NON-UNIFORM RANDOM VARIATE GENERATION

Luc Devroye
School of Computer Science
McGill University

ABSTRACT. This chapter provides a survey of the main methods in non-uniform random variate generation, and highlights recent research on the subject. Classical paradigms such as inversion, rejection, guide tables, and transformations are reviewed. We provide information on the expected time complexity of various algorithms, before addressing modern topics such as indirectly specified distributions, random processes, and Markov chain methods.

Authors' address: School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. The authors' research was sponsored by NSERC Grant A3456 and FCAR Grant 90-ER-0291.

1. The main paradigms

The purpose of this chapter is to review the main methods for generating random variables, vectors and processes. Classical workhorses such as the inversion method, the rejection method and table methods are reviewed in section 1. In section 2, we discuss the expected time complexity of various algorithms, and give a few examples of the design of generators that are uniformly fast over entire families of distributions. In section 3, we develop a few universal generators, such as generators for all log concave distributions on the real line. Section 4 deals with random variate generation when distributions are indirectly specified, e.g, via Fourier coefficients, characteristic functions, the moments, the moment generating function, distributional identities, infinite series or Kolmogorov measures. Random processes are briefly touched upon in section 5. Finally, the latest developments in Markov chain methods are discussed in section 6. Some of this work grew from Devroye (1986a), and we are carefully documenting work that was done since 1986. More recent references can be found in the book by Hörmann, Leydold and Derflinger (2004).

Non-uniform random variate generation is concerned with the generation of random variables with certain distributions. Such random variables are often discrete, taking values in a countable set, or absolutely continuous, and thus described by a density. The methods used for generating them depend upon the computational model one is working with, and upon the demands on the part of the output.

For example, in a RAM (random access memory) model, one accepts that real numbers can be stored and operated upon (compared, added, multiplied, and so forth) in one time unit. Furthermore, this model assumes that a source capable of producing an i.i.d. (independent identically distributed) sequence of uniform $[0,1]$ random variables is available. This model is of course unrealistic, but designing random variate generators based on it has several advantages: first of all, it allows one to disconnect the theory of non-uniform random variate generation from that of uniform random variate generation, and secondly, it permits one to plan for the future, as more powerful computers will be developed that permit ever better approximations of the model. Algorithms designed under finite approximation limitations will have to be redesigned when the next generation of computers arrives.

For the generation of discrete or integer-valued random variables, which includes the vast area of the generation of random combinatorial structures, one can adhere to a clean model, the pure bit model, in which each bit operation takes one time unit, and storage can be reported in terms of bits. Typically, one now assumes that an i.i.d. sequence of independent perfect bits is available. In this model, an elegant information-theoretic theory can be derived. For example, Knuth and Yao (1976) showed that to generate a random integer X described by the probability distribution

$$\mathbb{P}\{X = n\} = p_n, n \geq 1,$$

any method must use an expected number of bits greater than the binary entropy of the distribution,

$$\sum_n p_n \log_2(1/p_n).$$

They also showed how to construct tree-based generators that can be implemented as finite or infinite automata to come within three bits of this lower bound for any distribution. While this theory is elegant and theoretically important, it is somewhat impractical to have to worry about the individual bits in the binary expansions of the p_n's, so that we will, even for discrete distributions, consider only the RAM model. Noteworthy is that attempts have been made (see, e.g., Flajolet and Saheb, 1986) to extend the pure bit model to obtain approximate algorithms for random variables with densities.

1.1. The inversion method. For a univariate random variable, the inversion method is theoretically applicable: given the distribution function F, and its inverse F^{inv}, we generate a random variate X with that distribution as $F^{\text{inv}}(U)$, where U is a uniform $[0,1]$ random variable. This is the method of choice when the inverse is readily computable. For example, a standard exponential random variable (which has density $e^{-x}, x > 0$), can be generated as $\log(1/U)$. The following table gives some further examples.

Name	Density	Distribution function	Random variate
Exponential	$e^{-x}, x > 0$	$1 - e^{-x}$	$\log(1/U)$
Weibull (a), $a > 0$	$ax^{a-1}e^{-x^a}, x > 0$	$1 - e^{-x^a}$	$(\log(1/U))^{1/a}$
Gumbel	$e^{-x}e^{-e^{-x}}$	$e^{-e^{-x}}$	$-\log\log(1/U)$
Logistic	$\frac{1}{2+e^x+e^{-x}}$	$\frac{1}{1+e^{-x}}$	$-\log((1-U)/U)$
Cauchy	$\frac{1}{\pi(1+x^2)}$	$1/2 + (1/\pi)\arctan x$	$\tan(\pi U)$
Pareto (a), $a > 0$	$\frac{a}{x^{a+1}}$, $x > 1$	$1 - 1/x^a$	$1/U^{1/a}$

Table 1: Some densities with distribution functions that are explicitly invertible.

The fact that there is a monotone relationship between U and X has interesting benefits in the area of coupling and variance reduction. In simulation, one sometimes requires two random variates from the same distribution that are maximally anti-correlated. This can be achieved by generating the pair $(F^{\text{inv}}(U), F^{\text{inv}}(1-U))$. In coupling, one has two distribution functions F and G and a pair of (possibly dependent) random variates (X, Y) with these marginal distribution functions is needed so that some appropriate metric measuring distance between them is minimized. For example, the Wasserstein metric $d_2(F, G)$ is the minimal value over all couplings of (X, Y) of $\sqrt{\mathbb{E}\{(X-Y)^2\}}$. That minimal coupling occurs when $(X, Y) = (F^{\text{inv}}(U), G^{\text{inv}}(U))$ (see, e.g., Rachev, 1991). Finally, if we wish to simulate the maximum M of n i.i.d. random variables, each distributed as $X = F^{\text{inv}}(U)$, then noting that the maximum of n i.i.d. uniform $[0,1]$ random variables is distributed as $U^{1/n}$, we see that M can be simulated as $F^{\text{inv}}(U^{1/n})$.

1.2. Simple transformations. We call a simple transformation one that uses functions and operators that are routinely available in standard libraries, such as the trigonometric, exponential and logarithmic functions. For example, the inverse of the normal and stable distribution functions cannot be computed using simple transformations of one uniform random variate. For future reference, the standard normal density is given by $\exp(-x^2/2)/\sqrt{2\pi}$. However, the next step is to look at simple transformations of a k uniform $[0,1]$ random variates, where k is either a small fixed integer, or a random integer with a small mean. It is remarkable that one can obtain the normal and indeed all stable distributions using simple transformations with $k = 2$. In the Box-Müller method (Box and Müller, 1958), a pair of independent standard normal random variates is obtained by setting

$$(X, Y) = \left(\sqrt{\log(1/U_1)}\,\cos(2\pi U_2), \sqrt{\log(1/U_1)}\,\sin(2\pi U_2)\right),$$

where U_1, U_2 are independent uniform $[0,1]$ random variates. For the computational perfectionists, we note that the random cosine can be avoided: just generate a random point in the unit circle by rejection from the enclosing square (more about that later), and then normalize it so that it is of unit length. Its first component is distributed as a random cosine.

There are many other examples that involve the use of a random cosine, and for this reason, they are called polar methods. We recall that the beta (a, b) density is

$$\frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}, 0 \le x \le 1,$$

where $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a + b)$. A symmetric beta (a, a) random variate may be generated as

$$\frac{1}{2}\left(1 + \sqrt{1 - U_1^{\frac{2}{2a-1}}}\,\cos(2\pi U_2)\right)$$

(Ulrich, 1984), where $a \ge 1/2$. Devroye (1996) provided a recipe valid for all $a > 0$:

$$\frac{1}{2}\left(1 + \frac{S}{\sqrt{1 + \frac{1}{\left(U_1^{-\frac{1}{a}} - 1\right)\cos^2(2\pi U_2)}}}\right),$$

where S is a random sign. Perhaps the most striking result of this kind is due to Bailey (1994), who showed that

$$\sqrt{a\left(U_1^{-\frac{2}{a}} - 1\right)}\,\cos(2\pi U_2)$$

has the Student t density (invented by William S. Gosset in 1908) with parameter $a > 0$:

$$\frac{1}{\sqrt{a}B(a/2, 1/2)\left(1 + \frac{x^2}{a}\right)^{\frac{a+1}{2}}}, x \in \mathbb{R}.$$

Until Bailey's paper, only rather inconvenient rejection methods were available for the t density.

