Methodol Comput Appl Probab
DOI 10.1007/s11009-012-9292-9

Generating a Random Collection of Discrete Joint Probability Distributions Subject to Partial Information

Luis V. Montiel · J. Eric Bickel

Received: 4 October 2011 / Accepted: 7 May 2012

Abstract In this paper, we develop a practical and flexible methodology for generating a random collection of discrete joint probability distributions, subject to a specified information set, which can be expressed as a set of linear constraints (e.g., marginal assessments, moments, or pairwise correlations). Our approach begins with the construction of a polytope using this set of linear constraints. This polytope defines the set of all joint distributions that match the given information; we refer to this set as the "truth set." We then implement a Monte Carlo procedure, the Hit-and-Run algorithm, to sample points uniformly from the truth set. Each sampled point is a joint distribution that matches the specified information. We provide guidelines to determine the quality of this sampled collection. The sampled points can be used to solve optimization models and to simulate systems under different uncertainty scenarios.

Keywords Joint probability distributions · Simulation · Hit and run · Polytopes

AMS 2010 Subject Classifications 90C99 · 65C05 · 60H99

1 Introduction

Many decisions amenable to quantitative methods include significant uncertainty. In these cases, we often assume the underlying probability distribution is known. However, in many situations we have only partial information with which to constrain

L. V. Montiel · J. E. Bickel (✉)
Graduate Program in Operations Research/Industrial Engineering,
The University of Texas at Austin, 1 University Station C2200,
Austin, TX 78712, USA
e-mail: ebickel@mail.utexas.edu

L. V. Montiel
e-mail: lvmontiel@gmail.com, lvm92@mail.utexas.edu

Published online: 22 May 2012

this distribution. For example, we may know only the marginal distributions and some pairwise correlation coefficients. In these cases, the probability distribution is not unique, but rather is a member of a set of distributions with known properties. This uncertainty about the underlying distribution poses a significant challenge and solutions fall into two categories. First, robust optimization methods, seek a solution that is in some sense good over a wide range of possible scenarios (Ben-Tal and Nemirovski 2002). Second, approximation methods have been proposed, which produce a single joint distribution given partial information. The most popular of these approaches is maximum entropy (Jaynes 1957, 1968).

In this paper, we present a simulation procedure to create not one, but a collection of joint distributions uniformly sampled from a finite dimensional set consistent with the given information. Specifically, our procedure generates a collection of finite-dimensional, discrete, joint probability distributions whose marginals have finite support. This procedure can be used in conjunction with or as an alternative to the approximation and robust optimization methods discussed above.

As an example, consider a random vector $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$, with specified marginal distributions $F_i(X_i)$ and correlation matrix $\mathbf{\Sigma}_X$. There are an infinite number of joint distributions $G(\mathbf{X})$ that match these constraints. We refer to this set of distributions as the "truth set" ($\mathbb{T}$). By "truth" we mean that any distribution within this set is consistent with the stated constraints and therefore could be the true joint distribution. Our goal is to generate a collection of joint distributions $G_i(\mathbf{X})$, i = 1 to N, that are consistent with the given information, where N is the number of samples in our collection. As we detail below, we use the Hit-and-Run (HR) sampler to produce a collection of samples uniformly distributed in $\mathbb{T}$ (Smith 1984).

It is important to emphasize that the method we suggest here is fundamentally different from other methods of random variate generation such as NORTA (Ghosh and Henderson 2003) and chessboard techniques (Mackenzie 1994; Ghosh and Henderson 2001). These methods produce instantiations $\mathbf{x}$ of $\mathbf{X}$ based on a *single* distribution $G(\mathbf{X})$ that is consistent with a set of specified marginal distributions, correlation matrix, and in the case of NORTA, the assumption that the underlying dependence structure can be modeled with a normal copula. Thus, NORTA and the chessboard techniques produces random variates based on a single distribution contained within $\mathbb{T}$.

As the reader will see, in our discrete setting, we envision the sample space of $\mathbf{X}$ as being fixed and therefore seek to create a set of discrete *probabilities* that are consistent with the given information. Before proceeding, we should explain this focus on generating probabilities $G(\mathbf{X})$ rather than outcomes of $\mathbf{X}$. Within the decision analysis community, for example, the problem of specifying a probability distribution given partial information is well known (Jaynes 1968) and of great practical importance (Abbas 2006; Bickel and Smith 2006). For example, suppose one knows that the average number rolled on a six-sided die is 4.5. What probability should one assign to each of the six faces? As discussed above, one possible approach is the application of maximum entropy (Jaynes 1957, 1968). Maximum entropy would specify the (unique) probability mass function (pmf) that is closest to uniform, while having a mean of 4.5. The procedure described in this paper was originally developed to test the accuracy of maximum entropy and other approximations. Hence, it explores a larger number of probability distributions uniformly sampled from $\mathbb{T}$.

2 Literature Review

HR is an effective method to sample the interior of a polytope and is easy to implement. However, it is not the only possible sampling procedure. In the following, we present a brief review of alternative methods and discuss their shortcomings.

The first set of sampling procedures are acceptance-rejection methods (von Newmann 1963). These methods embed the region of interest S within a region D for which a uniform sampling algorithm is known. For example, one might embed S within the union of non-overlapping hyperrectangles or hyperspheres (Rubin 1984) and then uniformly sample from D, rejecting points that are not also in S. As pointed by Smith (1984), this method suffers from two significant problems as far as our work is concerned. First, embedding the region of interest within a suitable superset may be very difficult (Rubin 1984). Second, as the dimension of S increases, the number of rejections per accepted sample (i.e., the rejection rate) grows exponentially. For example, Smith (1984) shows that when S is a 100-dimensional hypercube and D is a circumscribed hypersphere, 10^{30} samples are required on average for every sample that is accepted. The polytopes that we consider are at least this large and more complex.

The second alternative, described by Devroye (1986), consists of generating random points within the polytope by taking random convex combinations of the polytope's vertices. This method is clearly infeasible for most problems of practical interest, since it requires specifying all of the polytope's vertices in advance. For high-dimensional polytopes this is very 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). While this is an upper bound, the number of vertices one is likely to encounter in real problems is still enormous (Schmidt and Mattheiss 1977).

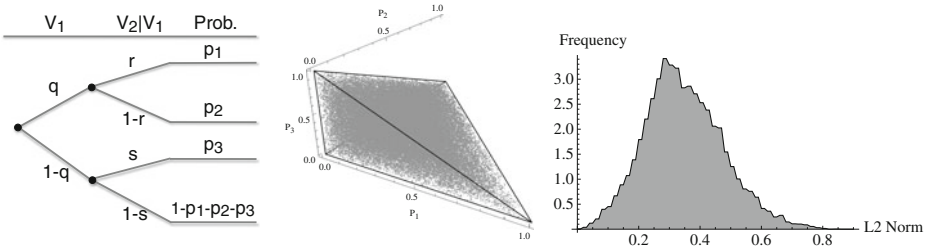
The final alternate method is based on decomposition, in which the area of interest is divided into non-overlapping segments for which uniform sampling is easy to perform. Again, this method requires full knowledge of all the extreme points of $\mathbb{T}$. Rubin (1984) provides a brief review of such methods and notes that they entail "significant computational overhead" and are practical only for low-dimensional polytopes.

