Chapter Five
UNIFORM AND EXPONENTIAL SPACINGS

1. MOTIVATION.

The goal of this book is to demonstrate that random variates with various distributions can be obtained by cleverly manipulating iid uniform [0,1] random variates. As we will see in this chapter, normal, exponential, beta, gamma and t distributed random variates can be obtained by manipulation of the order statistics or spacings defined by samples of iid uniform [0,1] random variates. For example, the celebrated polar method or Box-Muller method for normal random variates will be derived in this manner (Box and Muller, 1958).

There is a strong relationship between these spacings and radially symmetric distributions in R^d, so that with a little extra effort we will be able to handle the problem of generating uniform random variates in and on the unit sphere of R^d.

The polar method can also be considered as a special case of a more general method, the method of deconvolution. Because of this close relationship it will also be presented in this chapter.

We start with the fundamental properties of uniform order statistics and uniform spacings. This material is well-known and can be found in many books on probability theory and mathematical statistics. It is collected here for the convenience of the readers. In the other sections, we will develop various algorithms for univariate and multivariate distributions. Because order statistics and spacings involve sorting random variates, we will have a short section on fast expected time sorting methods. Just as chapter IV, this chapter is highly specialized, and can be skipped too. Nevertheless, it is recommended for new students in the fields of simulation and mathematical statistics.

2. PROPERTIES OF UNIFORM AND EXPONENTIAL SPACINGS.

2.1. Uniform spacings.

Let $U_1, \ldots, U_n$ be iid uniform [0,1] random variables with order statistics $U_{(1)} \le U_{(2)} \le \cdots \le U_{(n)}$. The statistics S_i defined by

$$S_i = U_{(i)} - U_{(i-1)} \quad (1 \le i \le n+1)$$

where by convention $U_{(0)}=0$, $U_{(n+1)}=1$, are called the uniform spacings for this sample.

Theorem 2.1.

$(S_1, \ldots, S_n)$ is uniformly distributed over the simplex

$$A_n = \{(x_1, \ldots, x_n) : x_i \ge 0, \sum_{i=1}^{n} x_i \le 1\} .$$

Proof of Theorem 2.1.

We know that $U_{(1)}, \ldots, U_{(n)}$ is uniformly distributed over the simplex

$$B_n = \{(x_1, \ldots, x_n) : 0 \le x_1 \le \cdots \le x_n \le 1\} .$$

The transformation

$$s_1 = u_1$$
$$s_2 = u_2 - u_1$$
$$\ldots$$
$$s_n = u_n - u_{n-1}$$

has as inverse

$$u_1 = s_1$$
$$u_2 = s_1 + s_2$$
$$\ldots$$
$$u_n = s_1 + s_2 + \cdots + s_n$$

and the Jacobian of the transformation, i.e. the determinant of the matrix formed by $\frac{\partial s_i}{\partial u_j}$ is 1. This shows that the density of $S_1, \ldots, S_n$ is uniformly distributed on the set A_n . ∎

Proofs of this sort can often be obtained without the cumbersome transformations. For example, when X has the uniform density on a set $A \subseteq R^d$, and B is a linear nonsingular transformation: $R^d \to R^d$, then $Y=BX$ is uniformly distributed on BA as can be seen from the following argument: for all Borel sets $C \subseteq R^d$,

$$P(Y \in C) = P(BX \in C) = P(X \in B^{-1}C)$$

$$= \frac{\int_{(B^{-1}C)\cap A} dx}{\int_A dx} = \frac{\int_{C\cap(BA)} dx}{\int_{BA} dx} .$$

Theorem 2.2.

$S_1, \ldots, S_{n+1}$ is distributed as

$$\frac{E_1}{\sum_{i=1}^{n+1} E_i}, \ldots, \frac{E_{n+1}}{\sum_{i=1}^{n+1} E_i}$$

where $E_1, \ldots, E_{n+1}$ is a sequence of iid exponential random variables. Furthermore, if G_{n+1} is independent of $(S_1, \ldots, S_{n+1})$ and is gamma $(n+1)$ distributed, then

$$S_1 G_{n+1}, \ldots, S_{n+1} G_{n+1} .$$

is distributed as $E_1, E_2, \ldots, E_{n+1}$.

The proof of Theorem 2.2 is based upon Lemma 2.1:

Lemma 2.1.

For any sequence of nonnegative numbers $x_1, \ldots, x_{n+1}$, we have

$$P(S_1 > x_1, \ldots, S_{n+1} > x_{n+1}) = \left(1 - \sum_{i=1}^{n+1} x_i\right)_+^n .$$

Proof of Lemma 2.1.

Assume without loss of generality that $\sum_{i=1}^{n+1} x_i \le 1$ (for otherwise the lemma is obviously true). We use Theorem 2.1. In the notation of Theorem 2.1, we start from the fact that $S_1, \ldots, S_n$ is uniformly distributed in A_n. Thus, our probability is equal to

$$P(S_1 > x_1, \ldots, S_n > x_n, 1 - \sum_{i=1}^{n} S_i > x_{n+1}) .$$

This is the probability of a set A_n^* which is a simplex just as A_n except that its top is not at (0,0, . . . , 0) but rather at $(x_1, \ldots, x_n)$, and that its sides are not of length 1 but rather of length $1-\sum_{i=1}^{n+1} x_i$. For uniform distributions, probabilities can be calculated as ratios of areas. In this case, we have

$$\frac{\int_{A_n^*} dx}{\int_{A_n} dx} = (1-\sum_{i=1}^{n+1} x_i)^n \quad . \blacksquare$$

Proof of Theorem 2.2.

Part one. Let $G = G_{n+1}$ be the random variable $\sum_{i=1}^{n+1} E_i$. Note that we need only show that

$$\frac{E_1}{G}, \ldots, \frac{E_n}{G}$$

is uniformly distributed in A_n. The last component $\frac{E_{n+1}}{G}$ is taken care of by noting that it equals 1 minus the sum of the first n components. Let us use the symbols e_i, y, x_i for the running variables corresponding to $E_i, G, \frac{E_i}{G}$. We first compute the joint density of $E_1, \ldots, E_n, G$:

$$f(e_1, \ldots, e_n, y) = \prod_{i=1}^{n} e^{-e_i} \, e^{-(y-e_1-\cdots-e_n)} = e^{-y} \, ,$$

valid when $e_i \geq 0$, all i, and $y \geq \sum_{i=1}^{n} e_i$. Here we used the fact that the joint density is the product of the density of the first n variables and the density of G given $E_1 = e_1, \ldots, E_n = e_n$. Next, by a simple transformation of variables, it is easily seen that the joint density of $\frac{E_1}{G}, \ldots, \frac{E_n}{G}, G$ is

$$y^n f(x_1 y, \ldots, x_n y, y) = y^n e^{-y} \quad (x_i y \geq 0 \, , \, \sum_{i=1}^{n} x_i y \leq y) \, .$$

This is easily obtained by the transformation $\{x_1 = \frac{e_1}{y}, \ldots, x_n = \frac{e_n}{y}, y = y\}$. Finally, the marginal density of $\frac{E_1}{G}, \ldots, \frac{E_n}{G}$ is obtained by integrating the last density with respect to dy, which gives us

$$\int_0^{\infty} y^n e^{-y} \, dy \; I_{A_n}(x_1, \ldots x_n) = n! \, I_{A_n}(x_1, \ldots, x_n) \, .$$

This concludes the proof of part one.

Part two. Assume that $x_1 \geq 0, \ldots, x_{n+1} \geq 0$. By Lemma 2.1, we have

$$P(GS_1 > x_1, \ldots, GS_{n+1} > x_{n+1})$$

$$= \int_0^\infty P(S_1 > \frac{x_1}{y}, \ldots, S_{n+1} > \frac{x_{n+1}}{y} \mid G = y) \frac{y^n e^{-y}}{n!} dy$$

$$= \int_{y : \sum \frac{x_i}{y} \leq 1} (1 - \sum_{i=1}^{n+1} \frac{x_i}{y})^n \frac{y^n e^{-y}}{n!} dy$$

$$= \int_c^\infty (y-c)^n \frac{e^{-y}}{n!} dy \qquad (\text{where } c = \sum_{i=1}^{n+1} x_i)$$

$$= e^{-c}$$

$$= \prod_{i=1}^{n+1} e^{-x_i} . \blacksquare$$

A myriad of results follow from Theorem 2.2. For example, if $U, U_1, \ldots, U_n$ are iid uniform [0,1] random variables, E is an exponential random variable, and G_n is a gamma (n) random variable, then the following random variables have identical distributions:

$$\min(U_1, \ldots, U_n)$$

$$1 - U^{\frac{1}{n}}$$

$$1 - e^{-\frac{E}{n}}$$

$$\frac{E}{E + G_n} \quad (E, G_n \text{ are independent })$$

$$(\frac{E}{n}) - \frac{1}{2!}(\frac{E}{n})^2 + \frac{1}{3!}(\frac{E}{n})^3 - \cdots .$$

It is also easy to show that $\frac{\max(U_1, \ldots, U_n)}{\min(U_1, \ldots, U_n)}$ is distributed as $1 + \frac{G_{n-1}}{E}$, that $\max(U_1, \ldots, U_n) - \min(U_1, \ldots, U_n)$ is distributed as $1 - S_1 - S_{n+1}$ (i.e. as $\frac{G_{n-1}}{G_{n-1} + G_2}$), and that $U_{(k)}$ is distributed as $\frac{G_k}{G_k + G_{n+1-k}}$ where G_k and G_{n+1-k} are independent. Since we already know from section I.4 that $U_{(k)}$ is beta $(k, n+1-k)$ distributed, we have thus obtained a well-known relationship between the gamma and beta distributions.

