# Chapter One
# INTRODUCTION

## 1. GENERAL OUTLINE.

Random number generation has intrigued scientists for a few decades, and a lot of effort has been spent on the creation of randomness on a deterministic (non-random) machine, that is, on the design of computer algorithms that are able to produce "random" sequences of integers. This is a difficult task. Such algorithms are called generators, and all generators have flaws because all of them construct the $n$-th number in the sequence in function of the $n-1$ numbers preceding it, initialized with a nonrandom seed. Numerous quantities have been invented over the years that measure just how "random" a sequence is, and most well-known generators have been subjected to rigorous statistical testing. However, for every generator, it is always possible to find a statistical test of a (possibly odd) property to make the generator flunk. The mathematical tools that are needed to design and analyze these generators are largely number theoretic and combinatorial. These tools differ drastically from those needed when we want to generate sequences of integers with certain non-uniform distributions, given that a perfect uniform random number generator is available. The reader should be aware that we provide him with only half the story (the second half). The assumption that a perfect uniform random number generator is available is now quite unrealistic, but, with time, it should become less so. Having made the assumption, we can build quite a powerful theory of non-uniform random variate generation.

The existence of a perfect uniform random number generator is not all that is assumed. Statisticians are usually more interested in continuous random variables than in discrete random variables. Since computers are finite memory machines, they cannot store real numbers, let alone generate random variables with a given density. This led us to the following assumptions:

Assumption 1. Our computer can store and manipulate real numbers.

Assumption 2. There exists a perfect uniform [0,1] random variate generator, i.e. a generator capable of producing a sequence $U_1, U_2, \ldots$ of independent random variables with a uniform distribution on [0,1].

The generator of assumption 2 is our fundamental building block. The sequence of $U_i$'s can be intelligently manipulated to give us random variables with specified distributions in $R^d$, $d$-dimensional Euclidean space. Occasionally, we mention the effect that the finite word-length of the computer has on the manipulated sequence. With the two assumptions given above, we demand that the random variables obtained by combining the $U_i$'s have the exact distribution that was asked. Algorithms or generators with this property is called exact. Exact algorithms approach reality if we use extended precision arithmetic (some languages allow users to work with integers of virtually unlimited length by linking words together in a linked list). Inexact algorithms, which are usually algorithms that are based upon a mathematical approximation of sorts, are forever excluded, because neither extended precision arithmetic nor improvements in the basic random number generator make them more exact.

A random variate generation algorithm is a program that halts with probability one and exits with a real number $X$. This $X$ is called a **random variate.** Because of our assumptions, we can treat random variates as if they were random variables! Note also that if we can produce one random variate $X$, then we are able to produce a sequence $X_1, X_2, \ldots$ of independent random variates distributed as $X$ (this follows from assumption 2). This facilitates our task a lot: rather than having to concentrate on infinite sequences, we just need to look at the properties of single random variates.

Simple, easy-to-understand algorithms will survive longer, all other things being roughly equal. Unfortunately, such algorithms are usually slower than their more sophisticated counterparts. The notion of time itself is of course relative. For theoretical purposes, it is necessary to equate time with the number of "fundamental" operations performed before the algorithm halts. This leads to our third assumption:

Assumption 3. The fundamental operations in our computer include addition, multiplication, division, compare, truncate, move, generate a uniform random variate, exp, log, square root, arc tan, sin and cos. (This implies that each of these operations takes one unit of time regardless of the size of the operand(s). Also, the outcomes of the operations are real numbers.)

The complexity of an algorithm, denoted by $C$, is the time required by the algorithm to produce one random variate. In many cases, $C$ itself is a random variable since it is a function of $U_1, U_2, \ldots$. We note here that we are mainly interested in generating independent sequences of random variables. The average complexity per random variate in a sequence of length $n$ is

$$\frac{1}{n}\sum_{i=1}^{n} C_i$$

where $C_i$ is the complexity for the $i$-th random variate. By the strong law of large numbers, we know that this average tends with probability one to the expected complexity, $E(C)$. There are examples of algorithms with infinite expected complexity, but for which the probability that $C$ exceeds a certain small constant is extremely small. These should not be a priori discarded.

We have now set the stage for the book. Our program is ambitious. In the remainder of this chapter, we introduce our notation, and define some distributions. By carefully selecting sections and exercises from the book, teachers could use it to introduce their students to the fundamental properties of distributions and random variables. Chapters II and III are crucial to the rest of the book: here, the principles of inversion, rejection, and composition are explained in all their generality. Less universal methods of random variate generation are developed in chapter IV. All of these techniques are then applied to generate random variates with specific univariate distributions. These include small families of densities (such as the normal, gamma or stable densities), small families of discrete distributions (such as the binomial and Poisson distributions), and families of distributions that are too large to be described by a finite number of parameters (such as all unimodal densities or all densities with decreasing hazard rate). The corresponding chapters are IX, X and VII. We devote chapter XI to multivariate random variate generation, and chapter VI to random process generation. In these chapters, we want to create dependence in a very specific way. This effort is continued in chapters XII and XIII on the generation of random subsets and the generation of random combinatorial objects such as random trees, random permutations and random partitions.

We do not touch upon the applications of random variate generation in Monte Carlo methods for solving various problems (see e.g. Rubinstein,1981): these problems include stochastic optimization, Monte Carlo integration, solving linear equations, deciding whether a large number is prime, etcetera. We will spend an entire section, however, on the important topic of discrete event simulation, driven by the beauty of some data structures used to make the simulation more efficient. As usual, we will not describe what happens inside some simulation languages, but merely give timeless principles and some analysis. Some of this is done in chapter XIV.

There are a few other chapters with specialized topics: the usefulness of order statistics is pointed out in chapter V. Shortcuts in simulation are highlighted in chapter XVI, and the important table methods are given special treatment in a chapter of their own (VIII). The reader will note that not a single experimental result is reported, and not one computer is explicitly named. The issue of programming in assembler language versus a high level language is not even touched (even though we think that assembler language implementations of many algorithms are essential). All of this is done to insure the universality of the text. Hopefully, the text will be as interesting in 1995 as in 1985 by not dwelling upon the shortcomings of today's computers. In fact, the emphasis is plainly upon complexity, the number of operations (instructions) needed to carry out certain tasks. Thus, chapter XV could very well be the most important chapter in the book for the future of the subject: here computers are treated as bit manipulating machines. This approach allows us to deduce lower bounds for the time needed to generate random variates with certain distributions.

We have taught some of the material at McGill University's School of Computer Science. For a graduate course on the subject for computer scientists, we recommend the material with a combinatorial and algorithmic flavor. One could

cover, not necessarily in the order given, parts of chapters I and II, all of chapter III, sections V.2 and V.3, selected examples from chapter X, all of chapters XII, XIII and XV, and section XIV.5. In addition, one could add chapter VIII. We usually cover I.1-3, II.1-2, II.3.1-2, II.3.6, II.4.1-2, III, V.1-3, V.4.1-4, VI.1, VIII.2-3, XII.1-2, XII.3.1, XII.4-5, XIII.1, XIII.2.1, XIII.3.3, XIII.4-5, and XIV.5.