HR is a random walk through the region of interest. As such, it avoids the problems detailed above since every sampled point is feasible and it does not require knowledge of the polytope's vertices. The drawback of this method is that the samples are only asymptotically uniformly distributed and it can take a large number of samples before the sample set is acceptably close to uniform (Rubin 1984; Smith 1984). We deal with this issue below.

This paper is organized as follows. Section 3 presents a motivational example to build the reader's intuition regarding the method we describe. Section 4 describes the general procedure for generating collection of joint distributions. Section 5 presents an illustrative example of the sampling procedure. Finally, in Section 6 we conclude and discuss future research.

3 Motivational Example

To illustrate and motivate our technique, suppose we are deciding whether or not to market a new product. At present, we are uncertain about our production costs and



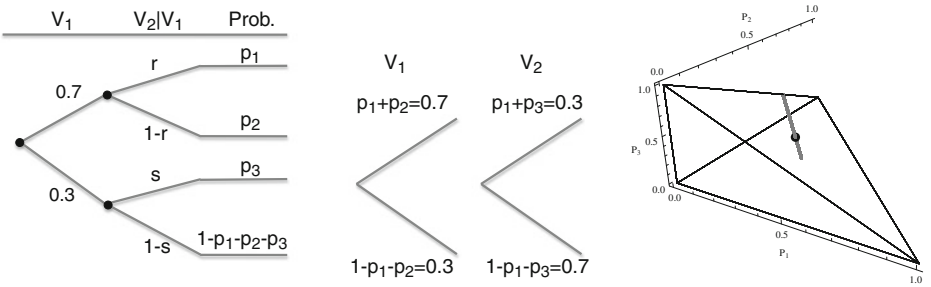
(a) Probability Tree. (b) Truth Set. (c) Euclidean Distance Distribution.

Fig. 1 Two binary variables with unknown information

whether or not a competitor will enter the market. Let V_1 represent the uncertainty regarding whether the competitor will enter the market ($V_1 = 1$) or not ($V_1 = 0$), and let V_2 represent our production costs being high ($V_2 = 1$) or low ($V_2 = 0$). Graphically, we can represent these scenarios using the binary probability tree in Fig. 1a.

We start by assuming we have no knowledge of the marginal distributions of V_1 and V_2 nor their dependence structure. In this case, $\mathbb{T}$ consists of all joint distributions of (V_1, V_2) with four outcomes $n = 4$ and probabilities $\mathbf{p} = (p_1, p_2, p_3, p_4)$. We can simplify the joint distributions using only three probabilities (p_1, p_2, p_3) since $p_4 = 1 - p_1 - p_2 - p_3$. The truth set, shown in Fig. 1b is a polytope in three dimensions and its vertices represent extreme joint distributions. In this case, the center of $\mathbb{T}$ is the joint pmf $\mathbf{p} = (0.25, 0.25, 0.25, 0.25)$, which assumes the random variables are independent and with uniform marginals. The small dots in Fig. 1b are samples (complete pmfs) generated by the HR procedure we describe below. By measuring the Euclidean distance ($L2$-Norm) from the center of $\mathbb{T}$ to all other joint distributions in $\mathbb{T}$ (Fig. 1c), we see that most of the samples are at 0.3 units from the center. The samples, which correspond to the distribution of volume in the truth set, are less concentrated close to the center and the corners.

Now, suppose we have information that there is a 70% chance the competitor will enter the market and that there is a 30% chance that production cost will be



(a) Probability Tree. (b) Marginal Variables. (c) New truth set $\mathbb{T}$.

Fig. 2 Two binary variables with marginal probability information

high (unconditioned on the entry of the competitor). The new information modifies the probability tree (Fig. 2a) and introduces two new constraints to restrict the joint distributions matching the marginal probabilities (Fig. 2b).

Each constraint is a hyperplane that cuts $\mathbb{T}$, reducing its dimension by one. As shown in Fig. 2c, $\mathbb{T}$ is now a line with extremes at (0.3, 0.4, 0.0) and (0.0, 0.7, 0.3). As a reference, the distribution that assumes independence (i.e., the maximum entropy approximation in this case) is located at (0.21, 0.49, 0.09) and is marked with a large black dot. Later, we will expand this example, increasing its dimension.

4 General Procedure

4.1 Problem Statement

The objective of our procedure is to create a collection of discrete joint probability distributions uniformly sampled from a finite dimensional, continuous, convex and compact set that contains all possible realizations of the joint distribution that are consistent with given information. We assume the joint distributions are discrete with finite support, as are the marginal distributions. To ensure that our truth set is convex, we only admit information that can be encoded with linear equality constraints. While this is certainty a limitation, we are still able to address a large class of problems that are of practical importance (Bickel and Smith 2006).

4.2 Notation

In order to describe the sampling procedure, we first establish the notation, which we illustrate with a simple example using a joint distribution with two variables; the first variable having "High," "Medium," and "Low" outcomes, and the second variable having "Up" and "Down" outcomes. In this case, we require one set including two random variables $\mathbb{V} = \{V_1, V_2\}$, plus two sets for the outcomes $\mathbb{O}^{V_1} = \{H, M, L\}$ and $\mathbb{O}^{V_2} = \{U, D\}$. Finally, we create a set including the cardinal product of all the outcomes:

$$\mathbb{U} = \{[H, U], [H, D], [M, U], [M, D], [L, U], [L, D]\}.$$

Additional sets are required to include more information. For example, sets that include joint outcomes where V_1 is set to "High." A formal definition of the notation is presented next.

Notation:

Indices and sets:

$\mathbb{I}$	Set of available information.
$\mathbb{V}$	Set of random variables.
$V_i \in \mathbb{V}$	Random variable i in $\mathbb{V}$.
$\mathbb{O}^{V_i}$	Set of possible outcomes for random variable V_i.
$\omega_r^{V_i} \in \mathbb{O}^{V_i}$	Realizations for random variable V_i indexed by $r = 1, 2 \ldots \lvert\mathbb{O}^{V_i}\rvert$.