2.2. Exponential spacings.

In this section, $E_{(1)} \le E_{(2)} \le \cdots \le E_{(n)}$ are the order statistics corresponding to a sequence of iid exponential random variables $E_1, E_2, \ldots, E_n$.

Theorem 2.3. (Sukhatme, 1937)

If we define $E_{(0)}=0$, then the normalized exponential spacings

$$(n-i+1)(E_{(i)}-E_{(i-1)}) \ , \ 1 \le i \le n \ ,$$

are iid exponential random variables. Also,

$$\frac{E_1}{n}, \frac{E_1}{n}+\frac{E_2}{n-1}, \ldots, \frac{E_1}{n}+\cdots+\frac{E_n}{1}$$

are distributed as $E_{(1)}, \ldots, E_{(n)}$.

Proof of Theorem 2.3.

The second statement follows from the first statement: it suffices to call the random variables of the first statement $E_1, E_2, \ldots, E_n$ and to note that

$$E_{(1)} = \frac{E_1}{n} \ ,$$
$$E_{(2)} = E_{(1)}+\frac{E_2}{n-1} \ ,$$
$$\cdots$$
$$E_{(n)} = E_{(n-1)}+\frac{E_n}{1} \ .$$

To prove the first statement, we note that the joint density of $E_{(1)}, \ldots, E_{(n)}$ is

$$n!\, e^{-\sum_{i=1}^{n} x_i} \qquad (0 \le x_1 \le x_2 \le \cdots \le x_n < \infty)$$
$$= n!\, e^{-\sum_{i=1}^{n}(n-i+1)(x_i-x_{i-1})} \qquad (0 \le x_1 \le x_2 \le \cdots \le x_n < \infty) \ .$$

Define now $Y_i=(n-i+1)(E_{(i)}-E_{(i-1)})$, $y_i=(n-i+1)(x_i-x_{i-1})$. Thus, we have

$$x_1 = \frac{y_1}{n} \ ,$$
$$x_2 = \frac{y_1}{n}+\frac{y_2}{n-1} \ ,$$
$$\cdots$$
$$x_n = \frac{y_1}{n}+\cdots+\frac{y_n}{1} \ .$$

The determinant of the matrix formed by $\frac{\partial x_i}{\partial y_j}$ is $\frac{1}{n!}$. Thus, $Y_1, \ldots, Y_n$ has density

$$e^{-\sum_{i=1}^{n} y_i} \quad (y_i \geq 0 \text{ ,all } i) ,$$

which was to be shown. ■

Theorem 2.3 has an important corollary: in a sample of two iid exponential random variates, $E_{(2)}-E_{(1)}$ is again exponentially distributed. This is basically due to the memoryless property of the exponential distribution: given that $E \geq x$, $E-x$ is again exponentially distributed. In fact, if we show the memoryless property (this is easy), and if we show that the minimum of n iid exponential random variables is distributed as $\frac{E}{n}$ (this is easy too), then we can prove Theorem 2.3 by induction.

Theorem 2.4. (Malmquist, 1950)

Let $0 \leq U_{(1)} \leq \cdots \leq U_{(n)} \leq 1$ be the order statistics of $U_1, U_2, \ldots, U_n$, a sequence of iid uniform [0,1] random variables. Then , if $U_{(n+1)}=1$,

A. $\{(\frac{U_{(i)}}{U_{(i+1)}})^i \ , 1 \leq i \leq n\}$ is distributed as $U_1, \ldots, U_n$.

B. $U_n^{\frac{1}{n}}, U_n^{\frac{1}{n}} U_{n-1}^{\frac{1}{n-1}}, \ldots, U_n^{\frac{1}{n}} \cdots U_1^{\frac{1}{1}}$ is distributed as $U_{(n)}, \ldots, U_{(1)}$.

Proof of Theorem 2.4.

In Theorem 2.3, replace U_i by e^{-E_i} and $U_{(i)}$ by $e^{-E_{(n-i+1)}}$. Then, in the notation of Theorems 2.3 and 2.4 we see that the following sequences are identically distributed:

$$(\frac{U_{(i)}}{U_{(i+1)}})^i \ , 1 \leq i \leq n \ ,$$

$$(e^{-E_{(n-i+1)}+E_{(n-i)}})^i \ , 1 \leq i \leq n \ ,$$

$$e^{-E_i} \ , 1 \leq i \leq n \ ,$$

$$U_i \ , 1 \leq i \leq n \ .$$

This proves part A. Part B follows without work from part A. ■

2.3. Exercises.

1. Give an alternative proof of Theorem 2.3 based upon the memoryless property of the exponential distribution (see suggestion following the proof of that theorem).
2. Prove that in a sample of n iid uniform [0,1] random variates, the maximum minus the minimum (i.e., the range) is distributed as
$$U^{\frac{1}{n}} V^{\frac{1}{n-1}}$$
where U,V are iid uniform [0,1] random variates.
3. Show that the minimum spacing in a uniform sample of size n is distributed as $\frac{1}{n+1}(1-U^{\frac{1}{n}})$ where U itself is uniformly distributed on [0,1].
4. Prove or disprove: $\frac{U}{U+V}$ is uniformly distributed on [0,1] when U,V are iid uniform [0,1] random variables.
5. Prove Whitworth's formula: if S_i , $1 \le i \le n+1$ are uniform spacings, then
$$P(\max_i S_i \ge x) = \binom{n}{1}(1-x)_+ - \binom{n}{2}(1-2x)_+^2 + \cdots .$$
(Whitworth, 1897)
6. Let E_1,E_2,E_3 be iid exponential random variables. Show that the following random variables are independent: $\frac{E_1}{E_1+E_2}$, $\frac{(E_1+E_2)}{E_1+E_2+E_3}$, $E_1+E_2+E_3$. Furthermore, show that their densities are the uniform [0,1] density, the triangular density on [0,1] and the gamma (3) density, respectively.

3. GENERATION OF ORDERED SAMPLES.

The first application that one thinks of when presented with Theorem 2.2 is a method for generating the order statistics $U_{(1)} \le \cdots \le U_{(n)}$ directly. By this we mean that it is not necessary to generate $U_1, \ldots, U_n$ and then apply some sorting method.

In this section we will describe several problems which require such ordered samples. We will not be concerned here with the problem of the generation of one order statistic such as the maximum or the median.

3.1. Generating uniform [0,1] order statistics.

The previous sections all suggest methods for generating uniform [0,1] order statistics:

A. Sorting

Generate iid uniform [0,1] random variates $U_1, \ldots, U_n$.
Obtain $U_{(1)}, \ldots, U_{(n)}$ by sorting the U_i's.

B. Via uniform spacings

Generate iid exponential random variates $E_1, \ldots, E_{n+1}$, and compute their sum G.
$U_{(0)} \leftarrow 0$
FOR $j := 1$ TO n DO
$\quad U_{(j)} \leftarrow U_{(j-1)} + \frac{E_j}{G}$

C. Via exponential spacings

$U_{(n+1)} \leftarrow 1$
FOR $j := n$ DOWNTO 1 DO
$\quad$ Generate a uniform [0,1] random variate U.
$\quad U_{(j)} \leftarrow U^{\frac{1}{j}} U_{(j+1)}$

Algorithm A is the naive approach. Sorting methods usually found in computer libraries are comparison-based. This means that information is moved around in tables based upon pairwise comparisons of elements only. It is known (see e.g. Knuth (1973) or Baase (1978)) that the worst-case and expected times taken by these algorithms are $\Omega(n \log n)$. Heapsort and mergesort have worst-case times that are $O(n \log n)$. Quicksort has expected time $O(n \log n)$, but worst-case time both $O(n^2)$ and $\Omega(n^2)$. For details, any standard textbook on data structures can be consulted (see e.g. Aho, Hopcroft and Ullman , 1983). What is different in the present case is that the U_i's are uniformly distributed on [0,1]. Thus, we can hope to take advantage of truncation. As we will see in the next section, we can bucket sort the U_i's in expected time $O(n)$.

Algorithms B and C are $O(n)$ algorithms in the worst-case. But only method C is a one-pass method. But because method C requires the computation of a power in each iteration, it is usually slower than either A or B. Storagewise, method A is least efficient since additional storage proportional to n is needed. Nevertheless, for large n, method A with bucket sorting is recommended. This is due to the accumulation of round-off errors in algorithms B and C.

Algorithms B and C were developed in a series of papers by Lurie and Hartley (1972), Schucany (1972) and Lurie and Mason (1973). Experimental comparisons can be found in Rabinowitz and Berenson (1974), Gerontides and Smith (1982), and Bentley and Saxe (1980). Ramberg and Tadikamalla (1978) consider the case of the generation of $U_{(k)}, U_{(k+1)}, \ldots, U_{(m)}$ where $1 \leq k \leq m \leq n$. This requires generating one of the extremes $U_{(k)}$ or $U_{(m)}$, after which a sequential method similar to algorithms B or C can be used, so that the total time is proportional to $m-k+1$.