In a statistics department, the needs are very different. A good sequence would be chapters II, III, V, VI, VII.2.1-3, selected examples from chapters IX,X, and chapter XII. In fact, this book can be used to introduce some of these students to the famous distributions in statistics, because the generators demand that we understand the connections between many distributions, that we know useful representations of distributions, and that we are well aware of the shape of densities and distribution functions. Some designs require that we disassemble some distributions, break densities up into parts, find tight inequalities for density functions.

The attentive reader notices very quickly that inequalities are ubiquitous. They are required to obtain efficient algorithms of all kinds. They are also useful in the analysis of the complexity. When we can make a point with inequalities, we will do so. A subset of the book could be used as the basis of a fun reading course on the development and use of inequalities: use parts of chapter I as needed, cover sections II.2, II.3, II.4.1, II.5.1, brush through chapter III, cover sections IV.5-7, include nearly all of chapter VII, and move on to sections VIII.1-2, IX.1.1-2, IX.3.1-3, IX.4, IX.6, X.1-4, XIV.3-4.

This book is intended for students in operations research, statistics and computer science, and for researchers interested in random variate generation. There is didactical material for the former group, and there are advanced technical sections for the latter group. The intended audience has to a large extent dictated the layout of the book. The introduction to probability theory in chapter I is not sufficient for the book. It is mainly intended to make the reader familiar with our notation, and to aid the students who will read the simpler sections of the book. A first year graduate level course in probability theory and mathematical statistics should be ample preparation for the entire book. But pure statisticians should be warned that we use quite a few ideas and "tricks" from the rich field of data structures and algorithms in computer science. Our short PASCAL programs can be read with only passing familiarity with the language.

Nonuniform random variate generation has been covered in numerous books. See for example Jansson (1966), Knuth (1969), Newman and Odell (1971), Yakowitz (1977), Fishman (1978), Kennedy and Gentle (1980), Rubinstein (1981), Payne (1982), Law and Kelton (1982), Bratley, Fox and Schrage (1983), Morgan (1984) and Banks and Carson (1984). In addition, there are quite a few survey articles (Zelen and Severo (1972), McGrath and Irving (1973), Patil, Boswell and Friday (1975), Marsaglia (1976), Schmeiser (1980), Devroye (1981), Ripley (1983) and Deak (1984)) and bibliographies (Sowey (1972), Nance and Overstreet (1972), Sowey (1978), Deak and Bene (1979), Sahal (1979)).

## 2. ABOUT OUR NOTATION.

In this section, we will briefly introduce the reader to the different formats that are possible for specifying a distribution, and to some of the most important densities in mathematical statistics.

### 2.1. Definitions.

A random variable $X$ has a density $f$ on the real line if for any Borel set $A$,

$$P(X \in A) = \int_A f(x)\, dx.$$

In other words, the probability that $X$ belongs to $A$ is equal to the area under the graph of $f$. The distribution function $F$ of $X$ is defined by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(y)\, dy \ , \qquad (x \in R) \ .$$

We have $F'(x) = f(x)$ for almost all $x$. The mean value of $X$ is

$$E(X) = \int x \ f(x) \ dx \ ,$$

provided that this integral exists. The $r$-th moment of $X$ is defined by $E(X^r)$. If the second moment of $X$ is finite, then its variance is defined by

$$Var(X) = E((X - E(X))^2) = E(X^2) - E^2(X) \ .$$

A mode of $X$, if it exists, is a point at which $f$ attains its maximal value. If $g$ is an arbitrary Borel measurable function and $X$ has density $f$, then $E(g(X)) = \int g(x)\, f(x)\ dx$. A $p$-th quantile of a distribution, for $p \in (0,1)$, is any point $x$ for which $F(x) = p$. The 0.5 quantile is also called the median. It is known that for nonnegative $X$,

$$E(X) = \int_0^{\infty} P(X \geq x)\, dx \ .$$

A distribution is completely specified when its distribution function is given. We recall that any nondecreasing function $F$, right-continuous, with limits 0 and 1 as $x \to -\infty$ and $x \to \infty$ respectively, is always the distribution function of some random variable. The distribution of a random variable is also completely known when the characteristic function

$$\phi(t) = E(e^{itX}) \ , \ t \in R \ ,$$

is given. For more details on the properties of distribution functions and characteristic functions, we refer to standard texts in probability such as Chow and Teicher (1978).

A random vector in $R^d$ has a distribution function

$$F(x_1, \ldots, x_d) = P(X_1 \leq x_1, \ldots, X_d \leq x_d) .$$

The random vector $(X_1, \ldots, X_d)$ has a density $f(x_1, \ldots, x_d)$ if and only if for all Borel sets $A$ of $R^d$,

$$P((X_1, \ldots, X_d) \in A) = \int_A f(x_1, \ldots, x_d)\, dx_1 \cdots dx_d .$$

The characteristic function of this random variable is

$$\phi(t_1, \ldots, t_d) = E(e^{it_1X_1 + \cdots + it_dX_d}) \qquad ((t_1, \ldots, t_d) \in R^d) .$$

The $X_i$'s are called marginal random variables. The marginal distribution function of $X_1$ is

$$F_1(x) = F(x, \infty, \ldots, \infty) \qquad (x \in R) .$$

Its marginal characteristic function is

$$\phi_1(t) = \phi(t, 0, \ldots, 0) , \qquad (t \in R) .$$

Another important notion is that of independence. Two random variables $X_1$ and $X_2$ are independent if and only if for all Borel sets $A$ and $B$,

$$P(X_1 \in A, X_2 \in B) = P(X_1 \in A)\, P(X_2 \in B) .$$

Thus, if $F$ is the distribution function of $(X_1, X_2)$, then $X_1$ and $X_2$ are independent if and only if

$$F(x_1, x_2) = F_1(x_1)\, F_2(x_2) , \text{ all } (x_1, x_2) \in R^2 ,$$

for some functions $F_1$ and $F_2$. Similarly, if $(X_1, X_2)$ has a density $f$, then $X_1$ and $X_2$ are independent if and only if this density can be written as the product of two marginal densities. Finally, $X_1$ and $X_2$ are independent if and only if for all bounded Borel measurable functions $g_1$ and $g_2$:

$$E(g_1(X_1) g_2(X_2)) = E(g_1(X_1))\, E(g_2(X_2)) .$$

In particular, the characteristic function of two independent random variables is the product of their characteristic functions:

$$\phi(t_1, t_2) = E(e^{it_1X_1} e^{it_2X_2}) = E(e^{it_1X_1})\, E(e^{it_2X_2}) = \phi_1(t_1)\, \phi_2(t_2) .$$

All the previous observations can be extended without trouble towards $d$ random variables $X_1, \ldots, X_d$.