$\mathbb{U}$	Set of all joint outcomes, $\mathbb{U} = \mathbb{O}^{V_1} \times \mathbb{O}^{V_2} \times \cdots \times \mathbb{O}^{V_{\lvert\mathbb{V}\rvert}}$.
$\omega_k \in \mathbb{U}$	Joint outcomes $\omega_k = \{\omega_r^{V_1}, \omega_s^{V_2}, \ldots, \omega_z^{V_{\lvert\mathbb{V}\rvert}}\}$ indexed by $k = 1, 2, \ldots \prod_{V_i} \lvert\mathbb{O}^{V_i}\rvert$.
$\mathbb{U}_{\omega_r^{V_i}}$	Set of joint outcomes for which random variable V_i obtains the value $\omega_r^{V_i}$.
$\mathbb{U}_{\omega_r^{V_i}\omega_s^{V_j}}$	Set of joint outcomes for which random variables V_i and V_j obtain values $\omega_r^{V_i}$ and $\omega_s^{V_j}$.

Data:

$q_{\omega_r^{V_i}}$	Probability that $V_i = \omega_r^{V_i}$.
$q_{\omega_r^{V_i}\omega_s^{V_j}}$	Probability that $V_i = \omega_r^{V_i}$ and $V_j = \omega_s^{V_j}$.
ρ_{V_i,V_j}	Moment correlation between V_i and V_j.
$\rho^r_{V_i,V_j}$	Rank correlation between V_i and V_j.
σ_{V_i,V_j}	Covariance between V_i and V_j.
$m^z_{V_i}$	The zth moment of random variable V_i.

Decision variables:

$\mathbf{p}$	Vector of decision variables defining the joint probability mass function.
$p_{\omega_k} \in \mathbf{p}$	Decision variables describing the probability of the joint event ω_k.

Table 1 applies the set notation to our example. The variables are V_1 and V_2 and their respective marginal outcomes are $\mathbb{O}^{V_1} = \{H, M, L\}$ and $\mathbb{O}^{V_2} = \{U, D\}$. $\omega_1^{V_1} = H$ is the first possible realization of V_1. The joint outcomes $\omega_k \in \mathbb{U}$ are defined as $\omega_1 = [H, U]$, $\omega_2 = [H, D]$, $\omega_3 = [M, U]$, $\ldots$, $\omega_6 = [L, D]$. The probabilities of these outcomes are $p_{\omega_1} = P(V_1 = H, V_2 = U)$, $p_{\omega_2} = P(V_1 = H, V_2 = D)$, $p_{\omega_3} = P(V_1 = M, V_2 = U)$, $\ldots$, $p_{\omega_6} = P(V_1 = L, V_2 = D)$.

We use dot-notation to marginalize the random variables. For example, $\mathbb{U}_{\omega_1^{V_1}} = \mathbb{U}_{H\cdot}$ and $\mathbb{U}_{\omega_2^{V_2}} = \mathbb{U}_{\cdot D}$, where "$\cdot$" implies marginalization over that random variable. Using the same index k for ω_k, we have $\mathbb{U}_{\cdot D} = \{\omega_2, \omega_4, \omega_6\}$.

The set of available information, denoted as $\mathbb{I} \equiv \{q_{\omega_r^{V_i}}, q_{\omega_r^{V_i}\omega_s^{V_j}}, \rho_{V_i,V_j}, \sigma_{V_i,V_j}, m^z_{V_i}\}$, includes all the information to be included in the model. $q_{\omega_r^{V_i}} \equiv P(V_i = \omega_r^{V_i})$ is

Table 1 Notation example

$\mathbb{V} = \{V_1, V_2\}$, $\mathbb{O}^{V_1} = \{H, M, L\}$, $\mathbb{O}^{V_2} = \{U, D\}$,
$\mathbb{U}_{\cdot\cdot} = \{[H, U], [H, D], [M, U], [M, D], [L, U], [L, D]\}$,
$\mathbb{U}_{H\cdot} = \{[H, U], [H, D]\}$, $\mathbb{U}_{\cdot U} = \{[H, U], [M, U], [L, U]\}$,
$\mathbb{U}_{M\cdot} = \{[M, U], [M, D]\}$, $\mathbb{U}_{\cdot D} = \{[H, D], [M, D], [L, D]\}$,
$\mathbb{U}_{L\cdot} = \{[L, U], [L, D]\}$,
$\mathbb{U}_{HU} = \{[H, U]\}$, $\mathbb{U}_{HD} = \{[H, D]\}$, $\mathbb{U}_{MU} = \{[M, U]\}$,
$\mathbb{U}_{MD} = \{[M, D]\}$, $\mathbb{U}_{LU} = \{[L, U]\}$, $\mathbb{U}_{LD} = \{[L, D]\}$.

the marginal distribution for variable V_i, and $q_{\omega_r^{V_i}\omega_s^{V_j}} \equiv P(V_i = \omega_r^{V_i}, V_j = \omega_s^{V_j})$ is the pairwise joint distribution for variables V_i, V_j. When marginal information is available, it is possible to describe the moment correlation ρ_{V_i,V_j}, the rank correlation $\rho^r_{V_i,V_j}$ and the covariance σ_{V_i,V_j} for the variables V_i and V_j. Additionally, if the marginals are unknown, we can make use of known moments $m^z_{V_i}$ to constrain the truth set. Our notation can be extended to more than two variables and to match three-way or four-way probabilities.

4.3 Constraints

We are now in position to constrain the truth set to match the information provided by $\mathbb{I}$. In this section, we present different families of equations that can be used to constrain $\mathbb{T}$, creating a system of m linear equations $\mathbf{Ap} = \mathbf{b}$ and n non-negativity constraints $\mathbf{p} \geq 0$, where $\mathbf{p} = \{p_{\omega_1}, p_{\omega_2}, \ldots, p_{\omega_{|\mathbb{U}|}}\}$. $\mathbf{A} \in \mathbb{R}^{m \times n}$ defines the properties we want to constrain and $\mathbf{b} \in \mathbb{R}^m$ represents the available information.

4.3.1 Matching Necessary Conditions

In all cases, the joint pmf must sum to one and each probability must be non-negative. We represent these constraints with Eqs. 1a and 1b.

$$\sum_{\omega_k \in \mathbb{U}} p_{\omega_k} = 1, \tag{1a}$$

$$p_{\omega_k} \geq 0, \qquad \forall \omega_k \in \mathbb{U}. \tag{1b}$$

Equations 1a and 1b give the necessary and sufficient conditions for $\mathbf{p}$ to be a pmf and are required in all cases. Notice that Eq. 1a reduces the dimension of the polytope $\mathbb{T}$ from n to $n-1$ and Eq. 1b limits $\mathbb{T}$ to positive quadrants. This constraint alone assures that $\mathbb{T}$ is a compact set.

