

Simulating Skewed Multivariate Distributions Using SAS®: Cases of Lomax, Mardia's Pareto (Type I), Logistic, Burr and F Distributions

Zhixin Lun, Oakland University, Rochester, MI
 Ravindra Khattree, Oakland University, Rochester, MI

ABSTRACT

By using various build-in functions in SAS software, it is easy to generate data from several common multivariate distributions such as multivariate normal (RANDNORMAL function) and multivariate Student's t (RANDMVT function). However, functions for generating data from other less common multivariate distributions are not readily available in SAS. We will illustrate the simulation and generation of random numbers from a multivariate Lomax distribution. Importance of the work lies in its wide applicability in reliability theory and many other situations where one needs to use a flexible nonnegative skewed multivariate distribution for modeling. Further, based on various useful properties of multivariate Lomax distribution, Mardia's multivariate Pareto of type I, multivariate Logistic, multivariate Burr, and multivariate F random variables can also be readily simulated. We develop and implement a SAS macro using SAS/IML to generate random numbers from all of these multivariate probability distributions.

INTRODUCTION

Multivariate Lomax (Pareto Type II) distribution was introduced by Nayak (1987) as a joint distribution of several skewed nonnegative random variables X_1, X_2, \dots, X_k with probability density function (pdf) given by,

$$f(x_1, x_2, \dots, x_k | a, \theta_1, \theta_2, \dots, \theta_k) = \frac{[\prod_{i=1}^k \theta_i] a(a+1) \dots (a+k-1)}{(1 + \sum_{i=1}^k \theta_i x_i)^{a+k}}, \quad x_i > 0, a > 0, \theta_i > 0, i = 1, \dots, k.$$

Nayak (1987) also indicated that the multivariate Lomax distribution can be obtained by mixing several independent exponential distributions with different failure rates via a random mixing parameter η . Specifically, with a given random variable η distributed as $gamma(a, b)$, we assume that conditional on η , the k random variables X_i 's, $i = 1, 2, \dots, k$, are independently and exponentially distributed with failure rates $\eta\lambda_i$ (that is, with respective means $1/\eta\lambda_i$), $i = 1, 2, \dots, k$. Then, the unconditional joint distribution of X_i 's, $i = 1, 2, \dots, k$, is a multivariate Lomax distribution with parameters a and $\theta_i = \lambda_i/b, i = 1, \dots, k$.

The above fact readily provides an approach to simulate multidimensional random vectors from a multivariate Lomax distribution. We further note later that the multivariate Lomax distribution is easily transformable to many other useful multivariate distributions and hence simulation of random numbers from any of these distributions is also easily accomplished.

The objective of this work is to accomplish the above tasks and to provide a ready to use SAS code for users to efficiently execute the simulations. This is what we achieve in next two sections.

SIMULATION FROM MULTIVARIATE LOMAX DISTRIBUTION

We use the relationship between the independent exponential random variables, a gamma variate and the multivariate Lomax distribution stated above to arrive at and then implement the following algorithm.

ALGORITHM

1. Generate a vector of random numbers η from gamma distribution using RANDGEN function with specified shape parameter a and scale parameter b .

2. For a given value of η , generate independent component X_i from exponential/gamma distribution using RANGAM function with specified parameter $\eta\lambda_i$. Note that exponential distribution is essentially gamma distribution with shape parameter=1.
3. For a chosen multivariate distribution among these mentioned earlier, transform the random vector $(X_1, X_2, \dots, X_k)'$ according to the appropriate transformation function (given in the next section).

