Proceedings of the 1994 Winter Simulation Conference
ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila

COMPOSITION FOR MULTIVARIATE RANDOM VARIABLES

Raymond R. Hill

AFSAA/SAGW
1570 A.F. Pentagon
Washington, D.C., 20330, U.S.A.

Charles H. Reilly

Department of Industrial, Welding
and Systems Engineering
The Ohio State University
1971 Neil Avenue
Columbus, Ohio, 43210, U.S.A.

ABSTRACT

We show how to find mixing probabilities, or weights, for composite probability mass functions (pmfs) for k-variate discrete random variables with specified marginal pmfs and a specified, feasible population correlation structure. We characterize a joint pmf that is a composition, or mixture, of 2^{k-1} extreme-correlation joint pmfs and the joint pmf under independence. Our composition method is also valid for multivariate continuous random variables. We consider the cases where all of the marginal distributions are discrete uniform, negative exponential, or continuous uniform.

1 INTRODUCTION

We consider the problem of generating samples of a k-variate discrete random variable via composition when the marginal probability mass functions (pmfs) and a feasible population correlation structure are specified. We are interested in this problem mainly because we want to be able to generate coefficients for synthetic optimization problems in which the dependence between each pair of coefficient types is controlled.

Many computational evaluations of solution procedures are conducted exclusively on synthetic optimization problems whose coefficients are generated independently. Results from other computational studies indicate that the statistical properties of the coefficients in synthetic optimization problems, e.g., the marginal distribution families and the population correlation structure, can affect the performance of solution methods (Loulou and Michaelides, 1979; Martello and Toth 1979, 1988; Balas and Martin, 1980; Balas and Zemel, 1980; Potts and Van Wassenhove, 1988; John, 1989; Moore, 1989; Reilly, 1991; Pollock, 1992; Rushmeier and Nemhauser, 1993; Moore and Reilly, 1993).

Our primary goal is to show how synthetic optimization problems with a prescribed population correlation structure among the coefficient types can be generated via composition. But, our work is not relevant to only one application. Our composition approach can be used for continuous random variables also, making this approach useful for many simulation applications. For example, the generation of values of multivariate continuous random variables can be crucial to realistic simulation models of manufacturing systems.

In this paper, we extend recent work on bivariate composite pmfs to the multivariate case. We pay particular attention to the case where all of the marginal pmfs are uniform because the discrete uniform distribution is very often used to represent the marginal distributions of coefficient values in synthetic optimization problems. We point out that our composition approach is valid for multivariate continuous random variables. We also consider the cases where all of the marginal distributions are negative exponential or continuous uniform because these cases may be important for simulations of tandem queueing systems and manufacturing systems.

2 BACKGROUND

In this section, we review the concepts of conventional mixtures, extreme mixtures, and parametric mixtures for bivariate discrete random variables (Y_1, Y_2). We also discuss methods for generating values of a multivariate random variable $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_k)$.

In §2.1 - 2.3, we assume that Y_i, $i = 1, 2$, is a finite discrete random variable distributed over the support $S_i = \{y_{i1}, y_{i2}, \ldots, y_{in_i}\}$ according to the pmf $f_i(y_i)$. We denote the maximum and minimum possible values of the Pearson product-moment correlation, $\rho = \text{Corr}(Y_1, Y_2)$, as ρ^+ and ρ^-, respectively. Also, we denote the minimum- and maximum-correlation pmfs for (Y_1, Y_2) as $g_1(y_1, y_2)$ and $g_2(y_1, y_2)$. We let ρ^0 be a specified value of ρ.

2.1 Conventional Mixtures for $(\mathbf{Y_1}, \mathbf{Y_2})$

We can generate values of (Y_1, Y_2) with population correlation ρ^0 such that $\rho^- \leq \rho^0 \leq \rho^+$ by mixing values of (Y_1, Y_2) generated under independence and values of (Y_1, Y_2) generated with extreme correlation. The pmf for (Y_1, Y_2) upon which such a generation method is based is

$$\left(1 - \frac{\rho^0}{\rho^-}\right) f_1(y_1) f_2(y_2) + \left(\frac{\rho^0}{\rho^-}\right) g_1(y_1, y_2), \quad (1)$$

if $\rho^0 < 0$, and

$$\left(1 - \frac{\rho^0}{\rho^+}\right) f_1(y_1) f_2(y_2) + \left(\frac{\rho^0}{\rho^+}\right) g_2(y_1, y_2), \quad (2)$$

if $\rho^0 \geq 0$. We refer to pmfs (1) and (2) as conventional mixtures.

Devroye (1986) defines a family of distributions for a bivariate random variable to be *comprehensive* if the family includes $f_1(y_1)f_2(y_2)$, $g_1(y_1, y_2)$, and $g_2(y_1, y_2)$. So conventional mixtures constitute a comprehensive family for (Y_1, Y_2).

Conventional mixtures are easy to use. There is a unique pmf (1) or (2) for every feasible value of ρ. As a result, when a conventional mixture is used to represent the distribution of (Y_1, Y_2) with $\rho^0 = 0$, Y_1 and Y_2 are independent. Conventional mixtures are useful for bivariate continuous random variables too.