4.3.2 Matching Marginal and Pairwise Probabilities

A second set of equations is used when we have information regarding the marginal and pairwise probabilities. Equation 2a requires that the joint probabilities match the marginal assessments. While, Eq. 2b requires that they match pairwise joint assessments.

$$\sum_{\omega_k \in \mathbb{U}_{\omega_r^{V_i}}} p_{\omega_k} = q_{\omega_r^{V_i}} \qquad \forall V_i \in \mathbb{V},\ \omega_r^{V_i} \in \mathbb{O}^{V_i}, \tag{2a}$$

$$\sum_{\omega_k \in \mathbb{U}_{\omega_r^{V_i}\omega_s^{V_j}}} p_{\omega_k} = q_{\omega_r^{V_i}\omega_s^{V_j}} \ \forall V_i, V_j \in \mathbb{V}, (\omega_r^{V_i}, \omega_s^{V_j}) \in \mathbb{O}^{V_i} \times \mathbb{O}^{V_j}. \tag{2b}$$

Equations 2a and 2b can be extended to cover three-way, four-way, or higher-order joint probability information.

4.3.3 Matching Moments

If the outcomes can be represented as numerical values, instead of categorical data, we can match moment information using Eq. 3.

$$\sum_{\omega_r^{V_i} \in \mathbb{O}^{V_i}} \left(\omega_r^{V_i}\right)^z \cdot \sum_{\omega_k \in \mathbb{U}_{\omega_r^{V_i}}} p_{\omega_k} = m^z_{V_i} \quad \forall\, V_i \in \mathbb{V}. \tag{3}$$

Equation 3 matches the zth moment of variable V_i. For $z = 1$ we can match the expected value of V_i and for $z = 2$ we can match the second raw moment. We note that $z = 0$ is simply a restatement of Eq. 1a and the requirement that the probabilities sum to one. Recall the outcomes $\omega_r^{V_i}$ are known, so the constraint is linear in the joint probabilities p_{ω_k}.

4.3.4 Matching Covariance and Correlation

If the first moments for variables V_i and V_j are known, it is possible to restrict the joint distribution to match a given covariance σ_{V_i,V_j}. Moreover, we can match the correlation if the variances for V_i and V_j are also known. Equations 4a and 4b match the covariance and moment correlation respectively.

$$\sum_{\omega_r^{V_i} \in \mathbb{O}^{V_i}} \sum_{\omega_s^{V_j} \in \mathbb{O}^{V_j}} \omega_r^{V_i} \cdot \omega_s^{V_j} \sum_{\omega_k \in \mathbb{U}_{\omega_r^{V_i} \omega_s^{V_j}}} p_{\omega_k} = \sigma_{V_i,V_j} + m^1_{V_i} \cdot m^1_{V_j} \quad \forall\, V_i, V_j \in \mathbb{V}, \tag{4a}$$

$$\sum_{\omega_r^{V_i} \in \mathbb{O}^{V_i}} \sum_{\omega_s^{V_j} \in \mathbb{O}^{V_j}} \omega_r^{V_i} \cdot \omega_s^{V_j} \sum_{\omega_k \in \mathbb{U}_{\omega_r^{V_i} \omega_s^{V_j}}} p_{\omega_k} = \rho_{V_i,V_j} \cdot \sqrt{\sigma^2_{V_i} \cdot \sigma^2_{V_j}} \quad \forall\, V_i, V_j \in \mathbb{V}. \tag{4b}$$

where $\sigma^2_{V_i}$ is the variance of variable V_i, and ρ_{V_i,V_j} is the moment correlation of variables V_i and V_j.

4.3.5 Matching Spearman's Correlation Coefficient

Another measure of variation that requires less information, and can be used with numerical as well as categorical information, is the rank correlation. The rank correlation is defined as

$$\rho^r_{V_i,V_j} = \frac{Cov\left(P\left(V_i \le \omega_r^{V_i}\right), P\left(V_j \le \omega_s^{V_j}\right)\right)}{\sqrt{Var\left(P\left(V_i \le \omega_r^{V_i}\right)\right) \cdot Var\left(P\left(V_j \le \omega_s^{V_j}\right)\right)}}. \tag{5}$$

Unlike the Pearson product-moment correlation, rank correlation is invariant with respect to the marginal outcomes. This and other characteristics make it a reliable measure of association (for more on rank correlation and assessment methods see Clemen and Reilly 1999 and Clemen et al. 2000). Rank correlation only requires information regarding the marginal probabilities for V_i and V_j and can be described as a linear function as follows.

Let $H : \overline{\mathbb{R}}^2 \to \overline{\mathbb{R}}$ be a two-place real function, and let $\mathbf{B} = [x_1, x_2] \times \left[y_1, y_2\right]$ be a rectangle whose vertices are in the domain of H. According to Nelsen (2005), the *H-volume* is defined as

$$\mathbf{V}_H[\mathbf{B}] = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1). \tag{6}$$

Let $\omega_k^+(V_i)$ be the outcome $\omega_r^{V_i}$ of variable V_i at the joint outcome ω_k and let $\omega_k^-(V_i)$ be the outcome $\omega_{r-1}^{V_i}$ of V_i. Notice that $\omega_{r-1}^{V_i}$ is the outcome that precedes $\omega_r^{V_i}$ in the marginal distribution of V_i. The cumulative probabilities that V_i is less than the outcomes $\omega_k^+(V_i)$ and $\omega_k^-(V_i)$ are $p_k^+(V_i) = P(V_i \leq \omega_k^+(V_i))$ and $p_k^-(V_i) = P(V_i \leq \omega_k^-(V_i))$, respectively. These cumulative probabilities define the interval $I_{\omega_k}(V_i)$ as follows

$$I_{\omega_k}(V_i) \equiv \left[p_k^+(V_i),\, p_k^-(V_i)\right]. \tag{7}$$

Using the intervals $I_{\omega_k}(V_i)$ and $I_{\omega_k}(V_j)$, we can define a rectangular area $I_{\omega_k}(V_i) \times I_{\omega_k}(V_j)$ equivalent to $\mathbf{B}$. Then, using the *H-volume* we can define the rank correlation between V_i and V_j as

$$\sum_{\omega_k \in \mathbb{U}} p_{\omega_k} \frac{\mathbf{V}_{x^2 * y^2}\left[I_{\omega_k}(V_i) \times I_{\omega_k}(V_j)\right]}{q_{\omega_k^+(V_i)} \cdot q_{\omega_k^+(V_j)}} = \frac{\rho^r_{V_i,V_j} + 3}{3}, \tag{8}$$

where $q_{\omega_k^+(V_i)} = P(V_i = \omega_k^+(V_i))$, which is the marginal probability of variable V_i having the outcome $\omega_r^{V_i}$ at the joint outcome ω_k. Additionally, the *H-volume* $\mathbf{V}_H$ is as defined for $H = x^2 \cdot y^2$, where $x \in I_{\omega_k}(V_i)$ and $y \in I_{\omega_k}(V_j)$.

It is important to recall that, the rank correlation ρ^r is bounded by a scalar such that $|a_{\hat{m}}| < 1$, where $\hat{m}$ is the maximum number of possible outcomes of variables V_i and V_j. The bounds were proven by Mackenzie (1994) for uniform discrete distributions. Mackenzie (1994) also proves that $\lim_{\hat{m}\to\infty} |a_{\hat{m}}| = 1$, meaning that using more outcomes in each marginal distribution provides a more refined rank correlation bounded by $[-1, 1]$.