IMPLEMENTATION

We develop the macro RANDGENMV which generates random numbers from a specified distribution by using the following format:

```
%RANDGENMV(distname, <a>, <paramList1>, <paramList2>, nSize, ndim, seed, out);
```

The macro takes the following input arguments:

| Argument | Description | Remark |
|--------------|---|-----------------------------------|
| distname | specifies the name of a probability distribution | |
| <a> | specifies a distribution parameter | optional with default value = 1 |
| <paramList1> | specifies a vector of distribution parameters | optional with default value = {1} |
| <paramList2> | specifies a vector of distribution parameters | optional with default value = {1} |
| nSize | specifies the sample size | |
| ndim | specifies the number of data dimension | |
| seed | specifies the seed for random numbers generation | |
| out | specifies a name of data set that is to be filled with random samples from the specified distribution | |

Table 1. Description of Arguments for macro RANDGENMV

| Distribution | distname | parm | parmList1 | parmList2 |
|--------------------------|------------|-------------------------|-------------------------------|---------------------|
| Lomax | 'Lomax' | $\langle a = 1 \rangle$ | $\{\theta_1 \dots \theta_k\}$ | N/A |
| Generalized Lomax | 'GLomax' | $\langle a = 1 \rangle$ | $\{\theta_1 \dots \theta_k\}$ | $\{l_1 \dots l_k\}$ |
| Mardia's Pareto (Type I) | 'MPareto1' | $\langle a = 1 \rangle$ | $\{\theta_1 \dots \theta_k\}$ | N/A |
| Logistic | 'Logistic' | N/A | N/A | N/A |
| Burr | 'Burr' | $\langle a = 1 \rangle$ | $\{d_1 \dots d_k\}$ | $\{c_1 \dots c_k\}$ |
| F | 'F' | $\langle a = 1 \rangle$ | $\{l_1 \dots l_k\}$ | N/A |

Table 2. Parameters Assignments for Distributions

AN EXAMPLE OF RANDOM NUMBERS GENERATION

To generate random variates from the multivariate Lomax distribution, specify 'LOMAX' for the distname argument. In the following example, we use our macro titled RANDGENMV to generate 10,000 samples for X_1, X_2, X_3 and X_4 from the multivariate Lomax distribution with parameters $a = 5, \theta_1 = 0.5, \theta_2 = 1, \theta_3 = 1.5, \theta_4 = 2.0$. For this, we take,

$$\eta \sim \text{gamma}(a = 5, b = 1),$$

$$X_1 \sim \exp(\lambda_1 \eta), X_2 \sim \exp(\lambda_2 \eta), X_3 \sim \exp(\lambda_3 \eta), X_4 \sim \exp(\lambda_4 \eta), \quad (\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0.5, 1.0, 1.5, 2.0).$$

Clearly then $\theta_i = \lambda_i/b = \lambda_i, i = 1, \dots, k$ for which $b = 1$ and hence $(\theta_1, \theta_2, \theta_3, \theta_4) = (0.5, 1.0, 1.5, 2.0)$. We assign θ_i 's to paramList1 argument as shown in the following code:

```
%include "Directory\randgenMV.sas";
```

```
%randgenMV(distname = 'LOMAX', a = 5, paramList1 = {0.5 1 1.5 2}, nSize = 10000, ndim = 4, seed = 123, out = TEST_Lomax);
```

We have shown the first five random vectors generated by above macro in Table 3. The descriptive statistics for 10,000 samples are summarized in Table 4. By comparing the descriptive statistics with the corresponding theoretical values, we can see that the summary of the simulated data matches the features of the theoretical multivariate Lomax distribution.

| Obs | x1 | x2 | x3 | x4 |
|-----|---------|---------|---------|---------|
| 1 | 0.75887 | 0.24062 | 0.00265 | 0.07670 |
| 2 | 1.46380 | 0.83515 | 0.04143 | 0.17602 |
| 3 | 0.75816 | 0.28828 | 0.14874 | 0.09205 |
| 4 | 0.19735 | 0.23063 | 0.07766 | 0.23814 |
| 5 | 2.62512 | 0.72608 | 0.79903 | 0.29616 |

Table 3. First five samples from multivariate Lomax distribution

| x_i | a | θ_i | Mean | Variance | Skewness |
|-------|-----|------------|--------------------|--------------------|--------------------|
| x_1 | 5 | 0.5 | 0.4840664 (0.5000) | 0.4093268 (0.4167) | 4.9902167 (4.6476) |
| x_2 | 5 | 1 | 0.2464027 (0.2500) | 0.1027076 (0.1047) | 4.1206686 (4.6476) |
| x_3 | 5 | 1.5 | 0.1676471 (0.1667) | 0.0462597 (0.0463) | 3.6823457 (4.6476) |
| x_4 | 5 | 2 | 0.1255278 (0.1250) | 0.0266295 (0.0260) | 4.1011152 (4.6476) |