There are many random variables that can be represented as $\psi(U)E^\alpha$, where ψ is a function, U is uniform $[0,1]$, α is a real number, and E is an independent exponential random variable. These lead to simple algorithms for a host of useful yet tricky distributions. A random variable $S_{\alpha,\beta}$ with characteristic function

$$\varphi(t) = \exp\left(-|t|^\alpha - \frac{i\pi\beta(\alpha - 2\mathbb{1}_{\alpha>1})\,\mathrm{sign}(t)}{2}\right)$$

is said to be stable with parameters $\alpha \in (0, 2]$ and $|\beta| \le 1$. Its parameter α determines the size of its tail. Using integral representations of distribution functions, Kanter (1975) showed that for $\alpha < 1$, $S_{\alpha,1}$ is distributed as

$$\psi(U)E^{1-\frac{1}{\alpha}},$$

where

$$\psi(u) = \left(\frac{\sin(\alpha\pi u)}{\sin(\pi u)}\right)^{\frac{1}{\alpha}} \times \left(\frac{\sin((1-\alpha)\pi u)}{\sin(\alpha\pi u)}\right)^{\frac{1-\alpha}{\alpha}}.$$

For general α, β, Chambers, Mallows and Stuck (1976) showed that it suffices to generate it as

$$\psi(U - 1/2)E^{1-\frac{1}{\alpha}},$$

where

$$\psi(u) = \left(\frac{\cos(\pi((\alpha-1)u + \alpha\theta)/2)}{\cos(\pi u/2)}\right)^{\frac{1}{\alpha}} \times \left(\frac{\sin(\pi\alpha(u+\theta)/2)}{\cos(\pi((\alpha-1)u + \alpha\theta)/2)}\right).$$

Zolotarev (1959, 1966, 1981, 1986) has additional representations and a thorough discussion on these families of distributions. The paper by Devroye (1990) contains other examples with $k = 3$, including

$$S_{\alpha,0}E^{\frac{1}{\alpha}},$$

which has the so-called Linnik distribution (Linnik, 1962) with characteristic function

$$\varphi(t) = \frac{1}{1 + |t|^\alpha}, 0 < \alpha \le 2.$$

See also Kotz and Ostrovskii (1996). It also shows that

$$S_{\alpha,1}E^{\frac{1}{\alpha}},$$

has the Mittag-Leffler distribution with characteristic function

$$\varphi(t) = \frac{1}{1 + (-it)^\alpha}, 0 < \alpha \le 1.$$

Despite these successes, some distributions seem hard to treat with any k-simple transformation. For example, to date, we do not know how to generate a gamma random variate (i.e., a variate with density $x^{a-1}e^{-x}/\Gamma(a)$ on $(0, \infty)$) with arbitrary parameter $a > 0$ using only a fixed finite number k of uniform random variates and simple transformations.

1.3. Simple methods for multivariate distributions. For random vectors $(X_1, \ldots, X_d)$, the inversion method proceeds by peeling off the dimensions. That is, we first generate X_d from its marginal distribution, then X_{d-1} given X_d, using a conditional distribution, and so forth. There is of course no reason to employ conditioning in that rigid manner. For example, if a distribution has a density that is a function of $\sum_i x_i^2$ only, then it is advantageous to find the marginal of $\sum_i X_i^2$, generate a random variate from this marginal distribution, and then generate a random variate uniformly distributed on the surface of the ball with the given random radius. And to generate a uniformly distributed point on the surface of the unit ball, just generate d i.i.d. standard normal random variates $X_1, \ldots, X_d$, and normalize the length to one.

Sometimes, one wishes to generate random vectors with a certain dependence structure. This can be captured by various measures of correlation. For a survey of random variate generators in this vein, see Devroye (1986a). An interesting new area of research is related to the development of generators with a given copula structure. The copula of two uniform $[0,1]$ random variables X and Y is given by the joint distribution function

$$C(x,y) = \mathbb{P}\{X \le x, Y \le y\}.$$

We say that two arbitrary continuous random variables X and Y (with distribution functions F and G, respectively) have the copula structure given by C if $F(X)$ and $G(Y)$ (which are both uniform $[0,1]$) have joint distribution function C. Thus, we need methods for generating random vectors with distribution function C. Various interesting families of copulas and some seedlings of algorithms, are given in Nelsen's book (1999) and in Chapter 5 of the present volume.

1.4. Inversion for integer-valued random variables. For integer-valued random variables with $\mathbb{P}\{X = n\} = p_n, n \ge 0$, the inversion method is always applicable:

$X \leftarrow 0$
`Generate` U `uniform` $[0,1]$
$S \leftarrow p_0$ `(`S `holds the partial sums of the` p_n`'s)`
`while` $U > S$ `do`
    $X \leftarrow X + 1$`,` $S \leftarrow S + p_X$
`return` X

The expected number of steps here is $\mathbb{E}\{X+1\}$. Improvements are possible by using data structures that permit one to invert more quickly. When there are only a finite number of values, a binary search tree may help. Here the leaves correspond to various outcomes for X, and the internal nodes are there to guide the search by comparing U with appropriately picked thresholds. If the cost of setting up this tree is warranted, then one could always permute the leaves to make this into a Huffman tree for the weights p_n (Huffman, 1952), which insures that the expected time to find a leaf is not more than one plus the binary entropy,

$$\sum_n p_n \log_2(1/p_n).$$

In any case, this value does not exceed $\log_2 N$, where N is the number of possible values X can take.

Another way of organizing this is to store an array with the partial sums $p_1, p_1 + p_2, \ldots, p_1 + \cdots + p_N = 1$, and to search for the interval in which U falls by binary search, which would take $O(\log_2 N)$ time. As mentioned above, the Huffman tree is optimal, and thus better, but requires more work to set up.

1.5. Guide tables. Hash tables, or guide tables (Chen and Asau, 1974), have been used to accelerate the inversion even further. Given the size of the universe, N, we create a hash table with N entries, and store in the i-th entry the value of X if U were i/N, $0 \le i < N$. Then U gets "hashed" to $Z = \lfloor NU \rfloor$. Then return to the table of partial sums and start the search at $p_1 + \cdots + p_Z$. It is easy to show that the expected time, table set-up excluded, is bounded by a constant uniformly over all distributions on N values. The table can be constructed in $O(N)$ time.

Walker (1974, 1977) showed that one can construct in time $O(N)$ a table (q_i, r_i), $1 \le i \le N$, such that the following method works: pick a uniform integer Z from $1, 2, \ldots, N$. Generate a uniform $[0,1]$ random variate U. If $U \le q_Z$, then return Z, else return r_Z. The values r_i are called the aliases, and the method is now known as the alias method. If a distribution is fixed once and for all, and N is such that the storage of the table is of no concern, then the alias method is difficult to beat.

1.6. Mixture methods. Densities that can be written as mixtures

$$\sum_n p_n f_n(x),$$

with nonnegative weights p_n summing to one, are trivially dealt with by first generating a random index Z with discrete distribution $\{p_n\}$, and then generating a random variate with density f_Z. They do, therefore, not require any special treatment.

Sometimes, we have more intricate mixtures, when, e.g.,

$$f(x) = \mathbb{E}\{g_Z(x)\},$$

where Z is a random variable, and g_Z is a density in which Z is a parameter. Clearly, it suffices to generate Z first and then to generate a random variate from g_Z. The textbook example here is $N/\sqrt{G_a}$, where N is standard normal, and G_a is a gamma random variable with parameter a, independent of N. The ratio has the Student t distribution with parameter a. Other examples relevant for further on are U/V, a ratio of two i.i.d. uniform $[0,1]$ random variates, which has density $(1/2)\min(1, 1/x^2)$, and UV, which has density $\log(1/x)$ on $(0,1]$. Gamma random variates of parameter less than one are often cumbersome, but we know that $G_a \stackrel{\mathcal{L}}{=} U^{1/a} G_{a+1}$ where the notation is as above and all variates are independent on the right. Finally, a Cauchy random variate can be obtained as N_1/N_2, the ratio of two independent standard normal random variates.

Mixtures are also useful in discrete settings. For example, the negative binomial distribution (the number of failures before the n-th success) with parameters (n,p), $n \ge 1$, $p \in (0,1)$, is given by the probabilities

$$p_k = \binom{n+k-1}{k} (1-p)^k p^n, k \ge 0.$$

One can check that this can be generated as a Poisson (Y) random variate where Y in turn is $(1-p)/p$ times a gamma (n) random variate. A special case is the geometric (p) distribution, which is negative binomial $(1,p)$. Here, though, it is better to use the inversion method which can be made explicit by the truncation operator: it suffices to take

$$\lfloor \log_{1-p} U \rfloor,$$

where U is uniform $[0,1]$.

1.7. The rejection method. Von Neumann (1951) proposed the rejection method, which uses the notion of a dominating measure. Let X have density f on $\mathbb{R}^d$. Let g be another density with the property that for some finite constant $c \geq 1$, called the rejection constant,

$$f(x) \leq cg(x), x \in \mathbb{R}^d.$$

For any nonnegative integrable function h on $\mathbb{R}^d$, define the body of h as $B_h = \{(x,y) : x \in \mathbb{R}^d, 0 \leq y \leq h(x)\}$. Note that if (X,Y) is uniformly distributed on B_h, then X has density proportional to h. Vice versa, if X has density proportional to h, then $(X, Uh(X))$, where U is uniform $[0,1]$ and independent of X, is uniformly distributed on B_h. These facts can be used to show the validity of the rejection method:

```
repeat
  Generate U uniformly on [0,1]
  Generate X with density g
until Ucg(X) ≤ f(X)
return X
```

The expected number of iterations before halting is c, so the rejection constant must be kept small. This method requires some analytic work, notably to determine c, but one attractive feature is that we only need the ratio $f(x)/(cg(x))$, and thus, cumbersome normalization constants often cancel out.