## 2.2. A few important univariate densities.

In the table shown below, several important densities are listed. Most of them have one or two parameters. From a random variate generation point of view, several of these parameters are unimportant. For example, if $X$ is a random variable with a distribution having three parameters, $a, b, c$, and when $kX+l$ has a distribution with parameters $ka+l, kb, c$, then $b$ is called a scale parameter, and $a$ is called a translation parameter. The shape of the distribution is only determined by the parameter $c$: since $c$ is invariant to changes in scale and to translations, it is called a shape parameter. For example, the normal distribution has no shape parameter, and the gamma distribution has one shape parameter.

**Some univariate densities.**

| $f(x)$ | $E(X)$ | $Var(X)$ | Mode$(X)$ | $F(x)$ |
|---|---|---|---|---|
| Normal$(\mu,\sigma^2)$ $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ | $\mu$ | $\sigma^2$ | $\mu$ | $\int_{-\infty}^{x} f(y)\,dy$ |
| Gamma$(a,b)$ $\frac{1}{\Gamma(a)b^a} x^{a-1} e^{-\frac{x}{b}}$ $(x>0)$ | $ab$ | $ab^2$ | $(a-1)b$ | $\int_{-\infty}^{x} f(y)\,dy$ |
| Exponential$(\lambda)$ $\lambda e^{-\lambda x}$ $(x>0)$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ | 0 | $1-e^{-\lambda x}$ |
| Cauchy$(\sigma)$ $\frac{\sigma}{\pi(x^2+\sigma^2)}$ | does not exist | does not exist | 0 | $\frac{1}{2}+\frac{1}{\pi}\arctan(\frac{x}{\sigma})$ |
| Pareto$(a,b)$ $\frac{ab^a}{x^{a+1}}$ $(x>b)$ | $\frac{ab}{a-1}$ $(a>1)$ | $\frac{ab^2}{(a-2)(a-1)^2}$ $(a>2)$ | $b$ | $1-\frac{b^a}{x}$ |
| Beta$(a,b)$ $\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$ $(x\in[0,1])$ | $\frac{a}{a+b}$ | $\frac{ab}{(a+b)^2(a+b+1)}$ | $\frac{a-1}{a+b-2}$ $(a,b>1)$ | $\int_{-\infty}^{x} f(y)\,dy$ |

A variety of shapes can be found in this table. For example, the beta family of densities on [0,1] has two shape parameters, and the shapes vary from standard unimodal forms to J-shapes and U-shapes. For a comprehensive description of most parametric families of densities, we refer to the two volumes by Johnson and Kotz (1970). When we refer to normal random variables, we mean normal random variables with parameters 0 and 1. Similarly, exponential random variables are exponential (1) random variables. The uniform [0,1] density is the density which puts its mass uniformly over the interval [0,1]:

$$f(x) = I_{[0,1]}(x) \qquad (x \in R) .$$

Here $I$ is the indicator function of a set. Finally, when we mention the gamma ($a$) density, we mean the gamma ($a$,1) density.

The strategy in this book is to build from simple cases: simple random variables and distributions are random variables and distributions that can easily be generated on a computer. The context usually dictates which random variables are meant. For example, the uniform [0,1] distribution is simple, and so are the exponential and normal distributions in most circumstances. At the other end of the scale we have the difficult random variables and distributions. Most of this book is about the generation of random variates with difficult distributions. To clarify the presentation, it is convenient to use the same capital letters for all simple random variables. We will use N, E and U for normal, exponential and uniform [0,1] random variables. The notations G and B are often used for gamma and beta random variables. For random variables in general, we will reserve the symbols X, Y, W, Z, V.

## 3. ASSESSMENT OF RANDOM VARIATE GENERATORS.

One of the most difficult problems in random variate generation is the choice of an appropriate generator. Factors that play an important role in this choice include:

1. Speed.
2. Set-up (initialization) time.
3. Length of the compiled code.
4. Machine independence, portability.
5. Range of the set of applications.
6. Simplicity and readability.

Of these factors, the last one is perhaps the most neglected in the literature. Users are more likely to work with programs they can understand. Five line programs are easily typed in, and the likelihood of making errors is drastically reduced. Even packaged generators can have subtle bugs in their conception or implementation. It is nearly impossible to certify that programs with dozens, let alone hundreds, of lines of code are correct. So, we will often spend more time on simple algorithms than on sophisticated ultra-fast ones.

Subprograms for random variate generation can be divided into three groups: (1) subprograms with no variable parameters, such as subprograms for the normal (0,1) density; (2) subprograms with a finite number of variable parameters (these are typically for parametric classes of densities such as the class of all beta densities); (3) subprograms that accept names of other subprograms as arguments, and can be applied for a wide class of distributions (the description of this class is of course not dependent upon parameters).

set-up time.

**An example.**

The admissibility of a method now depends upon the set-up time as well, as is seen from this example. Stadlober (1981) gave the following table of expected times per variate (in microseconds) and size of the program (in words) for several algorithms for the $t$ distribution:

| Algorithm: | TD | TROU | T3T |
|---|---|---|---|
| t a=3.5 | 65 | 66 | 78 |
| t a=5 | 70 | 67 | 81 |
| t a=10 | 75 | 68 | 84 |
| t a=50 | 78 | 69 | 88 |
| t a=1000 | 79 | 70 | 89 |
| s | 255 | 100 | 83 |
| u | 12 | 190 | 0 |

Here $t$ stands for the expected time, $a$ for the parameter of the distribution, $s$ for the size of the compiled code, and $u$ for the set-up time. TD, TROU and T3T refer to three algorithms in the literature. For any algorithm and any $a$, the expected time per random variate is $t+\lambda u$ where $\lambda\in[0,1]$ is the fraction of the variates that required a set-up. The most important cases are $\lambda=0$ (one set-up in a large sample for fixed $a$) and $\lambda=1$ (parameter changes at every call). Also, $1/\lambda$ is about equal to the waiting time between set-ups. Clearly, one algorithm dominates another timewise if $t+\lambda u$ considered as a function of $\lambda$ never exceeds the corresponding function for the other algorithm. One can do this for each $a$, and this leads to quite a complicated situation. Usually, one should either randomize the entries of $t$ over various values of $a$. Alternatively, one can compare on the basis of $t_{\max}=\max_a t$. In our example, the values would be 79, 70 and 89 respectively. It is easy to check that $t_{\max}+\lambda u$ is minimal for TROU when $0\leq\lambda\leq 9/178$, for TD when $9/178\leq\lambda\leq 5/6$, and for T3T when $5/6\leq\lambda\leq 1$. Thus, there are no inadmissible methods if we want to include all values of $\lambda$. For fixed values of $\lambda$ however, we have a given ranking of the $t_{\max}+\lambda u$ values and the discussion of the inadmissibility in terms of $t_{\max}+\lambda u$ and $s$ is as for the distributions without parameters. Thus, TD is inadmissible in this sense for $\lambda>5/6$ or $\lambda<9/178$, and TROU is inadmissible for $\lambda>1/10$. ■

### 3.1. Distributions with no variable parameters.

A frequently used subprogram for distributions with no variable parameters should be chosen very carefully: usually, speed is very important, while the length of the compiled code is less crucial. Clearly, the initialization time is zero, and in some cases it is worthwhile to write the programs in machine language. This is commonly done for distributions such as the normal distribution and the exponential distribution.