Table 4. Simulated and theoretical descriptive statistics (in parentheses) of multivariate Lomax distribution (with parameters $a = 5, \theta_1 = 0.5, \theta_2 = 1, \theta_3 = 1.5, \theta_4 = 2.0$) with sample size = 10,000.

For illustration, we also evaluate the distribution of the simulated data by comparing the histogram and the theoretical joint density plot in the bivariate case in Figure 1. The code of creating histogram and theoretical density plot is given as below. Consider the bivariate data with sample size = 10,000 from bivariate Lomax distribution with parameters ($a = 5, \theta_1 = 0.5, \theta_2 = 1$). The shape of histogram of simulated bivariate random variables clearly resembles the corresponding the plot of joint density.

```
/*----- Create the histogram for generated bivariate Lomax data -----*/
%randgenMV(distname = 'LOMAX', a = 5, paramList1 = {0.5 1}, nSize = 10000,
ndim = 2, seed = 123, out = Bivar_Lomax);
```

```
ods graphics on;
proc kde data = Bivar_Lomax;
    bivar x1 (gridl = 0 gridu = 5) x2 (gridl = 0 gridu = 5) / plots =
    HISTSURFACE noprint;
run;
ods graphics off;
```

```

/*----- Create the density plot of bivariate Lomax data -----*/
data density_Lomax;
  a = 5;
  k = 2;
  theta_1 = 0.5;
  theta_2 = 1.0;
  keep x1 x2 z;
  label z = Density;
  const = theta_1*theta_2*a*(a+1);
  do x1 = 0 to 5 by 0.10;
    do x2 = 0 to 5 by 0.10;
      z = const*(1/((1+theta_1*x1+theta_2*x2)**(a+k)));
      if z > .0001 then output;
    end;
  end;
run;

proc g3d data = density_Lomax;
  plot x2*x1 = z/ rotate = -30;
run;
quit;

```

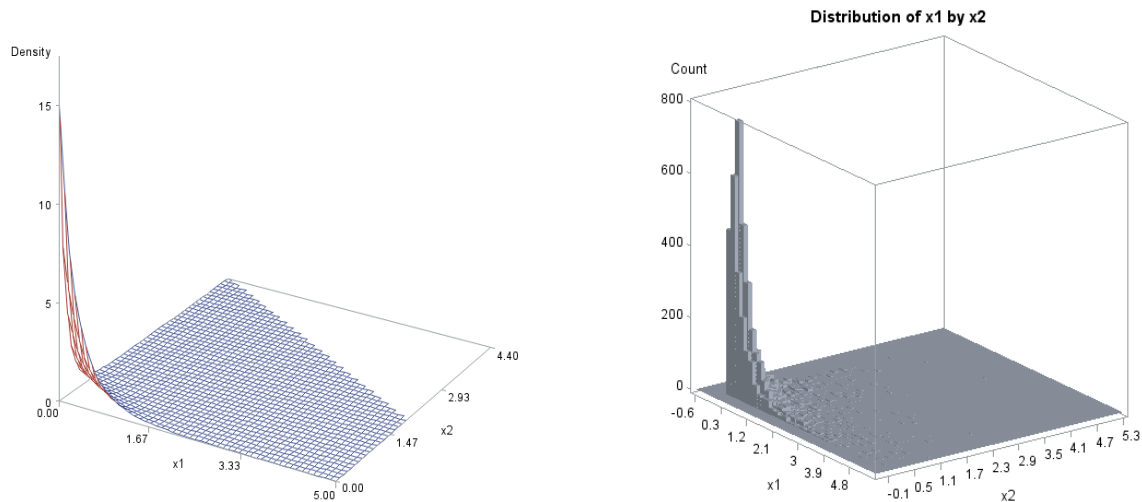


Figure 1. Plot of the joint density function (left) and histogram (right) of simulated bivariate Lomax random vector with parameters ($a = 5, \theta_1 = 0.5, \theta_2 = 1$)