1.8. Rejection: a simple example. The rejection principle also applies in the discrete setting, so a few examples follow to illustrate its use in all settings. We begin with the standard normal density. The start is an inequality such as

$$e^{-x^2/2} \leq e^{\alpha^2/2 - \alpha|x|}.$$

The area under the dominating curve is $e^{\alpha^2/2} \times 2/\alpha$, which is minimized for $\alpha = 1$. Generating a random variate with the Laplace density $e^{-|x|}$ can be done either as SE, where S is a random sign, and E is exponential, or as $E_1 - E_2$, a difference of two independent exponential random variables. The rejection algorithm thus reads:

```
repeat
  Generate U uniformly on [0,1]
  Generate X with with the Laplace density
until Ue^{1/2-|X|} ≤ e^{-X^2/2}
return X
```

However, taking logarithms in the last condition, and noting that $\log(1/U)$ is exponential, we can tighten the code using a random sign S, and two independent exponentials, E_1, E_2:

```
Generate a random sign S
repeat
  Generate E_1, E_2
until 2E_2 > (E_1 - 1)^2
return X ← SE_1
```

It is easy to verify that the rejection constant (the expected number of iterations) is $\sqrt{2e/\pi} \approx 1.35$.

1.9. Rejection: a more advanced example. Assume that f is a monotonically decreasing density on $[0,1]$, and define $M = f(0)$. We could apply rejection with M as a bound:

```
repeat
  Generate U uniformly on [0,1]
  Generate X uniformly on [0,1]
until UM ≤ f(X)
return X
```

The expected number of iterations is M. However, the algorithm does not make any use of the monotonicity. As we know that $f(x) \leq \min(M, 1/x)$, and the integral of the upper bound (or the rejection constant) is $1 + \log M$, we can exponentially speed up the performance by basing our rejection on the latter bound. It is a simple exercise to show that the indefinite integral of the upper bound is

$$\begin{cases} Mx & (0 \leq x \leq 1/M) \\ 1 + \log(Mx) & (1/M \leq x \leq 1). \end{cases}$$

A random variate can easily be generated by inversion of a uniform $[0,1]$ random variate U: if $U \leq 1/(1 + \log M)$, return $(1 + \log M)U/M$, else return $(1/M)\exp((1 + \log M)U - 1)$. Thus, we have the improved algorithm:

```
repeat
  Generate U,V independently and uniformly on [0,1].
  Set W = V(1 + log M).
  If W ≤ 1 then set X ← W/M
             else set X ← exp(WU − 1)/M
until U min(M, 1/X) ≤ f(X)
return X
```

It is possible to beat even this either by spending more time setting up (a one-time cost) or by using additional information on the density. In the former case, partition $[0,1]$ into k equal intervals, and invest once in the calculation of $f(i/k)$, $0 \le i < k$. Then the histogram function of height $f(i/k)$ on $[i/k, (i+1)/k)$ dominates f and can be used for rejection purposes. This requires the bin probabilities $p_i = f(i/k)/\sum_{j=0}^{k-1} f(j/k)$. Rejection can proceed as follows:

```
repeat
  Generate U,V independently and uniformly on [0,1]
  Generate Z on {0,1,...,k − 1} according to the probabilities {p_i}.
  Set X = (Z + V)/k.
until Uf(Z/k) ≤ f(X)
return X
```

Here the area under the dominating curve is at most one plus the area sandwiched between f and it. By shifting that area over to the left, bin by bin, one can bound it further by $1+M/k$. Thus, we can adjust k to reduce the rejection constant at will. For example, a good choice would be $k = \lceil M \rceil$. The expected time per variate would be $O(1)$ if we can generate the random integer Z in time $O(1)$. This can be done, but only at the expense of the set-up of a table of size k (see above). In one shot situations, this is a prohibitive expense. In Devroye (1986), methods based on partitioning the density in slabs are called strip methods. Early papers in this direction include Ahrens and Dieter (1989), Ahrens (1993, 1995), Marsaglia and Tsang (1984) and Marsaglia, Maclaren and Bray (1964).

Assume next that we also know the mean μ of f. Then, even dropping the requirement that f be supported on $[0,1]$, we have

$$\mu = \int_0^\infty y f(y) dy \ge f(x) \int_0^x y dy = f(x) x^2/2.$$

Hence,

$$f(x) \le \min\left(M, \frac{2\mu}{x^2}\right).$$

The integral of the upper bound from 0 to x is

$$\begin{cases} Mx & 0 \le x < \sqrt{2\mu/M} \\ 2\sqrt{2\mu M} - \frac{2\mu}{x} & x \ge \sqrt{2\mu/M}. \end{cases}$$

The area under the upper bound is therefore $2\sqrt{2\mu M}$. To generate a random variate with density proportional to the upper bound, we can proceed in various ways, but it is easy to check that $\sqrt{2\mu/M} U_1/U_2$ fits the bill, where U_1, U_2 are independent uniform $[0,1]$ random variables. Thus, we summarize:

```
repeat
  Generate U, V, W independently and uniformly on [0,1].
  Set X = sqrt(2μ/M) V/W.
until U M min(1, (W/V)^2) ≤ f(X)
return X
```

This example illustrates the development of universal generators, valid for large classes of densities, as a function of just a few parameters. The reader will be able to generalize the above example to unimodal densities with known r-th moment about the mode. The example given above applies with trivial modifications to all unimodal densities, and the expected time is bounded by a constant times $\sqrt{f(m)\mathbb{E}\{|X-m|\}}$, where m is the mode. For example, the gamma density with parameter $a \geq 1$ has mean a, and mode at $a-1$, with $f(a-1) \leq C/\sqrt{a}$ for some constant C, and $\mathbb{E}\{|X-(a-1)|\} \leq C'\sqrt{a}$ for another constant C'. Thus, the expected time taken by our universal algorithm takes expected time bounded by a constant, uniformly over $a \geq 1$. We say that the algorithm is uniformly fast over this class of densities.

1.10. The alternating series method. To apply the rejection method, we do not really need to know the ratio $f(x)/(cg(x))$ exactly. It suffices that we have an approximation $\phi_n(x)$ that tends, as $n \to \infty$, to $f(x)/(cg(x))$, and for which we know a monotone error bound $\epsilon_n(x) \downarrow 0$. In that case, we let n increase until for the first time, either

$$U \leq \phi_n(X) - \epsilon_n(X)$$

(in which case we accept X), or

$$U \geq \phi_n(X) + \epsilon_n(X)$$

(in which case we reject X). Equivalently, we have computable bounds $\xi_n(x)$ and $\psi_n(x)$ with the property that $\xi_n(x) \uparrow f(x)/(cg(x))$ and $\psi_n(x) \downarrow f(x)/(cg(x))$ as $n \to \infty$. This approach is useful when the precise computation of f is impossible, e.g., when f is known as infinite series or when f can never be computed exactly using only finitely many resources. It was first developed for the Kolmogorov-Smirnov limit distribution in Devroye (1981a). For another use of this idea, see Keane and O'Brien's Bernoulli factory (1994).

```
repeat
  Generate U uniformly on [0,1]
  Generate X with density g
  Set n = 0
  repeat n ← n + 1 until U ≤ ξ_n(X) or U ≥ ψ_n(X)
until U ≤ ξ_n(X)
return X
```

The expected number of iterations in the outer loop is still c, as in the rejection method. However, to take the inner loop into account, let N be the largest index n attained in the inner loop. Note that $N > t$ implies that $U \in [\xi_t(X), \psi_t(X)]$. Thus,

$$\mathbb{E}\{N|X\} = \sum_{t=0}^{\infty} \mathbb{P}\{N > t|X\} \leq \sum_{t=0}^{\infty} (\psi_t(X) - \xi_t(X)).$$

Unconditioning,

$$\mathbb{E}\{N\} \leq \sum_{t=0}^{\infty} \mathbb{E}\{\psi_t(X) - \xi_t(X)\}.$$

To be finite, this requires a rate of decrease in the approximation error that is at least $1/t$. One must also take into account that the computation of ξ_t and ψ_t may also grow with t. The bound above is indicative of the expected complexity only when, say, ξ_t can be computed from ξ_{t-1} in one time unit.

We cannot stress strongly enough how important the alternating series method is, as it frees us from having to compute f exactly. It is indeed the key to the solution of a host of difficult non-uniform random variate generation problems.