4.4 Sampling Procedure

After characterizing the truth set $\mathbb{T}$, the next step uses the HR sampler (Smith 1984) to uniformly sample distributions from $\mathbb{T}$. The HR procedure is the fastest known algorithm to sample the interior of an arbitrary polytope. The algorithm has been proven to mix in $\mathbf{O}(h^3)$ time, where $h = (n - m)$ is the dimension of the polytope. Although the mixing time is polynomial, as discussed above, the number of samples required to guarantee convergence to the uniform distribution can be large (Lovasz 1998). To overcome this problem, in the following sections we propose a practical definition for convergence that reduces the number of samples required to create a discrete representation of the truth set.

4.4.1 Hit-and-Run Sampler

The algorithm is described below and illustrated in two dimensions in Fig. 3.

Step 1: Set $i = 0$ and select an arbitrary point $x_i \in \mathbb{T}$.
Step 2: Generate a set $D \subseteq \mathbb{R}^n$ of directions.
Step 3: Choose a random direction d_i uniformly distributed over D.
Step 4: Find the line set $L = \mathbb{T} \cap \{x | x = x_i + \lambda d_i, \lambda \text{ a real scalar}\}$.
Step 5: Generate a random point uniformly distributed over $x_{i+1} \in L$.
Step 6: If $i = N$, stop. Otherwise, set $i = i + 1$ and return to Step 2.

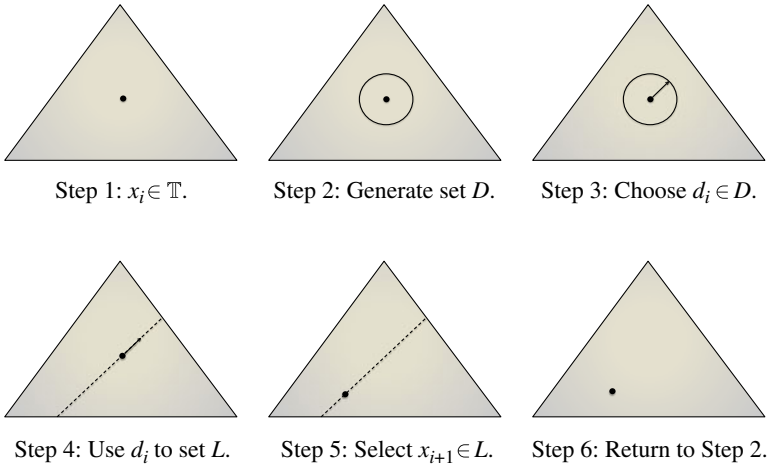


Fig. 3 Hit and run sampler. Illustration of the algorithm in $2d$

The HR was designed for full-dimensional polytopes, however, with minor modifications it can be adapted to sample efficiently from non-full-dimensional sets. These modifications are presented in the following section.

4.4.2 Sampling Non-full-dimensional Polytopes

As we noted in Section 4.3, the characterization of $\mathbb{T}$ describes the polytope as a system of m linear equations and n non-negative variables: $\mathbf{Ap} = \mathbf{b},\ \mathbf{p} \geq 0$. The HR sampler is designed to sample points in full-dimensional polytopes. However, the polytope $\mathbb{T}$ is not full-dimensional since $h = n - m < n$. To overcome this problem, we find the projection of $\bar{\mathbf{p}} \in \mathbb{R}^n$ into the hyperplane $\mathbf{Ap} = \mathbf{b}$ using Eq. 9, where $\mathbf{I}$ represents the identity matrix.

$$\mathbf{p} = (\mathbf{I} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A})\bar{\mathbf{p}} + \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b} \tag{9}$$

Then, we can create an hypersphere $D \in \mathbb{R}^n$ in the full-dimensional space by sampling independent vectors of size n from the multivariate standard normal and normalizing them so that each vector has equal magnitude. Using Eq. 9, we can project the direction set D into $\mathbb{T}$. With the proper scaling, the final result is a set of directions $D \in \mathbb{T}$ from which we can select directions uniformly distributed. The line L is created by extending the directions $\pm\mathbf{d}_i \in D$ until $\mathbf{p} \geq 0$ is violated. The rest of the implementation is straightforward.

This step removes the non full dimensional problem by reducing the dimension from n to $n - m$ for all the steps that require it. It is now possible to treat $\mathbb{T}$ as a full-dimensional polytope in $n - m$ dimensions.

4.5 Stopping Time

HR guarantees that the sampled collection eventually converges to the uniform distribution over $\mathbb{T}$ (Smith 1984). However, as pointed by Rubin (1984), the theoretical number of required samples to reach this convergence can be large. Yet, as we show

in this section, the number of samples required to achieve reasonable performance in practical applications in generally much smaller.

Measuring the rate of convergence to the uniform distribution, even in low-dimensional polytopes, is very difficult. Uniformity would imply that any possible partition of $\mathbb{T}$ contains a fraction of samples that is proportional to that partition's volume divided by the volume of the polytope. Computing the volume of arbitrary polytopes is a difficult task (Bárány and Füredi 1987). In fact, in many cases, the volume of the polytope can only be approximated by a random walk through the polytope (Dyer et al. 1991; Kannan et al. 1996), a procedure similar to HR. Therefore, we propose a measure of convergence that does not directly rely on global properties of the polytope and is easy to compute.

We begin by noting that for $\mathbf{p}_i$, a random vector sampled from $\mathbb{T}$ using HR, there exists unique vectors $\boldsymbol{\mu} = \{\mu_1, \ldots, \mu_n\}$ and $\boldsymbol{\sigma^2} = \{\sigma_1^2, \ldots, \sigma_n^2\}$ such that $\lim_{N\to\infty} \sum_{i=1}^{N} \frac{\mathbf{p}_i}{N} = \boldsymbol{\mu}$ and $\lim_{N\to\infty} \sum_{i=1}^{N} \frac{(\mathbf{p}_i - \boldsymbol{\mu})^2}{N-1} = \boldsymbol{\sigma^2}$, where all calculations over $\mathbf{p}_i$ are performed element-wise. Recall that since $\mathbf{p}_i$ has bounded support and the HR assures convergence in distribution, all the moments must converge (Casella and Berger 2002, p. 65). As discussed below, we measure convergence of HR by measuring the convergence of the sample mean and variance. These moments are of particular interest due to their intuitive interpretation. The sample mean describes how close the center of the collection is to the center of $\mathbb{T}$. The variance describes how the dispersion of the samples matches the dispersion of the $\mathbb{T}$'s volume. Hence, we now propose the following definitions for what we term *"fair-convergence"*.

