# Chapter Eleven
# MULTIVARIATE DISTRIBUTIONS

## 1. GENERAL PRINCIPLES.

### 1.1. Introduction.

In section V.4, we have discussed in great detail how one can efficiently generate random vectors in $R^d$ with radially symmetric distributions. Included in that section were methods for generating random vectors uniformly distributed in and on the unit sphere $C_d$ of $R^d$. For example, when $N_1, \ldots, N_d$ are iid normal random variables, then

$$\left(\frac{N_1}{N}, \ldots, \frac{N_d}{N}\right)$$

where $N = \sqrt{N_1^2 + \cdots + N_d^2}$, is uniformly distributed on the surface of $C_d$. This uniform distribution is the building block for all radially symmetric distributions because these distributions are all scale mixtures of the uniform distribution on the surface of $C_d$. This sort of technique is called a special property technique: it exploits certain characteristics of the distribution. What we would like to do here is give several methods of attacking the generation problem for $d$-dimensional random vectors, including many special property techniques.

The material has little global structure. Most sections can in fact be read independently of the other sections. In this introductory section several general principles are described, including the conditional distribution method. There is no analog to the univariate inversion method. Later sections deal with specific subclasses of distributions, such as uniform distributions on compact sets, elliptically symmetric distributions (including the multivariate normal distribution), bivariate uniform distributions and distributions on lines.

### 1.2. The conditional distribution method.

The conditional distribution method allows us to reduce the multivariate generation problem to $d$ univariate generation problems, but it can only be used when quite a bit of information is known about the distribution.

Assume that our random vector $\mathbf{X}$ has density

$$f(x_1, \ldots, x_d) = f_1(x_1) f_2(x_2 \mid x_1) \cdots f_d(x_d \mid x_1, \ldots, x_{d-1}),$$

where the $f_i$'s are conditional densities. Generation can proceed as follows:

**Conditional distribution method**

FOR $i:=1$ TO $d$ DO
    Generate $X_i$ with density $f_i(. \mid X_1, \ldots, X_{i-1})$. (For $i=1$, use $f_1(.)$.)
RETURN $\mathbf{X}=(X_1, \ldots, X_d)$

It is necessary to know all the conditional densities. This is equivalent to knowing all marginal distributions, because

$$f_i(x_i \mid x_1, \ldots, x_{i-1}) = \frac{f_i^*(x_1, \ldots, x_i)}{f_{i-1}^*(x_1, \ldots, x_{i-1})}$$

where $f_i^*$ is the marginal density of the first $i$ components, i.e. the density of $(X_1, \ldots, X_i)$.

### Example 1.1. The multivariate Cauchy distribution.

The multivariate Cauchy density $f$ is given by

$$f(x) = \frac{c}{(1+\|x\|^2)^{\frac{d+1}{2}}},$$

where $c=\Gamma(\frac{d+1}{2})/\pi^{(d+1)/2}$. Here $\| . \|$ is the standard $L_2$ Euclidean norm. It is known that $X_1$ is univariate Cauchy, and that given $X_1, \ldots, X_{i-1}$, the random variable $X_i$ is distributed as $T(1+\sum_{j=1}^{i-1} X_j)/\sqrt{i}$ where $T$ has the t distribution with $i$ degrees of freedom (Johnson and Kotz, 1970). ■

**Example 1.2. The normal distribution.**

Assume that $f$ is the density of the zero mean normal distribution on $R^2$, with variance-covariance matrix $\mathbf{A}=\{a_{ij}\}$ where $a_{ij}=E(X_i X_j)$:

$$f(x)=\frac{1}{2\pi\sqrt{|\mathbf{A}|}}e^{-\frac{1}{2}x'\mathbf{A}^{-1}x} .$$

In this case, the conditional density method yields the following algorithm:

**Conditional density method for normal random variates**

Generate $N_1,N_2$, iid normal random variates.

$X_1 \leftarrow N_1\sqrt{a_{11}}$

$X_2 \leftarrow \frac{a_{21}}{a_{11}}X_1+N_2\sqrt{\frac{a_{22}a_{11}-a_{21}^2}{a_{11}}}$

RETURN $(X_1,X_2)$

This follows by noting that $X_1$ is zero mean normal with variance $a_{11}$, and computing the conditional density of $X_2$ given $X_1$ as a ratio of marginal densities. ■

**Example 1.3.**

Let $f$ be the uniform density in the unit circle $C_2$ of $R^2$. The conditional density method is easily obtained:

Generate $X_1$ with density $f_1(x)=\frac{2}{\pi}\sqrt{1-x^2}$ $(|x|\leq 1)$.

Generate $X_2$ uniformly on $[-\sqrt{1-X_1^2},\sqrt{1-X_1^2}]$.

RETURN $(X_1,X_2)$ ■

In all three examples, we could have used alternative methods. Examples 1.1 and 1.2 deal with easily treated radially symmetric distributions, and Example 1.3 could have been handled via the ordinary rejection method.

## 1.3. The rejection method.

It should be clear that the rejection method is not tied to a particular space. It can be used in multivariate random variate generation problems, and is probably the most useful general purpose technique here. A few traps to watch out for are worth mentioning. First of all, rejection from a uniform density on a rectangle of $R^d$ often leads to a rejection constant which deteriorates quickly as $d$ increases. A case in point is the rejection method for generating points uniformly in the unit sphere of $R^d$ (see section V.4.3). Secondly, unlike in $R^1$, upper bounds for certain densities are not easily obtainable. For example, the information that $f$ is unimodal with a mode at the origin is of little use, whereas in $R^1$, the same information allows us to conclude that $f(x) \leq 1/|x|$. Similarly, combining unimodality with moment conditions is not enough. Even the fact that $f$ is log-concave is not sufficient to derive universally applicable upper bounds (see section VII.2).

In general, the design of an efficient rejection method is more difficult than in the univariate case.

## 1.4. The composition method.

The composition method is not tied to a particular space such as $R^1$. A popular technique for obtaining dependence from independence is the following: define a random vector $\mathbf{X}=(X_1, \ldots, X_d)$ as $(SY_1, \ldots, SY_d)$ where the $S_i$'s are iid random variables, and $S$ is a random scale. In such cases, we say that the distribution of $\mathbf{X}$ is a **scale mixture.** If $Y_1$ has density $f$, then $\mathbf{X}$ has a density given by

$$E\left\{\prod_{i=1}^{d}\left(\frac{1}{S} f\left(\frac{x_i}{S}\right)\right)\right\} .$$

If $Y_1$ has distribution function $F=1-G$, then

$$P(X_1>x_1, \ldots, X_d>x_d) = E\left(\prod_{i=1}^{d} G\left(\frac{x_i}{S}\right)\right) .$$

**Example 1.4. The multivariate Burr distribution.**

When $Y_1$ is Weibull with parameter $a$ (i.e. $G(y)=e^{-y^a}$ $(y>0)$), and $S$ is gamma $(b)$, then $(SY_1, \ldots, SY_d)$ has distribution function determined by

$$P(X_1>x_1, \ldots, X_d>x_d) = E(\prod_{i=1}^{d} e^{-(x_i/S)^a})$$

$$= \int_0^\infty \frac{s^{b-1}e^{-s}}{\Gamma(b)} e^{-s^{-a}(\sum_{i=1}^{d} x_i^a)} ds$$

$$= \frac{1}{(1+\sum_{i=1}^{d} x_i^a)^b} \quad (x_i>0 , i=1,2, \ldots, d) .$$

This defines the multivariate Burr distribution of Takahasi (1965). From this relation it is also easily seen that all univariate or multivariate marginals of a multivariate Burr distribution are univariate or multivariate Burr distributions. For more examples of scale mixtures in which $S$ is gamma, see Hutchinson (1981). ∎

**Example 1.5. The multinomial distribution.**

The conditional distribution method is not limited to continuous distributions. For example, consider the **multinomial distribution** with parameters $n, p_1, \ldots, p_d$ where the $p_i$'s form a probability vector and $n$ is a positive integer. A random vector $(X_1, \ldots, X_d)$ is multinomially distributed with these parameters when

$$P((X_1, \ldots, X_d)=(i_1, \ldots, i_d)) = \frac{n!}{\prod_{j=1}^{d} i_j!} \prod_{j=1}^{d} p_j^{i_j}$$

$$(i_j \geq 0 , j=1, \ldots, d ; \sum_{j=1}^{d} i_j=n) .$$

This is the distribution of the cardinalities of $d$ urns into which $n$ balls are thrown at random and independently of each other. Urn number $j$ is selected with probability $p_j$ by every ball. The ball-in-urn experiment can be mimicked, which leads us to an algorithm taking time $O(n+d)$ and $\Omega(n+d)$. Note however that $X_1$ is binomial $(n, p_1)$, and that given $X_1$, the vector $(X_2, \ldots, X_d)$ is multinomial $(n-X_1, q_2, \ldots, q_d)$ where $q_j=p_j/(1-p_1)$. This recurrence relation is nothing but another way of describing the conditional distribution method for this case. With a uniformly fast binomial generator we can proceed in expected time $O(d)$ uniformly bounded in $n$ :

**Multinomial random vector generator**

[NOTE: the parameters $n, p_1, \ldots, p_d$ are destroyed by this algorithm. Sum holds a cumulative sum of probabilities.]

Sum $\leftarrow 0$

FOR $i:=1$ TO $d$ DO

Generate a binomial $(n, \frac{p_i}{S})$ random vector $X_i$.

$n \leftarrow n - X_i$

Sum $\leftarrow$ Sum - $p_i$

For small values of $n$, it is unlikely that this algorithm is very competitive, mainly because the parameters of the binomial distribution change at every call. ■

## 1.5. Discrete distributions.

Consider the problem of the generation of a random vector taking only values on $d$-tuples of nonnegative integers. One of the striking differences with the continuous multivariate distributions is that the $d$-tuples can be put into one-to-one correspondence with the nonnegative integers on the real line. This one-to-one mapping can be used to apply the inversion method (Kemp, 1981; Kemp and Loukas, 1978) or one of the table methods (Kemp and Loukas, 1981). We say that the function which transforms $d$-tuples into nonnegative integers is a **coding function.** The inverse function is called the **decoding function.**

Coding functions are easy to construct. Consider $d=2$. Then we can visit all 2-tuples in the positive quadrant in cross-diagonal fashion. Thus, first we visit (0,0), then (0,1) and (1,0), then (0,2),(1,1) and (2,0), etcetera. Note that we visit all the integers $(i,j)$ with $i+j=k$ before visiting those with $i+j=k+1$. Since we visit $k(k-1)/2$ 2-tuples with $i+j<k$, we see that we can take as coding function

$$h(i,j) = \frac{(i+j)(i+j-1)}{2}+i \; .$$

This can be generalized to $d$-tuples (exercise 1.4), and a simple decoding function exists which allows us to recover $(i,j)$ from the value of $h(i,j)$ in time $O(1)$ (exercise 1.4). There are other orders of traversal of the 2-tuples. For example, we could visit 2-tuples in order of increasing values of $\max(i,j)$.

In general one cannot visit all 2-tuples in order of increasing values of $i$, its first component, as there could be an infinite number of 2-tuples with the same value of $i$. It is like trying to visit all shelves in a library, and getting stuck in the first shelf because it does not end. If the second component is bounded, as it often is, then the library traversal leads to a simple coding function. Let $M$ be the maximal value for $j$. Then we have

$$h(i,j) = (M+1)i+j.$$

One should be aware of some pitfalls when the univariate connection is exploited. Even if the distribution of probability over the $d$-tuples is relatively smooth, the corresponding univariate probability vector is often very oscillatory, and thus unfit for use in the rejection method. Rejection should be applied almost exclusively to the original space.

The fast table methods require a finite distribution. Even though on paper they can be applied to all finite distributions, one should realize that the number of possible $d$-tuples in such distributions usually explodes exponentially with $d$. For a distribution on the integers in the hypercube $\{1,2, \ldots, n\}^d$, the number of possible values is $n^d$. For this example, table methods seem useful only for moderate values of $d$. See also exercise 1.5.

Kemp and Loukas (1978) and Kemp (1981) are concerned with the inversion method and its efficiency for various coding functions. Recall that in the univariate case, inversion by sequential search for a nonnegative integer-valued random variate $X$ takes expected time (as measured by the expected number of comparisons) $E(X)+1$. Thus, with the coding function $h$ for $X_1, \ldots, X_d$, we see without further work that the expected number of comparisons is

$$E(h(X_1, \ldots, X_d)+1) .$$

**Example 1.6.**

Let us apply inversion for the generation of $(X_1,X_2)$, and let us scan the space in cross diagonal fashion (the coding function is $h(i,j) = \frac{(i+j)(i+j-1)}{2}+i$). Then the expected number of comparisons before halting is

$$E\left( \frac{(X_1+X_2)(X_1+X_2-1)}{2}+X_1+1 \right) .$$

This is at least proportional to either one of the marginal second moments, and is thus much worse than one would normally have expected. In fact, in $d$ dimensions, a similar coding function leads to a finite expected time if and only if $E(X_i^d)<\infty$ for all $i=1, \ldots, d$ (see exercise 1.6). ■

**Example 1.7.**

Let us apply inversion for the generation of $(X_1,X_2)$, where $0 \le X_2 \le M$, and let us perform a library traversal (the coding function is $h(i,j) = (M+1)i+j$). Then the expected number of comparisons before halting is

$$E\left((M+1)X_1+X_2+1\right) .$$

This is finite when only the first moments are finite, but has the drawback that $M$ figures explicitly in the complexity. ■

We have made our point. For large values of $d$, ordinary generation methods are often not feasible because of time or space inefficiencies. One should nearly always try to convert the problem into several univariate problems. This can be done by applying the conditional distribution method. For the generation of $X_1,X_2$, we first generate $X_1$, and then generate $X_2$ conditional on the given value of $X_1$. Effectively, this forces us to know the marginal distribution of $X_1$ and the joint two-dimensional distribution. The marginal distribution of $X_2$ is not needed. To see how this improves the complexities, consider using the inversion method in both stages of the algorithm. The expected number of comparisons in the generation of $X_2$ given $X_1$ is $E(X_2 \mid X_1)+1$. The number of comparisons in the generation of $X_1$ is $X_1+1$. Summing and taking expected values shows that the expected number of comparisons is

$$E(X_1+X_2+2)$$

(Kemp and Loukas, 1978). Compare with Examples 1.6 and 1.7.

In the conditional distribution method, we can improve the complexity even further by employing table methods in one, some or all of the stages. If $d=2$ and both components have infinite support, we cannot use tables. If only the second component has infinite support, then a table method can be used for $X_1$. This is the ideal situation. If both components have finite support, then we are tempted to apply the table method in both stages. This would force us to set up many tables, one for each of the possible values of $X_1$. In that case, we could as well have set up one giant table for the entire distribution. Finally, if the first component has infinite support, and the second component has finite support, then the incapability of storing an infinite number of finite tables forces us to set up the tables as we need them, but the time spent doing so is prohibitively large.

If a distribution is given in analytic form, there usually is some special property which can be used in the design of an efficient generator. Several examples can be found in section 3.

## 1.6. Exercises.

1. Consider the density $f(x_1,x_2)=5x_1e^{-x_1x_2}$ defined on the infinite strip $0.2 \le x_1 \le 0.4$, $0 \le x_2$. Show that the first component $X_1$ is uniformly distributed on $[0.2,0.4]$, and that given $X_1$, $X_2$ is distributed as an exponential random variable divided by $X_1$ (Schmeiser, 1980).

2. Show how you would generate random variates with density

$$\frac{6}{(1+x_1+x_2+x_3)^4} \quad (x_1,x_2,x_3 \ge 0) .$$

Show also that $X_1+X_2+X_3$ has density $3x^2/(1+x)^4$ $(x \ge 0)$ (Springer, 1979, p.87).