2.2 Extreme Mixtures for $(\mathbf{Y_1}, \mathbf{Y_2})$

Reilly (1994) describes how samples of (Y_1, Y_2) can be generated by mixing values of (Y_1, Y_2) generated based on $g_1(y_1, y_2)$ and values of (Y_1, Y_2) generated based on $g_2(y_1, y_2)$. Let $\rho^- \leq \rho^0 \leq \rho^+$. In this case, the composite pmf for (Y_1, Y_2) is

$$\left(\frac{\rho^+ - \rho^0}{\rho^+ - \rho^-}\right) g_1(y_1, y_2) + \left(\frac{\rho^0 - \rho^-}{\rho^+ - \rho^-}\right) g_2(y_1, y_2). \quad (3)$$

We refer to pmfs (3), which were first suggested by Fréchet (1951), as extreme mixtures.

Like conventional mixtures, extreme mixtures are quite easy to use. There is a unique extreme mixture for every feasible value of ρ. However, it is impossible to use an extreme mixture to generate values of (Y_1, Y_2) with Y_1 and Y_2 independent, so extreme mixtures do not form a comprehensive family for (Y_1, Y_2). Extreme mixtures are also useful for bivariate continuous random variables.

2.3 Parametric Mixtures for $(\mathbf{Y_1}, \mathbf{Y_2})$

Let θ be the smallest joint probability associated with any $(y_1, y_2) \in S_1 \times S_2$, and assume that $f_i(y_i) > 0$, $\forall y_i \in S_i$, $i = 1, 2$. Peterson and Reilly (1993) show that a piecewise linear curve that plots the maximal value of θ versus ρ can be constructed from the solution to a parametric linear program. They go on to show that, if $\max\{n_1, n_2\} \geq 3$, this curve outlines a parametric envelope that contains all points that correspond to feasible combinations of ρ and θ.

Let $\theta' = f_1(y_{1i^*}) f_2(y_{2j^*})$, where $i^* = \arg\min_i \{f_1(y_{1i})\}$ and $j^* = \arg\min_j \{f_2(y_{2j})\}$. Suppose that (ρ^0, θ^0) is a point in the parametric envelope such that $\theta^0 \leq \theta'$ and $(1 - \theta^0/\theta')\rho^- \leq \rho^0 \leq (1 - \theta^0/\theta')\rho^+$. Peterson and Reilly show that, if $g_1(y_{1i^*}, y_{2j^*}) = g_2(y_{1i^*}, y_{2j^*}) = 0$, then $\rho = \rho^0$ and $\theta = \theta^0$ for the pmf

$$\lambda_0 f_1(y_1) f_2(y_2) + \lambda_1 g_1(y_1, y_2) + \lambda_2 g_2(y_1, y_2), \quad (4)$$

where $\lambda_0 = \theta^0/\theta'$,

$$\lambda_1 = \left((1 - \theta^0/\theta')\rho^+ - \rho^0\right) / (\rho^+ - \rho^-),$$

and

$$\lambda_2 = \left(\rho^0 - (1 - \theta^0/\theta')\rho^-\right) / (\rho^+ - \rho^-).$$

Even if $g_1(y_{1i^*}, y_{2j^*}) \neq 0$ or $g_2(y_{1i^*}, y_{2j^*}) \neq 0$, $\rho = \rho^0$ for the pmf (4). We refer to pmfs (4) as parametric mixtures.

Parametric mixtures include conventional mixtures (1) and (2) and extreme mixtures (3). Therefore, parametric mixtures form a comprehensive family for (Y_1, Y_2). For all values of ρ except ρ^+ and ρ^-, there is an infinite number of parametric mixtures. Yet, there is only one parametric mixture for every point (ρ, θ) such that $\theta \leq \theta'$ and $(1 - \theta/\theta')\rho^- \leq \rho \leq (1 - \theta/\theta')\rho^+$.

2.4 Multivariate Variate Generation

One way to generate values of a multivariate random variable $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_k)$ with a specified population correlation structure is to generate a value for Y_1 based on its marginal pmf or probability density function (pdf), $f_1(y_1)$, then generate a value for Y_2 based on the conditional pmf or pdf $f_{2|1}(y_2|y_1)$ associated with $\rho_{12} = \mathrm{Corr}(Y_1, Y_2) = \rho_{12}^0$, and so on. Although certainly effective, this approach can be time consuming and tedious. Furthermore, a fully specified joint pmf or pdf is required to begin.

A method for generating values of multivariate normal random variables with a specified population correlation structure is well known. Johnson, Wang, and

Ramberg (1984) and Devroye (1986) describe methods for generating values from several multivariate distributions.

Iman and Conover (1982) present a method for generating values of a multivariate random variable with a specified Spearman rank correlation structure.

3 EXTREME-CORRELATION PMFs

Let $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_k)$ be a multivariate discrete random variable. We assume that each Y_i, $i = 1, 2, \ldots, k$, is distributed over the support $S_i = \{y_{i1}, y_{i2}, \ldots, y_{in_i}\}$ according to the pmf $f_i(y_i)$.