RANDOM NUMBER GENERATION FROM VARIOUS OTHER MULTIVARIATE DISTRIBUTIONS

Nayak (1987) has also pointed out the inter-relationships between many other nonnegative multivariate probability distributions and the multivariate Lomax distribution. In view of these relationships, once random number generation from multivariate Lomax distribution has been achieved, appropriate transformations allow us to generate random numbers from these distributions as well. This is a task which can be quite difficult to achieve directly by using the corresponding density functions for most of these random variables. Specifically, we have,

Mardia's Multivariate Pareto - Type I Distribution

Let $Y_i = X_i + 1/\theta_i, i = 1, \dots, k$, then the joint distribution of Y_1, \dots, Y_k is a multivariate Mardia's Pareto (Type I) with pdf

$$f(y_1, \dots, y_k | a, \theta_1, \dots, \theta_k) = \frac{[\prod_{i=1}^k \theta_i] a(a+1) \dots (a+k-1)}{(\sum_{i=1}^k \theta_i y_i - k + 1)^{a+k}}, \quad y_i > \frac{1}{\theta_i} > 0, a > 0, \theta_i > 0, i = 1, \dots, k.$$

Multivariate Logistic Distribution

Let $U_i = -\ln(\theta_i X_i), i = 1, \dots, k, a = 1$, then the joint density of U_1, \dots, U_k which follow multivariate logistic is given by

$$f(u_1, \dots, u_k) = \frac{k! \exp(-\sum_{i=1}^k u_i)}{[1 + \sum_{i=1}^k \exp(-u_i)]^{1+k}}, \quad -\infty < u_i < \infty.$$

Multivariate Burr Distribution

Let $B_i = (\theta_i X_i / d_i)^{1/c_i}, i = 1, \dots, k, c_i > 0, d_i > 0$, then the joint density of B_1, \dots, B_k is

$$f(b_1, \dots, b_k | a, c_1, \dots, c_k, d_1, \dots, d_k) = \frac{[\prod_{i=1}^k c_i d_i] a(a+1) \dots (a+k-1) [\prod_{i=1}^k b_i^{c_i-1}]}{(1 + \sum_{i=1}^k d_i b_i^{c_i})^{a+k}}, \quad b_i > 0, a > 0,$$

which corresponds to a multivariate Burr distribution.

RANDOM NUMBER GENERATION FROM GENERALIZED MULTIVARIATE LOMAX DISTRIBUTION

GENERALIZED MULTIVARIATE LOMAX DISTRIBUTION

A generalized multivariate Lomax distribution is defined essentially in the same way as the multivariate Lomax distribution but by replacing the conditional distribution of X_i given η by the more general distribution of gamma($l_i, \eta\lambda_i$).

Specifically, we let

$$\eta \sim \text{gamma}(a, b), \quad X_i \sim \text{gamma}(l_i, \eta\lambda_i), i = 1, \dots, k,$$

where X_1, \dots, X_k are conditionally independent and a and l_i are the shape parameters, b and $\eta\lambda_i$ are the rate parameters of corresponding *gamma* distributions. The joint density of X_1, \dots, X_k is, then

$$f(x_1, \dots, x_k | a, \theta_1, \dots, \theta_k, l_1, \dots, l_k) = \frac{[\prod_{i=1}^k \theta_i^{l_i}] \Gamma(\sum_{i=1}^k l_i + a) \prod_{i=1}^k x_i^{l_i-1}}{\Gamma(a) [\prod_{i=1}^k \Gamma(l_i)] (1 + \sum_{i=1}^k \theta_i x_i)^{\sum_{i=1}^k l_i + a}}, \quad x_i > 0,$$

where $\theta_i = \lambda_i/b, l_i > 0, i = 1, \dots, k$.