3.2. Bucket sorting. Bucket searching.

We start with the description of a data structure and an algorithm for sorting n [0,1] valued elements $X_1, \ldots, X_n$.

Bucket sorting

[SET-UP]

We need two auxiliary tables of size n called Top and Next. Top $[i]$ gives the index of the top element in bucket i (i.e. $[\frac{i-1}{n}, \frac{i}{n})$). A value of 0 indicates an empty bucket. Next $[i]$ gives the index of the next element in the same bucket as X_i. If there is no next element, its value is 0.

FOR $i:=1$ TO n DO Next $[i] \leftarrow 0$

FOR $i:=0$ TO $n-1$ DO Top $[i] \leftarrow 0$

FOR $i:=1$ TO n DO

  Bucket $\leftarrow \lfloor nX_i \rfloor$

  Next $[i] \leftarrow$ Top [Bucket]

  Top [Bucket] $\leftarrow i$

[SORTING]

Sort all elements within the buckets by ordinary bubble sort or selection sort, and concatenate the nonempty buckets.

The set-up step takes time proportional to n in all cases. The sort step is where we notice a difference between distributions. If each bucket contains one element, then this step too takes time proportional to n. If all elements on the other hand fall in the same bucket, then the time taken grows as n^2 since selection sort for that one bucket takes time proportional to n^2. Thus, for our analysis, we will have to make some assumptions about the X_i's. We will assume that the X_i's are iid with density f on [0,1]. In Theorem 3.1 we show that the expected time is $O(n)$ for nearly all densities f.

Theorem 3.1. (Devroye and Klincsek, 1981)

The bucket sort given above takes expected time $O(n)$ if and only if

$$\int f^2(x)\,dx < \infty .$$

Proof of Theorem 3.1.

Assume that the buckets receive $N_0, \ldots, N_{n-1}$ points. It is clear that each N_i is binomially distributed with parameters n and p_i where

$$p_i = \int_{\frac{i}{n}}^{\frac{i+1}{n}} f(x)\,dx .$$

By the properties of selection sort, we know that there exist finite positive constants c_1, c_2, such that the time T_n taken by the algorithm satisfies:

$$c_1 \le \frac{T_n}{n + \sum_{i=0}^{n-1} N_i^2} \le c_2 .$$

By Jensen's inequality for convex functions, we have

$$\sum_{i=0}^{n-1} E(N_i^2) = \sum_{i=0}^{n-1} (np_i(1-p_i)+n^2p_i^2)$$

$$\le \sum_{i=0}^{n-1} np_i + \sum_{i=0}^{n-1} \left(n \int_{\frac{i}{n}}^{\frac{i+1}{n}} f(x)\,dx \right)^2$$

$$\le n + \sum_{i=0}^{n-1} n \int_{\frac{i}{n}}^{\frac{i+1}{n}} f^2(x)\,dx$$

$$= n \left(1+\int_0^1 f^2(x)dx \right) .$$

This proves one implication. The other implication requires some finer tools, especially if we want to avoid imposing smoothness conditions on f. The key measure theoretical result needed is the Lebesgue density theorem, which (phrased in a form suitable to us) states among other things that for any density f on R, we have

$$\lim_{n \to \infty} n \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} | f(y)-f(x) | \, dy = 0 \quad \text{(for almost all } x \text{)} .$$

Consult for example Wheeden and Zygmund (1977).

If we define the density

$$f_n(x) = p_i \qquad (0 \le \frac{i}{n} \le x < \frac{i+1}{n} \le 1) ,$$

then it is clear that

$$| f_n(x)-f(x) | \le \int_{\frac{i}{n}}^{\frac{i+1}{n}} | f(y)-f(x) | \, dy \qquad (\frac{i}{n} \le x < \frac{i+1}{n})$$

$$\le n \int_{x-\frac{1}{n}}^{x+\frac{1}{n}} | f(y)-f(x) | \, dy ,$$

and this tends to 0 for for almost all x. Thus, by Fatou's lemma,

$$\lim\inf \int_0^1 f_n^2(x) \, dx \ge \int_0^1 \lim\inf f_n^2(x) \, dx = \int_0^1 f^2(x) \, dx .$$

But

$$\frac{1}{n} \sum_{i=0}^{n-1} E(N_i^2) \ge \sum_{i=0}^{n-1} n p_i^2 = \sum_{i=0}^{n-1} \int_{\frac{i}{n}}^{\frac{i+1}{n}} f_n^2(x) \, dx = \int_0^1 f_n^2(x) \, dx .$$

Thus, $\int f^2 = \infty$ implies $\lim\inf \frac{T_n}{n} = \infty$. ■

In selection sort, the number of comparisons of two elements is $(n-1)+(n-2)+ \cdots +1 = \frac{n(n-1)}{2}$. Thus, the total number of comparisons needed in bucket sort is, in the notation of the proof of Theorem 3.1,

$$\sum_{i=0}^{n-1} \frac{N_i(N_i-1)}{2} .$$

The expected number of comparisons is thus

$$\sum_{i=0}^{n-1} \frac{1}{2}(n^2 p_i^{\,2}+np_i(1-p_i)-np_i)$$
$$= \frac{n(n-1)}{2}\sum_{i=0}^{n-1} p_i^{\,2}$$
$$\leq \frac{n-1}{2}\int_0^1 f^{\,2}(x)\,dx \ .$$

This upper bound is, not unexpectedly, minimized for the uniform density on $[0,1]$, in which case we obtain the upper bound $\frac{n-1}{2}$. In other words, the expected number of comparisons is less than the total number of elements ! This is of course due to the fact that most of the sorting is done in the set-up step.

If selection sort is replaced by an $O(n \log n)$ expected time comparison-based sorting algorithm (such as quicksort, mergesort or heapsort), then Theorem 3.1 remains valid provided that the condition $\int f^{\,2} < \infty$ is replaced by

$$\int_0^\infty f(x)\log_+ f(x)\,dx < \infty \ .$$

See Devroye and Klincsek (1981). The problem with extra space can be alleviated to some extent by clever programming tricks. These tend to slow down the algorithm and won't be discussed here.

Let us now turn to searching. The problem can be formulated as follows. $[0,1]$-valued data $X_1, \ldots, X_n$ are given. We assume that this is an iid sequence with common density f. Let T_n be the time taken to determine whether X_Z is in the structure where Z is a random integer taken from $\{1, \ldots, n\}$ independent of the X_i's. This is called the successful search time. The time T_n* taken to determine whether X_{n+1} (a random variable distributed as X_1 but independent of the data sequence) is in the structure is called the unsuccessful search time. If we store the elements in an array, then linear (or sequential search) yields expected search times that are proportional to n. If we use binary search and the array is sorted, then it is proportional to $\log(n)$. Assume now that we use the bucket data structure, and that the elements within buckets are not sorted. Then, with linear search within the buckets, the expected number of comparisons of elements for successful search, given $N_0, \ldots, N_{n-1}$, is

$$\sum_{i=0}^{n-1} \frac{N_i}{n}\frac{N_i+1}{2} \ .$$

For unsuccessful search, we have

$$\sum_{i=0}^{n-1} \frac{N_i}{n} N_i \ .$$

Arguing now as in Theorem 3.1, we have:

Theorem 3.2.

When searching a bucket structure we have $E(T_n)=O(1)$ if and only if $\int f^2<\infty$. Also, $E(T_n^*)=O(1)$ of and only if $\int f^2<\infty$.

3.3. Generating exponential order statistics.

To generate a sorted sample of exponential random variables, there are two algorithms paralleling algorithms A and C for the uniform distribution.

A. Bucket sorting

Generate iid exponential random variates $E_1, \ldots, E_n$.
Obtain $E_{(1)} \leq \cdots \leq E_{(n)}$ by bucket sorting.

C. Via exponential spacings

$E_{(0)} \leftarrow 0$
FOR $i:=1$ TO n DO
  Generate an exponential random variate E.
  $E_{(i)} \leftarrow E_{(i-1)} + \frac{E}{n-i+1}$

Method C uses the memoryless property of the exponential distribution. It takes time $O(n)$. Careless bucket sorting applied to algorithm A could lead to a superlinear time algorithm. For example, this would be the case if we were to divide the interval $[0, \max E_i]$ up into n equi-sized intervals. This can of course be avoided if we first generate $U_{(1)} \leq \cdots \leq U_{(n)}$ for a uniform sample in expected time $O(n)$, and then return $-\log U_{(n)}, \ldots, -\log U_{(1)}$. Another possibility is to construct the bucket structure for $E_i \bmod 1$, $1 \leq i \leq n$, i.e. for the fractional parts only, and to sort these elements. Since the fractional parts have a bounded density,

$$\frac{e^{-x} I_{[0,1]}(x)}{1-\frac{1}{e}},$$

we know from Theorem 3.1 that a sorted array can be obtained in expected time $O(n)$. But this sorted array has many sorted sub-arrays. In one extra pass, we can untangle it provided that we have kept track of the unused integer parts of the data, $\lfloor E_i \rfloor$. Concatenation of the many sub-arrays requires another pass, but we still have linear behavior.