3. Prove that for any distribution function $F$ on $R^d$, there exists a measurable function $g:[0,1] \to R^d$ such that $g(U)$ has distribution function $F$, where $U$ is uniformly distributed on $[0,1]$. This can be considered as a generalization of the inversion method. Hint: from $U$ we can construct $d$ iid uniform $[0,1]$ random variables by skipping bits. Then argue via conditioning.

4. Consider the coding function for 2-tuples of nonnegative integers $(i,j)$ given by $h(i,j)=\frac{(i+j)(i+j-1)}{2}+i+1$.

   A. Generalize this coding function to $d$-tuples. The generalization should be such that all $d$-tuples with sum of the components equal to some integer $k$ are grouped together, and the groups are ordered according to increasing values for $k$. Within a group, this rule should be applied recursively to groups of $d-1$-tuples with constant sum.

   B. Give the decoding function for the two-dimensional $h$ shown above, and indicate how it can be evaluated in time $O(1)$ (independent of the size of the argument).

5. Consider the multinomial distribution with parameters $n, p_1, \ldots, p_d$, which assigns probability

$$\frac{n!}{i_1! \cdots i_d!} \prod_{j=1}^{d} p_j^{\,i_j}$$

to all $d$-tuples with $i_j \ge 0$, $\sum_{j=1}^{d} i_j = n$. Let the total number of possible values be $N(n,d)$. For fixed $n$, find a simple function $\psi(d)$ with the property that

$$\lim_{d \to \infty} \frac{N(n,d)}{\psi(d)} = 1 .$$

This gives some idea about how quickly $N(n,d)$ grows with $d$.

6. Show that when a cross-diagonal traversal is followed in $d$ dimensions for inversion by sequential search of a discrete probability distribution on the nonnegative integers of $R^d$, then the expected time required by the inversion is finite if and only if $E(X_i^{\,d}) < \infty$ for all $i=1, \ldots, d$ where $X_1, \ldots, X_d$ is a $d$-dimensional random vector with the given distribution.

7. **Relationship between multinomial and Poisson distributions.** Show that the algorithm given below in which the sample size parameter is used as a mixing parameter delivers a sequence of $d$ iid Poisson ($\lambda$) random variables.

> Generate a Poisson ($d\lambda$) random variate $N$.
>
> RETURN a multinomial $(N, \frac{1}{d}, \ldots, \frac{1}{d})$ random vector $(X_1, \ldots, X_d)$.

Hint: this can be proved by explicitly computing the probabilities, by working with generating functions, or by employing properties of Poisson point processes.

8. **A bivariate extreme value distribution.** Marshall and Olkin (1983) have studied multivariate extreme value distributions in detail. One of the distributions considered by them is defined by

$$P(X_1 > x_1, X_2 > x_2) = e^{-(e^{-x_1}+e^{-x_2}-(e^{x_1}+e^{x_2})^{-1})} \quad (x_1 \geq 0, x_2 \geq 0) .$$

How would you generate a random variate with this distribution?

9. Let $f$ be an arbitrary univariate density on $(0,\infty)$. Show that $f(x_1+x_2)/(x_1+x_2)$ $(x_1>0, x_2>0)$ is a bivariate density (Feller, 1971, p.100). Exploiting the structure in the problem to the fullest, how would you generate a random vector with the given bivariate density?

## 2. LINEAR TRANSFORMATIONS. THE MULTINORMAL DISTRIBUTION.

### 2.1. Linear transformations.

When an $R^d$-valued random vector $\mathbf{X}$ has density $f(\mathbf{x})$, then the random vector $\mathbf{Y}$ defined as the solution of $\mathbf{X}=\mathbf{HY}$ has density

$$g(\mathbf{y}) = |\mathbf{H}| f(\mathbf{Hy}), \ \mathbf{y} \in R^d ,$$

for all nonsingular $d \times d$ matrices $\mathbf{H}$. The notation $|\mathbf{H}|$ is used for the absolute value of the determinant of $\mathbf{H}$. This property is reciprocal, i.e. when $Y$ has density $g$, then $\mathbf{X}=\mathbf{HY}$ has density $f$.

The linear transformation $\mathbf{H}$ deforms the coordinate system. Particularly important linear deformations are rotations: these correspond to orthonormal transformation matrices $\mathbf{H}$. For random variate generation, linear transformations are important in a few special cases:

A. The generation of points uniformly distributed in $d$-dimensional simplices or hyperellipsoids.

B. The generation of random vectors with a given dependence structure, as measured by the covariance matrix.

These two application areas are now dealt with separately.

## 2.2. Generators of random vectors with a given covariance matrix.

The covariance matrix of an $R^d$-valued random vector $\mathbf{Y}$ with mean 0 is defined as $\Sigma = E(\mathbf{YY'})$ where $\mathbf{Y}$ is considered as a column vector, and $\mathbf{Y'}$ denotes the transpose of $\mathbf{Y}$. Assume first that we wish to generate a random vector $\mathbf{Y}$ with zero mean and covariance matrix $\Sigma$ and that we do not care for the time being about the form of the distribution. Then, it is always possible to proceed as follows: generate a random vector $\mathbf{X}$ with $d$ iid components $X_1, \ldots, X_d$ each having zero mean and unit variance. Then define $\mathbf{Y}$ by $\mathbf{Y}=\mathbf{HX}$ where $\mathbf{H}$ is a nonsingular $d \times d$ matrix. Note that

$$E(\mathbf{Y}) = \mathbf{H}E(\mathbf{X}) = 0 ,$$

$$E(\mathbf{YY'}) = \mathbf{H}E(\mathbf{XX'})\mathbf{H'} = \mathbf{HH'} = \Sigma .$$

We need a few facts now from the theory of matrices. First of all, we recall the definition of positive definiteness. A matrix $\mathbf{A}$ is positive definite (positive semidefinite) when $\mathbf{x'Ax} > 0$ $(\geq 0)$ for all nonzero $R^d$-valued vectors $\mathbf{x}$. But we have

$$\mathbf{x}'\Sigma\mathbf{x} \doteq E(\mathbf{x'YY'x}) = E(\,|\,|\mathbf{x'Y}|\,|\,) \geq 0$$

for all nonzero $\mathbf{x}$. Here $|\,|.|\,|$ is the standard $L_2$ norm in $R^d$. Equality occurs only if the $Y_i$'s are linearly dependent with probability one, i.e. $\mathbf{x'Y}=0$ with probability one for some $\mathbf{x} \neq 0$. In that case, $\mathbf{Y}$ is said to have dimension less than $d$. Otherwise, $\mathbf{Y}$ is said to have dimension $d$. Thus, all covariance matrices are positive semidefinite. They are positive definite if and only if the random vector in question has dimension $d$.

For symmetric positive definite matrices $\Sigma$, we can always find a nonsingular matrix $\mathbf{H}$ such that

$$\mathbf{HH'} = \Sigma .$$

In fact, such matrices can be characterized by the existence of a nonsingular $\mathbf{H}$. We can do even better. One can always find a lower triangular nonsingular $\mathbf{H}$ such that

$$\mathbf{HH'} = \Sigma .$$

We have now turned our problem into one of decomposing a symmetric positive definite matrix $\Sigma$ into a product of two lower triangular matrices. The algorithm can be summarized as follows:

**Generator of a random vector with given covariance matrix**

[SET-UP]

Find a matrix $\mathbf{H}$ such that $\mathbf{HH'}=\Sigma$.

[GENERATOR]

Generate $d$ independent zero mean unit variance random variates $X_1, \ldots, X_d$.

RETURN $\mathbf{Y}=\mathbf{HX}$

The set-up step can be done in time $O(d^3)$ as we will see below. Since $\mathbf{H}$ can have up to $\Omega(d^2)$ nonzero elements, there is no hope of generating $\mathbf{Y}$ in less than $\Omega(d^2)$. Note also that the distributions of the $X_i$'s are to be picked by the users. We could take them iid and biatomic: $P(X_1=1)=P(X_1=-1)=\frac{1}{2}$. In that case, $Y$ is atomic with up to $2^d$ atoms. Such atomic solutions are rarely adequate. Most applications also demand some control over the marginal distributions. But these demands restrict our choices for $X_1$. Indeed, if our method is to be universal, we should choose $X_1, \ldots, X_d$ in such a way that all linear combinations of these independent random variables have a given distribution. This can be assured in several ways, but the choices are limited. To see this, let us consider iid random variables $X_i$ with common characteristic function $\phi$, and assume that we wish all linear combinations to have the same distribution up to a scale factor. The sum $\sum a_j X_j$ has characteristic function

$$\prod_{j=1}^{d} \phi(a_j t) .$$

This is equal to $\phi(at)$ for some constant $a$ when $\phi$ has certain functional forms. Take for example

$$\phi(t) = e^{-|t|^\alpha}$$

for some $\alpha \in (0,2]$ as in the case of a symmetric stable distribution. Unfortunately, the only symmetric stable distribution with a finite variance is the normal distribution ($\alpha=2$). Thus, the property that the normal distribution is closed under the operation "linear combination" is what makes it so attractive to the user. If the user specifies non-normal marginals, the covariance structure is much more difficult to enforce. See however some good solutions for the bivariate case as developed in section XI.3.

A computational remark about $\mathbf{H}$ is in order here. There is a simple algorithm known as the **square root method** for finding a lower triangular $\mathbf{H}$ **with** $\mathbf{HH'} = \Sigma$ (Faddeeva, 1959; Moonan, 1957; Graybill, 1969). We give the relationship between the matrices here. The elements of $\Sigma$ are called $\sigma_{ij}$, and those of the lower triangular solution matrix $\mathbf{H}$ are called $h_{ij}$.

$$h_{i1}=\sigma_{i1}/\sqrt{\sigma_{11}} \quad (1 \le i \le d)$$

$$h_{ii}=\sqrt{\sigma_{ii}-\sum_{j=1}^{i-1} h_{ij}^{2}} \quad (1<i \le d)$$

$$h_{ij}=\frac{\sigma_{ij}-\sum_{k=1}^{j-1} h_{ik}\,h_{jk}}{h_{jj}} \quad (1<j<i \le d)$$

$$h_{ij}=0 \quad (i<j \le d)$$

## 2.3. The multinormal distribution.

The **standard multinormal distribution** on $R^d$ has density

$$f(\mathbf{x}) = (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}\mathbf{x}'\mathbf{x}}$$
$$= (2\pi)^{-\frac{d}{2}} e^{-\frac{1}{2}||\mathbf{x}||^2} \quad (\mathbf{x}\in R^d) .$$

This is the density of $d$ iid normal random variables. When $\mathbf{X}$ has density $f$, $\mathbf{Y}=\mathbf{H}\mathbf{X}$ has density

$$g(\mathbf{y}) = |\mathbf{H}^{-1}| f(\mathbf{H}^{-1}\mathbf{y}) , \mathbf{y}\in R^d .$$

But we know that $\Sigma=\mathbf{H}\mathbf{H}'$, so that $|\mathbf{H}^{-1}|=|\Sigma|^{-1/2}$. Also, $||\mathbf{H}^{-1}\mathbf{y}||^2=\mathbf{y}'\Sigma^{-1}\mathbf{y}$, which gives us the density

$$g(\mathbf{y}) = (2\pi)^{-\frac{d}{2}} |\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}\mathbf{y}'\Sigma^{-1}\mathbf{y}} \quad (\mathbf{y}\in R^d) .$$

This is the density of the multinormal distribution with zero mean and nonsingular covariance matrix $\Sigma$. We note without work that the $i$-th marginal distribution is zero mean normal with variance given by the $i$-th diagonal element of $\Sigma$. In the most general form of the normal distribution, we need only add a translation parameter (mean) to the distribution.

Random variate generation for the normal distribution can be done by the linear transformation of $d$ iid normal random variables described in the previous section. This involves decomposition of $\Sigma$ into a product of the form $\mathbf{H}\mathbf{H}'$. This method has been advocated by Scheuer and Stoller (1962) and Barr and Slezak (1972). Deak (1979) gives other methods for generating multinormal random vectors. For the conditional distribution method in the case $d=2$, we refer to Example 1.2. In the general case, see for example Scheuer and Stoller (1962).

An important special case is the bivariate multinormal distribution with zero mean, and covariance matrix

$$\begin{vmatrix} 1 & \rho \\ \rho & 1 \end{vmatrix} .$$

where $\rho \in [-1,1]$ is the correlation between the two marginal random variables. It is easy to see that if $(N_1,N_2)$ are iid normally distributed random variables, then

$$(N_1, \rho N_1 + \sqrt{1-\rho^2} N_2)$$

has the said distribution. The multinormal distribution can be used as the starting point for creating other multivariate distributions, see section XI.3. We will also exhibit many multivariate distributions with normal marginals which are not multinormal. To keep the terminology consistent throughout this book, we will refer to all distributions having normal marginals as multivariate normal distributions. Multinormal distributions form only a tiny subclass of the multivariate normal distributions.

### 2.4. Points uniformly distributed in a hyperellipsoid.

A hyperellipsoid in $R^d$ is defined by a symmetric positive definite $d \times d$ matrix $\mathbf{A}$: it is the collection of all points $\mathbf{y} \in R^d$ with the property that

$$\mathbf{y}'\mathbf{A}\mathbf{y} \leq 1 .$$

A random vector uniformly distributed in this hyperellipsoid can be generated by a linear transformation of a random vector $\mathbf{X}$ distributed uniformly in the unit hypersphere $C_d$ of $R^d$. Such random vectors can be generated quite efficiently (see section V.4). Recall that linear transformations cannot destroy uniformity. They can only alter the shape of the support of uniform distributions. The only problem we face is that of the determination of the linear transformation in function of $\mathbf{A}$.

Let us define $\mathbf{Y}=\mathbf{H}\mathbf{X}$ where $\mathbf{H}$ is our $d \times d$ transformation matrix. The set defined by

$$\mathbf{y}'\mathbf{A}\mathbf{y} \leq 1$$

corresponds to the set

$$\mathbf{x}'\mathbf{H}'\mathbf{A}\mathbf{H}\mathbf{x} \leq 1 .$$

But since this has to coincide with $\mathbf{x}'\mathbf{x} \leq 1$ (the definition of $C_d$), we note that

$$\mathbf{H}'\mathbf{A}\mathbf{H} = \mathbf{I}$$

where $\mathbf{I}$ is the unit $d \times d$ matrix. Thus, we need to take $\mathbf{H}$ such that $\mathbf{A}^{-1} = \mathbf{H}\mathbf{H}'$. See also Rubinstein (1982).

## 2.5. Uniform polygonal random vectors.

A **convex polytope** of $R^d$ with vertices $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is the collection of all points in $R^d$ that are obtainable as convex combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Every point $\mathbf{x}$ in this convex polytope can be written as

$$\mathbf{x} = \sum_{i=1}^{n} a_i \mathbf{v}_i$$

for some $a_1, \ldots, a_n$ with $a_i \geq 0$, $\sum_{i=1}^{n} a_i = 1$. The set $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is minimal for the convex polytope generated by it when all $\mathbf{v}_i$'s are distinct, and no $\mathbf{v}_i$ can be written as a strict convex combination of the $\mathbf{v}_j$'s. ( A strict convex combination is one which has at least one $a_i$ not equal to 0 or 1.)

We say that a set of vertices $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is in general position if no three points are on a line, no four points are in a plane, etcetera. Thus, if the set of vertices is minimal for a convex polytope $P$, then it is in general position.

A **simplex** is a convex polytope with $d+1$ vertices in general position. Note that $d$ points in general position in $R^d$ define a hyperplane of dimension $d-1$. Thus, any convex polytope with fewer than $d+1$ vertices must have zero $d$-dimensional volume. In this sense, the simplex is the simplest nontrivial object in $R^d$.