According to Whitt (1976), the maximum and minimum possible values of $\rho_{ij} = \text{Corr}(Y_i, Y_j)$ are

$$\rho_{ij}^{+} = \frac{K_{ij}^{+} - \text{E}(Y_i)\text{E}(Y_j)}{(\text{Var}(Y_i)\text{Var}(Y_j))^{\frac{1}{2}}}$$

and

$$\rho_{ij}^{-} = \frac{K_{ij}^{-} - \text{E}(Y_i)\text{E}(Y_j)}{(\text{Var}(Y_i)\text{Var}(Y_j))^{\frac{1}{2}}},$$

respectively, where

$$K_{ij}^{+} = \int_0^1 F_i^{-1}(u)F_j^{-1}(u)\,du,$$

$$K_{ij}^{-} = \int_0^1 F_i^{-1}(u)F_j^{-1}(1-u)\,du,$$

and $F_i^{-1}(u)$ and $F_j^{-1}(u)$ are the inverse cumulative distribution function (cdfs) for Y_i and Y_j. Peterson and Reilly (1993) show that K_{ij}^{+} and K_{ij}^{-} can be determined by solving one factored transportation problem with the Northwest Corner Rule and one with the Southwest Corner Rule.

We define the extreme-correlation pmfs for $\mathbf{Y}$ to be the joint pmfs for $\mathbf{Y}$ for which $\rho_{ij} = \rho_{ij}^{+}$ or $\rho_{ij} = \rho_{ij}^{-}$, $\forall i < j$. Each extreme-correlation pmf is associated with a possible assignment of extreme values for the ρ_{1j}, $j = 2, 3, \ldots, k$. Hence, there are 2^{k-1} extreme-correlation pmfs for $\mathbf{Y}$. Let $g_\ell(\mathbf{y})$, $\ell = 1, 2, \ldots, 2^{k-1}$, be the extreme-correlation pmfs. Table 1 shows the ρ_{1j} values associated with the eight extreme-correlation pmfs when $k = 4$.

4 MULTIVARIATE COMPOSITION

Define for each extreme-correlation pmf, $g_\ell(\mathbf{y})$, $\ell = 1, 2, \ldots, 2^{k-1}$,

$$\delta_{ij}^{\ell} = \begin{cases} 1 & \text{if } \rho_{ij} = \rho_{ij}^{+}; \\ 0 & \text{if } \rho_{ij} = \rho_{ij}^{-}; \end{cases}$$

Table 1: Possible ρ_{1j} combinations when $k = 4$.

ℓ	ρ_{1j}		
	ρ_{12}	ρ_{13}	ρ_{14}
1	ρ_{12}^{-}	ρ_{13}^{-}	ρ_{14}^{-}
2	ρ_{12}^{-}	ρ_{13}^{-}	ρ_{14}^{+}
3	ρ_{12}^{-}	ρ_{13}^{+}	ρ_{14}^{-}
4	ρ_{12}^{-}	ρ_{13}^{+}	ρ_{14}^{+}
5	ρ_{12}^{+}	ρ_{13}^{-}	ρ_{14}^{-}
6	ρ_{12}^{+}	ρ_{13}^{-}	ρ_{14}^{+}
7	ρ_{12}^{+}	ρ_{13}^{+}	ρ_{14}^{-}
8	ρ_{12}^{+}	ρ_{13}^{+}	ρ_{14}^{+}

Table 2: Values of δ_{ij}^{ℓ} when $k = 4$.

ℓ	1,j			i,j		
	1,2	1,3	1,4	2,3	2,4	3,4
1	0	0	0	1	1	1
2	0	0	1	1	0	0
3	0	1	0	0	1	0
4	0	1	1	0	0	1
5	1	0	0	0	0	1
6	1	0	1	0	1	0
7	1	1	0	1	0	0
8	1	1	1	1	1	1

to indicate whether $\rho_{ij} = \rho_{ij}^{+}$ or $\rho_{ij} = \rho_{ij}^{-}$. First, we assign values to δ_{1j}^{ℓ}, $j = 2, 3, \ldots, k$, $\ell = 1, 2, \ldots, 2^{k-1}$, that are consistent with the ρ_{1j} values found in a table like Table 1. Then, to find the remaining δ_{ij}^{ℓ}s, we use the following simple rule:

$$\delta_{ij}^{\ell} = 1 - |\delta_{1i}^{\ell} - \delta_{1j}^{\ell}|.$$

In other words, if $\delta_{1i}^{\ell} = \delta_{1j}^{\ell}$, then $\delta_{ij}^{\ell} = 1$; otherwise, $\delta_{ij}^{\ell} = 0$. Table 2 shows the δ_{ij}^{ℓ} values when $k = 4$.

Let $g_0(\mathbf{y}) = \prod_{i=1}^{k} f_i(y_i)$. The pmf

$$g(\mathbf{y}) = \sum_{\ell=0}^{2^{k-1}} \lambda_\ell g_\ell(\mathbf{y}), \tag{5}$$

where $\sum_{\ell=0}^{2^{k-1}} \lambda_\ell = 1$ and $\lambda_\ell \geq 0$, $\ell = 0, 1, \ldots, 2^{k-1}$, is a composite pmf for $\mathbf{Y}$. In fact, pmfs (5) could be said to constitute a comprehensive family of multivariate distributions for $\mathbf{Y}$.