2. Uniformly bounded times

If $\mathcal{F}$ is a class of distributions, it is useful to have generators for which the expected time is uniformly bounded over $\mathcal{F}$. In the case of the rejection method, this often (but not always, see, e.g., the alternating series method) means bounding the rejection constant uniformly. For example, pre-1970 papers routinely recommended generating gamma (k) random variables as a sum of k independent exponentials, and generating a binomial (n, p) random variate as a sum of n indepoendent Bernoulli (p) random variables. Even today, many may still simulate a sum of n i.i.d. random variables by generating all n of them and summing. So, whether $\mathcal{F}$ is the class of all gamma distributions, all binomial distributions, or all sums of distributions with a fixed density f, the idea of a uniformly bounded time is important, and has, in fact, given new life to the area of non-uniform random variate generation. In this section, we briefly survey some key classes $\mathcal{F}$, and provide some references to good generators.

Many methods in the literature are rejection methods in disguise. For example, for real-valued random variates, the ratio of uniforms method is based upon the use of bounding curves of the form

$$\min\left(A, \frac{B}{x^2}\right).$$

The area under the bounding curve is $4\sqrt{AB}$, and a random variate with density proportional to it can be generated as $\sqrt{B/A}\,SU/V$, where S is a random sign, and U, V are independent uniform $[0,1]$ random variates. The method is best applied after shifting the mode or mean of the density f to the origin. One could for example use

$$A = \sup_x f(x), B = \sup_x x^2 f(x).$$

In many cases, such as for mode-centered normals, gamma, beta, and Student t distributions, the rejection constant $4\sqrt{AB}$ is uniformly bounded over all or a large part of the parameter space. Discretized versions of it can be used for the Poisson, binomial and hypergeometric distributions.

Uniformly fast generators have been obtained for all major parametric families. Interesting contributions include the gamma generators of Cheng (1977), Le Minh (1988), Marsaglia (1977), Ahrens and Dieter (1982), Cheng and Feast (1979, 1980), Schmeiser and Lal (1980), Ahrens, Kohrt and Dieter (1983), and Best (1983), the beta generators of Ahrens and Dieter (1974), Zechner and Stadlober (1993), Best (1978), and Schmeiser and Babu (1980), the binomial methods of Stadlober (1988, 1989), Ahrens and Dieter (1980), Hörmann (1993), and Kachitvichyanukul and Schmeiser (1988, 1989), the hypergeometric generators of Stadlober (1988, 1990), and the code for Poisson variates by Ahrens and Dieter (1980), and Hörmann (1993). Some of these algorithms are described in Devroye (1986a), where some additional uniformly fast methods can be found.

All the distributions mentioned above are log-concave, and it is thus no surprise that we can find uniformly fast generators. The emphasis in most papers is on the details, to squeeze the last millisecond out of each piece of code. To conclude this section, we will just describe a recent gamma generator, due to Marsaglia and Tsang (2001). It uses almost-exact inversion. Its derivation is unlike anything in the present survey, and shows the careful planning that goes into the design of a nearly perfect generator. The idea is to consider a monotone transformation $h(x)$, and to note that if X has density

$$e^{g(x)} \stackrel{\text{def}}{=} \frac{h(x)^{a-1} e^{-h(x)} h'(x)}{\Gamma(a)},$$

then $Y = h(X)$ has the gamma (a) density. If $g(x)$ is close to $c - x^2/2$ for some constant c, then a rejection method from a normal density can be developed. The choice suggested by Marsaglia and Tsang is

$$h(x) = d(1+cx)^3 \ , \ -1/c < x < \infty,$$

with $d = a - 1/3$, $c = 1/\sqrt{9d}$. In that case, simple computations show that

$$g(x) = 3d \log(1+cx) - d(1+cx)^3 + d + C \ , \ -1/c < x < \infty,$$

for some constant C. Define $w(x) = -x^2/2 - g(x)$. Note that $w'(x) = -x - \frac{3dc}{1+cx} + 3dc(1+cx)^2$, $w''(x) = -1 + \frac{3dc^2}{(1+cx)^2} + 6dc^2(1+cx)$, $w'''(x) = -\frac{6dc^3}{(1+cx)^3} + 6dc^3$. The third derivative is zero at $(1+cx)^3 = 1$, or $x = 0$. Thus, w'' reaches a minimum there, which is $-1 + 9dc^2 = 0$. Therefore, by Taylor's series expansion, and the fact that $w(0) = -C$, $w'(0) = 0$, we have $w(x) \ge C$. Hence,

$$e^{g(x)} \le e^{C - x^2/2}$$

and the rejection method immediately yields the following:

```
d ← a − 1/3
c ← 1/√(9d)
repeat
    generate E exponential
    generate X normal
    Y ← d(1 + cX)^3
until −X^2/2 − E ≤ d log(Y) − Y + d
return Y
```

We picked d so that $\int e^{g-C}$ is nearly maximal. By the transformation $y = d(1+cx)^3$ and by Stirling's approximation, as $d \to \infty$,

$$\int_{-1/c}^{\infty} e^{g(x)-C} dx = \int_0^\infty \frac{y^{d-2/3}e^{d-y}}{3cd^{d+1/3}} dy = \frac{\Gamma(d+1/3)e^d}{3cd^{d+1/3}}$$
$$\sim \frac{(1+1/3d)^{d+1/3}\sqrt{2\pi}}{3ce^{1/3}\sqrt{d+1/3}} \sim \frac{\sqrt{18\pi d}}{3\sqrt{d+1/3}} \sim \sqrt{2\pi}.$$

Using $\int e^{-x^2/2}\, dx = \sqrt{2\pi}$, we note that the asymptotic (as $a \to \infty$) rejection constant is 1: we have a perfect fit! Marsaglia and Tsang recommend the method only for $a \geq 1$. For additional speed, a squeeze step (or quick acceptance step) may be added.

3. Universal generators

It is quite important to develop generators that work well for entire families of distributions. An example of such development is the generator for log concave densities given in Devroye (1984a), that is, densities for which $\log f$ is concave. On the real line, this has several consequences: these distributions are unimodal and have sub-exponential tails. Among the log concave densities, we find the normal densities, the gamma family with parameter $a \geq 1$, the Weibull family with parameter $a \geq 1$, the beta family with both parameters ≥ 1, the exponential power distribution with parameter ≥ 1 (the density being proportional to $\exp(-|x|^a)$), the Gumbel or double exponential distribution ($k^k \exp(-kx - ke^{-x})/(k-1)!$ is the density, $k \geq 1$ is the parameter), the generalized inverse Gaussian distribution, the logistic distribution, Perks' distribution (with density proportional to $1/(e^x + e^{-x} + a)$, $a > -2$), the hyperbolic secant distribution (with distribution function $(2/\pi)\arctan(e^x)$), and Kummer's distribution (for a definition, see Devroye, 1984a). Also, we should consider simple nonlinear transformations of other random variables. Examples include $\arctan X$ with X Pearson IV, $\log|X|$ (X being Student t for all parameters), $\log X$ (for X gamma, all parameters), $\log X$ (for X log-normal), $\log X$ (for X Pareto), $\log(X/(1-X))$ (for all beta variates X), and $\log X$ (for X beta (a,b) for all $a > 0, b \geq 1$).

In its most primitive version, assuming one knows that a mode occurs at m, we consider the generation problem for the normalized random variable $Y = f(m)(X - m)$. One may obtain X as $m + Y/f(m)$. The new random variable has a mode at 0 of value 1. Call its density g. Devroye (1984a) showed that

$$g(y) \leq \min\left(1, e^{1-|y|}\right).$$

The integral of the upper bound is precisely 4. Thus, the rejection constant will be 4, uniformly over this vast class of densities. A random variable with density proportional to the upper bound is generated by first flipping a perfect coin. If it is heads, then generate $(1+E)S$, where E is exponential, and S is another perfect coin. Otherwise generate US, where U is uniform $[0,1]$ and independent of S. The algorithm is as follows:

```
repeat
  Generate U (a uniform on [0,1]), E (exponential), and S (a fair random bit).
  Generate a random sign S'.
  If S = 0 then Y ← 1 + E
          else Y ← V, V uniform [0,1]
  Set Y ← YS'
until U min(1, e^{1-|Y|}) ≤ g(Y)
return Y
```

Various improvements can be found in the original paper. In adaptive rejection sampling, Gilks and Wild (1992) bound the density adaptively from above and below by a number of exponential pieces, to reduce the integral under the bounding curve. If we denote the log concave density by f, and if f' is available, then for all x, log-concavity implies that

$$\log f(y) \leq \log f(x) + (y-x)\frac{f'(x)}{f(x)}, y \in \mathbb{R}.$$

This is equivalent to

$$f(y) \leq f(x)e^{(y-x)f'(x)/f(x)}, y \in \mathbb{R}.$$

Assume that we have already computed f at k points. Then we can construct $k+1$ exponential pieces (with breakpoints at those k points) that bound the density from above, and in fact, sampling from the piecewise exponential dominating curve is easily carried out by inversion. The integral under a piece anchored at x and spanning $[x, x+\delta)$ is

$$\frac{f^2(x)\left(e^{\delta f'(x)/f(x)}-1\right)}{f'(x)}.$$

Gilks and Wild also bound f from below. Assume that $y \in [x, x']$, where x and x' are two of those anchor points. Then

$$\log(f(y)) \geq \log(f(x)) + \frac{y-x}{x'-x}\left(\log(f(x')) - \log(f(x))\right).$$

The lower bound can be used to avoid the evaluation of f most of the time when accepting or rejecting. It is only when f is actually computed that a new point is added to the collection to increase the number of exponential pieces.