We can define a basic simplex by the origin and $d$ points on the positive coordinate axes at distance one from the origin.

There are two distinct generation problems related to convex polytopes. We could be asked to generate a random vector uniformly distributed in a given polytope (see below), or we could be asked to generate a random collection of vertices defining a convex polytope. The latter problem is not dealt with here. See however Devroye (1982) and May and Smith (1982).

Random vectors distributed uniformly in an arbitrary simplex can be obtained by linear transformations of random vectors distributed uniformly in the basic simplex. Fortunately, we do not have to go through the agony of factorizing a matrix as in the case of a given covariance matrix structure. Rather, there is a surprisingly simple direct solution to the general problem.

**Theorem 2.1.**

Let $(S_1, \ldots, S_{d+1})$ be the spacings generated by a uniform sample of size $d$ on [0,1]. (Thus, $S_i \geq 0$ for all $i$, and $\sum S_i = 1$.) Then

$$\mathbf{X} = \sum_{i=1}^{d+1} S_i \mathbf{v}_i$$

is uniformly distributed in the polytope $P$ generated by $\mathbf{v}_1, \ldots, \mathbf{v}_{d+1}$, provided that $\mathbf{v}_1, \ldots, \mathbf{v}_{d+1}$ are in general position.

**Proof of Theorem 2.1.**

Let $\mathbf{S}$ be the column vector $S_1, \ldots, S_d$. We recall first that $\mathbf{S}$ is uniformly distributed in the basic simplex $B$ where

$$B = \{(x_1, \ldots, x_d) : x_i \geq 0 \ , \ \sum_i x_i \leq 1\} \ .$$

If all $\mathbf{v}_i$'s are considered as column vectors, and $\mathbf{A}$ is the matrix

$$\begin{bmatrix} \mathbf{v}_1-\mathbf{v}_{d+1} & \mathbf{v}_2-\mathbf{v}_{d+1} & \cdots & \mathbf{v}_d-\mathbf{v}_{d+1} \end{bmatrix} ,$$

then we can write $\mathbf{X}$ as follows:

$$\mathbf{X} = \mathbf{v}_{d+1} + \sum_{i=1}^{d} (\mathbf{v}_i - \mathbf{v}_{d+1}) S_i \ = \mathbf{v}_{d+1} + \mathbf{S}'\mathbf{A} \ .$$

It is clear that $\mathbf{X}$ is uniformly distributed, since it can be obtained by a linear transformation of $\mathbf{S}$. The support Supp ($\mathbf{X}$) of the distribution of $\mathbf{X}$ is the collection of all points which can be written as $\mathbf{v}_{d+1}+\mathbf{a}'\mathbf{A}$ where $\mathbf{a}\in B$ is a column vector. First, assume that $\mathbf{x}\in P$. Then, $\mathbf{x}=\sum_{i=1}^{d+1} a_i \mathbf{v}_i$ for some probability vector $a_1, \ldots, a_{d+1}$. This can be rewritten as follows:

$$\mathbf{x} = \mathbf{v}_{d+1} + \sum_{i=1}^{d} a_i (\mathbf{v}_i - \mathbf{v}_{d+1}) = \mathbf{v}_{d+1} + \mathbf{a}'\mathbf{A} \ ,$$

where $\mathbf{a}$ is the vector formed by $a_1, \ldots, a_d$. Thus, $P \subseteq$ Supp ($\mathbf{X}$). Next, assume $\mathbf{x}\in$ Supp ($\mathbf{X}$). Then, for some column vector $\mathbf{a}\in B$,

$$\mathbf{x} = \mathbf{v}_{d+1} + \mathbf{a}'\mathbf{A} = \mathbf{v}_{d+1} + \sum_{i=1}^{d} a_i (\mathbf{v}_i - \mathbf{v}_{d+1})$$
$$= \sum_{i=1}^{d+1} a_i \mathbf{v}_i \ ,$$

which implies that $\mathbf{x}$ is a convex combination of the $\mathbf{v}_i$'s, and thus $\mathbf{x}\in P$. Hence Supp ($\mathbf{X}$) $\subset P$, and hence Supp ($\mathbf{X}$)$=P$, which concludes the proof of Theorem 2.1. ■

**Example 2.1. Triangles.**

The following algorithm can be used to generate random vectors uniformly distributed in the triangle defined by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ of $R^2$:

**Generator for uniform distribution in triangle**

Generate iid uniform [0,1] random variates $U, V$.
IF $U > V$ then swap $U$ and $V$.
RETURN $(U\mathbf{v}_1+(V-U)\mathbf{v}_2+(1-V)\mathbf{v}_3)$

See also exercise 2.1. ■

### Example 2.2. Convex polygons in the plane.

Convex polygons on $R^2$ with $n > 3$ vertices can be partitioned into $n-2$ disjoint triangles. This can always be done by connecting all vertices with a designated root vertex. Triangulation of a polygon is of course always possible, even when the polygon is not convex. To generate a point uniformly in a triangulated polygon, it suffices to generate a point uniformly in the $i$-th triangle (see e.g. Example 2.1), where the $i$-th triangle is selected with probability proportional to its area. It is worth recalling that the area of a triangle formed by $(v_{11},v_{12}),(v_{21},v_{22}),(v_{31},v_{32})$ is

$$\frac{1}{2} \left| \sum_{i<j} (v_{i1}v_{j2}-v_{j1}v_{i2}) \right| . ■$$

We can deal with all simplices in all Euclidean spaces via Theorem 2.1. Example 2.2 shows that all polygons in the plane can be dealt with too, because all such polygons can be triangulated. Unfortunately, decomposition of $d$-dimensional polytopes into $d$-dimensional simplices is not always possible, so that Example 2.2 cannot be extended to higher dimensions. The decomposition is possible for all convex polytopes however. A decomposition algorithm is given in Rubin (1984), who also provides a good survey of the problem. Theorem 2.1 can also be found in Rubinstein (1982). Example 2.2 describes a method used by Hsuan (1982). The present methods which use decomposition and linear transformations are valid for polytopes. For sets with unusual shapes, the grid methods of section VIII.3.2 should be useful.

We conclude this section with the simple mention of how one can attack the decomposition of a convex polytope with $n$ vertices into simplices for general Euclidean spaces. If we are given an ordered polytope, i.e. a polytope with all its faces clearly identified, and with pointers to neighboring faces, then the partition is trivial: choose one vertex, and construct all simplices consisting of a face (each face has $d$ vertices) and the picked vertex. For selection of a simplex, we also need the area of a simplex with vertices $\mathbf{v}_i$, $i=1,2, \ldots, d+1$. This is given by

$$\frac{|\mathbf{A}|}{d!}$$

where $\mathbf{A}$ is the $d \times d$ matrix with as columns $\mathbf{v}_1-\mathbf{v}_{d+1}, \ldots, \mathbf{v}_d-\mathbf{v}_{d+1}$. The complexity of the preprocessing step (decomposition, computation of areas) depends upon $m$, the number of faces. It is known that $m=O(n^{\lfloor d/2 \rfloor})$ (McMullen, 1970). Since each area can be computed in constant time ($d$ is kept fixed, $n$ varies), the set-up time is $O(m)$. The expected generation time is $O(1)$ if a constant time selection algorithm is used.

The aforementioned ordered polytopes can be obtained from an unordered collection of $n$ vertices in worst-case time $O(n\log(n)+n^{\lfloor (d+1)/2 \rfloor})$ (Seidel, 1981), and this is worst-case optimal for even dimensions under some computational models.

## 2.6. Time series.

The generation of random time series with certain specific properties (marginal distributions, autocorrelation matrix, etcetera) is discussed by Schmeiser (1980), Franklin (1965), Price (1976), Hoffman (1979), Li and Hammond (1975), Lakhan (1981), Polge, Holliday and Bhagavan (1973), Mikhailov (1974), Fraker and Rippy (1974), Badel (1979), Lawrance and Lewis (1977, 1980, 1981), and Jacobs and Lewis (1977).

## 2.7. Singular distributions.

Singular distributions in $R^d$ are commonplace. Distributions that put all their mass on a line or curve in the plane are singular. So are distributions that put all their mass on the surface of a hypersphere of $R^d$. Computer generation of random vectors on such hyperspheres is discussed by Ulrich (1984), who in particular derives an efficient generator for the Fisher-von Mises distribution in $R^d$.

A line in $R^d$ can be given in many forms. Perhaps the most popular form is the parametric one, where $\mathbf{x}=\mathbf{h}(z)$ and $z \in R$ is a parameter. An example is the circle in $R^2$, determined by

$$x_1 = \cos(2\pi z) ,$$

$$x_2 = \sin(2\pi z) .$$

Now, if $Z$ is a random variable and $\mathbf{h}$ is a Borel measurable function, then $\mathbf{X}=\mathbf{h}(Z)$ is a random vector which puts all its mass on the line defined by $\mathbf{x}=\mathbf{h}(z)$. In other words, $\mathbf{X}$ has a line distribution. For a one-to-one mapping $\mathbf{h}:R \to R^d$, which is also continuous, we can define a line density $f(z)$ at the point $\mathbf{x}=\mathbf{h}(z)$ via the relationships

$$P(\mathbf{X}=\mathbf{h}(z) \text{ for some } z \in [a,b]) = \int_a^b f(z)\psi(z)\, dz \qquad (\text{all } [a,b]) ,$$

where $\psi(z)=\sqrt{\sum_{i=1}^{d} h'_i{}^2(z)}$ is the norm of the tangent of $\mathbf{h}$ at $z$, and $h_i$ is the $i$-th component of $\mathbf{h}$. But since this must equal $P(a \le Z \le b)=\int_a^b g(z)\, dz$ where $g$ is the density of $Z$, we see that

$$f(z) = \frac{g(z)}{\psi(z)} .$$

For a uniform line density, we need to take $g$ proportional to $\psi$.

As a first example, consider a function in the plane determined by the equation $y=\chi(x)$ $(0 \le x \le 1)$. A point with uniform line density can be obtained by considering the $x$-coordinate as our parameter $z$. This yields the algorithm

Generate a random variate $X$ with density $c\sqrt{1+\chi'^2(x)}$.
RETURN $(X,\chi(X))$

This could be called the projection method for obtaining random variates with certain line densities. The converse, projection from a line to the $x$-axis is much less useful, since we already have many techniques for generating real-line-valued random variates.

## 2.8. Exercises.

1. Consider a triangle with vertices $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$, and let $U,V$ be iid uniform [0,1] random variables.

   A. Show that if we set $\mathbf{Y} \leftarrow \mathbf{v}_2+(\mathbf{v}_3-\mathbf{v}_2)U$, and $\mathbf{X} \leftarrow \mathbf{v}_1+(\mathbf{Y}-\mathbf{v}_1)V$, then $\mathbf{X}$ is not uniformly distributed in the given triangle. This method is misleading, as $\mathbf{Y}$ is uniformly distributed on the edge $(\mathbf{v}_2,\mathbf{v}_3)$, and $\mathbf{X}$ is uniformly distributed on the line joining $\mathbf{v}_1$ and $\mathbf{Y}$.

B. Show that $\mathbf{X}$ in part A is uniformly distributed in the said triangle if we replace $V$ in the algorithm by $\max(V, V^*)$ where $V, V^*$ are iid uniform [0,1] random variables.

2. Define a simple boolean function which returns the value true if and only if $\mathbf{x}$ belongs to the a triangle in $R^2$ with three given vertices.

3. Consider a triangle ABC where AB has length one, BC has length $b$, and the angle ABC is $\theta$. Let $\mathbf{X}$ be uniformly distributed in the triangle, and let $\mathbf{Y}$ be the intersection of the lines $\mathbf{AX}$ and BC. Let $Z$ be the distance between $\mathbf{Y}$ and B. Show that $Z$ has density

$$\frac{1}{\sqrt{z^2-2z\cos(\theta)+1}} \quad (0<z<b) .$$

Compare the geometric algorithm for generating $Z$ given above with the inversion method.

## 3. DEPENDENCE. BIVARIATE DISTRIBUTIONS.

### 3.1. Creating and measuring dependence.

In many experiments, a controlled degree of dependence is required. Sometimes, users want distributions with given marginals and a given dependence structure as measured with some criterion. Sometimes, users know precisely what they want by completely specifying a multivariate distribution. In this section, we will mainly look at problems in which certain marginal distributions are needed together with a given degree of dependence. Usually, there are very many multivariate distributions which satisfy the given requirements, and sometimes there are none. In the former case, we should design generators which are efficient and lead to distributions which are not unrealistic.

For a clear treatment of the subject, it is best to emphasize bivariate distributions. A number of different measures of association are commonly used by practicing statisticians. First and foremost is the **correlation coefficient** $\rho$ (also called Pearson product moment correlation coefficient) defined by

$$\rho = \frac{E((X_1-\mu_1)(X_2-\mu_2))}{\sigma_1\sigma_2} ,$$

where $\mu_1,\mu_2$ are the means of $X_1,X_2$, and $\sigma_1,\sigma_2$ are the corresponding standard deviations. The key properties of $\rho$ are well-known. When $X_1,X_2$ are independent, $\rho=0$. Furthermore, by the Cauchy-Schwarz inequality, it is easy to see that $|\rho| \leq 1$. When $X_1=X_2$, we have $\rho=1$, and when $X_1=-X_2$, we have $\rho=-1$. Unfortunately, there are a few enormous drawbacks related to the correlation coefficient. First, it is only defined for distributions having marginals with finite variances. Furthermore, it is not invariant under monotone transformations of the coordinate axes. For example, if we define a bivariate uniform distribution with a given value for $\rho$ and then apply a transformation to get certain specific marginals, then the value of $\rho$ could (and usually does) change. And most importantly, the value of $\rho$ may not be a solid indicator of the dependence. For one thing, $\rho=0$ does not imply independence.

Measures of association which are invariant under monotone transformations are in great abundance. For example, there is **Kendall's tau** defined by

$$\tau = 2P((X_1-X_2)(X'_1-X'_2)>0)-1$$

where $(X_1,X_2)$ and $(X'_1,X'_2)$ are iid. The invariance under strictly monotone transformations of the coordinate axes is obvious. Also, for all distributions, $\tau$ exists and takes values in $[-1,1]$, and $\tau=0$ when the components are independent and nonatomic. The **grade correlation** (also called **Spearman's rho** or the **rank correlation**) $\rho_g$ is defined as $\rho(F_1(X_1),F_2(X_2))$ where $\rho$ is the standard correlation coefficient, and $F_1,F_2$ are the marginal distribution functions of $X_1,X_2$ (see for example Gibbons (1971)). $\rho_g$ always exists, and is invariant under monotone transformations. $\tau$ and $\rho_g$ are also called ordinal measures of association since they depend upon rank information only (Kruskal, 1958). Unfortunately, $\tau=0$ or $\rho_g=0$ do not imply independence (exercise 3.4). It would be desirable for a good measure of association or dependence that it be zero only when the components are independent.

The two measures given below satisfy all our requirements (universal existence, invariance under monotone transformations, and the zero value implying independence):

A. **The sup correlation** (or maximal correlation) $\rho*$ defined by Gebelein (1941) and studied by Sarmanov (1962,1963) and Renyi (1959):

$$\overline{\rho}(X_1,X_2) = \sup \rho(g_1(X_1),g_2(X_2))$$

where the supremum is taken over all Borel-measurable functions $g_1,g_2$ such that $g_1(X_1),g_2(X_2)$ have finite positive variance, and $\rho$ is the ordinary correlation coefficient.

B. **The monotone correlation** $\rho*$ introduced by Kimeldorf and Sampson (1978), which is defined as $\overline{\rho}$ except that the supremum is taken over monotone functions $g_1,g_2$ only.