Let ρ^0_{ij} be the desired value of ρ_{ij}, $\forall i < j$. A valid pmf (5) for $\mathbf{Y}$ with the desired population correlation structure exists if there is a solution to

$$\sum_{\ell=1}^{2^{k-1}} \lambda_\ell \left[\delta^\ell_{ij}\rho^+_{ij} + (1-\delta^\ell_{ij})\rho^-_{ij}\right] = \rho^0_{ij}, \quad \forall i < j, \quad (6)$$

$$\sum_{\ell=0}^{2^{k-1}} \lambda_\ell = 1, \quad (7)$$

$$\lambda_\ell \geq 0, \quad \ell = 0, 1, \ldots, 2^{k-1}. \quad (8)$$

In most cases, there is an infinite number of solutions to (6)-(8). Hence, we could use an appropriate criterion as an objective function, along with (6)-(8), to select a composite pmf (5) with desired characteristics. An interesting question is what criterion should be used to select a pmf for a given application.

Let $F_i^{-1}(u)$ be the inverse cdf for Y_i, $i = 1, 2, \ldots, k$. The procedure KVAR allows us to generate values of $\mathbf{Y}$ easily based on the joint pmfs (5).

Procedure KVAR

1. Generate $u_1, u_2, \ldots, u_{k+1} \sim U(0,1)$.
2. If $u_1 < \lambda_0$, then for $i = 1, 2, \ldots, k$, $y_i \leftarrow F_i^{-1}(u_{i+1})$ and go to Step 6. Otherwise, set $\ell = 1$, $\Lambda = \lambda_0 + \lambda_1$.
3. If $u_1 > \Lambda$, go to Step 4. Otherwise, go to Step 5.
4. $\ell \leftarrow \ell + 1$, $\Lambda \leftarrow \Lambda + \lambda_\ell$. Go to Step 3.
5. Generate $\mathbf{y}$ with u_2 based on $g_\ell(\mathbf{y})$.
   (a) $y_1 \leftarrow F_1^{-1}(u_2)$.
   (b) For y_i, $i = 2, 3, \ldots, k$, $y_i \leftarrow F_i^{-1}(u_2)$ if $\delta^\ell_{1i} = 1$, or $y_i \leftarrow F_i^{-1}(1 - u_2)$ if $\delta^\ell_{1i} = 0$.
6. Return $\mathbf{y}$.

5 SPECIAL CASES

In this section, we consider the special cases where $\mathbf{Y}$ is either a bivariate or a trivariate random variable.

5.1 Special Case 1: $k = 2$

Suppose we wish to find a pmf (5) with $\rho = \rho^0$ such that λ_0 is minimized, i.e., the frequency of independent sampling is minimized. We can show that $\lambda_0 = 0$, $\lambda_1 = (\rho^+ - \rho^0)/(\rho^+ - \rho^-)$, and $\lambda_2 = (\rho^0 - \rho^-)/(\rho^+ - \rho^-)$. In other words, the desired pmf is an extreme mixture (3) and a special case of a parametric mixture (4) where $\theta = 0$.

Suppose instead that we wish to find a pmf (5) for which $\rho = \rho^0$ such that λ_0 is maximized. Then, we can show that $\lambda_0 = 1 - \rho^0/\rho^+$, $\lambda_1 = 0$, and $\lambda_2 = \rho^0/\rho^+$ if $\rho^0 \geq 0$ and $\lambda_0 = 1 - \rho^0/\rho^-$, $\lambda_1 = \rho^0/\rho^-$, and $\lambda_2 = 0$ if $\rho^0 < 0$. Hence, the desired pmf is a conventional mixture (1) or (2).

5.2 Special Case 2: $k = 3$

Suppose that we seek a pmf (5) with a specified population correlation structure such that $\lambda_0 = 0$, or there is no independent sampling. Let $\bar{\rho}_{ij} = (\rho^+_{ij} + \rho^-_{ij})/2$, $\forall i < j$. Then the unique solution to (6)-(8) is

$$\lambda_\ell = \frac{1 + \sum_{i=1}^{2}\sum_{j=i+1}^{3} \frac{2(2\delta^\ell_{ij}-1)(\rho^0_{ij} - \bar{\rho}_{ij})}{\rho^+_{ij} - \rho^-_{ij}}}{4},$$

$\ell = 1, 2, 3, 4$. The factor $2\delta^\ell_{ij} - 1$ assures that each term in the numerator has the appropriate sign. It can be shown that (5) is a valid pmf for $\mathbf{Y}$ with the desired population correlation structure because $\lambda_\ell \geq 0$, $\ell = 1, 2, 3, 4$, $\sum_{\ell=1}^4 \lambda_\ell = 1$, and $\rho_{ij} = \rho^0_{ij}$, $\forall i < j$.