MULTIVARIATE F DISTRIBUTION

By replacing $\theta_i = l_i/a, i = 1, \dots, k$, the density of X_1, \dots, X_k becomes what is commonly called as the multivariate *F* density (Johnson and Kotz, 1972) with degrees of freedom $(2a, 2l_1, \dots, 2l_k)$. Its density function is given by,

$$f(x_1, \dots, x_k | a, l_1, \dots, l_k) = \frac{[\prod_{i=1}^k (l_i/a)^{l_i}] \Gamma(\sum_{i=1}^k l_i + a) \prod_{i=1}^k x_i^{l_i-1}}{\Gamma(a) [\prod_{i=1}^k \Gamma(l_i)] \left(1 + \sum_{i=1}^k \frac{l_i x_i}{a}\right)^{\sum_{i=1}^k l_i + a}}, \quad x_i > 0, a, l_i > 0, i = 1, \dots, k.$$

Thus, using the algorithm given earlier and by making the above change (that is, by replacing exponentiality by suitably chosen gamma), random number generation from the generalized multivariate Lomax distribution is straight forward. Further, rechristening the parameters as $\theta_i = l_i/a, i = 1, \dots, k$, we

also cover the case of random number generation from the multivariate F distribution. Implementation of above has been made part of previous algorithm.

SAMPLE OUTPUT

We simulate 10,000 samples and show first five of them for each of the multivariate distributions stated in the previous two sections. A comparison between simulated and theoretical descriptive statistics is also provided.

MARDIA'S MULTIVARIATE PARETO - TYPE I DISTRIBUTION

To generate the numbers from Mardia's multivariate Pareto of type I with parameters $a = 5, \theta_1 = 1, \theta_2 = 2, \theta_3 = 3$ and $\theta_4 = 4$, we specify 'MPareto1' for the `distname` and assign above parameters to the `a` and `paramList1` arguments. The Table 5 shows the first five samples and the Table 6 gives the comparison between simulated and theoretical descriptive statistics in terms of mean, variance and skewness.

```
%randgenMV(distname = 'MPareto1', a = 5, paramList1 = {1 2 3 4}, nSize = 10000, ndim = 4, seed = 123, out = TEST_MPareto1);
```

| Obs | x1 | x2 | x3 | x4 |
|-----|---------|---------|---------|---------|
| 1 | 1.37943 | 0.62031 | 0.33466 | 0.28835 |
| 2 | 1.73190 | 0.91758 | 0.35405 | 0.33801 |
| 3 | 1.37908 | 0.64414 | 0.40771 | 0.29603 |
| 4 | 1.09868 | 0.61532 | 0.37216 | 0.36907 |
| 5 | 2.31256 | 0.86304 | 0.73285 | 0.39808 |

Table 5. First five samples from Mardia's multivariate Pareto of type I distribution

| x_i | a | θ_i | Mean | Variance | Skewness |
|-------|-----|------------|--------------------|--------------------|--------------------|
| x_1 | 5 | 1 | 1.2420332 (1.2500) | 0.1023317 (0.1042) | 4.9902167 (4.2603) |
| x_2 | 5 | 2 | 0.6232013 (0.6250) | 0.0256769 (0.0260) | 4.1206686 (4.2603) |
| x_3 | 5 | 3 | 0.4171569 (0.4167) | 0.0115649 (0.0116) | 3.6823457 (4.2603) |
| x_4 | 5 | 4 | 0.3127639 (0.3125) | 0.0066574 (0.0065) | 4.1011152 (4.2603) |

Table 6. Simulated and theoretical descriptive statistics (in parentheses) of Mardia's multivariate Pareto of type I distribution (with parameters $a = 5, \theta_1 = 1, \theta_2 = 2, \theta_3 = 3$ and $\theta_4 = 4$) with sample size = 10,000.

