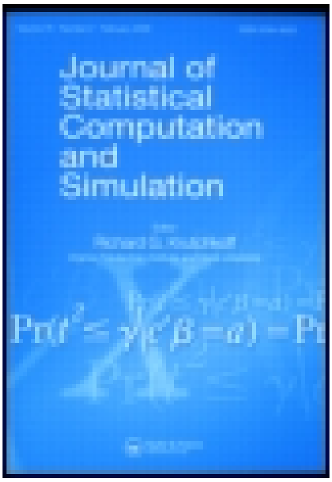

Journal of Statistical Computation and Simulation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gscs20

Constructing and simulating multivariate distributions using khintchine's theorem

Maurice C. Bryson [a] & Mark E. Johnson [a]
[a] Los Alamos National Laboratory , USA
Published online: 20 Mar 2007.

J. Statist. Comput. Simul., 1982, Vol. 16, pp. 129–137
0094-9655/82/1602-0129 $06.50/0

Constructing and Simulating Multivariate Distributions using Khintchine's Theorem†

MAURICE C. BRYSON and MARK E. JOHNSON
Los Alamos National Laboratory, U.S.A.

(*Received February 2, 1982*)

Khintchine's theorem is used to construct new multivariate distributions which are easy to simulate and which cover a broad range of dependence structures.

KEY WORDS: Pearson coefficient of skewness, bivariate uniform distributions, normal marginals, exponential marginals, coefficient of variation, unimodality, Morgenstern's distribution, Plackett's distribution, contour plots.

1. INTRODUCTION

Cook and Johnson (1981) recently discussed the paucity of continuous multivariate probability distributions. They noted that much of the current work in this area has been directed at multivariate generalizations of the "classical" univariate and bivariate distributions. These multivariate distributions tend to suffer many of the limitations of their lower-dimensional counterparts. The purpose of this paper is to offer a new approach to constructing multivariate distributions. We construct families which have several advantages: (1) wider ranges of dependence can be obtained than those found in commonly considered alternative distributions; (2) the distributions are extremely easy to simulate on digital computers; (3) many different multivariate distributions with the same marginal distributions can be constructed. Our primary interest here is in constructing useful families for simulation purposes. Other applications in simulation as well as theoretical aspects of Khintchine's theorem are given by Devroye (1981).

†This work was supported by an Institutional Supporting Research Grant from the Engineering Sciences Directorate of the Los Alamos National Laboratory.

In Section 2 we discuss Khintchine's theorem and some related results. Section 3 indicates the theorem's use with "classical" bivariate distributions in constructing new families. Section 4 introduces Khintchine mixtures—a general approach to constructing flexible families which are easy to simulate. Section 5 gives a related approach using Khintchine's theorem which leads to distributions with probability mass more suitably located for many simulation objectives.

2. KHINTCHINE'S THEOREM AND IMPLICATIONS

Our approach in constructing new multivariate distributions relies heavily on Khintchine's theorem (1938) which is discussed at length by Feller (1971, pp. 157–159). Olshen and Savage (1970) and de Silva (1978) consider further generalizations. Khintchine's theorem can be stated simply, as follows: Any continuous random variable X has a single mode at the origin, if and only if it can be expressed as the product $X=Z\cdot U$ where Z and U are independent continuous variables and U has a uniform distribution on the unit interval. It is readily shown that the density function of Z is given by

$$f_Z(z) = -z f'_X(z), \tag{2.1}$$

where f_X is the density of X. If X is normal $(0,1)$, then Z has the distribution of $\sqrt{\chi^2_{(3)}}$ with a random sign. More generally, if $|Z|$ is a power of a gamma variate, then X belongs to a distribution family developed by Johnson, Tietjen and Beckman (1980).

There are a number of convenient features of unimodal distributions readily derivable from this representation, which are not so easily found by more direct means. For example, if we take moments of $X=ZU$ and use the fact that $\operatorname{Var}(Z)\geqq 0$, we find immediately that

$$E(X^2) \geqq \frac{4}{3}\, E^2(X), \tag{2.2}$$

or equivalently, for any distribution unimodal at zero, the coefficient of variation must exceed $1/\sqrt{3}$.

By centering around the mode, we can also show immediately the relation for the Pearson coefficient of skewness (Sk) for any unimodal distribution

$$\text{Sk} = \frac{\text{mean} - \text{mode}}{\text{standard deviation}} \leqq \sqrt{3}, \tag{2.3}$$

a result that was derived with much more difficulty by Johnson and Rogers (1951).

A multivariate counterpart to Khintchine's theorem is developed by Olshen and Savage (1970). Here we are primarily interested in the reproduction of certain univariate marginals, so that our concern is with the univariate version of the theorem.