Suppose that $Y_i \sim U\{1, 2, \ldots, n_i\}$, $i = 1, 2, 3$. That is, suppose that Y_i is uniformly distributed over the integers from 1 to n_i. In this case, $\rho^+_{ij} = -\rho^-_{ij}$ and $\bar{\rho}_{ij} = 0$, $\forall i < j$. Therefore, the mixing probabilities λ_ℓ, $\ell = 1, 2, 3, 4$, become

$$\lambda_\ell = \frac{1 + \sum_{i=1}^{2}\sum_{j=i+1}^{3} \frac{(2\delta^\ell_{ij}-1)\rho^0_{ij}}{\rho^+_{ij}}}{4}.$$

There is less setup required to use the pmfs (5) when all of the marginal distributions are discrete uniform.

Example 1. Let $\mathbf{Y} = (Y_1, Y_2, Y_3)$, where $Y_i \sim U\{1, 2, \ldots, n\}$, $i = 1, 2, 3$, and R_1 be the desired population correlation matrix, where

$$R_1 = \begin{pmatrix} 1 & -0.4 & 0.3 \\ -0.4 & 1 & 0 \\ 0.3 & 0 & 1 \end{pmatrix}.$$

In this case, $\rho^+_{ij} = 1$, $\forall i < j$.

Suppose that $\lambda_0 = 0$. Then, $\lambda_1 = 0.275$, $\lambda_2 = 0.425$, $\lambda_3 = 0.075$, and $\lambda_4 = 0.225$. □

The mixtures (5) for which $\lambda_0 = 0$ have a shortcoming. For some values $\mathbf{y} \in S_1 \times S_2 \times S_3$, $g(\mathbf{y}) = 0$. Consequently, there are some values of $\mathbf{Y}$ that can never be generated. By including $g_0(\mathbf{y})$ in our composite pmf with a nonzero mixing probability, we can generate all of the possible values of $\mathbf{Y}$, i.e., all $\mathbf{y} \in S_1 \times S_2 \times S_3$.

Consider the following mixing probabilities: γ_0, where $\gamma_0 > 0$, and for $\ell = 1, 2, 3, 4$,

$$\gamma_\ell = \frac{1 - \gamma_0 + \sum_{i=1}^{2}\sum_{j=i+1}^{3} \frac{2(2\delta^\ell_{ij}-1)\left(\rho^0_{ij} - (1-\gamma_0)\bar{\rho}_{ij}\right)}{\rho^+_{ij} - \rho^-_{ij}}}{4},$$

Y_1	Y_2: 0	1	2	3
1	0.1657	0	0.0233	0.0110
2	0.0132	0.1607	0.0261	0
3	0	0.2000	0	0
4	0.0028	0.0713	0.1259	0
5	0.0343	0	0.1127	0.0530

Figure 1: Example pmf $g(\mathbf{y})$ with $\gamma_0 = 0$.

Y_1	Y_2: 0	1	2	3
1	0.1462	0.0296	0.0198	0.0044
2	0.0254	0.1504	0.0198	0.0044
3	0.0148	0.1610	0.0198	0.0044
4	0.0148	0.0614	0.1194	0.0044
5	0.0148	0.0296	0.1092	0.0464

Figure 2: Example pmf $h(\mathbf{y})$ with $\gamma_0 = \gamma^*$.

where

$$\gamma_0 \le \gamma^* = \min_\ell \left\{ \frac{4\lambda_\ell}{d_\ell} \right\}$$

and

$$d_\ell = 1 - \sum_{i=1}^{2} \sum_{j=i+1}^{3} \frac{(2\delta_{ij}^\ell - 1)(\rho_{ij}^+ + \rho_{ij}^-)}{\rho_{ij}^+ - \rho_{ij}^-}.$$

A valid pmf for $\mathbf{Y}$ is

$$h(\mathbf{y}) = \sum_{\ell=0}^{2^{k-1}} \gamma_\ell g_\ell(\mathbf{y}). \qquad (9)$$

If $\gamma_0 = \gamma^*$, then at least one of the other mixing probabilities, γ_ℓ, $\ell = 1,2,3,4$, will be zero.

If $Y_i \sim U\{1,2,\ldots,n_i\}$, $i = 1,2,3$, then $d_\ell = 1$ for $\ell = 1,2,3,4$, and $\gamma^* = 4\min_\ell\{\lambda_\ell\}$.

Example 2. Recall Example 1. In this case, $\gamma^* = 0.30$. If $\gamma_0 = \gamma^*$, then $\gamma_1 = 0.20$, $\gamma_2 = 0.35$, $\gamma_3 = 0$, and $\gamma_4 = 0.15$. □

Suppose that $k = 2$. Peterson and Reilly (1993) show that pmfs (9) give the user control over the smallest joint probability, θ, in parametric mixtures (4). If $\gamma_0 = \gamma^*$, then the pmf (9) is a conventional mixture (1) or (2). The next example illustrates how the value of γ_0 affects the nature of the pmf for (Y_1, Y_2).