MULTIVARIATE LOGISTIC DISTRIBUTION

To generate the numbers from multivariate logistic distribution, we specify 'Logistic' for the `distname` argument. The Table 7 shows the first five samples and the Table 8 gives the comparison between simulated and theoretical descriptive statistics.

```
%randgenMV(distname = 'Logistic', nSize = 10000, ndim = 4, seed = 123, out = TEST_Logistic);
```

| Obs | x1 | x2 | x3 | x4 |
|-----|----------|----------|----------|----------|
| 1 | -1.36314 | -0.90768 | 3.19451 | -0.45745 |
| 2 | 0.72953 | 0.59756 | 3.19576 | 1.46144 |
| 3 | 0.53348 | 0.80731 | 1.06354 | 1.25573 |
| 4 | 0.87686 | 0.02786 | 0.71089 | -0.69729 |
| 5 | -0.83880 | -0.24672 | -0.74793 | -0.04313 |

Table 7. First five samples from multivariate logistic distribution.

| x_i | Mean | Variance | Skewness |
|-------|----------------|----------------------|----------------|
| x_1 | 0.0149735 (0) | 3.3260940 (3.289868) | 0.0601667 (0) |
| x_2 | 0.0066787 (0) | 3.3277500 (3.289868) | 0.0354062 (0) |
| x_3 | -0.0167239 (0) | 3.4034982 (3.289868) | -0.0272500 (0) |
| x_4 | -0.0172631 (0) | 3.2073399 (3.289868) | -0.0295767 (0) |

Table 8. Simulated and theoretical descriptive statistics (in parentheses) of multivariate logistic distribution with sample size = 10,000.

MULTIVARIATE BURR DISTRIBUTION

To generate the numbers from multivariate Burr distribution with parameters $a = 3$, $d_1 = 1$, $d_2 = 2$, $d_3 = 2.5$, $d_4 = 3$ and $c_1 = 1$, $c_2 = 1.5$, $c_3 = 2$, $c_4 = 3$, we specify 'Burr' for the `distname` and assign above parameters to the `paramList1` and `paramList2` arguments. The Table 9 shows the first five samples and the Table 10 gives the comparison between simulated and theoretical descriptive statistics.

```
%randgenMV(distname = 'Burr', a = 3, paramList1 = {1 2 2.5 3}, paramList2 = {1 1.5 2 3}, nSize = 10000, ndim = 4, seed = 123, out = TEST_Burr);
```

| Obs | x1 | x2 | x3 | x4 |
|-----|---------|---------|---------|---------|
| 1 | 0.60908 | 0.33411 | 0.05055 | 0.43458 |
| 2 | 1.61558 | 0.94713 | 0.23424 | 0.63745 |
| 3 | 0.68515 | 0.40790 | 0.40163 | 0.48047 |
| 4 | 0.17758 | 0.35052 | 0.28958 | 0.65863 |
| 5 | 3.01316 | 0.88560 | 1.04908 | 0.76815 |

Table 9. First five samples from multivariate Burr distribution

| x_i | a | b_i | c_i | Mean | Variance | Skewness |
|-------|-----|-------|-------|-----------------------|---------------------|----------------------------|
| x_1 | 3 | 1 | 1 | 0.4867022 (0.5) | 0.6360966 (0.75) | 6.5802177 (no closed form) |
| x_2 | 3 | 2 | 1.5 | 0.3349340 (0.3385547) | 0.0908590 (0.09866) | 2.5023496 (no closed form) |
| x_3 | 3 | 2.5 | 2 | 0.3747323 (0.3725471) | 0.0639581 (0.06121) | 1.8628573 (no closed form) |
| x_4 | 3 | 3 | 3 | 0.4663825 (0.4657845) | 0.0417502 (0.04141) | 0.9077997 (no closed form) |

Table 10. Simulated and theoretical descriptive statistics (in parentheses) of multivariate Burr distribution (with parameters $a = 3, d_1 = 1, d_2 = 2, d_3 = 2.5, d_4 = 3$ and $c_1 = 1, c_2 = 1.5, c_3 = 2, c_4 = 3$) with sample size = 10,000.