3.4. Generating order statistics with distribution function F.

The order statistics $X_{(1)} \leq \cdots \leq X_{(n)}$ that correspond to $X_1, \ldots, X_n$, a sequence of iid random variables with absolutely continuous distribution function F on R^1 can be generated as

$$F^{-1}(U_{(1)}), \ldots, F^{-1}(U_{(n)})$$

or as

$$F^{-1}(1-e^{-E_{(1)}}), \ldots, F^{-1}(1-e^{-E_{(n)}})$$

starting from uniform or exponential order statistics. The exponential order statistics method based on C (see previous section) was proposed by Newby (1979). In general, the choice of one method over the other one largely depends upon the form of F. For example, for the Weibull distribution function

$$F(x) = 1-e^{-(\frac{x}{b})^a} \qquad (x \geq 0)$$

we have $F^{-1}(u) = b(-\log(1-u))^{\frac{1}{a}}$ and $F^{-1}(1-e^{-u}) = bu^{\frac{1}{a}}$, so that the exponential order statistics method seems better suited.

In many cases, it is much faster to just sort $X_1, \ldots, X_n$ so that the costly inversions can be avoided. If bucket sorting is used, one should make sure that the expected time is $O(n)$. This can be done for example by transforming the data in a monotone manner for the purpose of sorting to $[0,1]$ and insuring that the density f of the transformed data has a small value for $\int f^2$. Transformations that one might consider should be simple, e.g. $\frac{x}{a+x}$ is useful for transforming nonnegative data. The parameter $a > 0$ is a design parameter which should be picked such that the density after transformation has the smallest possible value for $\int f^2$.

The so-called grouping method studied by Rabonowitz and Berenson (1974) and Gerontides and Smith (1982) is a hybrid of the inversion method and the bucket sorting method. The support of the distribution is partitioned into k intervals, each having equal probability. Then one keeps for each interval a linked list. Intervals are selected with equal probability, and within each interval, random points are generated directly. In a final pass, all linked lists are sorted and concatenated. The sorting and concatenating take linear expected time when

$k=n$, because the interval cardinalities are as for the bucket method in case of a uniform distribution. There are two major differences with the bucket sorting method: first of all, the determination of the intervals requires $k-1$ implicit inversions of the distribution function. This is only worthwhile when it can be done in a set-up step and very many ordered samples are needed for the same distribution and the same n (recall that k is best taken proportional to n). Secondly, we have to be able to generate random variates with a distribution restricted to these intervals. Candidates for this include the rejection method. For monotone densities or unimodal densities and large n, the rejection constant will be close to one for most intervals if rejection from uniform densities is used.

But perhaps most promising of all is the **rejection method** itself for generating an ordered sample. Assume that our density f is dominated by cg where g is another density, and $c>1$ is the rejection constant. Then, exploiting properties of points uniformly distributed under f, we can proceed as follows:

Rejection method for generating an ordered sample

[NOTE: n is the size of the ordered sample; $m>n$ is an integer picked by the user. Its recommended value is $\left\lfloor nc+\sqrt{nc(c-1)\log\left(\frac{cn}{2\pi(c-1)}\right)}\right\rfloor$.]

REPEAT

Generate an ordered sample $X_1, \ldots, X_m$ with density g.

Generate m iid uniform [0,1] random variates $U_1, \ldots, U_m$.

Delete all X_i's for which $U_i > cg(X_i)/f(X_i)$.

UNTIL the edited (but ordered) sample has $N \geq n$ elements

Delete another $N-n$ randomly selected X_i's from this sample, and return the edited sample.

The main loop of the algorithm, when successful, gives an ordered sample of random size $N \geq n$. This sample is further edited by one of the well-known methods of selecting a random (uniform) sample of size $N-n$ from a set of size n (see chapter XII). The expected time taken by the latter procedure is $E(N-n \mid N \geq n)$ times a constant not depending upon N or n. The expected time taken by the global algorithm is $m/P(N \geq n)+E(N-n \mid N \geq n)$ if constants are omitted, and a uniform ordered sample with density g can be generated in linear expected time.

Theorem 3.3.

Let m, n, N, f, c, g keep their meaning of the rejection algorithm defined above. Then, if $m \geq cn$ and $m = O(n)$, the algorithm takes expected time $O(n)$. If in addition $m - cn = o(n)$ and $(m-cn)/\sqrt{n} \to \infty$, then

$$T_n = \frac{m}{P(N \geq n)} + E(N-n \mid N \geq n) \sim cn$$

as $n \to \infty$.

Proof of Theorem 3.3.

In order to analyze the success probability, we need some result about the closeness between the binomial and normal distributions. First of all, since N is binomial $(m, \frac{1}{c})$, we know from the central limit theorem that as $m \to \infty$,

$$P(N < n) \sim \Phi\left(\frac{n - \frac{m}{c}}{\sqrt{m\frac{1}{c}(1-\frac{1}{c})}}\right)$$

where Φ is the normal distribution function. If $m \geq cn$ at all times, then we see that $P(N < n)$ stays bounded away from 1, and oscillates asymptotically between 0 and 1/2. It can have a limit. If $(m-cn)/\sqrt{n} \to \infty$, then we see that $P(N < n) \to 0$.

We note that $E(N-n \mid N \geq n) = E((N-n)_+)/P(N \geq n)$. Since $N-n \leq m-n$, we see that $T_n \leq (2m-n)/P(N \geq n)$. The bound is $O(n)$ when $m = O(n)$ and $P(N \geq n)$ is bounded away from zero. Also, $T_n \sim cn$ when $P(N < n) \to 0$ and $m \sim cn$. ∎

Remark 3.1. Optimal choice of m.

The best possible value for T_n is cn because we cannot hope to accept n points with large enough probability of success unless the original sample is at least of size cn. It is fortunate that we need not take m much larger than cn. Detailed computations are needed to obtain the following recommendation for m: take m close to

$$nc + \sqrt{nc(c-1)\log\left(\frac{cn}{2\pi(c-1)}\right)} .$$

With this choice, T_n is $cn+O(\sqrt{n}\log(n))$. See exercise 3.7 for guidance with the derivation. ■

3.5. Generating exponential random variates in batches.

By Theorem 2.2, iid exponential random variates $E_1, \ldots, E_n$ can be generated as follows:

Exponential random variate generator

Generate an ordered sample $U_{(1)} \leq \cdots \leq U_{(n-1)}$ of uniform [0,1] random variates.
Generate a gamma (n) random variate G_n.
RETURN $(G_n U_{(1)}, G_n (U_{(2)}-U_{(1)}), \ldots, G_n(1-U_{(n-1)}))$.

Thus, one gamma variate (which we are able to generate in expected time $O(1)$) and a sorted uniform sample of size $n-1$ are all that is needed to obtain an iid sequence of n exponential random variates. Thus, the contribution of the gamma generator to the total time is asymptotically negligible. Also, the sorting can be done extremely quickly by bucket sort if we have a large number of buckets (exercise 3.1), so that for good implementations of bucket sorting, a super-efficient exponential random variate generator can be obtained. Note however that by taking differences of numbers that are close to each other, we loose some accuracy. For very large n, this method is not recommended.

One special case is worth mentioning here: UG_2 and $(1-U)G_2$ are iid exponential random variates.

3.6. Exercises.

1. In bucket sorting, assume that instead of n buckets, we take kn buckets where $k \geq 1$ is an integer. Analyze how the expected time is affected by the choice of k. Note that there is a time component for the set-up which increases as kn. The time component due to selection sort within the buckets is a decreasing function of k and f. Determine the asymptotically optimal value of k as a function of $\int f^2$ and of the relative weights of the two time components.
2. Prove the claim that if an $O(n \log n)$ expected time comparison-based sorting algorithm is used within buckets, then $\int_0^1 f \log_+ f < \infty$ implies that the

expected time is $O(n)$.

3. Show that $\int f \log_+ f < \infty$ implies $\int f^2 < \infty$ for any density f. Give an example of a density f on [0,1] for which $\int f \log_+ f < \infty$, yet $\int f^2 = \infty$. Give also an example for which $\int f \log_+ f = \infty$.

4. The randomness in the time taken by bucket sorting and bucket searching can be appropriately measured by $\sum_{i=0}^{n-1} N_i^2$, a quantity that we shall call T_n. It is often good to know that T_n does not become very large with high probability. For example, we may wish to obtain good upper bounds for $P(T_n > E(T_n)+\alpha)$, where $\alpha > 0$ is a constant. For example, obtain bounds that decrease exponentially fast in n for all bounded densities on [0,1] and all $\alpha > 0$. Hint: use an exponential version of Chebyshev's inequality and a Poissonization trick for the sample size.

5. Give an $O(n)$ expected time generator for the maximal uniform spacing in a sample of size n. Then give an $O(1)$ expected time generator for the same problem.

6. If a density f can be decomposed as $pf_1+(1-p)f_2$ where f_1, f_2 are densities and $p \in [0,1]$ is a constant, then an ordered sample $X_{(1)} \leq \cdots \leq X_{(n)}$ of f can be generated as follows:

   Generate a binomial (n, p) random variate N.

   Generate the order statistics $Y_{(1)} \leq \cdots \leq Y_{(N)}$ and $Z_{(1)} \leq \cdots \leq Z_{(n-N)}$ for densities f_1 and f_2 respectively.