Example 3. Let $Y_1 \sim U\{1,2,3,4,5\}$ and Y_2 be a binomial random variable with 3 independent trials and success probability 0.4. Suppose that the desired value of $\rho = 0.6$ and $\lambda_0 = 0$. Then, $\lambda_1 = 0.1715$ and $\lambda_2 = 0.8285$. The pmf shown in Figure 1 is the mixture (5), and the pmf shown in Figure 2 is the mixture (9) with $\gamma_0 = \gamma^* = 0.3431$, $\gamma_1 = 0$, $\gamma_2 = 0.6569$. We see in Figure 1 that $g(y_1, y_2) = 0$ for 7 of 20 members of $S_1 \times S_2$, while all members of $S_1 \times S_2$ have positive probability with the pmf in Figure 2. □

6 EXTENSIONS

In this section, we describe the limitations of our composition approach for $k \ge 4$. We also show how to use composition for continuous multivariate random variables.

6.1 General case: $k \ge 4$

It would be convenient if we could use mixing probabilities of the form

$$\lambda_\ell = \frac{1 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \frac{2(2\delta_{ij}^\ell - 1)(\rho_{ij}^0 - \bar{\rho}_{ij})}{\rho_{ij}^+ - \rho_{ij}^-}}{2^{k-1}}, \qquad (10)$$

$\ell = 1,2,\ldots,2^{k-1}$, to construct composite pmfs for general multivariate discrete random variables. However, there is no guarantee that mixing probabilities of this form will be nonnegative. For example, suppose that $k = 4$ and that $Y_i \sim U\{1,2,\ldots,n\}$, $i = 1,2,3,4$. Also suppose that the desired population correlation structure is that shown for $g_1(\mathbf{y})$ in Table 2. It should be that $\lambda_1 = 1$ and the remaining mixing probabilities are all 0. But, if we assume that $\lambda_0 = 0$, then we find that our formula yields $\lambda_1 = 0.875$, $\lambda_2, \lambda_3, \lambda_5, \lambda_8 = 0.125$, and $\lambda_4, \lambda_6, \lambda_7 = -0.125$.

The next example illustrates that the general formula (10) for the mixing probabilities can work in some cases.

Example 4. Let $\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)$, where Y_i, $i = 1,2,3,4$, is a discrete uniform random variable. Suppose that the desired correlation matrix is

$$R_2 = \begin{pmatrix} 1 & 0 & \rho_{13}^+/4 & \rho_{14}^-/16 \\ 0 & 1 & \rho_{23}^+/8 & 0 \\ \rho_{13}^+/4 & \rho_{23}^+/8 & 1 & \rho_{34}^-/8 \\ \rho_{14}^-/16 & 0 & \rho_{34}^-/8 & 1 \end{pmatrix}.$$

Then, if $\lambda_0 = 0$, the mixing probabilities for the extreme-correlation pmfs depicted in Table 1 are:

$\lambda_1 = 13/128$; $\lambda_2, \lambda_4 = 15/128$; $\lambda_3 = 21/128$; $\lambda_5 = 9/128$; $\lambda_6 = 11/128$; $\lambda_7 = 25/128$; and $\lambda_8 = 19/128$. We can generate values of $\mathbf{Y}$ with the population correlation structure given by R_2 using a joint pmf (9) with any γ_0 such that $0 < \gamma_0 \le \gamma^* = 9/16$. The mixing probabilities in this case are $\gamma_\ell = \lambda_\ell - \gamma_0/8$, $\ell = 1, 2, \ldots, 8$. □

In a case where the mixing probabilities (10) computed are invalid, we can resort to trying to find a solution to (6)-(8) with an appropriate criterion to guide our selection of a composite joint pmf. We are also investigating procedures for adjusting the invalid mixing probabilities so that they become nonnegative.

6.2 Mixtures for continuous random variables

Let $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ be a continuous random variable, and let $F_i(x_i)$, $i = 1, 2, \ldots, k$, be the cdf of X_i. We assume that $\text{Var}(X_i) < \infty$, $i = 1, 2, \ldots, k$. Then, if we refer to pdfs rather than pmfs, the mixtures (5) or (9) and the generation procedure KVAR can be used to generate values of $\mathbf{X}$.

Many opportunities to use the composite pdfs (5) and (9) will arise in simulations of manufacturing systems. Two of the important distributions in these simulations may be the negative exponential and the continuous uniform.

Suppose that X_i, $i = 1, 2, \ldots, k$, is a negative exponential random variable with arbitrary expectation. Then, $\rho_{ij}^+ = 1$ and $\rho_{ij}^- = 1 - \pi^2/6 = -0.6449$, $\forall i < j$ (Page, 1965). There is little setup required to use the composite pdfs (5) and (9) because the extreme correlation values are independent of the parameters of the marginal distributions of the X_i, $i = 1, 2, \ldots, k$.

