0.0801	0.0818	0.0779	0.0727
Event 53	0.0018	0.0008	0.0022	0.0023
Event 54	0.0081	0.0027	0.0044	0.0043
Event 55	0.0036	0.0035	0.0036	0.0037
Event 56	0.0164	0.0167	0.0182	0.0194
Event 57	0.0077	0.0004	0.0010	0.0010
Event 58	0.0350	0.0173	0.0072	0.0061
Event 59	0.0156	0.0032	0.0011	0.0011
Event 60	0.0710	0.1673	0.1956	0.2029
Event 61	0.0016	0.0001	0.0008	0.0008
Event 62	0.0072	0.0035	0.0044	0.0044
Event 63	0.0032	0.0007	0.0009	0.0009
Event 64 (all wells dry)	0.0146	0.0343	0.0237	0.0219

References

Abbas AE (2006) Entropy methods for joint distributions in decision analysis. *IEEE Trans. Engrg. Management* 53(1):146–159.

Bélisle CJP, Romeijn HE, Smith RL (1993) Hit-and-run algorithms for generating multivariate distributions. *Math. Oper. Res.* 18(2):255–266.

Bertsekas DP (1999) *Nonlinear Programming*, 2nd ed. (Athena Scientific, Cambridge, MA).

Bertsimas D, Tsitsiklis J (1997) *Introduction to Linear Optimization* (Athena Scientific, Belmont, MA).

Bickel JE, Smith JE (2006) Optimal sequential exploration: A binary learning model. *Decision Anal.* 3(1):16–32.

Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, New York).

Clemen RT, Reilly T (1999) Correlations and copulas for decision and risk analysis. *Management Sci.* 45(2):208–224.

Delfiner P (2003) Modeling dependencies between geologic risks in multiple targets. SPA 82659-PA. *SPE Reservoir Eval. Engrg.* 6:57–64.

Devroye L (1986) *Non-Uniform Random Variate Generation* (Springer-Verlag, New York).

Dyer JS, Farrell W, Bradley P (1973) Utility functions for test performance. *Management Sci.* 20(4-Part I):507–519.

Ireland CT, Kullback S (1968) Contingency tables with given marginals. *Biometrika* 55(1):179–188.

Jaynes ET (1957) Information theory and statistical mechanics. Part 1. *Physical Rev.* 106(4):620–630.

Jaynes ET (1968) Prior probabilities. *IEEE Trans. Systems Sci. Cybernetics* 4(3):227–241.

Keefer DL (2004) Underlying event model for approximating probabilistic dependence among binary events. *IEEE Trans. Engrg. Management* 51(2):173–182.

Korsan RJ (1990) Towards better assessments and sensitivity procedures. Oliver RM, Smith JQ, eds. *Influence Diagrams, Belief Nets and Decision Analysis* (Wiley & Sons, New York), 427–454.

Kullback S, Leibler RA (1951) On information sufficiency. *Ann. Math. Statist.* 22(1):79–86.

Lovasz L (1998) Hit-and-run mixes fast. *Math. Programming* 86(3):443–461.

Lowell DG (1994) Sensitivity to relevance in decision analysis. Ph.D. thesis, Engineering and Economic Systems, Stanford University, Stanford, CA.

MacKenzie GR (1994) Approximately maximum entropy multivariate distributions with specified marginals and pairwise correlations. Ph.D. thesis, University of Oregon, Eugene.

McMullen P (1970) The maximum number of faces of a convex polytope. *Mathematika* 17(2):179–184.

Montiel LV (2012) Approximations, simulation, and accuracy of multivariate discrete probability distributions in decision analysis. Ph.D. thesis, University of Texas at Austin, Austin.

Montiel LV, Bickel JE (2012a) Generating a random collection of discrete joint probability distributions subject to partial information. *Methodology Comput. Appl. Probab.*, ePub ahead of print, May 22, http://link.springer.com/article/10.1007%2Fs11009-012-9292-9.

Montiel LV, Bickel JE (2012b) A simulation-based approach to decision making with partial information. *Decision Anal.* 9(4): 329–347.

Moore LR, Mudford BS (1999) Probabilistic play analysis from geoscience to economics: An example from the Gulf of Mexico. *SPE 52955, SPE Hydrocarbon Econom. Evaluation Sympos., Dallas*.

Moskowitz H, Sarin RK (1983) Improving the consistency of conditional probability assessments for forecasting and decision making. *Management Sci.* 29(6):735–749.

Sarin RK (1978) A sequential approach to cross-impact analysis. *Futures* 10(1):53–62.

Sarin RK (1979) An approach for long term forecasting with an application to solar electric energy. *Management Sci.* 25(6):543–554.

Schmidt BK, Mattheiss TH (1977) The probability that a random polytope is bounded. *Math. Oper. Res.* 2(3):292–296.

Sklar A (1959) Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 8:229–231.

Smith RL (1984) Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Oper. Res.* 32(6):1296–1308.

Stabell CB (2000) Alternative approaches to modeling risks in prospects with dependent layers. *SPE 63204, SPE Annual Tech. Conf. Exhibition, Dallas*.

Luis V. Montiel is a postdoctoral member of the Decision Analysis Research Group at the University of Texas at Austin. His main research interest is mathematical modeling for optimization under uncertainty, with a special interest in decision analysis and simulation learning for optimization. His current research is dedicated to providing a theoretical and practical framework for approximating joint distributions under partial information. He has a Ph.D. in operations research from the University of Texas, an M.S. in financial engineering from Columbia University, and an M.S. in management science and engineering from Stanford University.

J. Eric Bickel is an assistant professor in the Graduate Program in Operations Research, Department of Mechanical Engineering, at the University of Texas at Austin. In addition, he is a fellow in both the Center for International Energy and Environmental Policy and the Center for Petroleum Asset Risk Management. He holds an M.S. and Ph.D. from the Department of Engineering-Economic Systems at Stanford University. His research interests include the theory and practice of decision analysis and its application in the energy and climate-change arenas. His research has addressed the modeling of probabilistic dependence, value of information, scoring rules, calibration, risk preference, education, decision making in sports, and climate engineering as a response to climate change. Prior to joining the University of Texas at Austin, he was a senior engagement manager for Strategic Decisions Group. He is president-elect for the Decision Analysis Society of INFORMS and an associate editor for *Decision Analysis*.