3. USE OF CLASSICAL BIVARIATE UNIFORM DISTRIBUTIONS

Bivariate continuous distributions which have uniform marginals have been surveyed in detail by Mardia (1970), Johnson and Kotz (1972) and Kowalski (1973). Barnett (1980) provides bivariate contour plots for Morgenstern (1956), Plackett (1965), Gumbel (1960), normal and Cauchy families of this type. Cook and Johnson (1981) also derived a bivariate uniform distribution. These distributions can be employed in the following construction scheme:

$$\begin{aligned} X_1 &= Z_1 U_1 \\ X_2 &= Z_2 U_2, \end{aligned} \tag{3.1}$$

where (U_1, U_2) follows one of these bivariate uniform distributions and the Z_i are selected using (2.1) to obtain desired X_i marginals. The paired random variables Z_i and $U_i, i=1, 2$ are assumed to be independent. Any correlation structure can be imposed on (Z_1, Z_2). We consider here the two simplest cases where Z_1 and Z_2 are (a) independent and identically distributed, or (b) identical ($Z_1 = Z_2$). For convenience in terminology we refer to the Z_i's as the "generator" variables. The resulting X_i's are in either case identically distributed, and we refer to them as having an X distribution.

a) *Independent generators.* The dependence structure of (X_1, X_2) arises primarily from the dependence structure already built into the (U_1, U_2) distribution. The (X_1, X_2) correlation is also affected by the form of the marginals being generated. Routine calculations from (3.1) show that the correlation of X_1 and X_2 is

$$\rho(X_1, X_2) = \rho(U_1, U_2) \cdot [3c_X^2]^{-1}, \tag{3.2}$$

where $\rho(U_1, U_2)$ is the correlation in the uniforms and c_X is the common coefficient of variation of X_1 and X_2. Note that the multiplier $[3c_X^2]^{-1}$ is necessarily less than or equal to 1 from (2.2). If X_1 and X_2 have normal or other symmetric distributions, then X_1 and X_2 are uncorrelated although they are independent only if U_1 and U_2 are. The density of (X_1, X_2) is given by

$$f(x_1,x_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f'_X(w_1)f'_X(w_2)\,g(x_1/w_1,x_2/w_2)\,dw_1\,dw_2,$$

where g is the density of (U_1, U_2) and f_X is the density common to X_1 and X_2.

b) *Identical generators.* Non-zero correlations between X_1 and X_2 can generally be realized if the generators are correlated—or in the most extreme case, identical. With the assumption $Z_1 = Z_2$ and (3.1), we have

$$\rho(X_1,X_2)=\tfrac{1}{4}[3-1/c_X^2+\rho(U_1,U_2)\cdot(1+1/c_X^2)],$$

using the same notation as above. For normal $(0, \sigma^2)$ marginals we have $\rho(X_1, X_2) = \frac{1}{4}[3 + \rho(U_1, U_2)]$, and for exponential marginals we have $\rho(X_1, X_2) = \frac{1}{2}[1 + \rho(U_1, U_2)]$. These formulae give a direct relationship between the dependence of (U_1, U_2) and that of (X_1, X_2).

The density of (X_1, X_2) under identical generators is given by

$$f(x_1,x_2)=\int_{-\infty}^{\infty}\frac{f'_X(w)}{w}\,g(x_1/w,x_2/w)\,dw,$$

where g is the density of (U_1, U_2) and f_X is the density common to X_1 and X_2.

In either case of the independent or identical generators structure, a method for generating the classical bivariate unfiorm distributions is required. Algorithms for this purpose are available in Johnson, Ramberg and Wang (1982). Contour and 3-D plots of these distributions for various correlation structures are given in Johnson, Bryson and Mills (1981).

4. KHINTCHINE MIXTURES

An alternative approach utilizing Khintchine's theorem involves mixture distributions where either the uniform variables or the generator variables depend on a mixing parameter. Specifically, we define as before $X_1 = Z_1 U_1$ and $X_2 = Z_2 U_2$, where the U_i's are uniform (0, 1), Z_i is the generator selected to give the desired marginals, and the pairs U_i and Z_i are independent. Two possible mixing formats are considered:

1) U_1 and U_2 are independent; Z_1 and Z_2 are identical with probability p and independent with probability $1-p$.

2) Z_1 and Z_2 are independent; U_1 and U_2 are identical with probability q and independent with probability $1-q$.