Example 5. Let X_1 and X_2 be negative exponential random variables with identical mean 1. Suppose that $\rho = 0.4$. Figure 3 shows a plot of 1000 points generated with the composite pdf (5) with $\lambda_0 = 0$, $\lambda_1 = 0.365$, and $\lambda_2 = 0.635$. Figure 4 shows a plot of 1000 points generated with the pdf (9) with $\gamma_0 = \gamma^* = 0.6$, $\gamma_1 = 0$, $\gamma_2 = 0.4$. Along the lines of our observation about Example 3, we see in Figure 4 that many more of the possible values of (X_1, X_2) are generated when $\gamma_0 = \gamma^*$ than when $\gamma_0 = 0$. □

Example 6. Let $\mathbf{X} = (X_1, X_2, X_3)$ be a trivariate random variable where each of the X_i, $i = 1, 2, 3$, is a negative exponential random variable. Suppose that the desired population correlation structure is given by

$$R_3 = \begin{pmatrix} 1 & 0.4 & -0.2 \\ 0.4 & 1 & -0.1 \\ -0.2 & -0.1 & 1 \end{pmatrix}.$$

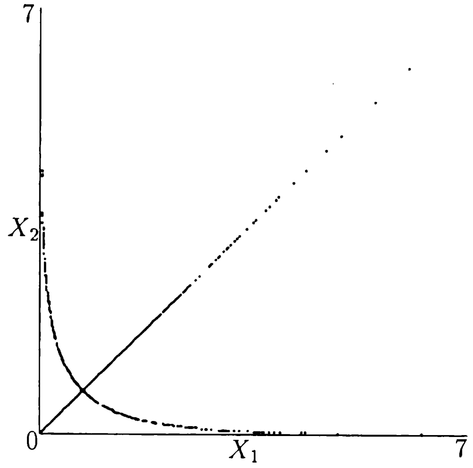

Figure 3: Exponential Marginals, $\rho = 0.4$, $\gamma_0 = 0$.

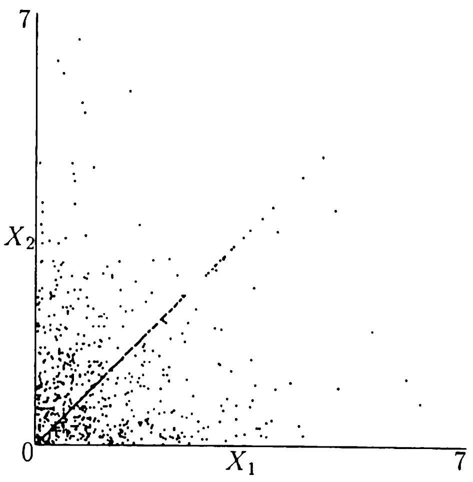

Figure 4: Exponential Marginals, $\rho = 0.4$, $\gamma_0 = \gamma^*$.

For a joint pdf of the form (5) with $\lambda_0 = 0$, we find that $\lambda_1 = 0.2128$, $\lambda_2 = 0.1520$, $\lambda_3 = 0.5167$, and $\lambda_4 = 0.1185$. Furthermore, $\gamma^* = 0.5$, and if $\gamma_0 = \gamma^*$, then $\gamma_1 = 0.0608$, $\gamma_2 = 0$, $\gamma_3 = 0.3648$, $\gamma_4 = 0.0744$. □

Suppose now that X_i, $i = 1, 2, \ldots, k$, is a continuous uniform random variable. For all $i < j$, $\rho_{ij}^+ = 1$ and $\rho_{ij}^- = -1$. It follows that $d_\ell = 1$, $\ell = 1, 2, \ldots, 2^{k-1}$. Like the case where the marginal distributions are all negative exponential, there is very little setup required for our composition method when the marginal distributions are all continuous uniform, and the parameters of the marginal distributions do not affect the value of the mixing probabilities λ_ℓ and γ_ℓ.

Example 7. Let $\mathbf{X} = (X_1, X_2, X_3)$ be a trivariate random variable where each of the X_i, $i = 1, 2, 3$, is a continuous uniform random variable. Suppose that the desired population correlation matrix is given by

$$R_4 = \begin{pmatrix} 1 & -0.25 & 0.2 \\ -0.25 & 1 & 0.4 \\ 0.2 & 0.4 & 1 \end{pmatrix}.$$

For a joint pdf of the form (5) with $\lambda_0 = 0$, we find that $\lambda_1 = 0.3625$, $\lambda_2 = 0.2625$, $\lambda_3 = 0.0375$, and $\lambda_4 = 0.3375$. Suppose we let $\gamma_0 = 0.10 < 0.15 = \gamma^*$, then $\gamma_1 = 0.3375$, $\gamma_2 = 0.2375$, $\gamma_3 = 0.0125$, $\gamma_4 = 0.3125$. □

7 CONCLUSIONS

Consider an optimization problem with $k - 1$ constraints. Suppose that A_i, $i = 1, 2, \ldots, k - 1$, is the random variable that represents the values of the coefficients in the ith constraint and that C is the random variable that represents the values of the objective function coefficients. We can generate coefficients with a specified population correlation structure for synthetic optimization problems if we use our composition approach as the basis for generating values of $(A_1, A_2, \ldots, A_{k-1}, C)$.

Although the motivating application for us is the generation of synthetic optimization problems, the composite pmfs and pdfs that we present can be useful in many practical simulation models of, for example, manufacturing systems.