Hörmann (1995) shows that it is best to add two exponential tails that touch g at the places where $g(x) = 1/e$, assuming that the derivative of g is available. In that case, the area under the bounding curve is reduced to $e/(e-1) \approx 1.582$. Hörmann and Derflinger (1994) and Hörmann (1995) (see also Leydold (2000a, 2000b, 2001) and Hörmann, Leydold and Derflinger (2004)) define the notion of T_c-concavity, related to the concavity of $-1/(f(x))^c$, where $0 < c < 1$. For example, the Student t distribution of parameter ≥ 1 is $T_{1/2}$-concave. Similar universal methods related to the one above are developed by them. Rejection is done from a curve with polynomial tails, and the rejection constant is $\left(1-(1-c)^{1/c-1}\right)^{-1}$. The notion of T-convexity proposed by Evans and Swartz (1996, 2000) is useful for unbounded densities.

Discrete distributions are called log concave if for all n,

$$p_n^2 \geq p_{n-1} p_{n+1}.$$

Some examples are shown in the table below.

Name	Probability vector p_k	Parameter(s)
Binomial (n,p)	$\binom{n}{k} p^k (1-p)^{n-k}$ $(0 \leq k \leq n)$	$n \geq 1$, $0 \leq p \leq 1$
Poisson (λ)	$\frac{\lambda^k e^{-\lambda}}{k!}$ $(k \geq 0)$	$\lambda > 0$
Negative binomial (n,p)	$\binom{n+k-1}{n-1}(1-p)^k p^n$ $(k \geq 0)$	$n \geq 1$, $0 \leq p \leq 1$
Geometric (p)	$p(1-p)^k$ $(k \geq 0)$	$0 \leq p \leq 1$
Logarithmic series (p)	$\frac{p^k}{k! \log(1/(1-p))}$ $(k \geq 1)$	$0 < p < 1$
Hypergeometric (b,w,n)	$\frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$ $(0 \leq k \leq \min(w,n))$	$n \geq 1$, $b \geq 0$, $w \geq 0$

Table 2: Some important discrete distributions and their parameter ranges.

Others not shown here included the Pólya-Eggenberger distribution (see Johnson and Kotz, 1969) and the hyper-Poisson distribution. For all discrete log-concave distributions having a mode at m, we have (Devroye, 1987a):

$$p_{m+k} \leq p_m \min\left(1, e^{1-p_m|k|}\right), \text{all } k.$$

Mimicking the development for densities, we may use rejection from a curve that is flat in the middle, and has two exponential tails. After taking care of rounding to the nearest integer, we obtain the following code, in which the expected number of iterations before halting is $4+p_m \leq 5$ for all distributions in the family.

```
Compute w ← 1 + p_m/2 (once)
repeat
  Generate U, V, W uniformly on [0,1], and let S be a random sign.
  If U ≤ w/(1+w) then Y ← Vw/p_m
                 else Y ← (w − log V)/p_m
  X ← S round(Y)
until W min(1, e^{w − p_m Y}) ≤ p_{m+X}/p_m
return m + X
```

We can tighten the rejection constant further to $2+p_m$ for log concave distributions that are monotonically decreasing. The disadvantage in the code is the requirement that p_m must be computed.

The development for log-concave distributions can be aped for other classes. For example, the adaptive rejection method of Gilks and Wild can be used for unimodal densities with known mode m. One just needs to replace the exponential upper and lower bounds by histograms. This is covered in an exercise in Devroye (1986a). General algorithms are known for all Lipschitz densities with known bounds on the Lipschitz constant C and variance σ^2, and all unimodal densities with mode at m, and variance bounded by σ^2, to give two examples (see Devroye, 1984b, 1986a).

4. Indirect problems

A certain amount of work is devoted to indirect problems, in which distributions are not described by their density, distribution function or discrete probability vector. A partial list follows, with brief descriptions on how the solutions may be constructed.

4.1. Characteristic functions. If the characteristic function

$$\varphi(t) = \mathbb{E}\left\{e^{itX}\right\}$$

is known in black box format, very little can be done in a universal manner. In particular cases, we may have enough information to be able to deduce that a density exists and that it is bounded in a certain way. For example, from the inversion formula

$$f(x) = \frac{1}{2\pi}\int \varphi(t)e^{-itx}dt,$$

we deduce the bounds

$$\sup_x f(x) \le M \stackrel{\text{def}}{=} \frac{1}{2\pi}\int |\varphi(t)|dt$$

and

$$\sup_x x^2 f(x) \le M' \stackrel{\text{def}}{=} \frac{1}{2\pi}\int |\varphi''(t)|dt$$

Thus, we have

$$f(x) \le \min(M, M'/x^2),$$

a bound that is easily dealt with by rejection. In principle, then, rejection can be used, with $f(x)$ approximated at will by integral approximations based on the inversion integral. This method requires bounds on the integral approximation, and these depend in turn on the smoothness of φ. Explicit bounds on some smoothness parameters must be available. There are descriptions and worked out examples in Devroye (1981b, 1986b, 1988). Devroye (1988) uses this method to simulate the sum S_n of n i.i.d. random variables with common characteristic function φ in expected time not depending upon n, as S_n has characteristic function φ^n.

Pólya showed that any convex function φ on the positive halfline that decreases from 1 to 0 is the characteristic function of a symmetric random variable if we extend it to the real line by setting $\varphi(-t) = \varphi(t)$. These distributions correspond to random variables distributed as Y/Z, where Y and Z are independent, Y has the de la Vallée-Poussin density

$$\frac{1}{2\pi}\left(\frac{\sin(x/2)}{x/2}\right)^2, x \in \mathbb{R}$$

(for which random variate generation by rejection is trivial), and Z has distribution function on the positive halfline given by $1 - \varphi(t) + t\varphi'(t)$. See Dugué and Girault (1955) for this property and Devroye (1984a) for its implications in random variate generation. Note that we have made a space warp from the complex to the real domain. This has unexpected corollaries. For example, if E_1, E_2 are independent exponential random variables, U is a uniform $[0, 1]$ random variable, and $\alpha \in (0, 1]$, then Y/Z with

$$Z = (E_1 + E_2 \mathbb{1}_{U<\alpha})^{\frac{1}{\alpha}}$$

is symmetric stable of parameter (α), with characteristic function $e^{-|t|^\alpha}$. The Linnik-Laha distribution (Laha, 1961) with parameter $\alpha \in (0, 1]$ is described by

$$\varphi(t) = \frac{1}{1 + |t|^\alpha}.$$

Here Y/Z with

$$Z = \left(\frac{\alpha + 1 - \sqrt{(\alpha + 1)^2 - 4\alpha U}}{2U} - 1\right)^{\frac{1}{\alpha}}$$

has the desired distribution.

4.2. Fourier coefficients. Assume that the density f, suitably scaled and translated, is supported on $[-\pi, \pi]$. Then the Fourier coefficients are given by

$$a_k = \mathbb{E}\left\{\frac{\cos(kX)}{\pi}\right\}, b_k = \mathbb{E}\left\{\frac{\sin(kX)}{\pi}\right\}, k \geq 0.$$

They uniquely describe the distribution. If the series absolutely converges,

$$\sum_k |a_k| + |b_k| < \infty,$$

then the following trigonometric series is absolutely and uniformly convergent to f:

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos(kx) + b_k \sin(kx)).$$

If we only use terms with index up to n, then the error made is not more than

$$R_{n+1} = \sum_{k=n+1}^{\infty} \sqrt{a_k^2 + b_k^2}.$$

If we know bounds on R_n, then one can use rejection with an alternating series method to generate random variates (Devroye, 1989).