For infrequently used subprograms, it is probably not worth to spend a lot of time developing a fast algorithm. Rather, a simple expedient method will often do. In many cases, the portability of a program is the determining factor: can we use the program in different installations under different circumstances? Portable programs have to be written in a machine-independent language. Furthermore, they should only use standard library subprograms and be compiler-independent. Optimizing compilers often lead to unsuspected problems. Programs should follow the universal conventions for giving names to variables, and be protected against input error. The calling program should not be told to use special statements (such as the COMMON statement in FORTRAN). Finally, the subprogram itself is not assumed to perform unasked tasks (such as printing messages), and all conventions for subprogram linkage must be followed.

Assume now that we have narrowed the competition down to a few programs, all equally understandable and portable. The programs take expected time $t_i$ per random variate where $i$ refers to the $i$-th program ($1 \leq i \leq K$). Also, they require $s_i$ bytes of storage. Among these programs, the $j$-th program is said to be **inadmissible** if there exists an $i$ such that $t_j \geq t_i$ and $s_j \geq s_i$ (with at least one of these inequalities strict). If no such $i$ exists, then the $j$-th program is admissible. If we measure the cost of the $i$-th program by some function $\psi(t_i, s_i)$, increasing in both its arguments, then it is obvious that the best program is an admissible program.

### 3.2. Parametric families.

The new ingredient for multi-parameter families is the set-up time, that is, the time spent computing constants that depend only upon the parameters of the distribution. We are often in one of two situations:

Case 1. The subprogram is called very often for fixed values of the parameters. The set-up time is unimportant, and one can only gain by initializing as many constants as possible.

Case 2. The parameters of the distribution change often between calls of the subprogram. The total time per variate is definitely influenced by the

**Speed versus size.**

It is a general rule in computer science that speed can be reduced by using longer more sophisticated programs. Fast programs are seldom short, and short programs are likely to be slow. But it is also true that long programs are often not elegant and more error-prone. Short smooth programs survive longer and are understood by a larger audience. This bias towards short programs will be apparent in chapters IV, IX and X where we must make certain recommendations to the general readership. ■

## 4. OPERATIONS ON RANDOM VARIABLES.

In this section we briefly indicate how densities and distribution functions change when random variables are combined or operated upon in certain ways. This will allow us to generate new random variables from old ones. We are specially interested in operations on simple random variables (from a random variate generation point of view) such as uniform [0,1] random variables. The actual applications of these operations in random variate generation are not discussed in this introductory chapter. Most of this material is well-known to students in statistics, and the chapter could be skipped without loss of continuity by most readers. For a unified and detailed treatment of operations on random variables, we refer to Springer(1979).

### 4.1. Transformations.

Transformations of random variables are easily taken care of by the following device:

**Theorem 4.1.**

Let $X$ have distribution function $F$, and let $h:R \to B$ be a strictly increasing function where $B$ is either $R$ or a proper subset of $R$. Then $h(X)$ is a random variable with distribution function $F(h^{-1}(x))$.

If $F$ has density $f$ and $h^{-1}$ is absolutely continuous, then $h(X)$ has density

$$(h^{-1})'(x)\ f\ (h^{-1}(x)), \quad \text{for almost all } x\ .$$

**Proof of Theorem 4.1.**

Observe first that for arbitrary $x$,

$$P(h(X) \leq x) = P(X \leq h^{-1}(x)) = F(h^{-1}(x)) .$$

This is thus the distribution function of $h(X)$. If this distribution function is absolutely continuous in $x$, then we know (Chow and Teicher (1978)) that $h(X)$ has a density that is almost everywhere equal to the derivative of the distribution function. This is the case for example when both $F$ and $h^{-1}$ are absolutely continuous, and the formal derivative is the one shown in the statement of the Theorem. ■

**Example 4.1. Linear transformations.**

If $F$ is the distribution function of a random variable $X$, then $aX+b$ has distribution function $F((x-b)/a)$ when $a>0$. The corresponding densities, if they exist, are $f(x)$ and $\frac{1}{a} f(\frac{x-b}{a})$. Verify that when $X$ is gamma $(a,b)$ distributed, then $cX$ is gamma $(a,cb)$, all $c>0$. ■

**Example 4.2. The exponential distribution.**

When $X$ has distribution function $F$ and $\lambda>0$ is a real number, then $-\frac{1}{\lambda}\log X$ has distribution function $1-F(e^{-\lambda x})$, which can be verified directly:

$$P(-\frac{1}{\lambda}\log X \leq x) = P(X \geq e^{-\lambda x}) = 1-F(e^{-\lambda x}) \qquad (x>0).$$

In particular, if $X$ is uniform [0,1], then $-\frac{1}{\lambda}\log X$ is exponential $(\lambda)$. Vice versa, when $X$ is exponential $(\lambda)$, then $e^{-\lambda X}$ is uniform [0,1]. ■

**Example 4.3. Power transformations.**

When $X$ has distribution function $F$ and density $f$, then $X^p$ ( $p>0$ is a real number, and the power is defined as a sign-preserving transformation ) has distribution function $F(x^{\frac{1}{p}})$ and density

$$\frac{1}{p}x^{\frac{1}{p}-1}f(x^{\frac{1}{p}}). \blacksquare$$

**Example 4.4. Non-monotone transformations.**

Non-monotone transformations are best handled by computing the distribution function first from general principles. To illustrate this, let us consider a random variable $X$ with distribution function $F$ and density $f$. Then, the random variable $X^2$ has distribution function

$$P(X^2 \leq x) = P(|X| \leq \sqrt{x}) = F(\sqrt{x}) - F(-\sqrt{x}) \qquad (x>0)$$

and density

$$\frac{1}{\sqrt{x}} \frac{f(\sqrt{x})+f(-\sqrt{x})}{2}.$$

In particular, when $X$ is normal (0,1), then $X^2$ is gamma distributed, as can be seen from the form of the density

$$\frac{1}{\sqrt{x}} \frac{1}{2\sqrt{2\pi}}(e^{-\frac{x}{2}}+e^{-\frac{x}{2}}) = \frac{1}{\sqrt{2\pi}}x^{-\frac{1}{2}}e^{-\frac{x}{2}} \qquad (x \geq 0).$$