   Merge the sorted tables into a sorted table $X_{(1)} \leq \cdots \leq X_{(n)}$.

   The acceleration is due to the fact that the method based upon inversion of F is sometimes simple for f_1 and f_2 but not for f; and that n coin flips needed for selection in the mixture are avoided. Of course, we need a binomial random variate. Here is the question: based upon this decomposition method, derive an efficient algorithm for generating an ordered sample from any monotone density on $[0,\infty)$.

7. This is about the optimal choice for m in Theorem 3.3 (the rejection method for generating an ordered sample). The purpose is to find an m such that for that choice of m, $T_n - cn \sim \inf_m (T_n - cn)$ as $n \to \infty$. Proceed as follows: first show that it suffices to consider only those m for which $T_n \sim cn$. This implies that $E((N-n)_+) = o(m-cn)$, $P(N<n) \to 0$, and $(m-cn)/\sqrt{n} \to \infty$. Then deduce that for the optimal m,

$$T_n = cn\,(1+(1+o(1))(\frac{m-cn}{cn}+P(N<n))) .$$

Clearly, $m \sim cn$, and $(m-cn)/cn$ is a term which decreases much slower than $1/\sqrt{n}$. By the Berry-Esseen theorem (Chow and Teicher (1978, p. 299) or Petrov (1975)), find a constant C depending upon c only such that

$$| P(N<n)-\Phi(\frac{n-\frac{m}{c}}{\sqrt{\frac{m}{c}(1-\frac{1}{c})}})| \leq \frac{C}{\sqrt{n}} .$$

Conclude that it suffices to find the m which minimizes $(m-cn)/(cn)+\Phi\left(\frac{n-\frac{m}{c}}{\sqrt{\frac{m}{c}(1-\frac{1}{c})}}\right)$. Next, using the fact that as $u \to \infty$,

$$1-\Phi(u) \sim \frac{1}{u\sqrt{2\pi}} e^{-\frac{u^2}{2}} ,$$

reduce the problem to that of minimizing

$$\rho\sqrt{\frac{c-1}{cn}}+\frac{1}{\rho\sqrt{2\pi}} e^{-\frac{\rho^2}{2}} ,$$

where $m-cn=\rho\sqrt{c(c-1)n}$ for some $\rho\to\infty$, $\rho=o(\sqrt{n})$. Approximate asymptotic minimization of this yields

$$\rho = \sqrt{\log\left(\frac{cn}{2\pi(c-1)}\right)} .$$

Finally, verify that for the corresponding value for m, the minimal value of T_n is asymptotically obtained (in the " $\sim$ " sense).

4. THE POLAR METHOD.

4.1. Radially symmetric distributions.

Here we will explain about the intimate connection between order statistics and random vectors with radially symmetric distributions in R^d. This connection will provide us with a wealth of algorithms for random variate generation. Most importantly, we will obtain the time-honored Box-Muller method for the normal distribution.

A random vector $X=(X_1, \ldots, X_d)$ in R^d is **radially symmetric** if AX is distributed as X for all orthonormal $d \times d$ matrices A. It is strictly radially symmetric if also $P(X=0)=0$. Noting that AX corresponds to a rotated version of X, radial symmetry is thus nothing else but invariance under rotations of the coordinate axes. We write C_d for the unit sphere in R^d. X is uniformly distributed on C_d when X is radially symmetric and $||X||=1$ with probability one. Here $||.||$ is the standard L_2 norm. Sometimes, a radially symmetric random vector has a density f, and then necessarily it is of the form

$$f(x_1, \ldots, x_d) = g(||x||) \quad (x=(x_1, \ldots, x_d) \in R^d)$$

for some function g. This function g on $[0,\infty)$ is such that

$$\int_0^\infty dV_d r^{d-1} g(r)\, dr = 1 ,$$

where

$$V_d = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2}+1)}$$

is the volume of the unit sphere C_d. We say that g defines or determines the radial density. Elliptical radial symmetry is not be treated in this early chapter, nor do we specifically address the problem of multivariate random variate generation. For a bibliography on radial symmetry, see Chmielewski (1981). For the fundamental properties of radial distributions not given below, see for example Kelker (1970).

Theorem 4.1. (Uniform distributions on the unit sphere.)

1. If X is strictly radially symmetric, then $\frac{X}{||X||}$ is uniformly distributed on C_d.
2. If X is uniformly distributed on C_d, then $(X_1^2, \ldots, X_d^2)$ is distributed as $(\frac{Y_1}{S}, \ldots, \frac{Y_d}{S})$, where $Y_1, \ldots, Y_d$ are independent gamma $(\frac{1}{2})$ random variables with sum S.
3. If X is uniformly distributed on C_d, then X_1^2 is beta $(\frac{1}{2}, \frac{d-1}{2})$. Equivalently, X_1^2 is distributed as $\frac{Y}{Y+Z}$ where Y,Z are independent gamma $(\frac{1}{2})$ and gamma $(\frac{d-1}{2})$ random variables. Furthermore, X_1 has density

$$\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{d-1}{2})}(1-x^2)^{\frac{d-3}{2}} \quad (|x| \leq 1) .$$

Proof of Theorem 4.1.

For all orthogonal $d \times d$ matrices A, $\frac{AX}{||X||}$ is distributed as $\frac{AX}{||AX||}$, which in turn is distributed as $\frac{X}{||X||}$ because X is strictly radially symmetric. Since $||\frac{X}{||X||}|| = \frac{||X||}{||X||} = 1$, statement 1 follows.

To prove statement 2, we define the iid normal random variables $N_1, \ldots, N_d$, and note that $N = (N_1, \ldots, N_d)$ is radially symmetric with density determined by

$$g(r) = \frac{1}{(2\pi)^{\frac{d}{2}}} e^{-\frac{r^2}{2}} \quad (r \geq 0) .$$

Thus, by part 1, the vector with components $\frac{N_i}{||N||}$ is uniformly distributed on C_d. But since N_i^2 is gamma $(\frac{1}{2}, 2)$, we deduce that the random vector with components $\frac{N_i^2}{||N||^2}$ is distributed as a random vector with components $\frac{2Y_i}{2S}$. This proves statement 2.

The first part of statement 3 follows easily from statement 2 and known properties of the beta and gamma distributions. The beta $(\frac{1}{2},\frac{d-1}{2})$ density is

$$c\frac{(1-x)^{\frac{d-3}{2}}}{\sqrt{x}} \quad (0<x<1) ,$$

where $c=\frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{d-1}{2})}$. Putting $Y=\sqrt{X}$, we see that Y has density

$$c\,(1-y^2)^{\frac{d-3}{2}}\frac{1}{y}2y \quad (0<y<1) ,$$

when X is beta $(\frac{1}{2},\frac{d-1}{2})$ distributed. This proves statement 3. ∎

Theorem 4.2. (The normal distribution.)

If $N_1, \ldots, N_d$ are iid normal random variables, then $(N_1, \ldots, N_d)$ is radially symmetric with density defined by

$$g(r) = \frac{1}{(2\pi)^{\frac{d}{2}}}e^{-\frac{r^2}{2}} \quad (r \geq 0) .$$

Furthermore, if $(X_1, \ldots, X_d)$ is strictly radially symmetric and the X_i's are independent, then the X_i's are iid normal random variables with nonzero variance.

Proof of Theorem 4.2.

The first part was shown in Theorem 4.1. The second part is proved for example in Kelker (1970). ∎

Theorem 4.3. (Radial transformations.)

1. If X is strictly radially symmetric in R^d with defining function g, then $R = ||X||$ has density

$$dV_d r^{d-1} g(r) \quad (r \geq 0) .$$

2. If X is uniformly distributed on C_d, and R is independent of X and has the density given above, then RX is strictly radially symmetric in R^d with defining function g.
3. If X is radially symmetric in R^d with defining function g, and if R is a random variable on $[0,\infty)$ with density h , independent of X, then RX is radially symmetric with defining function

$$g^*(r) = \int_0^\infty \frac{h(u)}{u^d} g\left(\frac{r}{u}\right) du .$$

Proof of Theorem 4.3.

For statement 1, we need the fact that the surface of C_d has $d-1$-dimensional volume dV_d. By a simple polar transformation,

$$P(R \leq r) = \int_{||x|| \leq r} g(||x||)\, dx = \int_{y \leq r} dV_d\, y^{d-1} g(y)\, dy \quad (r \geq 0) .$$

This proves statement 1.

RX is radially symmetric because for all orthogonal $d \times d$ matrices A, $A(RX)$ is distributed as $R(AX)$ and thus as RX. But such distributions are uniquely determined by the distribution of $||RX|| = R\,||X|| = R$, and thus, statement follows from statement 1.