Let us outline why these measures satisfy our requirements. If $\rho*=0$, and $X_1,X_2$ are nondegenerate, then $X_2$ is independent of $X_1$ (Kimeldorf and Sampson, 1978). This is best seen as follows. We first note that for all $s,t$,

$$\rho(I_{(-\infty,s]}(X_1),I_{(-\infty,t]}(X_2)) = 0$$

because the indicator functions are monotone and $\rho*=0$. But this implies

$$P(X_1\leq s,X_2\leq t) = P(X_1\leq s)P(X_2\leq t),$$

which in turn implies independence. For $\overline{\rho}$, we refer to exercise 3.6 and Renyi (1959). Good general discussions can be found in Renyi (1959), Kruskal (1958), Kimeldorf and Sampson (1978) and Whitt (1976). The measures of dependence are obviously interrelated. We have directly from the definitions,

$$|\rho| \leq \rho^* \leq \bar{\rho} \leq 1 .$$

There are examples in which we have equality between all correlation coefficients (multivariate normal distribution, exercise 3.5), and there are other examples in which there is strict inequality. It is perhaps interesting to note when $\rho^*$ equals one. This is for example the case when $X_2$ is monotone dependent upon $X_1$, i.e. there exists a monotone function $g$ such that $P(X_2=g(X_1))=1$, and $X_1,X_2$ are nonatomic (Kimeldorf and Sampson (1978)). This follows directly from the fact that $\rho^*$ is invariant under monotone transformations, so that we can assume without loss of generality that the distribution is bivariate uniform. But then $g$ must be the identity function, and the statement is proved, i.e. $\rho^*=1$. Unfortunately, $\rho^*=1$ does not imply monotone dependence.

For continuous marginals, there is yet another good measure of dependence, based upon the distance between probability measures. It is defined as follows:

$$L = \sup_A \ | P((X_1,X_2)\in A) - P((X_1,X'_2)\in A) |$$

$$= \frac{1}{2} \int | f(x_1,x_2) - f_1(x_1) f_2(x_2) | \ dx_1 dx_2 ,$$

where $A$ is a Borel set of $R^2$, $X'_2$ is distributed as $X_2$, but is independent of $X_1$, $f$ is the density of $(X_1,X_2)$, and $f_1,f_2$ are the marginal densities. The supremum in the definition of $L$ measures the distance between the given bivariate probability measure and the artificial bivariate probability measure constructed by taking the product of the two participating marginal probability measures. The invariance under strictly monotone transformations is clear. The integral form for $L$ is Scheffe's theorem in disguise (see exercise 3.9). It is only valid when all the given densities exist.

**Example 3.1.**

It is clearly possible to have uniform marginals and a singular bivariate distribution (consider $X_2=X_1$). It is even possible to find such a singular distribution with $\rho=\rho_g=0$ (consider a carefully selected distribution on the surface of the unit circle; or consider $X_2=SX_1$ where $S$ takes the values +1 and −1 with equal probability). However, when we take $A$ equal to the support of the singular distribution, then $A$ has zero Lebesgue measure, and therefore zero measure for any absolutely continuous probability measure. Hence, $L=1$. In particular, when $X_2$ is monotone dependent on $X_1$, then the bivariate distribution is singular, and therefore $L=1$. ■

**Example 3.2.**

$X_1,X_2$ are independent if and only if $L=0$. The if part follows from the fact that for all $A$, the product measure of $A$ is equal to the given bivariate probability measure of $A$. Thus, both probability measures are equal. The only if part is trivially true. ■

In the search for good measures of association, there is no clear winner. Probability theoretical considerations lead us to favor $L$ over $\rho_g$, $\rho*$ and $\overline{\rho}$. On the other hand, as we have seen, approximating the bivariate distribution by a singular distribution, always gives $L=1$. Thus, $L$ is extremely sensitive to even small local deviations. The correlation coefficients are much more robust in that respect.

We will assume that what the user wants is a distribution with given absolutely continuous marginal distribution functions, and a given value for one of the transformation-invariant measures of dependence. We can then construct a bivariate uniform distribution with the given measure of dependence, and then transform the coordinate axes as in the univariate inversion method to achieve given marginal distributions (Nataf, 1962; Kimeldorf and Sampson, 1975; Mardia, 1970). If we can choose between a family of bivariate uniform distributions, then it is perhaps possible to pick out the unique distribution, if it exists, with the given measure of dependence. In the next section, we will deal with bivariate uniform distributions in general.

## 3.2. Bivariate uniform distributions.

We say that a distribution is bivariate uniform (exponential, gamma, normal, Cauchy, etcetera) when the univariate marginal distributions are all uniform (exponential, gamma, normal, Cauchy, etcetera). Distributions of this form are extremely important in mathematical statistics in the context of testing for dependence between components. First of all, if the marginal distributions are continuous, it is always possible by a transformation of both axes to insure that the marginal distributions have any prespecified density such as the uniform [0,1] density. If after the transformation to uniformity the joint density is uniform on $[0,1]^2$, then the two component random variables are independent. In fact, the joint density after transformation provides a tremendous amount of information about the sort of dependence.

There are various ways of obtaining bivariate distributions with specified marginals from bivariate uniform distributions, which make these uniform distributions even more important. Good surveys are provided by Johnson (1976), Johnson and Tenenbein (1979) and Marshall and Olkin (1983). The following theorem comes closest to generalizing the univariate properties which lead to the inversion method.

**Theorem 3.1.**

Let $(X_1,X_2)$ be bivariate uniform with joint density $g$. Let $f_1,f_2$ be fixed univariate densities with corresponding distribution functions $F_1,F_2$. Then the density of $(Y_1,Y_2)=(F^{-1}{}_1(X_1),F^{-1}{}_2(X_2))$ is

$$f(y_1,y_2)=f_1(y_1)f_2(y_2)g(F_1(y_1),F_2(y_2)) .$$

Conversely, if $(Y_1,Y_2)$ has density $f$ given by the formula shown above, then $Y_1$ has marginal density $f_1$ and $Y_2$ has marginal density $f_2$. Furthermore, $(X_1,X_2)=(F_1(Y_1),F_2(Y_2))$ is bivariate uniform with joint density

$$g(x_1,x_2)=\frac{f(F^{-1}{}_1(x_1),F^{-1}{}_2(x_2))}{f_1(F^{-1}{}_1(x_1))f_2(F^{-1}{}_2(x_2))} \quad (0\le x_1,x_2\le 1) .$$

**Proof of Theorem 3.1.**

Straightforward. ■

There are many recipes for cooking up bivariate distributions with specified marginal distribution functions $F_1,F_2$. We will list a few in Theorem 3.2. It should be noted that if we replace $F_1(x_1)$ by $x_1$ and $F_2(x_2)$ by $x_2$ in these recipes, then we obtain bivariate uniform distribution functions. Recall also that the bivariate density, if it exists, can be obtained from the bivariate distribution function by taking the partial derivative with respect to $\partial x_1 \partial x_2$.

**Theorem 3.2.**

Let $F_1=F_1(x_1), F_2=F_2(x_2)$ be univariate distribution functions. Then the following is a list of bivariate distribution functions $F=F(x_1,x_2)$ having as marginal distribution functions $F_1$ and $F_2$:

A. $F = F_1F_2(1+a(1-F_1)(1-F_2))$. Here $a \in [-1,1]$ is a parameter (Farlie (1960), Gumbel (1958), Morgenstern (1956)). This will be called Morgenstern's family.

B. $F = \dfrac{F_1F_2}{1-a(1-F_1)(1-F_2)}$. Here $a \in [-1,1]$ is a parameter (Ali, Mikhail and Haq (1978)).

C. $F$ is the solution of $F(1-F_1-F_2+F) = a(F_1-F)(F_2-F)$ where $a \geq 0$ is a parameter (Plackett, 1965).

D. $F = a\max(0,F_1+F_2-1)+(1-a)\min(F_1,F_2)$ where $0 \leq a \leq 1$ is a parameter (Frechet, 1951).

E. $(-\log(F))^m = (-\log(F_1))^m + (-\log(F_2))^m$ where $m \geq 1$ is a parameter (Gumbel, 1960).

**Proof of Theorem 3.2.**

To verify that $F$ is indeed a distribution function, we must verify that $F$ is nondecreasing in both arguments, and that the limits as $x_1,x_2 \to -\infty$ and $\to \infty$ are 0 and 1 respectively. To verify that the marginal distribution functions are correct, we need to check that

$$\lim_{x_2 \to \infty} F(x_1,x_2) = F_1(x_1)$$

and

$$\lim_{x_1 \to \infty} F(x_1,x_2) = F_2(x_2) .$$

The latter relations are easily verified. ■

It helps to visualize these recipes. We begin with Frechet's inequalities (Frechet, 1951), which follow by simple geometric arguments in the plane:

**Theorem 3.3. Frechet's inequalities.**

For any two univariate distribution functions $F_1, F_2$, and any bivariate distribution function $F$ having these two marginal distribution functions,

$$\max(0, F_1(x)+F_2(y)-1) \leq F(x,y) \leq \min(F_1(x), F_2(y)) .$$

**Proof of Theorem 3.3.**

For fixed $(x_1,x_2)$ in the plane, let us denote by $Q_{SE}, Q_{NE}, Q_{SW}, Q_{NW}$ the four quadrants centered at $x,y$ where equality is resolved by including boundaries with the south and west halfplanes. Thus, $(x_1,x_2)$ belongs to $Q_{SW}$ while the vertical line at $x_1$ belongs to $Q_{SW} \cup Q_{NW}$. It is easy to see that at $x_1, x_2$,

$$F_1(x_1) = P(Q_{SW} \cup Q_{NW}) ,$$

$$F_2(x_2) = P(Q_{SW} \cup Q_{SE}) ,$$

$$F(x_1,x_2) = P(Q_{SW}) .$$

Clearly, $F \leq \min(F_1, F_2)$ and $1-F \leq 1-F_1+1-F_2$. ■

These inequalities are valid for all bivariate distribution functions $F$ with marginal distribution functions $F_1$ and $F_2$. Interestingly, both extremes are also valid distribution functions. In fact, we have the following property which can be used for the generation of random vectors with these distribution functions.

**Theorem 3.4.**

Let $U$ be a uniform $[0,1]$ random variable, and let $F_1$, $F_2$ be continuous univariate distribution functions. Then

$$(F^{-1}{}_1(U), F^{-1}{}_2(U))$$

has distribution function $\min(F_1, F_2)$. Furthermore,

$$(F^{-1}{}_1(U), F^{-1}{}_2(1-U))$$

has distribution function $\max(0, F_1+F_2-1)$.

**Proof of Theorem 3.4.**

We have

$$P(F^{-1}{}_1(U) \le x_1, F^{-1}{}_2(U) = x_2) = P(U \le \min(F_1(x_1), F_2(x_2))) .$$

Also,

$$P(F^{-1}{}_1(U) \le x_1, F^{-1}{}_2(1-U) = x_2) = P(U \le F_1(x_1), 1-U \le F_2(x_2)) . \blacksquare$$

Frechet's extremal distribution functions are those for which maximal positive and negative dependence are obtained respectively. This is best seen by considering the bivariate uniform case. The upper distribution function $\min(x_1,x_2)$ puts its mass uniformly on the 45 degree diagonal of the first quadrant. The bottom distribution function $\max(0,x_1+x_2-1)$ puts its mass uniformly on the -45 degree diagonal of $[0,1]^2$. Hoeffding (1940) and Whitt (1976) have shown that maximal positive and negative correlation are obtained for Frechet's extremal distribution functions (see exercise 3.1). Note also that maximally correlated random variable are very important in variance reduction techniques in Monte Carlo simulation. Theorem 3.4 shows us how to generate such random vectors. We have thus identified a large class of applications in which the inversion method seems essential (Fox, 1980). For Frechet's bivariate family (case D in Theorem 3.2), we note without work that it suffices to consider a mixture of Frechet's extremal distributions. This is often a poor way of creating intermediate correlation. For example, in the bivariate uniform case, all the probability mass is concentrated on the two diagonals of $[0,1]^2$.

The list of examples in Theorem 3.2 is necessarily incomplete. Other examples can be found in exercises 3.2 and 3.3. Random variate generation is usually taken care of via the conditional distribution method. The following example should suffice.

**Example 3.3. Morgenstern's family.**

Consider the uniform version of Morgenstern's bivariate family with parameter $|a| \le 1$ given by part A of Theorem 3.2. It is easy to see that for this family, there exists a density given by

$$f(x_1,x_2) = 1+a(2x_1-1)(2x_2-1) .$$

Here we can generate $X_1$ uniformly on [0,1]. Given $X_1$, $X_2$ has a trapezoidal density which is zero outside [0,1] and varies from $1-a(2X_1-1)$ at $x_2=0$ to $1+a(2X_1-1)$ at $x_2=1$. If $U,V$ are iid uniform [0,1] random variables, then $X_2$

can be generated as

$$\begin{cases} \min(U, -\frac{V}{a(2X_1-1)}) & X_1 < \frac{1}{2} \\ \max(U, 1-\frac{V}{a(2X_1-1)}) & X_1 \geq \frac{1}{2} \end{cases} \quad \blacksquare$$

There are other important considerations when shopping around for a good bivariate uniform family. For example, it is useful to have a family which contains as members, or at least as limits of members, Frechet's extremal distributions, plus the product of the marginals (the independent case). We will call such families **comprehensive**. Examples of comprehensive bivariate families are given in the table below. Note that the comprehensiveness of a family is invariant under strictly monotone transformations of the coordinate axes (exercise 3.11), so that the marginals do not really matter.

| Distribution function | Reference |
|---|---|
| $F$ is the solution of $F(1-F_1-F_2+F) = a(F_1-F)(F_2-F)$ where $a \geq 0$ is a parameter | Plackett (1965) |
| $F = \frac{a^2(1-a)}{2}\max(0,F_1+F_2-1) + \frac{a^2(1+a)}{2}\min(F_1,F_2)+(1-a^2)F_1F_2$ where $\lvert a \rvert \leq 1$ is a parameter | Frechet (1958) |
| $\frac{1}{2\pi\sqrt{1-r^2}} e^{-\frac{x_1^2+x_2^2-2rx_1x_2}{2(1-r^2)}}$ where $\lvert r \rvert \leq 1$ is a measure of association | Bivariate normal (see e.g. Mardia, 1970) |

From this table, one can create other comprehensive families either by monotone transformations, or by taking mixtures. Note that most families, including Morgenstern's family, are not comprehensive.

Another issue is that of the range spanned by the family in terms of the values of a given measure of dependence. For example, for Morgenstern's bivariate uniform family of Example 3.3, the correlation coefficient is $-a/3$. Therefore, it can take all the values in $[-\frac{1}{3},\frac{1}{3}]$, but no values outside this interval. Needless to say, full ranges for certain measures of association are an asset. Typically, this goes hand in hand with comprehensiveness.

**Example 3.4. Full correlation range families.**

Plackett's bivariate family with parameter $a \geq 0$ and arbitrary continuous marginal distribution functions has correlation coefficient

$$\rho = \frac{-(1-a^2)-2a\log(a)}{(1-a)^2},$$

which can be shown to take the values 1,0,–1 when $a \to \infty$, $a=1$ and $a=0$ respectively (see e.g. Barnett, 1980). Since $\rho$ is a continuous function of $a$, all values of $\rho$ can be achieved.