The latter density is known as the chi-square density with one degree of freedom (in shorthand: $\chi_1^2$). $\blacksquare$

**Example 4.5. A parametric form for the density.**

Let $X$ have density $f$ and let $h$ be as in Theorem 4.1. Then, putting $x=h(u)$ and $y=f(u)/h'(u)$, where $y$ stands for the value of the density of $h(X)$ at $x$, and $y$ and $x$ are related through the parameter $u$, we verify by elimination of $u$ that

$$y = f(h^{-1}(x)) \,/\, h'(h^{-1}(x)).$$

This is equal to $f(h^{-1}(x))h^{-1\prime}(x)$, which was to be shown. Thus, the parametric representation in terms of $u$ given above is correct, and will give us a plot of the density versus $x$. This is particularly useful when the inverse of $h$ is difficult to obtain in closed analytical form. For example, when $X$ is uniform [0,1], then for $a, b > 0$, $aX + bX^3$ has a density with parametric representation

$$x = au + bu^3, \quad y = \frac{1}{a + 3bu^2} \qquad (0 \leq au + bu^3 \leq 1).$$

By elimination of $u$, we obtain a simple formula of $x$ in terms of $y$:

$$x = \sqrt{\frac{1}{y} - a}\ \sqrt{\frac{1}{3b}}\left(\frac{2a}{3} + \frac{1}{3y}\right).$$

The plot of $y$ versus $x$ has the following general form: it vanishes outside [0,1], and decreases monotonically on this interval from $y = \frac{1}{a}$ at $x = 0$ to a nonzero value at $x = 1$. Furthermore, $\frac{\partial y}{\partial x}$ at $u = 0$ (i.e. at $x = 0$), is 0, so that the shape of the density resembles that of a piece of the normal density near 0. ■

Let us now look at functions of several random variables. We can obtain many distributions as relatively uncomplicated functions of simple random variables. Many cases can be handled by the following $d$-dimensional generalization of Theorem 4.1:

**Theorem 4.2.**

Let $X$ have a continuous density $f$ on $R^d$ and let $h: R^d \to R^d$ be a one-to-one and onto mapping to $T$, the image of $S$, the support set of $f$, under $h$. Thus, the inverse of the transformation $Y = h(X)$ exists: $X = h^{-1}(Y) = g(Y)$. If we write $y = (y_1, \ldots, y_d)$ and $g = (g_1, \ldots, g_d)$, then if the partial derivatives $g_{ij} = \frac{\partial g_i}{\partial y_j}$ exist and are continuous on $T$, $Y$ has density

$$f(g(y))\ |J| \qquad (y \in T),$$

where $J$ is the Jacobian of the transformation and is defined as the determinant of the matrix

$$\begin{bmatrix} g_{11} & \cdots & g_{1d} \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ g_{d1} & \cdots & g_{dd} \end{bmatrix}.$$

**Example 4.6. The t distribution.**

We will show here that when $X$ is normal (0,1) and $Y$ is independent of $X$ and gamma $(\frac{a}{2},2)$ distributed (this is called the chi-square distribution with $a$ degrees of freedom), then

$$Z = X/\sqrt{\frac{Y}{a}}$$

is $t$ distributed with $a$ degrees of freedom, that is, $Z$ has density

$$\frac{\Gamma(\frac{a+1}{2})}{\Gamma(\frac{a}{2})\sqrt{\pi a}} \frac{1}{(1+\frac{z^2}{a})^{\frac{a+1}{2}}} \qquad (z \in R) .$$

What one does in a situation like this is "invent" a 2-dimensional vector random variable (for example, $(Z,W)$) that is a function of $(X,Y)$, one of whose component random variables is $Z$. The obvious choice in our example is

$$\begin{aligned} Z &= X/\sqrt{\frac{Y}{a}} \\ W &= Y \end{aligned}$$

The inverse transformation is determined by $X = Z\sqrt{\frac{W}{a}}$, $Y = W$. This inverse transformation has a Jacobian $\sqrt{\frac{w}{a}}$ where we use $x,y,z,w$ for the running values that correspond to the random variables $X,Y,Z,W$. Thus, the density of $(Z,W)$ is

$$c\ e^{-\frac{wz^2}{2a}} w^{\frac{a}{2}-1} e^{-\frac{w}{2}} \sqrt{\frac{w}{a}}$$

where

$$c = \frac{1}{\Gamma(\frac{a}{2})2^{\frac{a}{2}}\sqrt{2\pi}} \qquad (w>0, z \in R)$$

is a normalization constant. From a joint density, we obtain a marginal density by taking the integral with respect to the non-involved variables (in this case with respect to $dw$). In $w$, we have for fixed $z$ a gamma $(\frac{a+1}{2}, \frac{2}{1+z^2/a})$ density times $\frac{c}{\sqrt{a}}$. After integration with respect to $dw$, we obtain

$$\frac{c}{\sqrt{a}}\Gamma(\alpha)\beta^{\alpha}$$

where $\alpha$ and $\beta$ are the parameters of the gamma density given above. This is precisely what we needed to show. ■

## 4.2. Mixtures.

### Discrete mixtures.

Let $Y$ be a positive integer valued random variable, and, given that $Y=i$, let $X$ have density $f_i$. Then the (unconditional) density of $X$ is

$$\sum_{i=1}^{\infty} P(Y=i)\, f_i(x)\,.$$

This device can be used to cut a given density $f$ up into simpler pieces $f_i$ that can be handled quite easily. Often, the number of terms in the mixture is finite. For example, if $f$ is a piecewise linear density with a finite number of breakpoints, then it can always be decomposed (rewritten) as a finite mixture of uniform and triangular densities.

### Continuous mixtures.

Let $Y$ have density $g$ on $R$, and given that $Y=y$, let $X$ have density $f_y$ (thus, $y$ can be considered as a parameter of the density of $X$), then the density $f$ of $X$ is given by

$$f(x) = \int f_y(x)\, g(y)\, dy\,.$$

As an example, we consider a mixture of exponential densities with parameter $Y$ itself exponentially distributed with parameter 1. Then $X$ has density

$$\begin{aligned} f(x) &= \int y e^{-yx} e^{-y}\, dy \\ &= \int y e^{-\frac{y}{(x+1)^{-1}}}\, dy \\ &= \frac{1}{(x+1)^2} \qquad (x>0)\,. \end{aligned}$$

Since the parameter of the exponential distribution is the inverse of the scale parameter, we see without work that when $E_1, E_2$ are independent exponential random variables, then $E_1/E_2$ has density $1/(x+1)^2$ on $[0,\infty)$.

### Mixtures of uniform densities.

If we consider a mixture of uniform $[0,y]$ densities where $y$ is the mixture parameter, then we obtain a density that is nonincreasing on $[0,\infty)$. The random variables $X$ thus obtained are distributed as the product $UY$ of a uniform $[0,1]$ random variable $U$ and an arbitrary (mixture) random variable $Y$. These distributions will be of great interest to us since $U$ is the fundamental random

variable in random variate generation.