More research is needed to find mixing probabilities for composite pmfs and pdfs for general multivariate random variables.

REFERENCES

Balas, E. and C.H. Martin. 1980. Pivot and Complement - A Heuristic for 0-1 Programming. *Management Science* **26**(1): 86–96.

Balas, E. and E. Zemel. 1980. An Algorithm for Large Zero-One Knapsack Problems. *Operations Research* **28**(5): 1130–1154.

Devroye, L. 1986. *Non-uniform Random Variate Generation*. New York: Springer-Verlag.

Fréchet, M. 1951. Sur les tableaux de corrélation dont les marges sont données. *Annales de l'Université de Lyon, Section A* **14**: 53–77.

Iman, R.L. and W.J. Conover. 1982. A Distribution-Free Approach to Inducing Rank Correlation Among Input Variables. *Communications in Statistics, Simulation and Computation* **11**: 311–334.

John, T.C. 1989. Tradeoff Solutions in Single Machine Production Scheduling for Minimizing Flow Time and Maximum Penalty. *Computers and Operations Research* **16**(5): 471–479.

Johnson, M.E., C. Wang, and J.S. Ramberg. 1984. Generation of Continuous Multivariate Distributions for Statistical Applications. *American Journal of Mathematical and Management Sciences* **4**(3–4): 225–248.

Loulou, R. and E. Michaelides. 1979. New Greedy Heuristics for the Multidimensional 0-1 Knapsack Problem. *Operations Research* **27**(6): 1101–1114.

Martello, S. and P. Toth. 1979. The 0-1 Knapsack Problem. *Combinatorial Optimization*, eds. N. Christofides, A. Mingozzi, and C. Sandi, 237–279. John Wiley and Sons, New York.

Martello, S. and P. Toth. 1988. A New Algorithm for the 0-1 Knapsack Problem. *Management Science* **34**(5): 633–644.

Moore, B.A. 1989. Correlated 0-1 Knapsack Problems. IND ENG 854 Course Project, Department of Industrial and Systems Engineering, The Ohio State University, Columbus, OH.

Moore, B.A. and C.H. Reilly. 1993. Randomly Generating Synthetic Optimization Problems with Explicitly Induced Correlation. Working Paper 1993-002, Department of Industrial and Systems Engineering, The Ohio State University, Columbus, OH.

Page, E.S. 1965. On Monte Carlo Methods in Congestion Problems: II. Simulation of Queuing Systems. *Operations Research* **13**(2): 300–305.

Peterson, J.A. and C.H. Reilly. 1993. Joint Probability Mass Functions for Coefficients in Synthetic Optimization Problems. Working Paper 1993-006, Department of Industrial and Systems Engineering, The Ohio State University, Columbus, OH.

Pollock, G.A. 1992. Evaluation of Solution Methods for Weighted Set Covering Problems Generated with Correlated Uniform Random Variables. Undergraduate Honors Thesis, Department of Industrial and Systems Engineering, The Ohio State University, Columbus, OH.

Potts, C.N. and L.N. Van Wassenhove. 1988. Algorithms for Scheduling a Single Machine to Minimize the Weighted Number of Late Jobs. *Management Science* 34(7): 843–858.

Reilly, C.H. 1991. Optimization Test Problems with Uniformly Distributed Coefficients. *Proceedings of the 1991 Winter Simulation Conference*, eds. B. L. Nelson, W. D. Kelton, G. M. Clark, 866–874. Institute of Electrical and Electronics Engineers, Phoenix, Arizona.

Reilly, C.H. 1994. Alternative Input Models for Generating Synthetic Optimization Problems: Analysis and Implications. Working Paper 1994-001, Department of Industrial and Systems Engineering, The Ohio State University, Columbus, OH.

Rushmeier, R.A. and G.L. Nemhauser. 1993. Experiments with parallel branch-and-bound algorithms for the set covering problem. *Operations Research Letters* **13**(5): 277–285.

Whitt, W. 1976. Bivariate Distributions with Given Marginals. *Annals of Statistics* **4**(6): 1280–1289.

AUTHOR BIOGRAPHIES

RAYMOND R. HILL is a Weapons and Tactics Analyst for the U.S. Air Force stationed at the Pentagon. He is also a Ph.D. candidate in the Department of Industrial, Welding and Systems Engineering at The Ohio State University. He earned his B.S. in mathematics from Eastern Connecticut State University and his M.S. in operations research from the U.S. Air Force Institute of Technology. Capt. Hill is a member of ORSA, TIMS, and SCS.

CHARLES H. REILLY is an Associate Professor and Acting Chair in the Department of Industrial, Welding and Systems Engineering at The Ohio State University. He earned a B.A. in mathematics and business administration at St. Norbert College in 1979 and M.S. and Ph.D. degrees in industrial engineering at Purdue University in 1980 and 1983, respectively. His current research interests are in the areas of empirical evaluation of solution methods for optimization problems and applications of optimization in manufacturing and telecommunications. Dr. Reilly is a member of ORSA, TIMS, IIE, and ASEE and Director-Elect of the Operations Research Division of IIE.