The bivariate normal family can also achieve all possible values of correlation. Since for this family, $\rho=\rho^*=\overline{\rho}$, we also achieve the full range for the sup correlation and the monotone correlation. ■

**Example 3.5. The Johnson-Tenenbein families.**

Johnson and Tenenbein (1981) proposed a general method of constructing bivariate families for which $\tau$ and $\rho_g$ can attain all possible values in (–1,1). The method consists simply of taking $(X_1,X_2)=(U,H(cU+(1-c)V))$, where $U,V$ are iid random variables with common distribution function $F$, $c \in [0,1]$ is a weight parameter, and $H$ is a monotone function chosen in such a way that $H(cU+(1-c)V)$ also has distribution function $F$. To take a simple example, let $U,V$ be iid normal random variables. Then we should take $H(u)=u/\sqrt{c^2+(1-c)^2}$. The resulting two-dimensional random vector is easily seen to be bivariate normal, as it is a linear combination of iid normal random variables. Its correlation coefficient is

$$\frac{c}{\sqrt{c^2+(1-c)^2}},$$

which can take all values in [0,1]. Moreover,

$$\rho_g = \frac{6}{\pi}\arcsin\left(\frac{c}{2\sqrt{c^2+(1-c)^2}}\right),$$

$$\tau = \frac{2}{\pi}\arcsin\left(\frac{c}{\sqrt{c^2+(1-c)^2}}\right).$$

It is easy to see that these measures of association can also take all values in [0,1] when we vary $c$. Negative correlations can be achieved by considering $(-U,H(cU+(1-c)V))$. Recall next that $\tau$ and $\rho_g$ are invariant under strictly monotone transformations of the coordinate axes. Thus, we can now construct bivariate families with specified marginals and given values for $\rho_g$ or $\tau$. ∎

### 3.3. Bivariate exponential distributions.

We will take the bivariate exponential distribution as our prototype distribution for illustrating just how we can construct such distributions directly. At the same time, we will discuss random variate generators. There are two very different approaches:

A. The analytic method: one defines explicitly a bivariate density or distribution function, and worries about generators later. An example is Gumbel's bivariate exponential family (1960) described below. Another example is the distribution of Nagao and Kadoya (1971) dealt with in exercise 3.10.

B. The empiric method: one constructs a pair of random variables known to have the correct marginals, and worries about the form of the distribution function later. Here, random variate generation is typically a trivial problem. Examples include distributions proposed by Johnson and Tenenbein (1981), Moran (1967), Marshall and Olkin (1967), Arnold (1967) and Lawrance and Lewis (1983).

The distinction between A and B is often not clear-cut. Families can also be partitioned based upon the range for given measures of association, or upon the notion of comprehensiveness. Let us start with Gumbel's family of bivariate exponential distribution functions:

$$1-e^{-x_1}-e^{-x_2}+e^{-x_1-x_2-ax_1x_2} \quad (x_1,x_2>0) .$$

Here $a \in [0,1]$ is the parameter. The joint density is

$$e^{-x_1-x_2-ax_1x_2}\left((1+ax_1)(1+ax_2)-a\right) .$$

Notice that the conditional density of $X_2$ given $X_1=x_1$ is

$$e^{-(1+ax_1)x_2}\left((1+ax_1)(1+ax_2)-a\right)$$

$$= \frac{a}{\theta}\left(\theta^2 x_2 e^{-\theta x_2}\right) + \frac{\theta-a}{\theta}\left(\theta e^{-\theta x_2}\right) ,$$

where $\theta=1+ax_1$. In this decomposition, we recognize a mixture of a gamma (2) and a gamma (1) density. Random variates can easily be generated via the conditional distribution method, where the conditional distribution of $X_2$ given $X_1$ can be handled by composition (see below). Unfortunately, the family contains only none of Frechet's extremal distributions, which suggests that extreme correlations cannot be obtained.

**Gumbel's bivariate exponential distribution with parameter a**

Generate iid exponential random variates $X_1, X_2$.
Generate a uniform [0,1] random variate $U$.
IF $U \leq \frac{a}{1+aX_1}$
THEN
Generate an exponential random variate $E$.
$X_2 \leftarrow X_2 + E$
RETURN $(X_1, \frac{X_2}{1+aX_1})$

Generalizations of Gumbel's distribution have been suggested by various authors. In general, one can start from a bivariate uniform distribution function $F$, and define a bivariate exponential distribution function by

$$F(1-e^{-x_1}, 1-e^{-x_2}) .$$

For a generator, we need only consider $(-\log(U), -\log(V))$ where $U, V$ is bivariate uniform with distribution function $F$. For example, if we do this for Morgenstern's family with parameter $|a| \leq 1$, then we obtain the bivariate exponential distribution function

$$(1-e^{-x_1})(1-e^{-x_2})(1+ae^{-x_1-x_2}) \quad (x_1, x_2 \geq 0).$$

This distribution has also been studied by Gumbel (1960). Both Gumbel's exponential distributions and other possible transformations of bivariate uniform distributions are often artificial.

In the empiric (or constructive) method, one argues the other way around, by first defining the random vector. In the table shown below, a sampling of such bivariate random vectors is given. We have taken what we consider are good didactical examples showing a variety of approaches. All of them exploit special properties of the exponential distribution, such as the fact that the sum of squares of iid normal random variables is exponentially distributed, or the fact that the minimum of independent exponential random variables is again exponentially distributed.

| $(X_1,X_2)$ | Reference |
|---|---|
| $\left\{\min(\frac{E_1}{\lambda_1},\frac{E_3}{\lambda_3}),\min(\frac{E_2}{\lambda_2},\frac{E_3}{\lambda_3})\right\}$ | Marshall and Olkin (1967) |
| $\left(\beta_1E_1+S_1E_2,\beta_2E_2+S_2E_1\right)$, $P(S_i=1)=1-P(S_i=0)=1-\beta_i$ $(i=1,2)$ | Lawrance and Lewis (1983) |
| $\left\{E_1,-\log((1-c)e^{-\frac{E_2}{1-c}}+ce^{-\frac{E_2}{c}})+\log(1-2c)\right\}$, $c\in[0,1]$ | Johnson and Tenenbein (1981) |
| $\left\{\frac{1}{2}(N_1^2+N_2^2),\frac{1}{2}(N_3^2+N_4^2)\right\}$, $(N_1,N_3),(N_2,N_4)$ iid multinormal with correlation $\rho$ | Moran (1967) |

In this table, $E_1,E_2,E_3$ are iid exponential random variates, and $\lambda_1,\lambda_2,\lambda_3\geq 0$ are parameters with $\lambda_1\lambda_2+\lambda_3>0$. The $N_i$'s are normal random variables, and $c,\beta_1,\beta_2$ are [0,1]-valued constants. A special property of the marginal distribution, closure under the operation **min**, is exploited in the definition. To see this, note that for $x>0$,

$$P(X_1>x)=P(E_1>\lambda_1x,E_3>\lambda_3x)$$
$$=e^{-(\lambda_1+\lambda_3)x}\qquad(x>0)\ .$$

Thus, $X_1$ is exponential with parameter $\lambda_1+\lambda_3$. The joint distribution function is uniquely determined by the function $G(x_1,x_2)$ defined by

$$G(x_1,x_2)=P(X_1>x_1,X_2>x_2)=e^{-\lambda_1x_1-\lambda_2x_2-\lambda_3\max(x_1,x_2)}\ .$$

The distribution is a mixture of a singular distribution carrying weight $\lambda_3/(\lambda_1+\lambda_2+\lambda_3)$, and an absolutely continuous part (exercise 3.6). Also, it is unfortunate that when $(X_1,X_2)$ has the given bivariate exponential distribution, then $(a_1X_1,a_2X_2)$ is bivariate exponential in the case $a_1=a_2$ only. On the positive side, we should note that the family includes the independent case ($\lambda_3=0$), and one of Frechet's extremal cases ($\lambda_1=\lambda_2=0$). In the latter case, note that

$$G(x_1,x_2)=P(X_1>x_1,X_2>x_2)=e^{-\lambda_3\max(x_1,x_2)}\ .$$

The Lawrance-Lewis bivariate exponential is just one of a long list of bivariate exponentials constructed by them. The one given in the table is particularly flexible. We can quickly verify that the marginals are exponential via characteristic functions. The characteristic function of $X_1$ is

$$\phi(t)=E(e^{itX_1})=E(e^{\beta_1itE_1})(\beta_1+(1-\beta_1)E(e^{itE_2}))$$
$$=\frac{1}{1-it\beta_1}(\beta_1+\frac{(1-\beta_1)}{1-it})=\frac{1}{1-it}\ .$$

The correlation $\rho=2\beta_1(1-\beta_2)+\beta_2(1-\beta_1)$, valid for $0\leq\beta_1\leq\beta_2\leq 1$, can take all values between 0 and 1. To create negative correlation, one can replace $E_1,E_2$ in the formulas for $X_2$ by two other exponential random variables, $h(E_1),h(E_2)$ where $h(x)=-\log(1-e^{-x})$ (Lawrance and Lewis, 1983).

The Johnson and Tenenbein construction is almost as simple as the Lawrance-Lewis construction. Interestingly, by varying the parameter $c$, all possible nonnegative values for $\rho_g$, $\tau$ and $\rho$ are achievable.

Finally, in Moran's bivariate distribution, good use is made of yet another property of exponential random variables. His distribution has correlation $\rho^2$ where $\rho$ is the correlation of the underlying bivariate normal distribution. Again, random variate generation is extremely simple, and the correlation spans the full nonnegative range. Difficulties arise only when one needs to compute the exact value of the density at some points, but then again, these same difficulties are shared by most empiric methods.

## 3.4. A case study: bivariate gamma distributions.

We have seen how bivariate distributions with any given marginals can be constructed from bivariate uniform distributions or bivariate distributions with other continuous marginals, via transformations of the coordinate axes. These transformations leave $\rho_g$, $\tau$ and other ordinal measures of association invariant, but generally speaking not $\rho$. Furthermore, the inversion of the marginal distribution functions $(F_1,F_2)$ required to apply these transformations is often unfeasible. Such is the case for the gamma distribution. In this section we will look at these new problems, and provide new solutions.

To clarify the problems with inversion, we note that if $X_1,X_2$ is bivariate gamma $(a_1,a_2)$, where $a_i$ is the parameter for $X_i$, then maximum and minimum correlation are obtained for the Frechet bounds, i.e.

$$X_2 = F_2^{-1}(F_1(X_1)) ,$$
$$X_2 = F_2^{-1}(1-F_1(X_1))$$

respectively (Moran (1967), Whitt (1976)). Direct use of Frechet's bounds is possible but not recommended if generator efficiency is important. In fact, it is not recommended to start from any bivariate uniform distribution. Also, the method of Johnson and Tenenbein (1981) illustrated on the bivariate uniform, normal and exponential distributions in the previous sections requires an inversion of a gamma distribution function if it were to be applied here.

We can also obtain help from the composition method, noting that the random vector $(X_1,X_2)$ defined by

$$(X_1,X_2) = \begin{cases} (Y_1,Y_2) & \text{,with probability } p \\ (Z_1,Z_2) & \text{,with probability } 1-p \end{cases}$$

has the right marginal distributions if both random vectors on the right hand side have the same marginals. Also, $(X_1,X_2)$ has correlation coefficient $p\rho_Y+(1-p)\rho_Z$ where $\rho_Y,\rho_Z$ are the correlation coefficients of the two given random vectors. One typically chooses $\rho_Y$ and $\rho_Z$ at the extremes, so that the entire range of $\rho$ values is covered by adjusting $p$. For example, one could take $\rho_Y=0$ by considering iid random variables $Y_1,Y_2$. Then $\rho_Z$ can be taken maximal by using the Frechet maximal dependence as in $(Z_1,Z_2)=(Z_1,F_2^{-1}(1-F_1(Z_1))$ where $Z_1$ is gamma $(a_1)$. Doing so leads to a mixture of a continuous distribution (the product measure) and a singular distribution, which is not desirable.

The gamma distribution shares with many distributions the property that it is closed under additions of independent random variables. This has led to inversion-free methods for generating bivariate gamma random vectors, now known as **trivariate reduction methods** (Cherian, 1941; David and Fix, 1961; Mardia, 1970; Johnson and Ramberg, 1977; Schmeiser and Lal, 1982). The name is borrowed from the principle that two dependent random variables are constructed from three independent random variables. The application of the principle is certainly not limited to the gamma distribution, but is perhaps best illustrated here. Consider independent gamma random variables $G_1,G_2,G_3$ with parameters $a_1,a_2,a_3$. Then the random vector

$$(X_1,X_2)=(G_1+G_3,G_2+G_3)$$

is bivariate gamma. The marginal gamma distributions have parameters $a_1+a_3$ and $a_2+a_3$ respectively. Furthermore, the correlation is given by

$$\rho=\frac{a_3}{\sqrt{(a_1+a_3)(a_2+a_3)}} .$$

If $\rho$ and the marginal gamma parameters are specified beforehand, we have one of two situations: either there is no possible solution for $a_1,a_2,a_3$, or there is exactly one solution. The limitation of this technique, which goes back to Cherian (1941) (see Schmeiser and Lal (1980) for a survey), is that

$$0\le\rho\le\frac{\min(\alpha_1,\alpha_2)}{\sqrt{\alpha_1\alpha_2}}$$

where $\alpha_1,\alpha_2$ are the marginal gamma parameters. Within this range, trivariate reduction leads to one of the fastest algorithms known to date for bivariate gamma distributions.

**Trivariate reduction for bivariate gamma distribution**

[NOTE: $\rho$ is a given correlation, $\alpha_1,\alpha_2$ are given parameters for the marginal gamma distributions. It is assumed that $0 \leq \rho \leq \frac{\min(\alpha_1,\alpha_2)}{\sqrt{\alpha_1\alpha_2}}$.]

[GENERATOR]

Generate a gamma $(\alpha_1-\rho\sqrt{\alpha_1\alpha_2})$ random variate $G_1$.

Generate a gamma $(\alpha_2-\rho\sqrt{\alpha_1\alpha_2})$ random variate $G_2$.

Generate a gamma $(\rho\sqrt{\alpha_1\alpha_2})$ random variate $G_3$.

RETURN $(G_1+G_3,G_2+G_3)$

Ronning (1977) generalized this principle to higher dimensions, and suggested several possible linear combinations to achieve desired correlations. Schmeiser and Lal (1982) (exercise 3.19) fill the void by extending the trivariate reduction method in two dimensions, so that all theoretically possible correlations can be achieved in bivariate gamma distributions. But we do not get something for nothing: the algorithm requires the inversion of the gamma distribution function, and the numerical solution of a set of nonlinear equations in the set-up stage.

### 3.5. Exercises.

1. Prove that over all bivariate distribution functions with given marginal univariate distribution functions $F_1,F_2$, the correlation coefficient $\rho$ is minimized for the distribution function $\max(0,F_1(x)+F_2(y)-1)$. It is maximized for the distribution function $\min(F_1(x),F_2(y))$ (Whitt, 1976; Hoeffding, 1940).

2. **Plackett's bivariate uniform family** (Plackett (1965). Consider the bivariate uniform family defined by part C of Theorem 3.2, with parameter $a \geq 0$. Show that on $[0,1]^2$, this distribution has a density given by

$$f(x_1,x_2) = \frac{a(a-1)(x_1+y_1-2x_1x_2)+a}{\left(((a-1)(x_1+x_2)+1)^2-4a(a-1)x_1x_2\right)^{3/2}} .$$

   For this distribution, Mardia (1970) has proposed the following generator:

**Mardia's generator for Plackett's bivariate uniform family**

Generate two iid uniform [0,1] random variables $U, V$.
$X_1 \leftarrow U$
$Z \leftarrow V(1-V)$
$X_2 \leftarrow \frac{2Z(a^2X_1+1-X_1)+a(1-2Z)-(1-2V)\sqrt{a(a+4ZX_1(1-X_1)(1-a)^2)}}{a+Z(1-a)^2}$
RETURN $(X_1, X_2)$

Show that this algorithm is valid.