GENERALIZED MULTIVARIATE LOMAX DISTRIBUTION

To generate the numbers from generalized multivariate Lomax distribution with parameters $a = 5, \theta_1 = 1, \theta_2 = 3, \theta_3 = 5, \theta_4 = 7$ and $l_1 = 3, l_2 = 6, l_3 = 9, l_4 = 12$, we specify 'GLOMAX' for the `distname` argument and assign above parameters to `a`, `paramList1`, and `paramList2`, respectively. The Table 11 shows the first five samples and the Table 12 gives the comparison between simulated and theoretical descriptive statistics.

```
%randgenMV(distname = 'GLOMAX', a = 5, paramList1 = {1 3 5 7}, paramList2 = {3 6 9 12}, nSize = 10000, ndim = 4, seed = 123, out = TEST_GLomax);
```

| Obs | x1 | x2 | x3 | x4 |
|-----|---------|---------|---------|---------|
| 1 | 0.22082 | 0.23608 | 0.51410 | 0.23713 |
| 2 | 1.17087 | 1.13377 | 0.93032 | 0.46419 |
| 3 | 0.27848 | 0.17701 | 0.44490 | 0.36987 |
| 4 | 1.49210 | 0.19961 | 0.40918 | 0.51605 |
| 5 | 3.73857 | 2.34994 | 0.78979 | 0.91950 |

Table 11. First five samples from generalized multivariate Lomax distribution

| x_i | a | θ_i | l_i | Mean | Variance | Skewness |
|-------|-----|------------|-------|-----------------------|------------------------|----------------------|
| x_1 | 5 | 1 | 3 | 0.7527525 (0.75) | 0.4352055 (0.4375) | 3.1878871 (3.779645) |
| x_2 | 5 | 3 | 6 | 0.4948338 (0.5) | 0.1275871 (0.13888889) | 2.7736863 (3.577709) |
| x_3 | 5 | 5 | 9 | 0.4456667 (0.45) | 0.0946357 (0.0975) | 3.1771275 (3.522819) |
| x_4 | 5 | 7 | 12 | 0.4258841 (0.4285714) | 0.0775375 (0.08163265) | 2.8155959 (3.5) |

Table 12. Simulated and theoretical descriptive statistics (in parentheses) of generalized multivariate Lomax distribution (with parameters $a = 5, \theta_1 = 1, \theta_2 = 3, \theta_3 = 5, \theta_4 = 7$ and $l_1 = 3, l_2 = 6, l_3 = 9, l_4 = 12$) with sample size = 10,000.

MULTIVARIATE F DISTRIBUTION

To generate the numbers from multivariate F distribution with degrees of freedom (10, 2, 6, 10, 14), i.e., $a = 5$ and $l_1 = 1, l_2 = 3, l_3 = 5, l_4 = 7$, we specify 'F' for the `distname` argument and assign above parameters to `a` and `paramList1`. The Table 13 shows the first five samples and the Table 14 gives the comparison between simulated and theoretical descriptive statistics.


```
%randgenMV(distname = 'F', a = 5, paramList1 = {1 3 5 7}, nSize = 10000, ndim = 4, seed = 123, out = TEST_F);
```

| Obs | x1 | x2 | x3 | x4 |
|-----|---------|---------|---------|---------|
| 1 | 1.89716 | 0.56261 | 0.80985 | 0.56524 |
| 2 | 1.76018 | 1.30660 | 2.02943 | 2.78502 |
| 3 | 0.58985 | 0.58212 | 0.86527 | 1.02004 |
| 4 | 0.64897 | 1.19425 | 0.96708 | 1.91616 |
| 5 | 7.90548 | 1.84270 | 1.39392 | 1.58767 |

Table 13. First five samples from multivariate F distribution.

| x_i | a | l_i | Mean | Variance | Skewness |
|-------|-----|-------|------------------|-----------------------|----------------------|
| x_1 | 5 | 1 | 1.2422795 (1.25) | 2.5050634 (2.6041667) | 3.9292800 (4.647580) |
| x_2 | 5 | 3 | 1.2217160 (1.25) | 1.0810380 (1.2152778) | 2.8596516 (3.779645) |
| x_3 | 5 | 5 | 1.2423353 (1.25) | 0.8831846 (0.9375000) | 2.6852025 (3.614784) |
| x_4 | 5 | 7 | 1.2413076 (1.25) | 0.7484897 (0.8184524) | 2.4953732 (3.552939) |

Table 14. Simulated and theoretical descriptive statistics (inside parentheses) of multivariate F distribution with degrees of freedom (10, 2, 6, 10, 14) and sample size = 10,000.