A particularly simple situation occurs when only the cosine coefficients are nonzero, as occurs for symmetric distributions. Assume furthermore that the Fourier cosine series coefficients a_k are convex and $a_k \downarrow 0$ as $k \to \infty$. Even without absolute convergence, the Fourier cosine series converges, and can in fact be rewritten as follows:

$$f(x) = \sum_{k=0}^{\infty} \pi(k+1)\Delta^2 a_k K_k(x),$$

where $\Delta^2 a_k = a_{k+2} - 2a_{k+1} + a_k$ (a positive number, by convexity), and

$$K_k(x) = \frac{1}{2\pi(k+1)} \left(\frac{\sin((k+1)x/2)}{\sin(x/2)} \right)^2, |x| \leq \pi,$$

is the Fejer kernel. As $K_k(x) \leq \min((k+1)/4, \pi/(2(k+1)x^2)$, random variates from it can be generated in expected time uniformly bounded in k by rejection. Thus, the following simple algorithm does the job:

```
Generate U uniformly on [0,1].
Z ← 0, S ← πΔ²a₀.
while U > S do
  Z ← Z + 1.
  S ← S + π(Z + 1)Δ²a_Z.
Generate X with Fejer density K_Z.
return X
```

4.3. The moments are known. Let μ_n denote the n-th moment of a random variable X. One could ask to generate a random variate when for each n, μ_n is computable (in a black box). Again, this is a nearly impossible problem to solve, unless additional information is available. For one thing, there may not be a unique solution. A sufficient condition that guarantees the uniqueness is Carleman's condition

$$\sum_{n=0}^{\infty} |\mu_{2n}|^{-\frac{1}{2n}} = \infty$$

(see Akhiezer, 1965). Sufficient conditions in terms of a density f exist, such as Krein's condition

$$\int \frac{-\log(f(x))}{1+x^2}\, dx = \infty,$$

combined with Lin's condition, applicable to symmetric and differentiable densities, which states that $x|f'(x)|/f(x) \uparrow \infty$ as $x \to \infty$ (Lin, 1997, Krein, 1944, Stoyanov, 2000). Whenever the distribution is of compact support, the moments determine the distribution. In fact, there are various ways for reconstructing the density from the moments. An example is provided by the series

$$f(x) = \sum_{k=0}^{\infty} a_k \phi_j(x), |x| \le 1,$$

where ϕ_k is the Legendre polynomial of degree k, and a_k is a linear function of all moments up to the k-th. The series truncated at index n has an error not exceeding

$$\frac{C_r \int |f^{(r+1)}|}{(1-x^2)^{1/4} n^{r-1/2}}$$

where r is an integer and C_r is a constant depending upon r only (Jackson, 1930). Rejection with dominating curve of the form $C'/(1-x^2)^{1/4}$ (a symmetric beta) and with the alternating series method can thus lead to a generator (Devroye, 1989).

The situation is much more interesting when X is supported on the positive integers. Let

$$M_r = \mathbb{E}\{X(X-1)\cdots(X-r+1)\}$$

denote the r-th factorial moment. Then, approximate $p_j = \mathbb{P}\{X = j\}$ by

$$p_{nj} = \frac{1}{j!} \sum_{i=0}^{n} (-1)^i \frac{M_{j+i}}{i!}.$$

Note that $p_{nj} \ge p_j$ for n even, $p_{nj} \le p_j$ for n odd, and $p_{nj} \to p_j$ as $n \to \infty$ provided that $\mathbb{E}\{(1+u)^X\} < \infty$ for some $u > 0$. Under this condition, an alternating series rejection method can be implemented, with a dominating curve suggested by the crude bound

$$p_j \le \min\left(1, \frac{\mu_r}{j^r}\right).$$

In fact, we may even attempt inversion, as the partial sums $p_0 + p_1 + \cdots + p_j$ are available with any desired accuracy: it suffices to increase n until we are absolutely sure in which interval a given uniform random variate U lies. For more details, see Devroye (1991).

4.4. The moment generating function. For random variables on the positive integers, as in the previous section, we may have knowledge of the moment generating function

$$k(s) = p_0 + p_1 s + p_2 s^2 + \cdots = \mathbb{E}\left\{s^X\right\}.$$

Here we have trivially,

$$p_j \le \frac{k(s)}{s^j},$$

where $s > 0$ can be picked at will. Approximations for p_j can be constructed based on differences of higher orders. For example, if we take $t > 0$ arbitrary, then we can define

$$p_{nj} = \frac{\sum_{i=0}^{j} (-1)^{j-i} \binom{j}{i} k(it)}{j! t^j}$$

and note that

$$0 \le p_{nj} - p_j \le \frac{1}{(1-jt)^{j+1}} - 1.$$

This is sufficient to apply the alternating series method (Devroye, 1991).

4.5. An infinite series. Some densities are known as infinite series. Examples include the theta distribution (Rényi and Szekeres, 1967) with distribution function

$$F(x) = \sum_{j=-\infty}^{\infty} (1 - 2j^2x^2)e^{-j^2x^2} = \frac{4\pi^{5/2}}{x^3} \sum_{j=1}^{\infty} j^2 e^{-\pi^2 j^2/x^2}, x > 0,$$

and the Kolmogorov-Smirnov distribution (Feller, 1948) with distribution function

$$F(x) = 1 - 2\sum_{j=1}^{\infty} (-1)^j e^{-2j^2x^2}, x > 0.$$

In the examples above, it is relatively straightforward to find tight bounding curves, and to apply the alternating series method (Devroye, 1981a and 1997).

4.6. Hazard rates. Let X be a positive random variable with density f and distribution function F. Then the hazard rate, the probability of instantaneous death given that one is still alive, is given by

$$h(x) = \frac{f(x)}{1 - F(x)}.$$

Assume that one is given h (note that $h \ge 0$ must integrate to ∞). Then f can be recovered as follows:

$$f(x) = h(x)e^{-\int_0^x h(y)dy}.$$

This may pose a problem for random variate generation. Various algorithms for simulation in the presence of hazard rates are surveyed by Devroye (1986c), the most useful among which is the thinning method of Lewis and Shedler (1979): assume that $h \le g$, where g is another hazard rate. If $0 < Y_1 < Y_2 < \cdots$ is a nonhomogeneous Poisson point process with rate function g, and $U_1, U_2, \ldots$ is a sequence of independent uniform $[0,1]$ random variables, independent of the Y_i's, and if i is the smallest index for which $U_i g(Y_i) \le f(Y_i)$, then $X = Y_i$ has hazard rate h.

This is not a circular argument if we know how to generate a nonhomogeneous Poisson point process with rate function g. Define the cumulative rate function $G(x) = \int_0^x g(y)dy$. Let $E_1, E_2, \ldots$ be i.i.d. exponential random variables with partial sums $S_i = \sum_{j=1}^i E_j$. Then the sequence $Y_i = G^{\mathrm{inv}}(S_i)$, $i \ge 1$, is a nonhomogeneous Poisson point process with rate g. Thus, the thinning method is a Poisson process version of the rejection method.

If $h(0)$ is finite, and X has the given hazard rate, then for a DHR (decreasing hazard rate) distribution, we can use $g = h(0)$. In that case, the expected number of iterations before halting is $\mathbb{E}\{h(0)X\}$. However, we can dynamically thin (Devroye, 1986c) by lowering g as values $h(Y_i)$ trickle in. In that case, the expected time is finite when X has a finite logarithmic moment, and in any case, it is not more than $4 + \sqrt{24\mathbb{E}\{h(0)X\}}$.

4.7. A distributional identity. The theory of fixed points and contractions can be used to derive many limit laws in probability (see, e.g., Rösler and Rüschendorf, 2001). These often are described as distributional identities known to have a unique solution. For example, the identity

$$X \stackrel{\mathcal{L}}{=} W(X+1)$$

where W is a fixed random variable on $[0,1]$ and $X \geq 0$ sometimes has a unique solution for X. By iterating the identity, we see that if the solution exists, it can be represented as

$$X \stackrel{\mathcal{L}}{=} W_1 + W_1 W_2 + W_1 W_2 W_3 + \cdots,$$

where the W_i's are i.i.d. These are known as perpetuities (Vervaat, 1979, Goldie and Grübel, 1996). Some more work in the complex domain can help out: for example, when $W = U^{1/\alpha}$, where U is uniform $[0,1]$ and $\alpha > 0$ leads to the characteristic function

$$\varphi(t) = e^{\alpha \int_0^1 \frac{e^{itx}-1}{x}\, dx}.$$

For the particular case $\alpha = 1$, we have Dickman's distribution (Dickman, 1930), which is the limit law of

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

where Z_i is Bernoulli $(1/i)$. None of the three representations above (infinite series, Fourier transform, limit of a discrete sum) gives a satisfactory path to an exact algorithm (although approximations are trivial). The solution in Devroye (2001) uses the characteristic function representation in an alternating series method, after first determining explicit bounds on the density of f that can be used in the rejection method. A simpler approach to this particular distributional identity is still lacking, and a good exact solution for distributional identities in general is sorely needed.

Consider distributional identities of the form

$$X \stackrel{\mathcal{L}}{=} f_1(U)X + f_2(U)X' + f_3(U)$$

where the f_i's are known functions, U is a uniform random variate, and X, X' are i.i.d. and independent of U. Exact generation here is approached by discretization in Devroye and Neininger (2002): suitably discretize the distributions of $f_i(U)$, and apply the map n times, starting with arbitrary values (say, zero) for X, X'. If X_n is the random variable thus obtained, and F_n is its distribution function, note that F_n can be calculated as a sum, due to the discretization. Some theoretical work is needed (which involves parameters of the f_i's) to obtain inequalities of the type

$$\left| \frac{F_n(x+\delta) - F_n(x)}{\delta} - f(x) \right| \leq R_n(x),$$

where $R_n(x)$ is explicitly known and can be made as small as desired by appropriate choice of all the parameters (the discretization, δ and n). This suffices to apply the alternating series method, yet again.