3. Suggest generators for the following bivariate uniform families of distributions:

| Density | Parameter(s) | Reference |
|---|---|---|
| $1+a((m+1)x_1^m-1)((n+1)x_2^n-1)$ | $\frac{1}{mn} \le a \le \max(m,n)$ , $m,n \ge 0$ | Farlie (1960) |
| $\frac{a(x_1^{1-a}+x_2^{1-a}-1)^{\frac{2a-1}{1-a}}}{(x_1x_2)^a}$ | $a > 1$ | (derived from multivariate Pareto) |
| $\frac{\pi(1+u^2)(1+v^2)}{2+u^2+v^2}$ where $u=1/\tan^2(\pi x_1)$ , $v=1/\tan^2(\pi x_2)$ | | Mardia (1970) (derived from multivariate Cauchy) |
| $1+a(2x_1-1)(2x_2-1)+b(3x_1^2-1)(3x_2^2-1)$ | $\mid a \mid \le \frac{1}{2}$ , $\mid b \mid \le \frac{1}{8}$ | Kimeldorf and Sampson (1975) |

4. This is about various measures of association. Construct a bivariate uniform distribution for which $\rho=\rho_g=\tau=0$, and $X_2=g(X_1)$ for some function $g$ (i.e. $X_2$ is completely dependent on $X_1$, see e.g. Lancaster, 1963).

5. Show that for the normal distribution in $R^2$, $\mid \rho \mid = \rho * = \overline{\rho}$.

6. Prove that $\overline{\rho}=0$ implies independence of components (Renyi, 1959).

7. Recall the definition of complete dependence of exercise 3.4. Construct a sequence of bivariate uniform distributions in which for every $n$, the second coordinate is completely dependent on the first coordinate. The sequence should also tend in distribution to the independent bivariate uniform distribution (Kimeldorf and Sampson, 1978). Conclude that the notion of complete dependence is peculiar.

8. The phenomenon described in exercise 7 cannot happen for monotone dependent sequences. If a sequence of random bivariate uniform random vectors in

which the second component is monotone dependent on the first component for all $n$, tends in distribution to a random vector, then this new random vector is bivariate uniform, and the second component is monotone dependent on the first component (Kimeldorf and Sampson, 1978).

9. One measure of association for bivariate distributions is

$$L = \sup_A \; | P((X_1,X_2)\in A) - P((X_1,X'_2)\in A) |$$

$$= \frac{1}{2} \int | f(x_1,x_2) - f_1(x_1) f_2(x_2) | \; dx_1 dx_2 ,$$

where $A$ is a Borel set of $R^2$, $X'_2$ is distributed as $X_2$, but is independent of $X_1$, $f$ is the density of $(X_1,X_2)$ and $f_1,f_2$ are the marginal densities. The second equality is valid only if the densities involved in the right-hand-side exist. Prove the second equality (Scheffe, 1947).

10. Nagao and Kadoya (1971) studied the following bivariate exponential density:

$$f(x_1,x_2) = \frac{e^{-\frac{1}{1-r}\left(\frac{x_1}{\sigma_1}+\frac{x_2}{\sigma_2}\right)} I_0\left(\frac{2}{1-r}\sqrt{\frac{r x_1 x_2}{\sigma_1\sigma_2}}\right)}{\sigma_1\sigma_2(1-r)}$$

where $r \in [0,1)$ is a measure of dependence, $\sigma_1,\sigma_2>0$ are constants (parameters), and $I_0$ is a modified Bessel function of the first kind. Obtain the parameters of the marginal exponential distributions. Compute the correlation coefficient $\rho$. Finally indicate how you would generate random vectors in uniformly bounded expected time.

11. Show that the property of comprehensiveness of a bivariate family is invariant under strictly monotone transformations of the coordinate axes (Kimeldorf and Sampson, 1975).

12. Show that Plackett's bivariate family with parameter $a \geq 0$ is comprehensive. Show in particular that Frechet's extremal distributions are attained for $a=0$ and $a\rightarrow\infty$, and that the product of the marginals is obtained for $a=1$.

13. Show that the standard bivariate normal family (i.e., the normal distribution in the plane) with variable correlation is comprehensive.

14. Show that Morgenstern's bivariate family is not comprehensive.

15. Consider the Johnson-Tenenbein family of Example 3.4, with parameter $c \in [0,1]$. Let $U$ and $V$ have uniform $[0,1]$ densities.

    A. Find $H$ such that the distribution is bivariate uniform. Hint: $H$ is parabolic on $[0,b]$ and $[1-b,1]$, and linear in between, where $b=\min(c,1-c)$.

    B. Find $\rho,\tau$ and $\rho_g$ as a function of $c$. In particular, prove that

$$\tau = \begin{cases} \dfrac{4c-5c^2}{6(1-c)^2} & 0<c<\dfrac{1}{2} \\[2ex] \dfrac{11c^2-6c+1}{6c^2} & \dfrac{1}{2}<c<1 \end{cases} ,$$

$$\rho_g = \begin{cases} \dfrac{10c-13c^2}{10(1-c)^2} & 0<c<\dfrac{1}{2} \\ \dfrac{3c^3+16c^2-11c+2}{10c^3} & \dfrac{1}{2}<c<1 \end{cases} .$$

Conclude that all nonnegative values for $\rho$, $\tau$ and $\rho_g$ are achievable by adjusting $c$ (Johnson and Tenenbein, 1981).

16. Show that for Gumbel's bivariate exponential family with parameter $a \in [0,1]$, the correlation reaches a minimum for $a=1$, and this minimum is $-0.40365...$. Show that the correlation is a decreasing function of $a$, taking the maximal value 0 at $a=0$.

17. Consider the following pair of random variables: $\beta_1 E_1+S_1 E_2, \beta_2 E_2+S_2 E_1$ where $P(S_i=1)=1-P(S_i=0)=1-\beta_i$ $(i=1,2)$ and $E_1,E_2$ are iid exponential random variables (Lawrance and Lewis (1983)). Does this family contain one of Frechet's extremal distributions?

18. Compute $\rho,\rho_g$ and $\tau$ for the bivariate exponential distribution of Johnson and Tenenbein (1981), defined as the distribution of $E_1, -\log((1-c)e^{-\frac{E_2}{1-c}}+ce^{-\frac{E_2}{c}})+\log(1-2c)$ where $c \in [0,1]$ and $E_1,E_2$ are iid exponential random variables.

19. Schmeiser and Lai (1982) proposed the following method for generating a bivariate gamma random vector: let $G_1,G_2,G_3$ be independent gamma random variables with respective parameters $a_1,a_2,a_3$, let $U,V$ be an independent bivariate uniform random vector with $V=U$ or $V=1-U$, let $F_b$ denote the gamma distribution function with parameter $b$, and let $b_1,b_2$ be two nonnegative numbers. Define

$$(X_1,X_2) = (F_{b_1}^{-1}(U)+G_1+G_3, F_{b_2}^{-1}(V)+G_2+G_3) .$$

A. Show that this random vector is bivariate gamma.

B. Show constructively that the five-parameter family is comprehensive, i.e. for every possible combination of specified marginal gamma distributions, give the values of the parameters needed to obtain the Frechet extremal distributions and the product distribution. Indicate also whether $V=U$ or $V=1-U$ is needed each time.

C. Show that by varying the five parameters, we can cover all theoretically possible combinations for the correlation coefficient and the marginal gamma parameters.

D. Consider the simplified three parameter model

$$(X_1,X_2) = (F_{b_1}^{-1}(U)+G_1, F_{\alpha_2}^{-1}(V))$$

for generating a bivariate gamma random vector with marginal parameters $(\alpha_1,\alpha_2)$ and correlation $\rho$. Show that this family is still comprehensive. There are two equations for the two free parameters ($b_1$ and $a_1$).

Suggest a good numerical algorithm for finding these parameters.

20. **A bivariate Poisson distribution.** $(X_1,X_2)$ is said to be bivariate Poisson with parameters $\lambda_1,\lambda_2,\lambda_3$, when it has characteristic function

$$\phi(t_1,t_2) = e^{\lambda_1(e^{it_1}-1)+\lambda_2(e^{it_2}-1)+\lambda_3(e^{it_1+it_2}-1)} .$$

A. Show that this is indeed a bivariate Poisson distribution.

B. Apply the trivariate reduction principle to generate a random vector with the given distribution.

C. (Kemp and Loukas, 1978). Show that we can generate the random vector as $(Z+W,X_2)$ where $X_2$ is Poisson $(\lambda_1+\lambda_3)$, and given $X_2$, $Z,W$ are independent Poisson $(\lambda_2)$ and binomial $(X_2,\lambda_3/(\lambda_1+\lambda_3))$ random variables. Hint: prove this via generating functions.

21. **The Johnson-Ramberg bivariate uniform family.** Let $U_1,U_2,U_3$ be iid uniform [0,1] random variables, and let $b \geq 0$ be a parameter of a family of bivariate uniform random vectors defined by

$$(X_1,X_2) = \left(\frac{U_1U_3{}^b - bU_1{}^{\frac{1}{b}}U_3}{1-b} , \frac{U_2U_3{}^b - bU_2{}^{\frac{1}{b}}U_3}{1-b}\right) .$$

This construction can be considered as trivariate reduction. Show that the full range of nonnegative correlations is possible, by first showing that the correlation is

$$\frac{b^2(2b^2+9b+6)}{(1+b)^2(1+2b)(2+b)} .$$

Show also that one of the Frechet extremal distributions can be approximated arbitrarily closely from within the family. For $b=1$, the defining formula is invalid. By what should it be replaced? (Johnson and Ramberg, 1977)

22. Consider a family of univariate distribution functions $\{1-(1-F)^a , a>0\}$, where $F$ is a distribution function. Families of this form are closed under the operation $\min(X_1,X_2)$ where $X_1,X_2$ are independent random variables with parameters $a_1,a_2$: the parameter of the minimum is $a_1+a_2$. Use this to construct a bivariate family via trivariate reduction, and compute the correlations obtainable for bivariate exponential, geometric and Weibull distributions obtained in this manner (Arnold, 1967).

23. **The bivariate Hermite distribution.** A univariate Hermite distribution $\{p_i , i \geq 0\}$ with parameters $a,b>0$ is a distribution on the nonnegative integers which has generating function (defined as $\sum_i p_i s^i$)

$$e^{a(s-1)+b(s^2-1)} .$$

The bivariate Hermite distribution with parameters $a_i > 0$, $i=1,2,\ldots,5$, is defined on all pairs of nonnegative integers and has bivariate generating function (defined as $E(s_1^{X_1}s_2^{X_2})$ where $(X_1,X_2)$ is a bivariate Hermite random vector)

$$e^{a_1(s_1-1)+a_2(s_1^2-1)+a_3(s_2-1)+a_4(s_2^2-1)+a_5(s_1s_2-1)}$$

(Kemp and Kemp (1965,1966); Kemp and Papageorgiou (1976)).

A. How can you generate a univariate Hermite $(a,b)$ random variate using only Poisson random variates in uniformly bounded expected time?

B. Give an algorithm for the efficient generation of bivariate Hermite random variates. Hint: derive first the generating function of $(X_1+X_3,X_2+X_3)$ where $X_1,X_2,X_3$ are independent random variables with generating functions $g_1,g_2,g_3$.

This exercise is adapted from Kemp and Loukas (1978).

24. Write an algorithm for computing the probabilities of a bivariate discrete distribution on $\{1,2,\ldots,K\}^2$ with specified marginal distributions, and achieving Frechet's inequality. Repeat for both of Frechet's extremal distributions.

## 4. THE DIRICHLET DISTRIBUTION.

### 4.1. Definitions and properties.

Let $a_1,\ldots,a_{k+1}$ be positive numbers. Then $(X_1,\ldots,X_k)$ has a **Dirichlet distribution** with parameters $(a_1,\ldots,a_{k+1})$, denoted $(X_1,\ldots,X_k) \sim D(a_1,\ldots,a_{k+1})$, if the joint distribution has density

$$f(x_1,\ldots,x_k) = cx_1^{a_1-1}\cdots x_k^{a_k-1}(1-x_1-\cdots-x_k)^{a_{k+1}-1}$$

over the $k$-dimensional simplex $S_k$ defined by the inequalities $x_i>0$ $(i=1,2,\ldots,k)$, $\sum_{i=1}^{k} x_i<1$. Here $c$ is a normalization constant. Basically, the $X_i$'s can be thought of as $a_i$-spacings in a uniform sample of size $\sum a_j$ if the $a_i$'s are all positive integers. The only novelty is that the $a_i$'s are now allowed to take non-integer values. The interested reader may want to refer back to section V.2 for the properties of spacings and to section V.3 for generators. The present section is only a refinement of sorts.

**Theorem 4.1.**

Let $Y_1, \ldots, Y_{k+1}$ be independent gamma random variables with parameters $a_i > 0$ respectively. Define $Y = \sum Y_i$ and $X_i = Y_i / Y$ $(i = 1,2, \ldots, k)$. Then $(X_1, \ldots, X_k) \sim D(a_1, \ldots, a_{k+1})$ and $(X_1, \ldots, X_k)$ is independent of $Y$.

Conversely, if $Y$ is gamma $(\sum a_i)$, and $Y$ is independent of $(X_1, \ldots, X_k) \sim D(a_1, \ldots, a_{k+1})$, then the random variables $YX_1, \ldots, YX_k, Y(1-\sum_{i=1}^{k} X_i)$ are independent gamma random variables with parameters $a_1, \ldots, a_{k+1}$.

**Proof of Theorem 4.1.**

The joint density of the $Y_i$'s is

$$f(y_1, \ldots, y_{k+1}) = c \prod_{i=1}^{k+1} y_i^{a_i - 1} e^{-\sum_{i=1}^{k+1} y_i}$$

where $c$ is a normalization constant. Consider the transformation $y = \sum_{y_i}, x_i = y_i / y$ $(i \le k)$, which has as reverse transformation $y_i = yx_i$ $(i \le k), y_{k+1} = y(1-\sum_{i=1}^{k} x_i)$. The Jacobian of the transformation is $y^k$. Thus, the joint density of $Y, X_1, \ldots, X_k$ is

$$g(y, x_1, \ldots, x_k) = c \prod_{i=1}^{k} x_i^{a_i - 1} (1-\sum_{i=1}^{k} x_i)^{a_{k+1}-1} y^{\sum_{i=1}^{k+1} a_i - 1} e^{-y} .$$

This proves the first part of the Theorem. The proof of the second part is omitted. ■

Theorem 4.1 suggests a generator for the Dirichlet distribution via gamma generators. There are important relationships with the beta distribution as well, which are reviewed by Wilks (1962), Aitchison (1963) and Basu and Tiwari (1982). Here we will just mention the most useful of these relationships.

**Theorem 4.2.**

Let $Y_1, \ldots, Y_k$ be independent beta random variables where $Y_i$ is beta $(a_i, a_{i+1}+ \cdots +a_{k+1})$. Then $(X_1, \ldots, X_k) \sim D(a_1, \ldots, a_{k+1})$ where the $X_i$'s are defined by

$$X_i = Y_i \prod_{j=1}^{i-1} Y_j .$$

Conversely, when $(X_1, \ldots, X_k) \sim D(a_1, \ldots, a_{k+1})$, then the random variables $Y_1, \ldots, Y_k$ defined by

$$Y_i = \frac{X_i}{1-X_1- \cdots -X_{i-1}}$$

are independent beta random variables with parameters given in the first statement of the Theorem.

**Theorem 4.3.**

Let $Y_1, \ldots, Y_k$ be independent random variables, where $Y_i$ is beta $(a_1+ \cdots +a_i, a_{i+1})$ for $i<k$ and $Y_k$ is gamma $(a_1+ \cdots +a_k)$. Then the following random variables are independent gamma random variables with parameters $a_1, \ldots, a_k$:

$$X_i = (1-Y_{i-1}) \prod_{j=i}^{k} Y_j \qquad (i=1,2, \ldots, k) .$$

To avoid trivialities, set $Y_0=0$.