CONCLUSION

We have an approach to simulate and generate random numbers from a multivariate Lomax distribution. Further, we have shown how to simulate from many other multivariate distributions due to their relationships with the multivariate Lomax distribution. The %RANDGENMV macro is useful to readily generate random numbers from these multivariate distributions. The empirical distributions of the simulated random numbers clearly match the theoretical features of corresponding distributions and hence it is reassuring that our macro offers a very useful tool to support statistical modeling and data simulation.

REFERENCES

- Johnson, N. L. and Kotz, S. 1972. *Distribution in Statistics: Continuous Multivariate Distributions*. NJ: John Wiley & Sons.
- Nayak, Tapan Kumar. 1987. "Multivariate Lomax Distribution: Properties and Usefulness in Reliability Theory." *Journal of Applied Probability*, 24(1):170-177.

APPENDIX

```
/* *****;
*** This SAS program generates multivariate data according to specified
*** distribution and parameters.
*****/
```

```
%macro randgenMV(distname = , a = 1, paramList1 = {1}, paramList2 = {1},
nSize = , ndim = , seed = , out = );
```

```
proc iml;
```

```

call randseed(&seed);

eta = j(&nSize, 1);
eta_a = &a;
eta_b = 1;

mvGLomax = j(&nSize, &ndim);

call randgen(eta, 'GAMMA', eta_a, 1/eta_b);

if &distname='F' then do;
  GLalpha = &paramList1;
end;

else if &distname='GLOMAX' then do;
  GLalpha = &paramList2;
end;

else GLalpha = j(&ndim, 1);

if &distname = 'Logistic' | &distname='Burr' then do;
  x_lambda = j(&ndim, 1);
end;

else if &distname = 'F' then do;
  x_lambda = &paramList1/eta_a;
end;

else x_lambda = &paramList1;

do i = 1 to &nSize;
  do j = 1 to &ndim;
    GLbeta = eta[i]*x_lambda[j];
    mvGLomax[i,j] = rangam((&seed+i+j),GLalpha[j])/GLbeta;
  end;
end;

if &distname = 'LOMAX' | &distname = 'GLOMAX' then do;
  mvData = mvGLomax;
end;

if &distname = 'MPareto1' then do;
  mvMPareto1 = j(&nSize, &ndim);
  theta = &paramList1/eta_b;
  do k = 1 to &nSize;
    mvMPareto1[k,] = mvGLomax[k,] + 1/theta;
  end;
  mvData = mvMPareto1;
end;

if &distname = 'Logistic' then do;
  mvLogistic = j(&nSize, &ndim);
  do k=1 to &nSize;
    mvLogistic[k,] = -log(mvGLomax[k,]);
  end;
  mvData = mvLogistic;
end;

```

```

end;

if &distname = 'Burr' then do;
  mvBurr = j(&nSize, &ndim);
  do k=1 to &nSize;
    mvBurr[k,] = (mvGLomax[k,] # 1/&paramList1) ## (1/&paramList2);
  end;
  mvData = mvBurr;
end;

if &distname = 'F' then do;
  mvData = mvGLomax;
end;

create &out from mvData[c = ("x1":"x&ndim")];
append from mvData;
close &out;
quit;

%mend randgenMV;

```

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Zhixin Lun, PhD Student
 Department of Mathematics and Statistics, Oakland University
zlun@oakland.edu; zxulun9@gmail.com
 ORCID: 0000-0002-8980-1554

Ravindra Khattree, Professor/Co-director
 Department of Mathematics and Statistics/Center for Data Science and Big Data Analytics, Oakland University
khattree@oakland.edu
 ORCID: 0000-0002-9305-2365

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.