Definition 1 A collection of joint distributions of size N is called *fair-in-mean*, if the average vectors of the joint distributions in a collection for the first $\frac{N}{2}$ and N samples of the HR algorithm are within an ε-ball of diameter α.

Definition 2 A collection of joint distributions of size N is called *fair-in-dispersion*, if the standard deviation vectors of the joint distributions in a collection for the first $\frac{N}{2}$ and N samples of the HR algorithm are within an ε-ball of diameter β.

Definition 3 A collection of joint distributions of size N is called *fair*, if it is fair-in-mean and fair-in-dispersion for selected small parameters $\alpha, \beta > 0$.

We implemented these definitions using Eqs. 10 and 11, where $\mathbf{p}_i$ is the ith sampled discrete probability distribution with n joint elements. To make notation easier we use $\mathbf{p}_i$ and assume all calculations are performed element-wise except for $||\cdot||_2$. Equation 10 computes the average of the collection sampled after N iterations $(\sum_{i=1}^{N} \frac{\mathbf{p}_i}{N})$ and compares it to the average after $\frac{N}{2}$ iterations. If after N iterations the vector of averages is within an ε-*ball* if diameter α of the previous vector (for some small $\alpha > 0$), we assume the sample is fair-in-mean.

Equation 11 is the equivalent version for the variance, where

$$\left(\sum_{j=1}^{N} \mathbf{p}_i - \mathbf{p}_j\right)^2 = \left(N\mathbf{p}_i - \sum_{j=1}^{N} \mathbf{p}_j\right)^2 = N^2 \left(\mathbf{p}_i - \frac{\sum_{j=1}^{N} \mathbf{p}_j}{N}\right)^2 = N^2(\mathbf{p}_i - \boldsymbol{\mu})^2,$$

and where $\boldsymbol{\mu}$ is the vector of averages for each joint element of the sample. In a similar way, if after N iterations the new vector of variances is within a ε-*Ball* of

diameter β of the previous vector (for some small $\beta > 0$), we assume the sample is fair-in-dispersion.

$$\left\| \sum_{i=1}^{N} \frac{\mathbf{p}_i}{N} - \sum_{i=1}^{\frac{N}{2}} \frac{2 \cdot \mathbf{p}_i}{N} \right\|_2 \leq \alpha \tag{10}$$

$$\left\| \sum_{i=1}^{N} \frac{\left(\sum_{j=1}^{N} \mathbf{p}_i - \mathbf{p}_j\right)^2}{N-1} - \sum_{i=1}^{\frac{N}{2}} \frac{4 \cdot \left(\sum_{j=1}^{\frac{N}{2}} \mathbf{p}_i - \mathbf{p}_j\right)^2}{\left(\frac{N}{2}-1\right)} \right\|_2 \leq \beta \cdot N^2 \tag{11}$$

The implementation of this stopping time for the mean can be performed by keeping track of $\sum_{i=1}^{N} \mathbf{p}_i$ at each iteration and dividing it by the number of samples at each check point. Additionally, using the recursion in Eq. 12, we can also keep track of the variance of each joint element at each iteration.

$$(i-1)\sigma^2_{w,(i)} = (i-2)\sigma^2_{w,(i-1)} + \left(\frac{f(i)}{i} \cdot \frac{f(i)}{i-1}\right), \tag{12a}$$

$$f(i) = \sum_{j=1}^{i-1} p^j_w - (i-1)p^i_w \quad \forall\, i = 2, 3, \ldots, N.$$

$$\sigma^2_{w,(1)} = 0, \quad \forall\, w \in \{1, 2, \ldots, n\}. \tag{12b}$$

Our experience suggests that the number of samples required for fair convergence is considerably smaller (reductions are in the order of $\approx 10^9$) than the Lovasz (1998) theoretical lower bound. As an example, Fig. 4 provides illustrative results for fair convergence in six unconstrained polytopes (Eq. 1 only) of different dimensions.

If $\mathbb{T}$ is unconstrained ($h = n - 1$), the truth set is symmetric and the center of mass of $\mathbb{T}$ is known to be the discrete uniform distribution. Therefore, we can test the convergence of HR by starting the algorithm at a point close to a corner and monitor the number of samples needed for the mean to arrive within an ε-Ball of radius $\alpha > 0$ with center at the discrete uniform distribution. For these collections, the algorithm will stop once the sample is fair-in-mean, and using Eq. 11 we check for fair-in-dispersion. This is a strong test because we are selecting the worst possible point to initialize the algorithm.

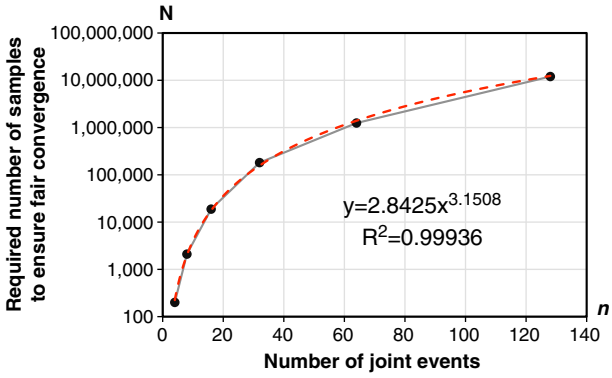


Fig. 4 Minimum required number of samples (N) to generate a fair sample vs the number of events (n) in the distribution for six unconstrained polytopes. A *solid line* connects the empirical data and the *dashed line* present the best fit

In particular, we initialize the algorithm by measuring the distance from the center to a corner of $\mathbb{T}$: $\sqrt{\frac{n-1}{n}}$. We then use $\delta = \frac{1-\tau}{n}$ and $\tau = .9$ to define the initial point $\mathbf{p}_0 = \{1 - \delta \cdot (n-1), \delta, \ldots, \delta\}$ where $||\mathbf{p}_0 - \mathbf{p}_\infty||_2 = \tau\sqrt{\frac{n-1}{n}}$. After the initial point is set, we look for the smallest N such that $||\sum_{i=1}^{N} \frac{\mathbf{p}_i}{N} - \mathbf{p}_\infty||_2 < \alpha = \varphi\sqrt{\frac{n-1}{n}}$ for $\varphi = 0.05$. Finally, we check for convergence every $K = 100$ iterations. For the sample sizes proposed the collections are also fair-in-dispersion.