These possibilities are explored with exponential and normal marginals.

a) *Exponential marginals.* To achieve exponential marginals, the generator variables must be gamma with shape parameter 2, scale parameter 1. With identical generators and independent uniforms the density is

$$f_1(x_1, x_2) = -Ei(-x_{\max}),$$

where $x_{\max}$ = maximum (x_1, x_2), and $Ei(x)$ is the exponential integral function. With identical uniforms and independent generators the density is

$$f_2(x_1, x_2) = \frac{x_1 x_2}{(x_1+x_2)^3}[2+2(x_1+x_2)+(x_1+x_2)^2]\,e^{-(x_1+x_2)}. \tag{4.1}$$

These densities can be mixed with the totally independent case with density $f_3(x_1, x_2) = e^{-(x_1+x_2)}$ to give density

$$f(x_1, x_2) = pf_1(x_1, x_2) + qf_2(x_1, x_2) + (1-p-q)f_3(x_1, x_2) \tag{4.2}$$

having correlation equal to $p/2 + q/3$.

b) *Normal marginals.* As noted in Section 2, normal marginals are obtained from $X = ZU$ if Z has $\pm\sqrt{\chi^2_{(3)}}$ distribution where "$\pm$" reflects a random sign. The distribution with independent uniforms and identical generators has density function

$$f_1(x_1, x_2) = \begin{cases} 1-\Phi(x_{\max}) & \text{for } x_1 > 0 \text{ and } x_2 > 0 \\ \Phi(x_{\min}) & \text{for } x_1 < 0 \text{ and } x_2 < 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

where Φ is the standard normal distribution function. Note that the support of this density is restricted to the first and third quadrants. With independent generators and identical uniforms, the density function is

$$f_2(x_1, x_2) = \frac{3}{\sqrt{\pi}\,2^{3/2}} \frac{x_1^2 x_2^2}{(x_1^2+x_2^2)^{5/2}}\left[1 - H\left(\frac{2}{x_1^2+x_2^2}\right)\right], \tag{4.4}$$

where H is the gamma distribution function with shape parameter 5/2 and scale parameter 1. This density has support in all quadrants but has zero correlation.

Most generally we can consider the density function

$$f(x_1,x_2)=pf_1(x_1,x_2)+qf_2(x_1,x_2)+(1-p-q)f_3(x_1,x_2),$$

where f_3 is the usual bivariate normal density with zero correlation and standard marginals. If $q=0$, we have case (1); if $p=0$, we have case (2). For any $p+q=1$, this density has normal $(0,1)$ marginals and full support in the plane if $p\neq 1$. The correlation is given by $\rho(X_1,X_2)=\frac{3}{4}p$. Note that although q does not enter into the correlation explicitly, it has a strong effect on the shape of f. Again, plots of f for various parameter values of p and q are given in Johnson, Bryson and Mills (1981).

5. AN ALTERNATIVE FORMULATION

For the common situation of normal marginals, the simple cases of section 3 may be altered with the construction:

$$\begin{aligned} Y_1&=X_1(2U_1-1)\\ Y_2&=X_2(2U_2-1), \end{aligned} \tag{5.1}$$

where X_i is $\sqrt{\chi^2_{(3)}}$, U_1 is uniform $0-1$ and X_i and U_i are independent. The Y_i's are normal $(0,1)$. This construction has the effect of giving smoother densities throughout the plane—in particular, discontinuities on the axes are avoided.

The joint density of Y_1 and Y_2 using (5.1) is (for the case $X_1=X_2$)

$$g(y_1,y_2)=\int_{\max(|y_1|,|y_2|)}^{\infty}\frac{1}{\Gamma(3/2)\,2^{5/2}}\,e^{-t^2/2}\,h\left(\frac{y_1}{2t}+\frac{1}{2},\frac{y_2}{2t}+\frac{1}{2}\right)dt, \tag{5.2}$$

where h is a bivariate uniform density on the unit square. For example, if h is constant, then we have

$$g(y_1,y_2)=\tfrac{1}{2}\{1-\Phi[\max(|y_1|,|y_2|)]\},$$

which is a bivariate density with normal $(0,1)$ marginals. This form can be contrasted with (4.3) which had mass restricted to the first and third quadrants.

If h has the Morgenstern bivariate uniform density, then the following density function is obtained:

$$g(y_1, y_2) = \frac{\alpha y_1 y_2}{2A} \phi(A) + \frac{1 - \alpha y_1 y_2}{2} [1 - \Phi(A)], \tag{5.3}$$

where $A = \max |y_1|, |y_2|$ and ϕ is the standard normal density function. Contour plots for $\alpha = -1$, 0 and 1 are given in Figure 1. For the case $\alpha = 0$, the random variables are uncorrelated but not independent.

A similar set of bivariate distributions were devised by Patil (1969). He also obtained normally distributed random variables that are uncorrelated but not independent.