Consider finally part 3. Clearly, RX is radially symmetric. Given R, $R\,||X||$ has density

$$\frac{1}{R} dV_d \left(\frac{r}{R}\right)^{d-1} g\left(\frac{r}{R}\right) \quad (r \geq 0) .$$

Thus, the density of $||X||$ is the expected value of the latter expression with respect to R, which is seen to be g^*. ■

Let us briefly discuss these three theorems. Consider first the marginal distributions of random vectors that are uniformly distributed on C_d:

d	Density of X_1 (on [-1,1])	Name of density
2	$\frac{1}{\pi\sqrt{1-x^2}}$	Arc sine density
3	$\frac{1}{2}$	Uniform [-1,1] density
4	$\frac{2}{\pi}\sqrt{1-x^2}$	
5	$\frac{3}{4}(1-x^2)$	
6	$\frac{8}{3\pi}(1-x^2)^{\frac{3}{2}}$	

Since all radially symmetric random vectors are distributed as the product of a uniform random vector on C_d and an independent random variable R, it follows that the first component X_1 is distributed as R times a random variable with densities as given in the table above or in part 3 of Theorem 4.1. Thus, for $d \geq 2$, X_1 has a marginal density whenever X is strictly radially symmetric. By Khinchine's theorem, we note that for $d \geq 3$, the density of X_1 is unimodal.

Theorem 4.2 states that radially symmetric distributions are virtually useless if they are to be used as tools for generating independent random variates $X_1, \ldots, X_n$ unless the X_i's are normally distributed. In the next section, we will clarify the special role played by the normal distribution.

4.2. Generating random vectors uniformly distributed on C_d.

The following two algorithms can be used to generate random variates with a uniform distribution on C_d:

Via normal random variates

Generate iid normal random variates, $N_1, \ldots, N_d$, and compute $S \leftarrow \sqrt{N_1^2 + \cdots + N_d^2}$.

RETURN $(\frac{N_1}{S}, \ldots, \frac{N_d}{S})$.

Via rejection from the enclosing hypercube

REPEAT

Generate iid uniform [-1,1] random variates $X_1, \ldots, X_d$, and compute $S \leftarrow X_1^2 + \cdots + X_d^2$.

UNTIL $S \leq 1$

$S \leftarrow \sqrt{S}$

RETURN $(\frac{X_1}{S}, \ldots, \frac{X_d}{S})$

In addition, we could also make good use of a property of Theorem 4.1. Assume that d is even and that a d-vector X is uniformly distributed on C_d. Then,

$$(X_1^2+X_2^2, \ldots, X_{d-1}^2+X_d^2)$$

is distributed as

$$(\frac{E_1}{S}, \ldots, \frac{E_{\frac{d}{2}}}{S})$$

where the E_i's are iid exponential random variables and $S = E_1 + \cdots + E_{\frac{d}{2}}$. Furthermore, given $X_1^2+X_2^2=r^2$, $(\frac{X_1}{r}, \frac{X_2}{r})$ is uniformly distributed on C_2. This leads to the following algorithm:

Via uniform spacings

Generate iid uniform [0,1] random variates $U_1, \ldots, U_{\frac{d}{2}-1}$.

Sort the uniform variates (preferably by bucket sorting), and compute the spacings $S_1, \ldots, S_{\frac{d}{2}}$.

Generate independent pairs $(V_1, V_2), \ldots, (V_{d-1}, V_d)$, all uniformly distributed on C_2.

RETURN $(V_1\sqrt{S_1}, V_2\sqrt{S_1}, V_3\sqrt{S_2}, V_4\sqrt{S_2}, \ldots, V_{d-1}\sqrt{S_{\frac{d}{2}}}, V_d\sqrt{S_{\frac{d}{2}}})$.

The normal and spacings methods take expected time $O(d)$, while the rejection method takes time increasing faster than exponentially with d. By Stirling's formula, we observe that the expected number of iterations in the rejection method is

$$\frac{2^d}{V_d} = \frac{2^d\Gamma(\frac{d}{2}+1)}{\pi^{\frac{d}{2}}} \sim \left(\frac{2d}{\pi e}\right)^{\frac{d}{2}}\sqrt{\pi d} \ ,$$

which increases very rapidly to ∞. Some values for the expected number of iterations are given in the table below.

d	Expected number of iterations
1	1
2	$\frac{4}{\pi}\approx 1.27$
3	$\frac{6}{\pi}\approx 1.91$
4	$\frac{32}{\pi^2}\approx 3.24$
5	$\frac{60}{\pi^2}\approx 6.06$
6	$\frac{384}{\pi^3}\approx 12.3$
7	$\frac{840}{\pi^3}\approx 27.0$
8	$\frac{6144}{\pi^4}\approx 62.7$
10	$\frac{122880}{\pi^5}\approx 399$

The rejection method is not recommended except perhaps for $d \le 5$. The normal and spacings methods differ in the type of operations that are needed: the normal method requires d normal random variates plus one square root, whereas the spacings method requires one bucket sort , $\frac{d}{2}$ square roots and $\frac{d}{2}-1$ uniform random variates. The spacings method is based upon the assumption that a very fast method is available for generating random vectors with a uniform distribution on C_2. Since we work with spacings, it is also possible that some accuracy is lost for large values of d. For all these reasons, it seems unlikely that the spacings method will be competitive with the normal method. For theoretical and experimental comparisons, we refer the reader to Deak (1979) and Rubinstein (1982). For another derivation of the spacings method, see for example Sibuya (1962), Tashiro (1977), and Guralnik, Zemach and Warnock (1985).

4.3. Generating points uniformly in and on C_2.

We say that a random vector is uniformly distributed in C_d when it is radially symmetric with defining function $g(r)=\frac{1}{V_d}$ $(0\leq r \leq 1)$. For $d=2$, such random vectors can be conveniently generated by the rejection method:

Rejection method

REPEAT
  Generate two iid uniform $[-1,1]$ random variates U_1,U_2.
UNTIL $U_1^2+U_2^2\leq 1$
RETURN (U_1,U_2)

On the average, $\frac{4}{\pi}$ pairs of uniform random variates are needed before we exit. For each pair, two multiplications are required as well. Some speed-up is possible by squeezing:

Rejection method with squeezing

REPEAT
  Generate two iid uniform $[-1,1]$ random variates U_1,U_2, and compute $Z\leftarrow |U_1|+|U_2|$.
  Accept $\leftarrow [Z\leq 1]$
  IF NOT Accept THEN IF $Z\geq\sqrt{2}$
    THEN Accept $\leftarrow [U_1^2+U_2^2\leq 1]$
UNTIL Accept
RETURN (U_1,U_2)

In the squeeze step, we avoid the two multiplications precisely 50% of the time.

The second, slightly more difficult problem is that of the generation of a point uniformly distributed on C_2. For example, if (X_1,X_2) is strictly radially symmetric (this is the case when the components are iid normal random variables, or when the random vector is uniformly distributed in C_2), then it suffices to take $(\frac{X_1}{S},\frac{X_2}{S})$ where $S=\sqrt{X_1^2+X_2^2}$. At first sight, it seems that the costly square root is unavoidable. That this is not so follows from the following key theorem:

Theorem 4.4.

If (X_1,X_2) is uniformly distributed in C_2, and $S=\sqrt{X_1^2+X_2^2}$, then:

1. S and $(\frac{X_1}{S},\frac{X_2}{S})$ are independent.
2. S^2 is uniformly distributed on [0,1].
3. $\frac{X_2}{X_1}$ is Cauchy distributed.
4. $(\frac{X_1}{S},\frac{X_2}{S})$ is uniformly distributed on C_2.
5. When U is uniform [0,1], then $(\cos(2\pi U),\sin(2\pi U))$ is uniformly distributed on C_2.
6. $(\frac{X_1^2-X_2^2}{S^2},\frac{2X_1X_2}{S^2})$ is uniformly distributed on C_2.

Proof of Theorem 4.4.

Properties 1,3 and 4 are valid for all strictly radially symmetric random vectors (X_1,X_2). Properties 1 and 4 follow directly from Theorem 4.3. From Theorem 4.1, we recall that S has density $dV_d\, r^{d-1}=2r$ $(0 \le r \le 1)$. Thus, S^2 is uniformly distributed on [0,1]. This proves property 2. Property 5 is trivially true, and will be used to prove properties 3 and 6. From 5, we know that $\frac{X_2}{X_1}$ is distributed as $\tan(2\pi U)$, and thus as $\tan(\pi U)$, which in turn is Cauchy distributed (property 3). Finally, in view of

$$\cos(4\pi U) = \cos^2(2\pi U)-\sin^2(2\pi U) ,$$

$$\sin(4\pi U) = 2\sin(2\pi U)\cos(2\pi U) ,$$

we see that $(\frac{X_1^2-X_2^2}{S^2},\frac{2X_1X_2}{S^2})$ is uniformly distributed on C_2, because it is distributed as $(\cos(4\pi U),\sin(4\pi U))$. This concludes the proof of Theorem 4.4. ■

Thus, for the generation of a random vector uniformly distributed on C_2, the following algorithm is fast:

REPEAT

Generate iid uniform [-1,1] random variates X_1, X_2.

Set $Y_1 \leftarrow X_1^2, Y_2 \leftarrow X_2^2, S \leftarrow Y_1 + Y_2$.

UNTIL $S \leq 1$

RETURN $\left(\frac{Y_1 - Y_2}{S}, \frac{2X_1X_2}{S}\right)$

4.4. Generating normal random variates in batches.

We begin with the description of the polar method for generating d iid normal random variates:

Polar method for normal random variates

Generate X uniformly on C_d.

Generate a random variate R with density $dV_d r^{d-1} e^{-\frac{r^2}{2}}$ $(r \geq 0)$. (R is distributed as $\sqrt{2G}$ where G is gamma $(\frac{d}{2})$ distributed.)