5 Illustrative Example

We now demonstrate the procedure by applying it to an extension of the motivational example introduced earlier. Recall that we are unsure if a competitor will enter the market (V_1) and about our production costs (V_2). Additionally, assume we are in negotiations to acquire another company that would allow us to add a new

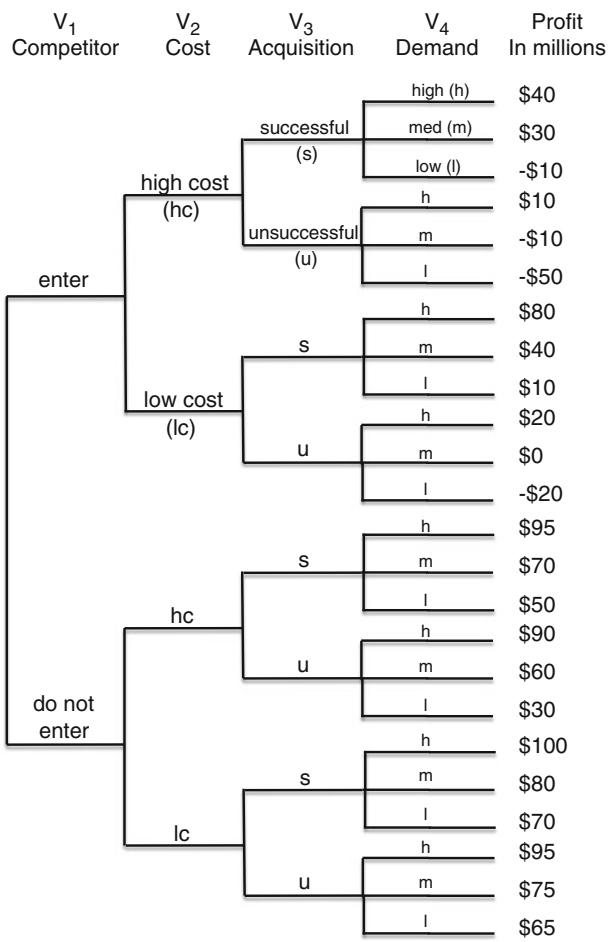


Fig. 5 Possible profits of the economic model

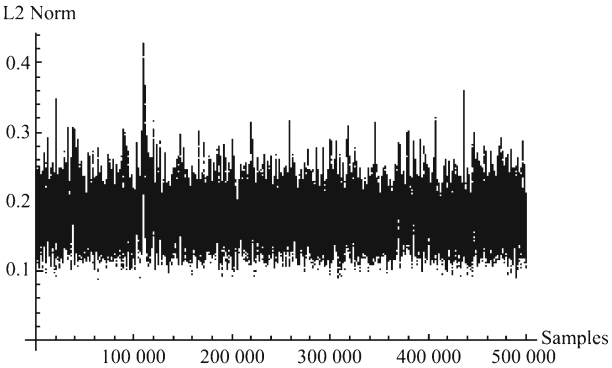


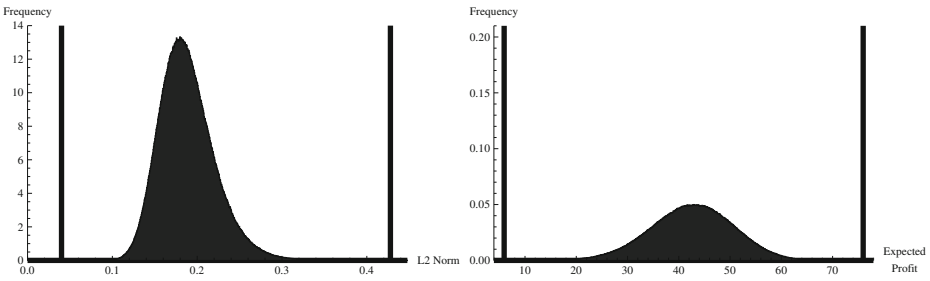
Fig. 6 Euclidean distance from each sampled point to center

product feature, boosting sales. The acquisition will either be "successful" ($V_3 = s$) or "unsuccessful" ($V_3 = u$). In addition, we are uncertain about the demand for our product (V_4), which could be "high", "medium", or "low". In this case, there are 24 possible scenarios. Assume the company has estimated the profit associated with each scenario, which is shown in Fig. 5.

We simulate 10 million joint pmfs, under an assumption of no information regarding the likelihood of the various scenarios. For each scenario, we measure the Euclidean distance (i.e., the $L2$-Norm) from the sample to the center of $\mathbb{T}$, which we take to be the mean of all sampled distributions. Figure 6 shows a scatterplot for a portion of the simulation results, where each point is a possible true joint pmf given the state of information.

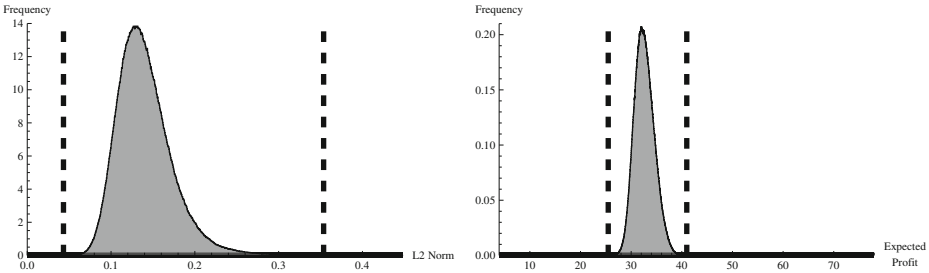
Figure 7a presents the distribution of the $L2$-Norm, where the solid lines mark the extremes in the sample. As in Section 3, we observe that the samples are less concentrated near the center and the corners of $\mathbb{T}$ and more concentrated at distances where the volume is more abundant. For each sampled joint distribution, we can calculate the expected profits of our model. Figure 7b presents the distribution of the expected profit when no information is available.

Now, suppose our research team has assessed the marginal probabilities for the four random variables. Specifically, assume the probability that the competitor will enter is 70%, the probability that production cost will be high is 30%, the probability that the acquisition will be successful is 35%, and the probabilities for the demand



(a) Euclidean distance pdf. (b) Expected profit pdf (in millions).

Fig. 7 Simulation results for a set with no partial information



(a) Euclidean distance pdf. (b) Expected profit pdf (in millions).

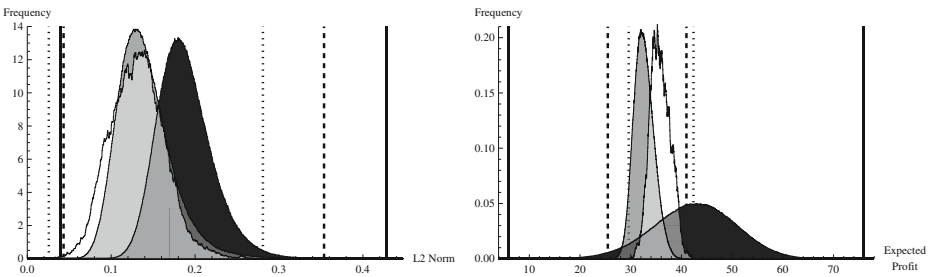
Fig. 8 Simulation results for a set with marginal assessments

being high, medium, and low are 50, 20 and 30%, respectively. The new results for the distributions of the Euclidean distance ($L2$-Norm) and the expected profit are shown in Fig. 8a and b, respectively. Again, the dashed lines mark the extremes of the sample.