### 4.3. Order statistics.

If $U_1, \ldots, U_n$ are iid uniform [0,1] random variables, then the order statistics for this sample are $U_{(1)}, \ldots, U_{(n)}$, where

$$U_{(1)} \leq U_{(2)} \leq \cdots \leq U_{(n)}$$

and $U_{(1)}, \ldots, U_{(n)}$ is a permutation of $U_1, \ldots, U_n$. We know that $(U_1, \ldots, U_n)$ is uniformly distributed in the unit cube $[0,1]^n$. Thus, $(U_{(1)}, \ldots, U_{(n)})$ is uniformly distributed in the simplex $S_n$:

$$S_n = \{(x_1, \ldots, x_n) : 0<x_1<x_2< \cdots <x_n <1\} .$$

**Theorem 4.3.**

The joint density of $(U_{(1)}, \ldots, U_{(n)})$ is

$$n!I_{S_n}(x_1, \ldots, x_n) .$$

The $i$-th order statistic $U_{(i)}$ has the beta density with parameters $i$ and $n-i+1$, i.e. its density is

$$\frac{\Gamma(n+1)}{\Gamma(i)\Gamma(n-i+1)} x^{i-1}(1-x)^{n-i} \qquad (x \in [0,1]) .$$

**Proof of Theorem 4.3.**

The first part is shown by a projection argument: there are $n!$ points in $[0,1]^n$ that map to a given point in $S_n$ when we order them. This can be formalized as follows. Let $A$ be an arbitrary Borel set contained in $S_n$. Writing $x_{(1)}< \cdots <x_{(n)}$ for the ordered permutation of $x_1, \ldots, x_n$, we have

$$\begin{aligned}
&\int_A dx_1 \cdots dx_n \\
&= \sum_\sigma \int_{A,(x_1=x_{\sigma(1)}, \ldots, x_n=x_{\sigma(n)})} dx_1 \cdots dx_n \\
&(\sigma=\sigma(1), \ldots, \sigma(n) \text{ is a permutation of } 1, \ldots, n) \\
&= \sum_\sigma \int_A dx_{(1)} \cdots dx_{(n)} \\
&= \int_A n!\, dx_{(1)} \cdots dx_{(n)} .
\end{aligned}$$

The first part of the Theorem follows by the arbitrariness of $A$. For the second part, we choose $x$ in [0,1], and compute the marginal density of $U_{(i)}$ at $x$ by integrating the density with respect to all variables $x_j$ , $j \neq i$. This yields

$$n! \int_0^{x_2} \cdots \int_0^{x} \int_x^{1} \cdots \int_{x_{n-1}}^{1} dx_n \cdots dx_{i+1} dx_{i-1} \cdots dx_1 .$$

This gives the beta density with parameters $i$ and $n-i+1$. ■

Of particular importance will be the distribution of $\max(U_1, \ldots, U_n)$: the distribution function is easily obtained by a direct argument because

$$\begin{aligned} & P(\max(U_1, \ldots, U_n) \leq x) \qquad (x \in [0,1]) \\ &= P(U_1 \leq x) \cdots P(U_n \leq x) \\ &= x^n \\ &= P(U_1 \leq x^n) \\ &= P(U_1^{\frac{1}{n}} \leq x) . \end{aligned}$$

Thus, the distribution function is $x^n$ on [0,1], and the density is $nx^{n-1}$ on [0,1]. We have also shown that $\max(U_1, \ldots, U_n)$ is distributed as $U_1^{1/n}$.

Another important order statistic is the median. The median of $U_1, \ldots, U_{2n+1}$ is $U_{(n)}$. We have seen in Theorem 4.3 that the density is

$$\frac{(2n+1)!}{n!^2} (x(1-x))^n \qquad (x \in [0,1]) .$$

**Example 4.7.**

If $U_{(1)}, U_{(2)}, U_{(3)}$ are the order statistics of three independent uniform [0,1] random variables, then their densities on [0,1] are respectively,

$$3(1-x)^2 ,$$

$$6x(1-x)$$

and

$$3x^2 . \blacksquare$$

The generalizations of the previous results to other distributions are straightforward. If $X_1, \ldots, X_n$ are iid random variables with density $f$ and distribution function $F$, then the maximum has distribution function $F^n$. From Theorem 4.3, we can also conclude that the $i$-th order statistic $X_{(i)}$ has density

$$\frac{n!}{(i-1)!(n-i)!} F(x)^{i-1}(1-F(x))^{n-i} f(x).$$

## 4.4. Convolutions. Sums of independent random variables.

The distribution of the sum $S_n$ of $n$ random variables $X_1, \ldots, X_n$ is usually derived by one of two tools, convolution integrals or characteristic functions. In this section, we will write $f_i, F_i, \phi_i$ for the density, distribution function and characteristic function of $X_i$, and we will use the notation $f, F, \phi$ for the corresponding functions for the sum $S_n$. In the convolution method, we argue as follows:

$$F(x) = P(X_1 + \cdots + X_n \leq x)$$
$$= \int \prod_{i<n} f_i(y_i)\, F_n(x - y_1 - \cdots - y_{n-1}) \prod_{i<n} dy_i .$$

Also,

$$f(x) = \int \prod_{i<n} f_i(y_i)\, f_n(x - y_1 - \cdots - y_{n-1}) \prod_{i<n} dy_i .$$

Except in the simplest cases, these convolution integrals are difficult to compute. In many instances, it is more convenient to derive the distribution of $S_n$ by finding its characteristic function. By the independence of the $X_i$'s, we have

$$\phi(t) = E(e^{it(X_1 + \cdots + X_n)})$$
$$= \prod_{j=1}^{n} E(e^{itX_j})$$
$$= \prod_{j=1}^{n} \phi_j(t) .$$

If the $X_i$'s are iid, then $\phi = \phi_1^n$.