RETURN RX

In particular, for $d=2$, two independent normal random variates can be obtained by either one of the following methods:

$\sqrt{2E}\,\left(\frac{X_1}{S}, \frac{X_2}{S}\right)$
$\sqrt{2E}\,(\cos(2\pi U), \sin(2\pi U))$
$\sqrt{2E}\,\left(\frac{X_1^2 - X_2^2}{X_1^2 + X_2^2}, \frac{2X_1X_2}{X_1^2 + X_2^2}\right)$
$\sqrt{-4\log(S)}\left(\frac{X_1}{S}, \frac{X_2}{S}\right)$

Here (X_1, X_2) is uniformly distributed in C_2, $S = \sqrt{X_1^2 + X_2^2}$, U is uniformly distributed on [0,1] and E is exponentially distributed. Also, E is independent of the other random variables. The validity of these methods follows from Theorems 4.2, 4.3 and 4.4. The second formula is the well-known Box-Muller method

(1958). Method 4, proposed by Marsaglia, is similar to method 1, but uses the observation that S^2 is a uniform [0,1] random variate independent of $(\frac{X_1}{S},\frac{X_2}{S})$ (see Theorem 4.4), and thus that $-2\log(S)$ is exponentially distributed. If the exponential random variate in E is obtained by inversion of a uniform random variate, then it cannot be competitive with method 4. Method 3, published by Bell (1968), is based upon property 6 of Theorem 4.4, and effectively avoids the computation of the square root in the definition of S. In all cases, it is recommended that (X_1,X_2) be obtained by rejection from the enclosing square (with an accelerating squeeze step perhaps). A closing remark about the square roots. Methods 1 and 4 can always be implemented with just one (not two) square roots, if we compute, respectively,

$$\sqrt{\frac{2E}{S^2}}$$

and

$$\sqrt{\frac{-2\log(S^2)}{S^2}} .$$

In one of the exercises, we will investigate the polar method with the next higher convenient choice for d, $d=4$. We could also make d very large , in the range $100 \cdots 300$, and use the spacings method of section 4.2 for generating X with a uniform distribution on C_d (the normal method is excluded since we want to generate normal random variates). A gamma $(\frac{d}{2})$ random variate can be generated by one of the fast methods described elsewhere in this book.

4.5. Generating radially symmetric random vectors.

Theorem 4.3 suggests the following method for generating radially symmetric random vectors in R^d with defining function g:

Generate a random vector X uniformly distributed on C_d.

Generate a random variate R with density $dV_d r^{d-1} g(r)$ $(r \geq 0)$.

RETURN RX

Since we already know how to generate random variates with a uniform distribution on C_d, we are just left with a univariate generation problem. But in the multiplication with R, most of the information in X is lost. For example, to insure that X is on C_d, the rejection method generates X uniformly in C_d and divides then by $||X||$. But when we multiply the result with R, this division by $||X||$ seems somehow wasteful. Johnson and Ramberg (1977) observed that it is sometimes better to start from a random vector with a uniform distribution in C_d:

The Johnson-Ramberg method for generating radially symmetric random vectors

Generate a random vector X uniformly in C_d (preferably by rejection from the enclosing hypercube).

Generate a random variate R with density $-V_d r^d g'(r)$ $(r \geq 0)$, where g is the defining function of the radially symmetric distribution.

RETURN RX

This method only works when $-V_d r^d g'(r)$ is indeed a density in r on $[0,\infty)$. A sufficient condition for this is that g is continuously differentiable on $(0,\infty)$, $g'(r)<0$ $(r>0)$, and $r^d g(r) \to 0$ as $r \downarrow 0$ and $r \uparrow \infty$.

Example 4.1. The multivariate Pearson II density.

Consider the multivariate Pearson II density with parameter $a \geq 1$, defined by

$$g(r) = c(1-r^2)^{a-1} \quad (0 \leq r \leq 1),$$

where

$$c = \frac{\Gamma(a+\frac{d}{2})}{\pi^{\frac{d}{2}}\Gamma(a)} .$$

The density of R in the standard algorithm is the density of $\sqrt{B}$ where B is a beta $(\frac{d}{2},a)$ random variable:

$$g(r) = cdV_d r^{d-1}(1-r^2)^{a-1} \quad (0 \leq r \leq 1) .$$

For $d=2$, R can thus be generated as $\sqrt{1-U^{\frac{1}{a}}}$ where U is a uniform [0,1] random variate. We note further that in this case, very little is gained by using the Johnson-Ramberg method since R must have density

$$g(r) = 2cV_d r^{d+1}(a-1)(1-r^2)^{a-2} \quad (0 \leq r \leq 1) .$$

This is the density of the square root of a beta $(\frac{d}{2}+1,a-1)$ random variable. ■

Example 4.2. The multivariate Pearson VII density.

The multivariate Pearson VII density with parameter $a>\frac{d}{2}$ is defined by the function

$$g(r)=\frac{c}{(1+r^2)^a},$$

where

$$c=\frac{\Gamma(a)}{\pi^{\frac{d}{2}}\Gamma(a-\frac{d}{2})}.$$

The densities of R for the standard and Johnson-Ramberg methods are respectively,

$$\frac{cdV_d r^{d-1}}{(1+r^2)^a}$$

and

$$\frac{2cV_d r^{d+1}a}{(1+r^2)^{a+1}}.$$

In both cases, we can generate random R as $\sqrt{\frac{B}{1-B}}$ where B is beta $(\frac{d}{2},a-\frac{d}{2})$ in the former case, and beta $(\frac{d}{2}+1,a-\frac{d}{2})$ in the latter case. Note here that for the special choice $a=\frac{d+1}{2}$, the multivariate Cauchy density is obtained. ■

Example 4.3.

The multivariate radially symmetric distribution determined by

$$g(r)=\frac{1}{V_d(1+r^d)^2}$$

leads to a density for R given by

$$\frac{dr^{d-1}}{(1+r^d)^2} .$$

This is the density of $(\frac{U}{1-U})^{\frac{1}{d}}$ where U is a uniform [0,1] random variable.

4.6. The deconvolution method.

Assume that we know how to generate Z, a random variable which is distributed as the sum $X+Y$ of two iid random variables X,Y with density f. We can then generate the pair X,Y by looking at the conditional density of X given the value of Z. The following algorithm can be used:

The deconvolution method

Generate a random variate Z with the density $h(z)=\int f(x)f(z-x)\,dx$.

Generate X with density $\frac{f(x)f(Z-x)}{h(Z)}$.

RETURN $(X,Z-X)$

First, we notice that h is indeed the density of the sum of two iid random variables with density f. Also, given Z, X has density $\frac{f(x)f(Z-x)}{h(Z)}$. Thus, the algorithm is valid.

To illustrate this, recall that if X,Y are iid gamma $(\frac{1}{2})$, then $X+Y$ is exponentially distributed. In this example, we have therefore,

$$f(x) = \frac{1}{\sqrt{\pi x}} e^{-x} \quad (x \geq 0) ,$$

$$h(z) = e^{-z} \quad (z \geq 0) .$$

Furthermore, the density $\frac{f(x)f(Z-x)}{h(Z)}$ can be written as

$$\frac{1}{\pi\sqrt{x(Z-x)}} \quad (x \in (0,Z)) ,$$

which is the arc sine density. Thus, applying the deconvolution method shows the following: if E is an exponential random variable, and W is a random variable with the standard arc sine density

$$\frac{1}{\pi\sqrt{x(1-x)}} \quad (x \in (0,1)) ,$$

then $(EW, E(1-W))$ is distributed as a pair of iid gamma $(\frac{1}{2})$ random variables. But this leads precisely to the polar method because the following pairs of random variables are identically distributed:

(N_1, N_2)(two iid normal random variables);

$(\sqrt{2EW}, \sqrt{2E(1-W)})$;

$(\sqrt{2E}\cos(2\pi U), \sqrt{2E}\sin(2\pi U))$.

Here U is a uniform [0,1] random variable. The equivalence of the first two pairs is based upon the fact that a normal random variable is distributed as the square root of 2 times a gamma $(\frac{1}{2})$ random variable. The equivalence of the first and the third pair was established in Theorem 4.4. As a side product, we observe that W is distributed as $\cos^2(2\pi U)$, i.e. as $\frac{X_1^2}{X_1^2+X_2^2}$ where (X_1, X_2) is uniformly distributed in C_2.

4.7. Exercises.

1. Write one-line random variate generators for the normal, Cauchy and arc sine distributions.
2. If N_1, N_2 are iid normal random variables, then $\frac{N_1}{N_2}$ is Cauchy distributed, $N_1^2+N_2^2$ is exponentially distributed, and $\sqrt{N_1^2+N_2^2}$ has the Rayleigh distribution (the Rayleigh density is $xe^{-\frac{x^2}{2}}$ $(x \geq 0)$).
3. Show the following. If X is uniformly distributed on C_d and R is independent of X and generated as $\max(U_1, \ldots, U_d)$ where the U_i's are iid uniform [0,1] random variates, then RX is uniformly distributed in C_d.
4. Show that if X is uniformly distributed on C_d, then $Y/||Y||$ is uniformly distributed on C_k where $k \leq d$ and $Y=(X_1, \ldots, X_k)$.
5. Prove by a geometrical argument that if (X_1, X_2, X_3) is uniformly distributed on C_3, then X_1, X_2 and X_3 are uniform [-1,1] random variables.
6. If X is radially symmetric with defining function g, then its first component, X_1, has density

$$\frac{2\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d-1}{2})} \int_r^{\infty} u(u^2-r^2)^{\frac{d-3}{2}} g(u)\,du \qquad (r \geq 0) .$$

7. Show that two independent gamma $(\frac{1}{2})$ random variates can be generated

as $(-S\log(U_2),-(1-S)\log(U_2))$, where $S=\sin^2(2\pi U_1)$ and U_1,U_2 are independent uniform [0,1] random variates.