Figure 8a shows a slight reduction in the distance, compared to Fig. 7a, from the center to the sampled points. This is the result of constraining the truth set with new information regarding the marginal probabilities. Figure 8b shows a substantial reduction in the possible range for the expected profit.

Taking our example one step further, assume now that our research team has determined that the rank correlations between V_1 (entry of the competitor) and V_2 (prodution cost), V_1 and V_4 (demand), and V_2 and V_4 are -0.2, 0.3, and 0.2, respectively. Figure 9a and b show the results of the distributions for the Euclidean distance and the expected profits for the three scenarios: no information (black), only marginal assessments (grey), and marginal assessments and three rank correlations (white).

The extremes of each case are shown with vertical solid (no information), dashed (marginals only), and dotted lines (marginals with three pairwise correlations). We see that additional information further constrains the truth set and reduces the range for the possible expected profits.



(a) Euclidean distance pdf. (b) Expected profit pdf (in millions).

Fig. 9 Simulation results for three information scenarios

6 Conclusions

The general simulation procedure we have described provides a flexible and powerful tool to analyze stochastic models when the joint distribution is incompletely specified. The methodology is easy to implement, develops a collection of joint distributions, and represents an alternative to previous approaches such as robust optimization and approximations such as maximum entropy. We demonstrated the procedure with a simple example based on marginal and pairwise rank correlation coefficients. The methodology can be extended to higher number of random variables, random variables with more than three possible outcomes, and higher-order conditioning such as three-way assessments. Future research will explore different applications of this sampling methodology and test the accuracy of existing distribution approximations such as maximum entropy, among others.

Appendix A: Spearman's Correlation

The derivation of Eq. 8 starts from basic principles as follows:

$$\rho^r_{V_i,V_j} + 3 = \frac{Cov(F_{V_i}, F_{V_j})}{\sqrt{Var(F_{V_i}) * Var(F_{V_j})}} + 3$$

$$= \frac{E(F_{V_i}, F_{V_j}) - \frac{1}{4}}{\frac{1}{12}} + 3 = 12 \cdot E(F_{V_i}, F_{V_j}) \tag{13a}$$

$$= 12 \int_0^1 \int_0^1 V_i \cdot V_j \cdot c(V_i, V_j)\, dV_i\, dV_j$$

$$= 12 \sum_{\omega_k \in \mathbb{U}} c_{\omega_k}(V_i, V_j) \cdot \int_{p_k^-(V_i)}^{p_k^+(V_i)} \int_{p_k^-(V_j)}^{p_k^+(V_j)} V_i \cdot V_j\, dV_i\, dV_j \tag{13b}$$

$$= 3 \sum_{\omega_k \in \mathbb{U}} c_{\omega_k}(V_i, V_j) \Big[\left[p_k^+(V_i) p_k^+(V_j)\right]^2 - \left[p_k^+(V_i) p_k^-(V_j)\right]^2$$

$$- \left[p_k^-(V_i) p_k^+(V_j)\right]^2 + \left[p_k^-(V_i) p_k^-(V_j)\right]^2 \Big] \tag{13c}$$

$$= 3 \sum_{\omega_k \in \mathbb{U}} c_{\omega_k}(V_i, V_j) \mathbf{V}_{x^2 * y^2} \left[I_{\omega_k}(V_i) \times I_{\omega_k}(V_j) \right]$$

$$= 3 \sum_{\omega_k \in \mathbb{U}} p_{\omega_k} \frac{\mathbf{V}_{x^2 * y^2} \left[I_{\omega_k}(V_i) \times I_{\omega_k}(V_j) \right]}{q_{\omega_k^+(V_i)} \cdot q_{\omega_k^+(V_j)}}. \tag{13d}$$

In Eq. 13a Transitions are given by $F_{V_i} \sim U[0, 1]$, for which mean and variance are well known.

In Eq. 13b We expand the expectation, partition the integrals in rectangles, and sets $c_{\omega_k}(V_i, V_j)$ as constant inside each rectangle (Mackenzie 1994).

In Eq. 13c We solve the integrals and evaluate each rectangle area.

In Eq. 13d We use the definition for $H\text{-}Volume$ and take $c_{\omega_k}(V_i, V_j)$ as in Eq. 14.

For each rectangle area in **B**, the probability p_{ω_k} is the volume of a body with base area $q_{\omega_k^+(V_i)} \cdot q_{\omega_k^+(V_j)}$ and height $c_{\omega_k}(V_i, V_j)$. Hence, we have:

$$c_{\omega_k}(V_i, V_j) = \frac{p_{\omega_k}}{q_{\omega_k^+(V_i)} \cdot q_{\omega_k^+(V_j)}}. \tag{14}$$

Acknowledgement This material is based upon work supported by the National Science Foundation under CAREER Grant No. SES-0954371.

References

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

Bárány I, Füredi Z (1987) Computing the volume is difficult. Discrete Comput Geom 2(1):319–326

Ben-Tal A, Nemirovski A (2002) Robust optimization—methodology and applications. Math Program 92:453–480

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

Casella G, Berger RL (2002) Statistical inference, 2nd edn. Duxbury Advanced Series, Thomson Learning, California

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

Clemen RT, Fischer GW, Winkler RL (2000) Assessing dependence: some experimental results. Manage Sci 46(8):1100–1115

Devroye L (1986) Non-uniform random variate generation. Springer-Verlag, New York

Dyer M, Frieze A, Kannan R (1991) A random polynomial time algorithm for approximating the volume of convex bodies. Journal of the ACM 38(1):1–17

Ghosh S, Henderson G (2001) Chessboard distributions and random vectors with specified marginals and covariance matrix. Oper Res 50(5):820–834

Ghosh S, Henderson SG (2003) Behavior of the NORTA method for correlated random vector-generation as the dimension increases. ACM Trans Model Comput Simul 13(3):276–294

Jaynes ET (1957) Inform. theory and statist. mechanics. part 1. Phys Rev 106(4):620–630

Jaynes ET (1968) Prior probabilities. IEEE Trans Syst Sci Cybern 4(3):227–241

Kannan R, Lovsz L, Simonovits M (1996) Random walks and an O*(n^5) volume algorithm for convex bodies. Random Struct Algorithms 11(1):1–50

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

Mackenzie GR (1994) Approximately maximum entropy multivariate distributions with specified marginals and pairwise correlations. Disertation, Univ. of Oregon, Oklahoma

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

Nelsen RB (2005) An introduction to Copulas, 2nd edn. Springer Series in Statist., Springer

von Newmann J (1963) Various techniques used in connection with random digits. Collected Works, vol 5. Pergamon Press, New York, NY, pp 768–770

Rubin PA (1984) Generating random points in a polytope. Commun Stat Simul Comput 13(3): 375–396

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

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