### Example 4.8. Sums of normal random variables.

First, we show that the characteristic function of a normal (0,1) random variable is $e^{-t^2/2}$. To see this, note that it can be computed as follows for $t \in R$:

$$\int \frac{1}{\sqrt{2\pi}} e^{ity - y^2/2}\, dy$$

$$= e^{-t^2/2} \int \frac{1}{\sqrt{2\pi}} e^{-(y-it)^2/2}\, dy$$

$$= e^{-t^2/2} .$$

From the definition of the characteristic function we see that if $X$ has characteristic function $\phi(t)$, then $aX+b$ has characteristic function $e^{ibt}\phi(at)$. Thus, a normal $(\mu,\sigma^2)$ random variable has characteristic function

$$e^{it\mu}\,\phi(\sigma t) .$$

If $X_i$ is normal $(\mu_i,\sigma_i^2)$, then $S_n$ has characteristic function

$$\prod_{j=1}^{n} e^{it\mu_j}\, e^{-\sigma_j^2 t^2/2}$$

$$= e^{it\sum_{j=1}^{n}\mu_j}\, e^{-\sum_{j=1}^{n}\sigma_j^2 t^2/2} ,$$

which is the characteristic function of a normal random variable with parameters $\sum\mu_j$ and $\sum\sigma_j^2$. ■

**Example 4.9. Sums of gamma random variables.**

In this example too, it is convenient to first obtain the characteristic function of a gamma $(a,b)$ random variable. It can be computed as follows:

$$\int_0^\infty \frac{y^{a-1}e^{-y/b}}{\Gamma(a)b^a} e^{ity}\, dy \qquad \text{(by definition)}$$

$$= \int_0^\infty \frac{y^{a-1}e^{-y(1-itb)/b}}{\Gamma(a)b^a}\, dy$$

$$= \int_0^\infty \frac{z^{a-1}e^{-z/b}}{(1-itb)^a\Gamma(a)b^a}\, dz \qquad (\text{use } z=y(1-itb))$$

$$= \frac{1}{(1-itb)^a} .$$

Thus, if $X_1, \ldots, X_n$ are independent gamma random variables with parameters $a_i$ and $b$, then the sum $S_n$ is gamma with parameters $\sum a_i$ and $b$. ■

It is perhaps worth to mention that when the $X_i$'s are iid random variables, then $S_n$, properly normalized, is nearly normally distributed when $n$ grows large.

If the distribution of $X_1$ has mean $\mu$ and variance $\sigma^2>0$, then $(S_n - n\mu)/(\sigma\sqrt{n})$ tends in distribution to a normal (0,1) random variable, i.e,

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad \text{all } x .$$

This is called the central limit theorem (Chow and Teicher, 1978). This will be exploited further on in the design of algorithms for families of distributions that are closed under additions, such as the gamma or Poisson families. If the variance is not finite, then the limit law is no longer normal. See for example exercise 4.17, where an example is found of such non-normal attraction.

## 4.5. Sums of independent uniform random variables.

In this section we consider the distribution of

$$\sum_{i=1}^{n} a_i U_i$$

where the $a_i$'s are positive constants and the $U_i$'s are independent uniform [0,1] random variables. We start with the main result of this section.

**Theorem 4.4.**

The distribution function of $\sum_{i=1}^{n} a_i U_i$ (where $a_i>0$, all $i$, and the $U_i$'s are independent uniform [0,1] random variables) is given by

$$F(x) = \frac{1}{a_1 a_2 \cdots a_n \, n!}\left(x_+^{\,n} - \sum_i (x-a_i)_+^{\,n} + \sum_{i\neq j} (x-a_i-a_j)_+^{\,n} - \cdots\right).$$

Here $(.)_+$ is the positive part of (.). The density is obtained by taking the derivative with respect to $x$.

**Proof of Theorem 4.4.**

Consider the simplex $S$ formed by the origin and the vertices on the $n$ coordinate axes at distances $x/a_1, \ldots, x/a_n$, where $x>0$ is the point at which we want to calculate $F(x)$. Let us define the sets $B_i$ as

$$B_i = [0,\infty)^{i-1} \times (1,\infty) \times [0,\infty)^{n-i}$$

where $1 \leq i \leq n$. Note now that the first quadrant minus the unit cube $[0,1]^n$ can be decomposed by the inclusion/exclusion principle as follows:

$$[0,\infty)^n - [0,1]^n$$
$$= \sum_i B_i - \sum_{i\neq j} B_i \cap B_j + \cdots .$$

Now, since $F(x) = \text{area}\,(S \cap [0,1]^n) = \text{area}\,(S) - \text{area}\,(S \cap ([0,\infty)^n - [0,1]^n))$, we obtain

$$F(x) = \text{area}\,(S) - \sum_i \text{area}\,(S \cap B_i) + \sum_{i \neq j} \text{area}\,(S \cap B_i \cap B_j) - \cdots .$$

This is all we need, because for any subset $J$ of $1, \ldots, n$, we have

$$\text{area}\,(S \cap_{i \in J} B_i) = \frac{(x - \sum_{i \in J} a_i)_+^n}{a_1 \cdots a_n \, n!} .$$

This concludes the proof of Theorem 4.4. ■

It is instructive to do the proof of Theorem 4.4 for the special case $n = 2$, and to draw the simplex and the various sets used in the geometric proof. For the important case $a_1 = a_2 = \cdots = a_n = 1$, the distribution function is

$$F(x) = \frac{1}{n!}\left( x_+^n - \binom{n}{1}(x-1)_+^n + \binom{n}{2}(x-2)_+^n - \cdots \right) .$$

In particular, for $n = 2$, obtaining the density by taking the derivative of the distribution function, we have

$$f(x) = x_+ - 2(x-1)_+ + (x-2)_+$$

$$= \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x \leq 1 \\ 2-x & \text{if } 1 \leq x \leq 2 \\ 0 & \text{if } 2 < x \end{cases} .$$

In other words, the density has the shape of an isosceles triangle. In general, the density of $U_1 + U_2 + \cdots + U_n$ consists of pieces of polynomials of degree $n-1$ with breakpoints at the integers. The form approaches that of the normal density as $n \to \infty$.

### 4.6. Exercises.

1. If $h$ is strictly monotone, $h'$ exists and is continuous, $g$ is a given density, and $X$ is a random variable with density $h'(x)g(h(x))$, then $h(X)$ has density $g$. (This is the inverse of Theorem 4.1.)
2. If $X$ has density $1/(x^2\sqrt{\pi \log x})$ $(x \geq 1)$, then $\sqrt{2\log X}$ is distributed as the absolute value of a normal random variable. (Use exercise 1.)
3. If $X$ is a gamma $(\frac{1}{2},1)$ random variable, i.e. $X$ has density $e^{-x}/\sqrt{\pi x}$ $(x>0)$, then $\sqrt{2X}$ is distributed as the absolute value of a normal random variable. (Use exercise 1.)
4. Let $A$ be a $d \times d$ matrix with nonzero determinant. Let $Y=AX$ where both $X$ and $Y$ are $R^d$-valued random vectors. If $X$ has density $f$, then $Y$ has density

$$f(A^{-1}y)\,|\det A^{-1}| \qquad (y \in R^d).$$

Thus, if $X$ has a uniform density on a set $B$ of $R^d$, then $Y$ is uniformly distributed on a set $C$ of $R^d$. Also, determine $C$ from $B$ and $A$.