6. POSSIBLE MULTIVARIATE EXTENSIONS

The multivariate extension of the construction procedure is straightforward:

$$\begin{aligned} X_1 &= Z_1 U_1 \\ X_2 &= Z_2 U_2 \\ &\cdots \\ X_n &= Z_n U_n, \end{aligned} \tag{6.1}$$

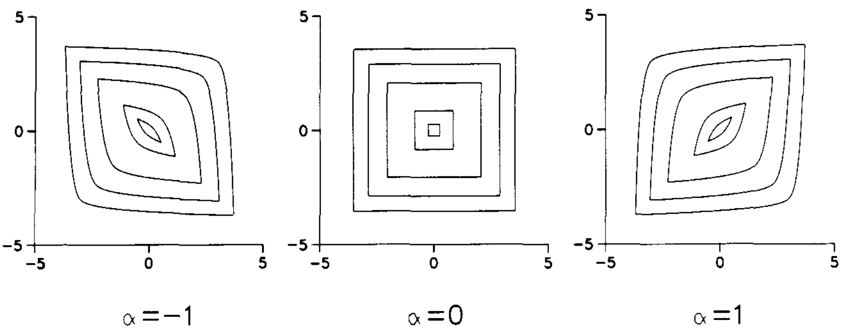

FIGURE 1 Typical contours for the density function of (5.3).

where the U_i's are uniform $(0, 1)$, the Z_i's have the distribution necessary to attain the specified X_i distributions, and (Z_i, U_i) are independent for $i = 1, 2, \ldots, n$. The practical exigencies of attaining "interesting" $\boldsymbol{X}$ distributions primarily rest with models for U. In particular, the Morgenstern, Plackett and Gumbel bivariate uniform distributions do not generalize to multivariate uniforms which are suitable (either they do not generalize to higher dimensions or the resultant multivariate family models only weak dependence). The multivariate Cauchy could be used to attain a multivariate uniform but the pairwise correlations would be zero. The two leading candidates for the $\boldsymbol{U}$ distributions are based on the multivariate normal and the multivariate uniform developed by Cook and Johnson (1981). In either case we can use the mixture approach and select the Z_i's identical or independent as desired. For certain choices of marginals (exponential or normal) it would be possible to consider non-degenerate, dependent distributions for $\boldsymbol{Z}$ (select a suitable multivariate gamma-type distribution). Analytical tractibility would then be expedited by considering identical or independent uniforms.

References

Barnett, V. (1980). Some bivariate uniform distributions. *Communications in Statistics* **A9(4)**, 453–461.

Cook, R. D. and Johnson, M. E. (1981). A family of distributions for modelling non-elliptically symmetric multivariate data. *Journal of the Royal Statistical Society* **43,** 210–218.

de Silva, B. M. (1978). A class of multivariate symmetric stable distributions. *Journal of Multivariate Analysis* **8,** 335–345.

Devroye, L. (1981). Random variate generation for unimodal and monotone densities. School of Computer Science, McGill University (personal communication: submitted for publication).

Feller, W. (1971). *An Introduction of Probability Theory and Its Applications, Vol. II.* John Wiley and Sons, New York.

Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association* **55,** 698–707.

Johnson, M. E., Bryson, M. C. and Mills, C.F. (1981). Some new multivariate distributions with enhanced comparisons via contour and three-dimensional plots. Los Alamos National Laboratory report, LA-8903-MS.

Johnson, M. E., Ramberg, J. S. and Wang, C. (1981). Generation of multivariate distributions with a slant towards statistical applications, in preparation.

Johnson, M. E., Tietjen, G. L. and Beckman, R. J. (1980). A new family of probability distributions with applications to Monte Carlo studies. *Journal of the American Statistical Association* **75,** 276–279.

Johnson, N. L. and Kotz, S. (1972). *Distributions is Statistics: Continuous Multivariate Distributions.* John Wiley & Sons, New York.

Johnson, N. L. and Rogers, C. A. (1951). The moment problem for unimodal distributions. *Annals of Mathematical Statistics* **22,** 433–439.

Khintchine, A. Y. (1938). On unimodal distributions. *Izvch. Nauchno-Issled. Inst. Mat. Mekh. Tomsk. Gos. Univ.*, **2,** 1–7.

Kowalski, C. J. (1973). Non-normal bivariate distributions with normal marginals. *The American Statistician* **27,** 103–106.

Mardia, K. V. (1970). Families of bivariate distributions, Hafner, Darien, Connecticut.

Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitteilungs. Math. Stat.*, **8,** 234–235.

Olshen, R. and Savage, L. J. (1970). A generalized unimodality. *Journal of Applied Probability* **7,** 21–34.

Patil, S. A. (1969). A bivariate distribution of product of beta variables and square root of chi-square variable, *Sankhya, Series B* **31,** 25–28.

Plackett, R. L. (1965). A class of bivariate distributions. *Journal of the American Statistical Association* **60,** 516–522.