8. Consider the pair of random variables defined by

$$(\sqrt{2E}\frac{2S}{1+S},\sqrt{2E}\frac{1-S}{1+S})$$

where E is an exponential random variable, and $S\leftarrow\tan^2(\pi U)$ for a uniform [0,1] random variate U. Prove that the pair is a pair of iid absolute normal random variables.

9. Show that when (X_1,X_2,X_3,X_4) is uniformly distributed on C_4, then (X_1,X_2) is uniformly distributed in C_2.

10. Show that both $\frac{N}{\sqrt{N^2+2E}}$ and $\sqrt{\frac{G}{G+E}}$ are uniformly distributed on [0,1] when N,E and G are independent normal, exponential and gamma $(\frac{1}{2})$ random variables, respectively.

11. **Generating uniform random vectors on C_4.** Show why the following algorithm is valid for generating random vectors uniformly on C_4:

> Generate two iid random vectors uniformly in C_2, $(X_1,X_2),(X_3,X_4)$ (this is best done by rejection).
>
> $S\leftarrow X_1^2+X_2^2, W\leftarrow X_3^2+X_4^2$
>
> RETURN $(X_1,X_2,X_3\sqrt{\frac{1-S}{W}},X_4\sqrt{\frac{1-S}{W}})$

(Marsaglia, 1972).

12. **Generating random vectors uniformly on C_3.** Prove all the starred statements in this exercise. To obtain a random vector with a uniform distribution on C_3 by rejection from $[-1,1]^3$ requires on the average $\frac{18}{\pi}=5.73...$ uniform $[-1,1]$ random variates, and one square root per random vector. The square root can be avoided by an observation due to Cook (1957): if (X_1,X_2,X_3,X_4) is uniformly distributed on C_4, then

$$\frac{1}{X_1^2+X_2^2+X_3^2+X_4^2}(2(X_2X_4+X_1X_3),2(X_3X_4-X_1X_2),X_1^2-X_2^2-X_3^2+X_4^2)$$

is uniformly distributed on C_3 (*). Unfortunately, if a random vector with a uniform distribution on C_4 is obtained by rejection from the enclosing hypercube, then the expected number of uniform random variates needed is $4(\frac{32}{\pi^2})\approx 13$. Thus, both methods are quite expensive. Using Theorem 4.4 and exercises 4 and 5, one can show (*) that

$$\left(\frac{X_1\sqrt{1-Z^2}}{\sqrt{S}},\frac{X_2\sqrt{1-Z^2}}{\sqrt{S}},Z\right)$$

is uniformly distributed on C_3 when (X_1,X_2) is uniformly distributed in C_2, $S=X_1^2+X_2^2$, and Z is independent of $(\frac{X_1}{\sqrt{S}},\frac{X_2}{\sqrt{S}})$ and uniformly distributed on $[-1,1]$. But $2S-1$ itself is a candidate for Z (*). Replacing Z by $2S-1$, we conclude that

$$(2X_1\sqrt{1-S}\ ,2X_2\sqrt{1-S}\ ,2S-1)$$

is uniformly distributed on C_3 (this method was suggested by Marsaglia (1972)). If the random vector (X_1,X_2) is obtained by rejection from $[-1,1]^2$, the expected number of uniform $[-1,1]$ random variates needed per three-dimensional random vector is $\frac{8}{\pi}\approx 2.55$ (*).

13. **The polar methods for normal random variates, d=4.** Random vectors uniformly distributed on C_4 can be obtained quite efficiently by Marsaglia's method described in exercise 11. To apply the polar method for normal random variates, we need an independent random variate R distributed as $\sqrt{2(E_1+E_2)}$ where E_1,E_2 are independent exponential random variates. Such an R can be generated in a number of ways:
    (i) As $\sqrt{2(E_1+E_2)}$.
    (ii) As $\sqrt{-2\log(U_1U_2)}$ where U_1,U_2 are independent uniform $[0,1]$ random variates.
    (iii) As $\sqrt{-2\log(WU_2)}$ where U_2 is as in (ii) and W is an independent random variate as in exercise 11.

    Why is method (iii) valid ? Compare the three methods experimentally. Compare also with the polar method for $d=2$.
14. Implement the polar method for normal random variates when d is large. Generate random vectors on C_d by the spacings method when you do so. Plot the average time per random variate versus d.
15. **The spacings method for uniform random vectors on C_d when d is odd.** Show the validity of the following method for generating a uniform random vector on C_d:

Generate $\frac{d-1}{2}-1$ iid uniform [0,1] random variates.

Obtain the spacings $S_1, \ldots, S_{\frac{d-1}{2}}$ by bucket sorting the uniform random variates.

Generate independent gamma $(\frac{d-1}{2})$ and gamma $(\frac{1}{2})$ random variates G,H.

$R \leftarrow \sqrt{\frac{G}{G+H}}$, $R* \leftarrow \sqrt{1-R^2} = \sqrt{\frac{H}{H+G}}$

Generate iid random vectors $(V_1,V_2), \ldots, (V_{d-2},V_{d-1})$ uniformly on C_2.

RETURN $(RV_1\sqrt{S_1}, RV_2\sqrt{S_1}, RV_3\sqrt{S_2}, \ldots, RV_{d-1}\sqrt{S_{\frac{d-1}{2}}}, R*)$.

16. Let X be a random vector uniformly distributed on C_{d-1}. Then the random vector Y generated by the following procedure is uniformly distributed on C_d:

Generate independent gamma $(\frac{d-1}{2})$ and gamma $(\frac{1}{2})$ random variates G,H.

$R \leftarrow \sqrt{\frac{G}{G+H}}$

RETURN $Y \leftarrow (RX, \pm\sqrt{1-R^2})$ where $\pm$ is a random sign.

Show this. Notice that this method allows one to generate Y inductively by starting with $d=1$ or $d=2$. For $d=1$, X is merely ± 1. For $d=2$, R is distributed as $\sin(\frac{\pi U}{2})$. For $d=3$, R is distributed as $\sqrt{1-U^2}$ where U is a uniform [0,1] random variable. To implement this procedure, a fast gamma generator is required (Hicks and Wheeling, 1959; see also Rubinstein, 1982).

17. In a simulation it is required at one point to obtain a random vector (X,Y) uniformly distributed over a star on R^2. A star S_a with parameter $a>0$ is defined by four curves, one in each quadrant and centered at the origin. For example, the curve in the positive quadrant is a piece of a closed line satisfying the equation

$$|1-x|^a + |1-y|^a = 1 .$$

The three other curves are defined by symmetry about all the axes and about the origin. For $a=\frac{1}{2}$, we obtain the circle, for $a=1$, we obtain a diamond, and for $a=2$, we obtain the complement of the union of four circles. Give an algorithm for generating a point uniformly distributed in S_a, where the expected time is uniformly bounded over a.

18. **The Johnson-Ramberg method for normal random variates.** Two methods for generating normal random variates in batches may be competitive with the ordinary polar method because they avoid square roots. Both are based upon the Johnson-Ramberg technique:

> Generate X uniformly in C_2 by rejection from $[-1,1]^2$.
>
> Generate R, which is distributed as $\sqrt{2G}$ where G is a gamma $(\frac{3}{2})$ random variable. (Note that R has density $\frac{r^2}{2}e^{-\frac{r^2}{2}}$.)
>
> RETURN RX

> Generate X uniformly in C_3 by rejection from $[-1,1]^3$.
>
> Generate R, where R is distributed as $\sqrt{2G}$ and G is a gamma (2) random variable. (Note that R has density $(\frac{r}{\sqrt{2}})^3\Gamma^{-1}(\frac{5}{2})e^{-\frac{r^2}{2}}$.)
>
> RETURN RX

These methods can only be competitive if fast direct methods for generating R are available. Develop such methods.

19. Extend the entire theory towards other norms, i.e. C_d is now defined as the collection of all points for which the p-th norm is less than or equal to one. Here $p>0$ is a parameter. Reprove all theorems. Note that the role of the normal density is now inherited by the density

$$f(x)=ce^{-|x|^p},$$

where $c>0$ is a normalization constant. Determine this constant. Show that a random variate with this density can be obtained as $X^{\frac{1}{p}}$ where X is gamma $(\frac{1}{p})$ distributed. Find a formula for the probability of acceptance when random variates with a uniform distribution in C_d are obtained by rejection from $[-1,1]^2$. (To check your result, the answer for $d=2$ is $\Gamma^2(\frac{1}{p})\Gamma^{-1}(\frac{2}{p})$ (Beyer, 1968, p. 630).) Discuss various methods for generating random vectors uniformly distributed on C_d, and deduce the marginal density of such random vectors .