5. If $Y$ is gamma $(a,1)$ and $X$ is exponential $(Y)$, then the density of $X$ is

$$f(x) = \frac{a}{(x+1)^{a+1}} \qquad (x \geq 0).$$

6. A random variable is said to have the F distribution with $a$ and $b$ degrees of freedom when its density is

$$f(x) = \frac{cx^{\frac{a}{2}-1}}{(1+\frac{ax}{b})^{\frac{a+b}{2}}}, \qquad (x>0).$$

Here, $c$ is the constant $\Gamma(\frac{a+b}{2})(\frac{a}{b})^{\frac{a}{2}}/\Gamma(\frac{a}{2})\Gamma(\frac{b}{2})$. Show that when $X$ and $Y$ are independent chi-square random variables with parameters $a$ and $b$ respectively, then $(\frac{X}{a})/(\frac{Y}{b})$ is $F(a,b)$. Show also that when $X$ is $F(a,b)$, then $\frac{1}{X}$ is $F(b,a)$. Show finally that when $X$ is $t$-distributed with $a$ degrees of freedom, $X^2$ is $F(1,a)$. Draw the curves of the densities of $F(2,2)$ and $F(3,1)$ random variables.

7. When $N_1$ and $N_2$ are independent normal random variables, the random variables $N_1^2+N_2^2$ and $N_1/N_2$ are independent.
8. Let $f$ be the triangular density defined by

$$f(x) = \begin{cases} 1-\frac{x}{2} & \text{if } 0 \leq x \leq 2 \\ 0 & \text{elsewhere} \end{cases}$$

When $U_1$ and $U_2$ are independent uniform [0,1] random variables, then the following random variables all have density $f$ :

$2 \min (U_1, U_2)$ ;

$2 \mid U_1 + U_2 - 1 \mid$ ;

$2(1-\sqrt{U_1})$ .

9. Show that the density of the product $\prod_{i=1}^{n} U_i$ of $n$ iid uniform [0,1] random variables is

$$f(x) = \begin{cases} \frac{1}{(n-1)!} \log(\frac{1}{x})^{n-1} & 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

10. When $X$ is gamma $(a, 1)$, then $1/X$ has density

$$f(x) = (\frac{1}{x})^{a+1} \frac{e^{-\frac{1}{x}}}{\Gamma(a)} \qquad (x \geq 0) .$$

11. Let $Y = \prod_{i=1}^{k} X_i$ where $X_1, \ldots, X_k$ are iid random variables each distributed as the maximum of n iid uniform [0,1] random variables. Then $Y$ has density

$$f(x) = \frac{n^k}{\Gamma(k)} x^{n-1} (-\log(x))^{k-1} \qquad (0 \leq x \leq 1) .$$

(Rider, 1955; Rahman, 1964).

12. Let $X_1, \ldots, X_n$ be iid uniform [-1,1] random variables, and let $Y$ be equal to $(\min(X_1, \ldots, X_n) + \max(X_1, \ldots, X_n))/2$. Show that $Y$ has density

$$f(x) = \frac{n}{2}(1 - \mid x \mid)^{n-1} \qquad (\mid x \mid \leq 1) ,$$

and variance $\frac{2}{(n+1)(n+2)}$ (Neyman and Pearson, 1928; Carlton, 1946).

13. We say that the power distribution with parameter $a > -1$ is the distribution corresponding to the density

$$f(x) = (a+1)x^a \qquad (0 < x < 1) .$$

If $X_1, \ldots$ are iid random variables having the power distribution with parameter $a$, then show that

A. $X_1/X_2$ has density

$$\begin{cases} \frac{a+1}{2} x^a & 0 < x < 1 \\ \frac{a+1}{2} x^{-(a+2)} & 1 \leq x \end{cases}$$

B. $\prod_{i=1}^{n} X_i$ has density

$$\frac{(a+1)^n}{\Gamma(n)} x^a (\log \frac{1}{x})^{n-1} \qquad (0<x<1) .$$

(Springer, 1979, p. 161).

14. The ratio $G_a / G_b$ of two independent gamma random variables with parameters $(a,1)$ and $(b,1)$ respectively has density

$$\frac{1}{B(a,b)} \frac{x^{a-1}}{(1+x)^{a+b}} \qquad (x>0) .$$

Here $B(a,b)$ is the standard abbreviation for the constant in the beta integral, i.e. $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. This is called the beta density of the second kind. Furthermore, $G_a/(G_a+G_b)$ has the beta density with parameters $a$ and $b$.

15. Let $U_1, \ldots, U_4$ be iid uniform [0,1] random variables. Show that $(U_1+U_2)/(U_3+U_4)$ has density

$$\begin{cases} \frac{7x}{6} & 0<x<\frac{1}{2} \\ \frac{8}{3}-\frac{3x}{2}-\frac{2}{3x^2}+\frac{1}{6x^3} & \frac{1}{2}\le x<1 \\ -\frac{2}{3}+\frac{x}{6}+\frac{8}{3x^2}-\frac{3}{2x^3} & 1\le x<2 \\ \frac{7}{6x^3} & 2\le x \end{cases}$$

16. Show that $N_1N_2+N_3N_4$ has the Laplace density (i.e., $\frac{1}{2}e^{-|x|}$), whenever the $N_i$'s are iid normal random variables (Mantel, 1973).

17. Show that the characteristic function of a Cauchy random variable is $e^{-|t|}$. Using this, prove that when $X_1, \ldots, X_n$ are iid Cauchy random variables, then $\frac{1}{n}\sum_{i=1}^{n} X_i$ is again Cauchy distributed, i.e. the average is distributed as $X_1$.

18. Use the convolution method to obtain the densities of $U_1+U_2$ and $U_1+U_2+U_3$ where the $U_i$'s are iid uniform [-1,1] random variables.

19. In the oldest FORTRAN subroutine libraries, normal random variates were generated as

$$X_n = \frac{1}{\sqrt{n/3}} \sum_{j=1}^{n} (U_j - \frac{1}{2})$$

where the $U_j$'s are iid uniform [0,1] random variates. Usually $n$ was equal to 12. This generator is of course inaccurate. Verify however that the mean and variance of such random variables are correct. Bolshev (1959) later

proposed the corrected random variate

$$Y = X_5 - \frac{3X_5 - X_5^3}{100} .$$

Define a notion of closeness between densities, and verify that $Y$ is closer to a normal random variable than $X_5$.

20. Let $U_1, \ldots, U_n, V_1, \ldots, V_m$ be iid uniform [0,1] random variables. Define $X = \max(U_1, \ldots, U_n)$, $Y = \max(V_1, \ldots, V_m)$. Then $X/Y$ has density

$$f(x) = \begin{cases} cx^{n-1} & ,0 \le x \le 1 \\ \dfrac{c}{x^{m+1}} & ,x \ge 1 \end{cases}$$

where $c = \dfrac{nm}{n+m}$ (Murty, 1955).

21. Show that if $X \le Y \le Z$ are the order statistics for three iid normal random variables, then

$$\frac{\min(Z-Y, Y-X)}{Z-X}$$

has density

$$f(x) = \frac{3\sqrt{3}}{\pi(1-x+x^2)} , \quad (0 \le x \le \frac{1}{2}) .$$

See e.g. Lieblein (1952).