4.8. Kolmogorov and Lévy measures. Infinitely divisible distributions (Sato, 2000) have several representations. One of them is Kolmogorov's canonical representation for the logarithm of the characteristic function:

$$\log \varphi(t) = ict + \int \frac{e^{itx} - 1 - itx}{x^2} \, dK(x),$$

where $K(-\infty) = 0$, $K \uparrow$, and $K(\infty) - K(-\infty) = \sigma^2$. Thus, K can be viewed as some sort of measure. When K puts mass σ^2 at the origin, then we obtain the normal distribution. When K puts mass λ at 1, then the Poisson (λ) distribution is obtained. However, in general, we only have representations as integrals. For example, with $c = 1$, $K(x) = \min(x^2/2, 1/2)$ on the positive halfline, we obtain Dickman's distribution,

$$\log \varphi(t) = \int_0^1 \frac{e^{itx} - 1}{x} \, dx.$$

Bondesson (1982) presents approximations when one is given the measure K. Basically, if K is approximated by n point masses each of total weight σ^2/n, then the characteristic function can be written as a product, where each term in the product, assuming $c = 0$, is of the form

$$e^{\alpha(e^{it\beta} - 1 - it\beta)}$$

for fixed values of α and β. This is the characteristic function of $\beta(X - \alpha)$ where X is Poisson (α). Thus, approximations consist of sums of such random variables. However, if the true distribution has a density, such discrete approximations are not acceptable.

5. Random processes

A random process $X_t, t \in \mathbb{R}^d$, can not be produced using a finite amount of resources, as the domain of the index is uncountable. So, it is customary to say that we can generate a random process exactly if for any k, and for any set of indices $t_1, \ldots, t_k$, we can generate the random vector $(X_{t_1}, \ldots, X_{t_k})$ exactly. Often, these points are nicely spaced out. In many processes, we have particularly elegant independence properties. For example, for all Lévy processes and setting $d = 1$, we have $X_0 = 0$, $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_k} - X_{t_{k-1}}$ independent, and furthermore, less important to us, the distribution of $X_{s+t} - X_s$ depends upon t only and we have continuity: $\mathbb{P}\{|X_{s+t} - X_s| > \epsilon\} \to 0$ as $t \downarrow 0$ for all $\epsilon > 0$. With that independence alone, we can thus generate each piece independently. When k is large, to protect against a proliferation of errors, it is wise to use a dyadic strategy: first generate the process at $t_{k/2}$, then at $t_{k/4}$ and $t_{3k/4}$, and so on. The dyadic trick requires that we know the distribution of $X_{s+u} - X_s$ conditional on $X_{s+t} - X_s$ for all $0 \le u \le t$.

An example suffices to drive home this point. In Brownian motion $(S(t), t \ge 0)$ (a Lévy process), we have differences $X_{s+t} - X_s$ that are distributed as $\sqrt{t}N$, where N is a standard normal random variable. Furthermore, setting $u = t/2$, we see that

$$X_{s+t} - X_s = (X_{s+t} - X_{s+u}) + (X_{s+u} - X_s)$$

is a sum of two independent normal random variables with variances $t/2$ each. Conditional on

$X_{s+t} - X_s$, $X_{s+u} - X_s$ thus is distributed as

$$\frac{X_{s+t} - X_s}{2} + \sqrt{\frac{t}{2}} N.$$

Therefore, the dyadic trick applies beautifully. In the same manner, one can simulate the Brownian bridge on $[0,1]$, $B(t) = S(t) - tS(1)$. For the first systematic examples of the use of this splitting trick, see Caflish, Morokoff and Owen (1997), and Fox (1999).

In a more advanced example, consider a gamma process $S(t)$, where $S(t)$ is distributed as a gamma random variable with parameter αt. This is not a Lévy process, but rather a monotonically increasing process. We have the following recursive property: given $S(t)$, $S(t/2)/S(t)$ is distributed as a symmetric beta with parameter $\alpha t/2$. This allows one to apply the dyadic trick.

Perhaps the most basic of all processes is the Poisson point process. These processes are completely described by an increasing sequence of occurrences $0 < X_1 < X_2, \cdots$. In a homogeneous point process of density λ, the inter-occurrence distances are i.i.d. and distributed as exponentials of mean $1/\lambda$. In a nonhomegenoeous Poisson point process of density $\lambda(t)$, a nonnegative function, the integrals $\int_{X_i}^{X_{i+1}} \lambda(t)\,dt$ are i.i.d. exponential. If $\Lambda(t) = \int_0^t \lambda(t)dt$, then it is easy to see that $X_{i+1} - X_i$ is distributed as $\Lambda^{\mathrm{inv}}(\Lambda(X_i) + E) - X_i$, where E is exponential. We already discussed sampling by thinning (rejection) for this process elsewhere. This process view also yields a simple but not uniformly fast generator for the Poisson distribution: if $E_1, E_2, \ldots$ are i.i.d. exponential random variables, and X is the smallest integer such that

$$\sum_{i=1}^{X+1} E_i > \lambda,$$

then X is Poisson (λ), as we are just counting the number of occurrences of a unit density homogeneous Poisson point process in the interval $[0, \lambda]$. By noting that an exponential is distributed as the logarithm of one over a uniform, we see that we may equivalently generate i.i.d. uniform $[0,1]$ random variables $U_1, U_2, \ldots$ and let X be the smallest integer such that

$$\prod_{i=1}^{X+1} U_i < e^{-\lambda}.$$

This is the so-called product method for Poisson random variates.

6. Markov chain methodology

We say that we generate X by the Markov chain method if we can generate a Markov chain $X_n, n \geq 1$ with the property that

$$\lim_{n\to\infty} \mathbb{P}\{X_n \leq z\} = \mathbb{P}\{X_\infty \leq z\},$$

at all z at which the function $\mathbb{P}\{X_\infty \leq z\}$ is continuous (here $\leq$ is to be taken componentwise). We write $X_n \xrightarrow{\mathcal{L}} X_\infty$, and say that X_n converges in distribution to X_∞. This is a weak notion, as X_n may be discrete and X_∞ continuous. For example, we may partition $\mathbb{R}^d$ into a grid of cells of sides $1/n^{1/(2d)}$, and let p_u be the probability content of cell u. Let X_n be the midpoint of cell u, where cell u is selected with probability p_u. A simple exercise shows that $X_n \xrightarrow{\mathcal{L}} X_\infty$ for any distribution of X_∞. Yet, X_n is discrete, and X_∞ may have any type of distribution. One needs of course a way of computing the p_u's. For this, knowledge of the distribution function F of X_∞ at each point suffices, as p_u can be written as a finite linear combination of the values of F at the vertices of the cell u. If X_n needs to have a density, one can define X_n as a uniform variate over the cell u, picked as above. In the latter case, if X_∞ has any density, we have a stronger kind of convergence, namely convergence in total variation (see below), by virtue of the Lebesgue density theorem (Wheeden and Zygmund, 1977; Devroye, 1987b). A good general reference for this section is Häggström (2002).

The paradigm described in the previous paragraph is a special case of what one could call the limit law method. Many processes, such as averages, maxima, and so forth, have asymptotic distributions that are severely restricted (e.g., averages of independent random variables have a stable limit law). Markov chains can easily be molded to nearly all distributions we can imagine, hence the emphasis on them as a tool.

In the exact Markov chain method, we can use properties of the Markov chain to generate a random variate X_T, where $X_T \stackrel{\mathcal{L}}{=} X_\infty$, where we recall that $\stackrel{\mathcal{L}}{=}$ denotes "is distributed as". There are a certain number of tricks that one can use to define a computable stopping time T. The first one was the so-called coupling from the past method of Propp and Wilson (1996), which has been shown to work well for discrete Markov chains, especially for the generation of random combinatorial objects. For general continuous distributions, a sufficiently general construction of stopping times is still lacking.

6.1. The Metropolis-Hastings chain. The Metropolis-Hastings chain (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, 1953, and Hastings, 1970) can be used for the generation of a random variate with an arbitrary density, provided that some care is taken in its choice of Markov transition probabilities. It requires transition probabilities $q(x, y)$ representing the density of Y on $[0,1]^d$ given $X = x$. Thus, for every fixed $x \in [0,1]^d$, $q(x, y)$ is a valid density in y from which we can generate random variates at will. The following algorithm produces a chain on random variates $\{X_n, n \geq 1\}$, which we call the Metropolis-Hastings chain:

Set $X \leftarrow x \in [0,1]^d$.
Repeat forever:
    Generate U uniformly on $[0,1]$.
    Generate Y with density $q(X, \cdot)$ on $[0,1]^d$.
    If $U < \frac{f(Y)q(Y,X)}{f(X)q(X,Y)}$, then $X \leftarrow Y$.