Conversely, when $X_1, \ldots, X_k$ are independent gamma random variables with parameters $a_1, \ldots, a_k$, then the $Y_i$'s defined by

$$Y_i = \frac{X_1+ \cdots +X_i}{X_1+ \cdots +X_{i+1}} \qquad (i=1,2, \ldots, k-1)$$

and

$$Y_k = X_1+ \cdots +X_k$$

are independent. Here $Y_i$ is beta $(a_1+ \cdots +a_i, a_{i+1})$ for $i<k$ and $Y_k$ is gamma $(a_1+ \cdots +a_k)$.

The proofs of Theorems 4.2 and 4.3 do not differ substantially from the proof of Theorem 4.1, and are omitted. See however the exercises. Theorem 4.2 tells us how to generate a Dirichlet random vector by transforming a sequence of beta random variables. Typically, this is more expensive than generating a Dirichlet random vector by transforming a sequence of gamma random variables, as is suggested by Theorem 4.1.

Theorem 4.2 also tells us that the marginal distributions of the Dirichlet distribution are all beta. In particular, when $(X_1, \ldots, X_k) \sim D(a_1, \ldots, a_{k+1})$, then $X_i$ is beta $(a_i, \sum_{j \neq i} a_j)$.

Theorem 4.1 tells us how to relate independent gammas to a Dirichlet random vector. Theorem 4.2 tells us how to relate independent betas to a Dirichlet distribution. These two connections are put together in Theorem 4.3, where independent gammas and betas are related to each other. This offers the exciting possibility of using simple transformations to transform long sequences of gamma random variables into equally long sequences of beta random variables. Unfortunately, the beta random variables do not have equal parameters. For example, consider $k$ iid gamma $(a)$ random variables $X_1, \ldots, X_k$. Then the second part of Theorem 4.3 tells us how to obtain independent random variables distributed as beta $(a, a)$, beta $(2a, a)$, ..., beta $((k-1)a, a)$ and gamma $(ka)$ random variables respectively. When $a=1$, this reduces to a well-known property of spacings given in section V.2.

We also deduce that $BG, (1-B)G$ are independent gamma $(a)$, gamma $(b)$ random variables when $G$ is gamma $(a+b)$ and $B$ is beta $(a, b)$ and independent of $G$. In particular, we obtain Stuart's theorem (Stuart, 1962), which gives us a very fast method for generating gamma $(a)$ random variates when $a<1$: a gamma $(a)$ random variate can be generated as the product of a gamma $(a+1)$ random variate and an independent beta $(a, 1)$ random variate (the latter can be obtained as $e^{-E/a}$ where $E$ is exponentially distributed).

## 4.2. Liouville distributions.

Sivazlian (1981) introduced the class of Liouville distributions, which generalizes the Dirichlet distributions. These distributions have a density on $R^k$ given by

$$c\,\psi\left(\sum_{i=1}^{k} x_i\right) \prod_{i=1}^{k} x_i^{a_i-1} \quad (x_i \geq 0,\ i=1,2, \ldots, k),$$

where $\psi$ is a Lebesgue measurable nonnegative function, $a_1, \ldots, a_k$ are positive constants (parameters), and $c$ is a normalization constant. The functional form of $\psi$ is not fixed. Note however that not all nonnegative functions $\psi$ can be substituted in the formula for the density because the integral of the unnormalized density has to be finite. A random vector with the density given above is said to be Liouville $L_k(\psi, a_1, \ldots, a_k)$. Sivazlian (1981) calls this distribution a **Liou-**

**ville distribution of the first kind.**

**Example 4.1. Independent gamma random variables.**

When $X_1, \ldots, X_k$ are independent gamma random variables with parameters $a_1, \ldots, a_k$, then $(X_1, \ldots, X_k)$ is $L_k(e^{-x}, a_1, \ldots, a_k)$. ■

**Example 4.2.**

A random variable $X$ with density $c\,\psi(x)x^{a-1}$ on $[0,\infty)$ is $L_1(\psi,a)$. This family of distributions contains all densities on the positive halfline. ■

We are mainly interested in generating random variates from multivariate Liouville distributions. It turns out that two key ingredients are needed here: a Dirichlet generator, and a generator for univariate Liouville distributions of the form given in Example 4.2. The key property is given in Theorem 4.4.

**Theorem 4.4. (Sivazlian, 1981)**

The normalization constant $c$ for the Liouville $L_k(\psi, a_1, \ldots, a_k)$ density is given by

$$\frac{\Gamma(\sum_{i=1}^{k} a_i)}{\prod_{i=1}^{k} \Gamma(a_i) \int_0^{\infty} \psi(x) x^{a-1}\, dx}$$

where $a = \sum_{i=1}^{k} a_i$.

Let $(X_1, \ldots, X_k)$ be $L_k(\psi, a_1, \ldots, a_k)$, and let $(Y_1, \ldots, Y_k)$ be defined by

$$Y_i = \frac{X_i}{X_1 + \cdots + X_k} \quad (1 \leq i < k),$$

$$Y_k = X_1 + \cdots + X_k .$$

Then $(Y_1, \ldots, Y_{k-1})$ is Dirichlet $(a_1, \ldots, a_k)$, and $Y_k$ is independent of this Dirichlet random vector and $L_1(\psi, \sum_{i=1}^{k} a_i)$.

Conversely, if $(Y_1, \ldots, Y_{k-1})$ is Dirichlet $(a_1, \ldots, a_k)$, and $Y_k$ is independent of this Dirichlet random vector and $L_1(\psi, \sum_{i=1}^{k} a_i)$, then the random vector $(X_1, \ldots, X_k)$ defined by

$$X_i = Y_i Y_k \quad (1 \leq i < k),$$

$$X_k = (1 - Y_1 - \cdots - Y_{k-1}) Y_k .$$

is $L_k(\psi, a_1, \ldots, a_k)$.

**Proof of Theorem 4.4.**

The constant $c$ is given by

$$\frac{1}{c} = \int_0^{\infty} \cdots \int_0^{\infty} \psi(\sum_{i=1}^{k} x_i) \prod_{i=1}^{k} x_i^{a_i - 1} dx_1 \cdots dx_k$$

$$= \frac{\prod_{i=1}^{k} \Gamma(a_i)}{\Gamma(\sum_{i=1}^{k} a_i)} \int_0^{\infty} \psi(x) x^{a-1}\, dx ,$$

where a property of Liouville multiple integrals is used (Sivazlian, 1981). This proves the first part of the Theorem.

Assume next that $(X_1, \ldots, X_k)$ is $L_k(\psi, a_1, \ldots, a_k)$, and that $(Y_1, \ldots, Y_k)$ is obtained via the transformation given in the statement of the Theorem. This transformation has Jacobian $Y_k^{\,k-1}$. The joint density of $(Y_1, \ldots, Y_k)$ is

$$c y_k^{\,k-1} \psi(y_k) \prod_{i=1}^{k-1} (y_i y_k)^{a_i - 1} ((1-\sum_{i=1}^{k-1} y_i) y_k)^{a_k - 1}$$

$$= c\, \psi(y_k) y_k^{\,a-1} \prod_{i=1}^{k-1} y_i^{\,a_i-1} (1-\sum_{i=1}^{k-1} y_i)^{a_k-1} \qquad (y_i \geq 0 \ (i=1,2, \ldots, k-1)\ ,\ \sum_{i=1}^{k-1} y_i \leq 1)\ .$$

In this we recognize the product of an $L_1(\psi, a)$ density (for $Y_k$), and a Dirichlet $(a_1, \ldots, a_k)$ density (for $(Y_1, \ldots, Y_{k-1})$). This proves the second part of the Theorem.

For the third part, we argue similarly, starting from the last density shown above. After the transformation to $(X_1, \ldots, X_k)$, which has Jacobian $(\sum_{i=1}^{k} X_i)^{k-1}$, we obtain the $L_k(\psi, a_1, \ldots, a_k)$ density again. ∎

Dirichlet generators are described in section 4.1, while $L_1(\psi, a)$ generators can be handled individually based upon the particular form for $\psi$. Since this is a univariate generation problem, we won't be concerned with the associated problems here.

## 4.3. Exercises.

1. Prove Theorems 4.2 and 4.3.
2. Prove the following fact: when $(X_1, \ldots, X_k) \sim D(a_1, \ldots, a_{k+1})$, then $(X_1, \ldots, X_i) \sim D(a_1, \ldots, a_i, \sum_{j=i+1}^{k+1} a_j)$, $i < k$.
3. **The generalized Liouville distribution.** A random vector $(X_1, \ldots, X_k)$ is generalized Liouville (Sivazlian, 1981) when it has a density which can be written as

$$c\, \psi(\sum_{i=1}^{k} (\frac{x_i}{c_i})^{b_i}) \prod_{i=1}^{k} x_i^{\,a_i-1} \qquad (x_i \geq 0)\ .$$

Here $a_i, b_i, c_i > 0$ are parameters, $\psi$ is a nonnegative Lebesgue measurable function, and $c$ is a normalization constant. Generalize Theorem 4.4 to this distribution. In particular, show how you can generate random vectors with this distribution when you have a Dirichlet generator and an $L_1(\psi, a)$ generator at your disposal.

4. In the proof of Theorem 4.4, prove the two statements made about the Jacobian of the transformation.

## 5. SOME USEFUL MULTIVARIATE FAMILIES.

### 5.1. The Cook-Johnson family.

Cook and Johnson (1981) consider the multivariate uniform distribution defined as the distribution of

$$(X_1, \ldots, X_d) = ((1+\frac{E_1}{S})^{-a}, \ldots, (1+\frac{E_d}{S})^{-a}),$$

where $E_1, \ldots, E_d$ are iid exponential random variables, $S$ is an independent gamma ($a$) random variable, and $a>0$ is a parameter. This family is interesting from a variety of points of view:

A. Random variate generation is easy.

B. Many multivariate distributions can be obtained by appropriate monotone transformations of the components, such as the multivariate logistic distribution (Satterthwaite and Hutchinson, 1978; Johnson and Kotz, 1972, p. 291), the multivariate Burr distribution (Takahasi, 1965; Johnson and Kotz, 1972, p. 289), and the multivariate Pareto distribution (Johnson and Kotz, 1972, p. 286).

C. For $d=2$, the full range of nonnegative correlations can be achieved. The independent bivariate uniform distribution and one of Frechet's extremal distributions (corresponding to the case $X_2=X_1$) are obtainable as limits.

**Theorem 5.1.**

The Cook-Johnson distribution has distribution function

$$F(x_1, \ldots, x_d) = \left( \sum_{i=1}^{d} x_i^{-\frac{1}{a}} - (d-1) \right)^{-a} \qquad (0 < x_i \le 1 \ , \ i=1,2, \ldots, d)$$

and density

$$f(x_1, \ldots, x_d) = \frac{\Gamma(a+d)}{\Gamma(a) a^d} \prod_{i=1}^{d} x_i^{-\frac{1}{a}-1} \left( \sum_{i=1}^{d} x_i^{-\frac{1}{a}} - (d-1) \right)^{-(a+d)}$$

$$(0 < x_i \le 1 \ , \ i=1,2, \ldots, d) \ .$$

The distribution is invariant under permutations of the coordinates, and is multivariate uniform. Furthermore, as $a \to \infty$, the distribution function converges to $\prod_{i=1}^{d} x_i$ (the independent case), and as $a \downarrow 0$, it converges to $\min(x_1, \ldots, x_d)$ (the totally dependent case).

**Proof of Theorem 5.1.**

The distribution function is derived without difficulty. The density is obtained by differentiation. The permutation invariance follows by inspection. The marginal distribution function of the first component is $F(x_1,1, \ldots, 1)=x_1$ for $0<x_1 \le 1$. Thus, the distribution is multivariate uniform. The limit of the distribution function as $a \downarrow 0$ is $\min(x_1, \ldots, x_d)$. Similarly, for $0<\min(x_1, \ldots, x_d) \le \max(x_1, \ldots, x_d) < 1$, as $a \to \infty$,

$$F(x_1, \ldots, x_d) = \left( \sum_{i=1}^{d} e^{-\frac{\log(x_i)}{a}} - (d-1) \right)^{-a}$$

$$= \left( \sum_{i=1}^{d} \left(1 - \frac{\log(x_i)}{a} + O(a^{-2})\right) - (d-1) \right)^{-a}$$

$$= \left( 1 - \frac{\log(\prod_{i=1}^{d} x_i)}{a} + O(a^{-2}) \right)^{-a}$$

$$\sim e^{\log(\prod_{i=1}^{d} x_i)} = \prod_{i=1}^{d} x_i \ . \blacksquare$$

Let us now turn to a collection of other distributions obtainable from the Cook-Johnson family with parameter $a$ by simple transformations of the $X_i$'s. Some transformations to be applied to each $X_i$ are shown in the next table.

| Transformation for $X_i$ | Parameters | Resulting distribution | Reference |
|---|---|---|---|
| $-\log(X_i^{-\frac{1}{a}}-1)$ | | Gumbel's bivariate logistic ($d=2$) and the multivariate logistic ($a=1$) | Satterthwaite and Hutchinson (1978), Johnson and Kotz (1972, p. 291) |
| $(d_i(X_i^{-\frac{1}{a}}-1))^{\frac{1}{c_i}}$ | $c_i, d_i > 0$ | multivariate Burr | Takahasi (1965), Johnson and Kotz (1972, p. 286) |
| $a_i X_i^{-\frac{1}{a}}$ | $a_i > 0$ | multivariate Pareto | Johnson and Kotz (1972, p. 286) |
| $\Phi^{-1}(X_i)$ | None. $\Phi$ is the normal distribution function. | multivariate normal without elliptical contours | Cook and Johnson (1981) |

## Example 5.1. The multivariate logistic distribution.

In 1961, Gumbel proposed the bivariate logistic distribution, a special case of the generalized multivariate logistic distribution with distribution function

$$\left(1+\sum_{i=1}^{d} e^{-x_i}\right)^{-a} \qquad (x_i > 0 \,,\, i=1,2, \ldots, d) \,.$$

For $a=1$ this reduces to the multivariate logistic distribution given by Johnson and Kotz (1972, p. 293). Note that from the form of the distribution function, we can deduce immediately that all univariate and multivariate marginals are again multivariate logistic. Transformation of a Cook-Johnson random variate leads to the following simple recipe for generating multivariate logistic random variates:

**Multivariate logistic generator**

Generate iid exponential random variates $E_1, \ldots, E_{d+1}$.

RETURN $(\log(\frac{E_1}{E_{d+1}}), \ldots, \log(\frac{E_d}{E_{d+1}}))$ ■

**Example 5.2.**

The multivariate normal distribution in the table has nonelliptical contours. Kowalski (1973) provides other examples of multivariate normal distributions with nonnormal densities. ■

## 5.2. Multivariate Khinchine mixtures.

Bryson and Johnson (1982) proposed the family of distributions defined constructively as the distributions of random vectors in $R^d$ which can be written as

$$(Z_1 U_1, \ldots, Z_d U_d)$$

where the $Z_1, \ldots, Z_d$ is independent of the multivariate uniform random vector $U_1, \ldots, U_d$, and has a distribution which is such that certain given marginal distributions are obtained. Recalling Khinchine's theorem (section IV.6.2), we note that all marginal distributions have unimodal densities.

Controlled dependence can be introduced in many ways. We could introduce dependence in $U_1, \ldots, U_d$ by picking a multivariate uniform distribution based upon the multivariate normal density or the Cook-Johnson distribution. Two models for the $Z_i$'s seem natural:

A. The identical model: $Z_1 = \cdots = Z_d$.

B. The independent model: $Z_1, \ldots, Z_d$ are iid.

These models can be mixed by choosing the identical model with probability $p$ and the independent model with probability $1-p$.

**Example 5.3.**

To achieve exponential marginals, we can take all $Z_i$'s gamma (2). In the identical bivariate model, the joint bivariate density is

$$\int_{\max(x_1,x_2)}^{\infty} \frac{e^{-t}}{t}\, dt \ .$$

In the independent bivariate model, the joint density is

$$\frac{x_1 x_2}{(x_1+x_2)^3}(2+2(x_1+x_2)+(x_1+x_2)^2)e^{-(x_1+x_2)} \ .$$

Unfortunately, the correlation in the first model is $\frac{1}{2}$, and that of the second model is $\frac{1}{3}$. By probability mixing, we can only cover correlations in the small range $[\frac{1}{3},\frac{1}{2}]$. Therefore, it is useful to replace the independent model by the totally independent model (with density $e^{-(x_1+x_2)}$), thereby enlarging the range to $[0,\frac{1}{2}]$. ■

**Example 5.4. Nonnormal bivariate normal distributions.**

For symmetric marginals, it is convenient to take the $U_i$'s uniform on $[-1,1]$. It is easy to see that in order to obtain normal marginals, the $Z_i$'s have to be distributed as the square roots of chi-square random variables with 3 degrees of freedom. If $(U_1,U_2)$ has bivariate density $h$ on $[-1,1]^2$, then $(Z_1U_1,Z_1U_2)$ has joint density

$$\int_{\max(|x_1|,|x_2|)}^{\infty} \Gamma^{-1}(\frac{3}{2})2^{-\frac{5}{2}} e^{-\frac{t^2}{2}} h(\frac{1}{2}+\frac{y_1}{2t},\frac{1}{2}+\frac{y_2}{2t})\,dt \ .$$

This provides us with a rich source of examples of bivariate distributions with normal marginals, zero correlations and non-normal densities. At the same time, random variate generation for these examples is trivial (Bryson and Johnson, 1982). ■

## 5.3. Exercises.

1. **The multivariate Pareto distribution.** The univariate Pareto density with parameter $a>0$ is defined by $a/x^{a+1}$ $(x \geq 1)$. Johnson and Kotz (1972, p. 286) define a multivariate Pareto density on $R^d$ with parameter $a$ by

$$\frac{a(a+1)\cdots(a+d-1)}{(\sum_{i=1}^{d} x_i-(d-1))^{a+d}} \qquad (x_i \geq 1 \ , \ i=1,2, \ldots, d) \ .$$

A. Show that the marginals are all univariate Pareto with parameter $a$.

B. In the bivariate case, show that the correlation is $\frac{1}{a}$. Since the marginal variance is finite if and only if $a>2$, we see that all correlations between 0 and $\frac{1}{2}$ can be achieved.

C. Prove that a random vector can be generated as $(X_1^{-\frac{1}{a}}, \ldots, X_d^{-\frac{1}{a}})$ where $(X_1, \ldots, X_d)$ has the Cook-Johnson distribution with parameter $a$. Equivalently, it can be generated as $(1+\frac{E_1}{S}, \ldots, 1+\frac{E_d}{S})$, where $E_1, \ldots, E_d$ are iid exponential random variables, and $S$ is an independent gamma $(a)$ random variable.

## 6. RANDOM MATRICES.

### 6.1. Random correlation matrices.

To test certain statistical methods, one should be able to create random test problems. In several applications, one needs a random correlation matrix. This problem is equivalent to that of the generation of a random covariance matrix if one asks that all variances be one. Unfortunately, posed as such, there are infinitely many answers. Usually, one adds structural requirements to the correlation matrix in terms of expected value of elements, eigenvalues, and distributions of elements. It would lead us too far to discuss all the possibilities in detail. Instead, we just kick around a few ideas to help us to better understand the problem. For a recent survey, consult Marsaglia and Olkin (1984).

A correlation matrix is a symmetric positive semi-definite matrix with ones on the diagonal. It is well known that if $\mathbf{H}$ is a $d \times n$ matrix with $n \geq d$, then $\mathbf{HH}'$ is a symmetric positive semi-definite matrix. To make it a correlation matrix, it is necessary to make the rows of $\mathbf{H}$ of length one (this forces the diagonal elements to be one). Thus, we have the following property, due to Marsaglia and Olkin (1984):

**Theorem 6.1.**

$\mathbf{HH}'$ is a random correlation matrix if and only if the rows of $\mathbf{H}$ are random vectors on the unit sphere of $R^n$.

Theorem 6.1 leads to a variety of algorithms. One still has the freedom to choose the random rows of $\mathbf{H}$ according to any recipe. It seems logical to take the rows as independent uniformly distributed random vectors on the surface of $C_n$, the unit sphere of $R^n$, where $n \geq d$ is chosen by the user. For this case, one can actually compute the explicit form of the marginal distributions of $\mathbf{HH}'$. Marsaglia and Olkin suggest starting from any $d \times n$ matrix of iid random variables, and to normalize the rows. They also suggest in the case $n = d$ starting from lower triangular $\mathbf{H}$, thus saving about 50% of the variates.

The problem of the generation of a random correlation matrix with a given set of eigenvalues is more difficult. The diagonal matrix $\mathbf{D}$ defined by

$$\begin{vmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_d \end{vmatrix}$$

has eigenvalues $\lambda_1, \ldots, \lambda_d$. Also, eigenvalues do not change when $\mathbf{D}$ is pre and post multiplied with an orthogonal matrix. Thus, we need to make sure that there exist many orthogonal matrices $\mathbf{H}$ such that $\mathbf{HDH}'$ is a correlation matrix. Since the trace of our correlation matrix must be $d$, we have to start with a matrix $\mathbf{D}$ with trace $d$. For the construction of random orthogonal $\mathbf{H}$ that satisfy the given collection of equations, see Chalmers (1975), Bendel and Mickey (1978) and Marsaglia and Olkin (1984). See also Johnson and Welch (1980), Bendel and Afifi (1977) and Ryan (1980).

In a third approach, designed to obtain random correlation matrices with given mean $\mathbf{A}$, Marsaglia and Olkin (1984) suggest forming $\mathbf{A}+\mathbf{H}$ where $\mathbf{H}$ is a perturbation matrix. We have

**Theorem 6.2.**

Let $\mathbf{A}$ be a given $d \times d$ correlation matrix, and let $\mathbf{H}$ be a random symmetric $d \times d$ matrix whose elements are zero on the diagonal, and have zero mean off the diagonal. Then $\mathbf{A}+\mathbf{H}$ is a random correlation matrix with expected value $\mathbf{A}$ if and only if the eigenvalues of $\mathbf{A}+\mathbf{H}$ are nonnegative.

**Proof of Theorem 6.2.**

The expected value is obviously correct. Also, $\mathbf{A}+\mathbf{H}$ is symmetric. Furthermore, the diagonal elements are all one. Finally, $\mathbf{A}+\mathbf{H}$ is positive semi-definite when its eigenvalues are nonnegative. ■

We should also note that the eigenvalues of $\mathbf{A}+\mathbf{H}$ and those of $\mathbf{A}$ differ by at most

$$\Delta = \max\left(\sqrt{\sum_{i,j} h_{ij}^{\,2}}\ ,\ \max_i \sum_j |\, h_{ij}\,|\right),$$

where $h_{ij}$ is an element of $\mathbf{H}$. Thus, if $\Delta$ is less than the smallest eigenvalue of $\mathbf{A}$, then $\mathbf{A}+\mathbf{H}$ is a correlation matrix. Marshall and Olkin (1984) use this fact to suggest two methods for generating $\mathbf{H}$:

A. Generate all $h_{ij}$ for $i<j$ with zero mean and support on $[-b_{ij}, b_{ij}]$ where the $b_{ij}$'s form a zero diagonal symmetric matrix with $\Delta$ smaller than the smallest eigenvalue of $\mathbf{A}$. Then for $i>j$, define $h_{ij}=h_{ji}$. Finally, $h_{ii}=0$.

B. Generate $h_{12}, h_{13}, \ldots, h_{d-1,d}$ with a radially symmetric distribution in or on the $d(d-1)/2$ sphere of radius $\lambda/\sqrt{2}$ where $\lambda$ is the smallest eigenvalue of $\mathbf{A}$. Define the other elements of $\mathbf{H}$ by symmetry.

### 6.2. Random orthogonal matrices.

An orthonormal $d \times d$ matrix can be considered as a rotation of the coordinate axes in $R^d$. In such a rotation, there are $d(d-1)/2$ degrees of freedom. To see this, we look at where the points $(1,0,0, \ldots, 0), \ldots, (0,0, \ldots, 0,1)$ are mapped to by the orthonormal transformation. These points are mapped to other points on the unit sphere. In turn, the mapped points define the rotation. We can choose the first point ($d$ coordinates). Given the first point, the second point should be in a hyperplane perpendicular to the line joining the origin and the first point. Here we have only $d-1$ degrees of freedom. Continuing in this fashion, we see that there are $d(d-1)/2$ degrees of freedom in all.

Heiberger (1978) (correction by Tanner and Thisted (1982)) gives an algorithm for generating an orthonormal matrix which is uniformly distributed. This means that the first point is uniformly distributed on the unit sphere of $R^d$, that the second point is uniformly distributed on the unit sphere of $R^d$ intersected with the hyperplane which is perpendicular to the line from the origin to the first point, and so forth.

His algorithm requires $d(d+1)/2$ independent normal random variables, while the total time is $O(d^3)$. It is perhaps worth noting that no heavy matrix computations are necessary at all if one is willing to spend a bit more time. To illustrate this, consider performing $\binom{d}{2}$ random rotations of two axes, each rotation keeping the $d-2$ other axes fixed. A random rotation of two axes is easy to carry out, as we will see below. The global random rotation boils down to $\binom{d}{2}$ matrix multiplications. Luckily, each matrix is nearly diagonal: there are four random elements on the intersections of two given rows and columns. The remainder of each matrix is purely diagonal with ones on the diagonal. This structure implies that the time needed to compute the global (product) rotation matrix is $O(d^3)$.

**A random uniform rotation** of $R^2$ can be generated as

$$\begin{vmatrix} X & Y \\ -SY & SX \end{vmatrix}$$

where $(X,Y)$ is a point uniformly distributed on $C_2$, and $S$ is a random sign. A random rotation in $R^3$ in which the $z$-axis remains fixed is

$$\begin{vmatrix} X & Y & 0 \\ -SY & SX & 0 \\ 0 & 0 & 1 \end{vmatrix} .$$

Thus, by the threefold combination (i.e., product) of matrices of this type, we can obtain a random rotation in $R^3$. If $\mathbf{A}_{12}, \mathbf{A}_{23}, \mathbf{A}_{13}$ are three random rotations of two axes with the third one fixed, then the product

$$\mathbf{A}_{12}\mathbf{A}_{23}\mathbf{A}_{13}$$

is a random rotation of $R^3$.

### 6.3. Random R×C tables.

A two-way contingency table with $r$ rows and $c$ columns is a matrix of non-negative integer-valued numbers. It is also called an $R \times C$ table. Typically, the integers represent the frequencies with which a given pair of integers is observed in a sample of size $n$. The purpose of this section is to explore the generation of a random $R \times C$ table with given sample size (sum of elements) $n$. Again, this is an ill-posed problem unless we impose more structure on it. The standard restrictions are:

A. Generate a random table for sample size $n$, such that all tables are equally likely.

B. Generate a random table for sample size $n$, with given row and column totals. The row totals are called $r_i$ ,$1 \leq i \leq r$. The column totals are $c_i$ ,$1 \leq i \leq c$.

Let us just consider problem B. In a first approach, we take a ball-in-urn strategy. Consider balls numbered $1,2, \ldots, n$. Of these, the first $c_1$ are class one balls, the next $c_2$ are class two balls, and so forth. Think of classes as different colors. Generate a random permutation of the balls, and put the first $r_1$ balls in row 1, the next $r_2$ balls in row 2, and so forth. Within a given row, class $i$ balls should all be put in column $i$. This ball-in-urn method, first suggested by Boyett (1979), takes time proportional to $n$, and is not recommended when $n$ is much larger than $rc$, the size of the matrix.

**Ball-in-urn method**

```
[NOTE: N is an r × c array to be returned. B[1], . . . , B[n] is an auxiliary array.]
Sum ←0
FOR j:=1 TO c DO
      FOR i:= Sum+1 TO Sum+c_j DO B[i]←j
      Sum ← Sum + c_j
Randomly permute the array B.
Set N to all zeroes.
Sum ←0
FOR j:=1 TO r DO
      FOR i:= Sum+1 TO Sum+r_j DO N[j,B[i]]←N[j,B[i]]+1
      Sum ← Sum + r_j
RETURN N
```

Patefield (1980) uses the conditional distribution method to reduce the dependence of the performance upon $n$. The conditional distribution of an entry $N_{ij}$ given the entries in the previous rows, and the previous entries in the same row $i$ is given by

$$P(N_{ij}=k) = \frac{\alpha\beta\gamma\delta}{\epsilon\eta\varsigma\theta k!}$$

where

$$\alpha = (r_i - \sum_{l<j} N_{il})! ,$$

$$\beta = (n - \sum_{m \le i} r_m - \sum_{m<j} c_m + \sum_{l<j, m \le i} N_{ml})! ,$$

$$\gamma = (c_j - \sum_{m<i} N_{mj})! ,$$

$$\delta = \left( \sum_{l>j} (c_l - \sum_{m<i} N_{ml}) \right)! ,$$

$$\epsilon = (r_i - \sum_{l<j} N_{il} - k)! ,$$

$$\eta = (n - \sum_{m \le i} r_m - \sum_{m \le j} c_m + \sum_{l<j, m \le i} N_{ml} + \sum_{m<i} N_{mj} + k)! ,$$

$$\varsigma = (c_j - \sum_{m<i} N_{mj} - k)! ,$$

$$\theta = \left( \sum_{l \ge j} (c_l - \sum_{m<i} N_{ml}) \right)! .$$

The range for $k$ is such that all factorial terms are nonnegative. Although the expression for the conditional probabilities appears complicated, we note that quite a bit of regularity is present, which makes it possible to adjust the partial sums "on the fly". As we go along, we can quickly adjust all terms. More precisely, the constants needed for the computation of the probabilities of the next entry in the same row can be computed from the previous one and the value of the current element $N_{ij}$ in constant time. Also, there is a simple recurrence relation for the probability distribution as a function of $k$, which makes the distribution tractable by the sequential inversion method (as suggested by Patefield, 1980). However, the expected time of this procedure is not bounded uniformly in $n$ for fixed values of $r, c$.

### 6.4. Exercises.

1. Let $\mathbf{A}$ be a $d \times d$ correlation matrix, and let $\mathbf{H}$ be a symmetric matrix. Show that the eigenvalues of $\mathbf{A}+\mathbf{H}$ differ by at most $\Delta$ from the eigenvalues of $\mathbf{A}$, where

$$\Delta = \max\left(\sqrt{\sum_{i,j} h_{ij}^2}\,,\ \max_i \sum_j \mid h_{ij} \mid\right) .$$

2. Generate $h_{12}, h_{13}, \ldots, h_{d-1,d}$ with a radially symmetric distribution in or on the $d(d-1)/2$ sphere of radius $\lambda/\sqrt{2}$ where $\lambda$ is the smallest eigenvalue of $\mathbf{A}$. Define the other elements of $\mathbf{H}$ by symmetry. Put zeroes on the diagonal of $\mathbf{H}$. Then $\mathbf{A}+\mathbf{H}$ is a correlation matrix when $\mathbf{A}$ is. Show this.
3. Consider Patefield's conditional distribution method for generating a random $R \times C$ table. Show the following:
   - A. The conditional distribution as given in the text is correct.
   - B. (Difficult.) Design a constant expected time algorithm for generating one element in the $r \times c$ matrix. The expected time should be uniformly bounded over all conditions, but with $r$ and $c$ fixed.