The limit distribution of X_n is not necessarily a distribution with density f. For example, if $d = 1$, and $q(x, y) = 2\mathbb{1}_{(x,y)\in B}$, where $B = [0, 1/2]^2 \cup [1/2, 1]^2$, then if $x < 1/2$ is the starting point, regardless of what f is, $X_n < 1/2$ for all n, and thus, universal convergence is excluded. A sufficient condition for convergence is that for every $x, y \in [0,1]^d$, $q(x, y) > 0$. Here X_∞ is a random variable with density f and $\leq$ is to be taken componentwise in $\mathbb{R}^d$. In particular, in the absence of any information on f, we may well just take $q(x, y) = 1$ for all x, y. We then obtain the following simple version of the chain:

```
Set X ← x ∈ [0,1]^d.
Repeat forever:
    Generate U uniformly on [0,1].
    Generate Y uniformly on [0,1]^d.
    If U < f(Y)/f(X), then X ← Y.
```

We would have obtained the same algorithm if we had the symmetry condition $q(x,y) = q(y,x)$, so we call this the symmetric Metropolis-Hastings chain. The algorithm that uses a uniform Y thus produces a sequence of X's that are selected from among uniforms. In this sense, they are "pure" and amenable to theoretical analysis. In the subsection below, we recall a lower bound for such pure methods.

6.2. The independence sampler. Tierney (1994) proposed $q(x,y) \equiv q(y)$ for all x, y. The independence sampler thus reads:

```
Set X ← x ∈ [0,1]^d.
Repeat forever:
    Generate U uniformly on [0,1].
    Generate Y with density q on [0,1]^d.
    If U < f(Y)q(X) / (f(X)q(Y)), then X ← Y.
```

If $q = f$, then no rejection takes place, and this algorithm produces an i.i.d. sequence of random variates with density f. Otherwise, if $q > 0$ there is convergence (in total variation, see below), and even geometric convergence at the rate $1 - \inf_x q(x)/f(x)$ (Liu, 1996). One thus should try and match q as well as possible to f.

6.3. The discrete Metropolis chain. The chains given above all remain valid if the state space is finite, provided that the density $f(x)$ and conditional density $q(x,y)$ are replaced by a stationary probability π_x and a transition probability $p(x,y)$ (summing to one with respect to y). In the discrete case, there is an important special case. A graph model for the uniform generation of random objects is the following. Let each node in the finite graph (V, E) correspond to an object. Create a connected graph by joining nodes that are near. For example, if a node represents an n-bit vector, then connect it to all nodes at Hamming distance one. The degree $\delta(x)$ of a node x is its number of neighbors, and $N(x)$ is the set of its neighbors. Set

$$q(x,y) = \frac{\mathbb{1}_{y \in N(x)}}{\delta(x)}.$$

If we wish to have a uniform stationary vector, that is, each node x has asymptotic probability $1/|V|$, then the Metropolis-Hastings chain method reduces to this:

```
Set X ← x ∈ V.
Repeat forever:
    Generate Y uniformly in N(X).
    Generate U uniformly on [0,1].
    If U < δ(X)/δ(Y), then X ← Y.
```

If the degrees are the same, then we always accept. Note also that we have a uniform limit only if the graph thus obtained is aperiodic, which in this case is equivalent to asking that it is not bipartite.

In some situations, there is a positive energy function $H(x)$ of the variable x, and the desired asymptotic density is

$$f(x) = Ce^{-H(x)},$$

where C is a normalization constant. As an example, consider x is a n-bit pixel vector for an image, where $H(x)$ takes into account occurrences of patterns in the image. We select $q(x, y)$ as above by defining neighborhoods for all x. As x is a bit vector, neighborhoods can be selected based on Hamming distance, and in that case, $q(x, y) = q(y, x)$ for all x, y. The symmetric Metropolis-Hastings chain then applies (in a discrete version) and we have:

```
Set X ← x ∈ V.
Repeat forever:
    Generate U uniformly on [0,1].
    Generate Y uniformly in N(X).
    If U < exp(H(X) − H(Y)), then X ← Y.
```

For further discussion and examples, see Besag (1986), Geman and Geman (1984) and Geman and McClure (1985).

6.4. Letac's lower bound. Generating a random variate with density f on the real line can always be done by generating a random variate with some density g on $[0, 1]$ via a monotone transformation. Thus, restricting oneself to $[0, 1]$ is reasonable and universal. Assume thus that our random variate X is supported on $[0, 1]$. Letac (1975) showed that any generator which outputs $X \leftarrow U_T$, where T is a stopping time and $U_1, U_2, \ldots$ is an i.i.d. sequence of uniform $[0, 1]$ random variables must satisfy

$$\mathbb{E}\{T\} \geq \sup_x f(x).$$

Applying a simple rejection method with the bound $\sup_x f(x)$ requires a number of iterations equal to $\sup_x f(x)$, and is thus optimal. (In fact, two uniforms are consumed per iteration, but there are methods of recirculating the second uniform.) Since some of the symmetric Metropolis-Hastings chains and independence samplers are covered by Letac's model, we note that even if they were stopped at some stopping time T tuned to return an exact random variate at that point, Letac's bound indicates that they cannot do better than the rejection method. There are two caveats though: firstly, the rejection method requires knowledge of $\sup_x f(x)$; secondly, for the Markov chains, we do not know when to stop, and thus have only an approximation. Thus, for fair comparisons, we need lower bounds in terms of approximation errors.

6.5. Rate of convergence. The rate of convergence can be measured by the total variation distance

$$V_n \stackrel{\text{def}}{=} \sup_A |\mathbb{P}\{X_n \in A\} - \mathbb{P}\{X_\infty \in A\}|,$$

where A ranges over all Borel sets of $\mathbb{R}^d$, and X_∞ is the limit law for X_n. Note Scheffé's identity (Scheffé, 1947)

$$V_n = \frac{1}{2}\int |f_n(x) - f(x)|dx,$$

where f_n is the density of X_n (with the above continuous versions of the chains, started at X_0, we know that X_n has a density for $n > 0$.)

Under additional restrictions on q, e.g.,

$$\inf_x \inf_y \frac{q(x,y)}{f(x)} > 0,$$

we know that $V_n \leq \exp(-\rho n)$ for some $\rho > 0$ (Holden, 1998; Jarner and Hansen, 1998). For more on total variation convergence of Markov chains, see Meyn and Tweedie (1993).

6.6. The Metropolis random walk. The Metropolis random walk is the Metropolis chain obtained by setting

$$q(x,y) = q(y,x) = g(y-x)$$

where g is symmetric about 0. On the real line, we could take, e.g., $g(u) = (1/2)e^{-|u|}$. The algorithm thus becomes:

```
Set X ← x ∈ R^d.
Repeat forever:
    Generate U uniformly on [0,1].
    Generate Z with density g.
    Set Y ← X + Z.
    If U < f(Y)/f(X), then X ← Y.
```

If g has support over all of $\mathbb{R}^d$, then the Metropolis random walk converges.

6.7. The Gibbs sampler. Among Markov chains that do not use rejection, we cite the Gibbs sampler (Geman and Geman, 1984) which uses the following method to generate X_{n+1} given $X_n = (X_{n,1}, \ldots, X_{n,d})$: first generate

$$X_{n+1,1} \sim f(\cdot|X_{n,2}, \ldots, X_{n,d}).$$

Then generate

$$X_{n+1,2} \sim f(\cdot|X_{n+1,1}, X_{n,3} \ldots, X_{n,d}),$$

and continue on until

$$X_{n+1,d} \sim f(\cdot|X_{n+1,1}, \ldots, X_{n+1,d-1}).$$

Under certain mixing conditions, satisfied if each conditional density is supported on all of the real line, this chain converges. Of course, one needs to be able to generate from all the $d-1$-dimensional conditional densities. It is also possible to group variables and generate random vectors at each step, holding all the components not involved in the random vector fixed.

6.8. Universal generators. A universal generator for densities generates a random variate X with density f without any a priori knowledge about f. The existence of such a generator is discussed in this section. It is noteworthy that a universal discrete generator (where X is supported on the positive integers) is easily obtained by inversion (see above). In fact, given any distribution function F on the reals, an approximate inversion method can be obtained by constructing a convergent iterative solution of $U = F(X)$, where U is uniform $[0,1]$. For example, binary search could be employed.

Assume that we can compute f as in a black box, and that X is supported on $[0,1]^d$ (see the previous section for a discussion of this assumption). We recall that if (X,Y) is uniformly distributed on the set $A = \{(x,y) : x \in [0,1]^d, 0 \le y \le f(x)\}$, then X has density f. Thus, the purpose is to generate a uniform random vector on A. This can be achieved by the Gibbs sampler, starting from any $x \in [0,1]^d$:

```
Set X ← x ∈ [0,1]^d.
Repeat forever:
    Generate U uniformly on [0,1].
    Set Y ← Uf(X).
    Repeat
          Generate X uniformly on [0,1]^d
    Until f(X) ≥ Y.
```

This algorithm is also known as the slice sampler (Swendsen and Wang, 1987; Besag and Green, 1993). In the first step of each iteration, given X, Y is produced with the conditional distribution required, that is, the uniform distribution on $[0, f(X)]$. Given Y, a new X is generated uniformly on the part of $[0,1]^d$ that is tucked under the curve of f. We chose to do this by rejection, because nothing is known about f. If some information is known about f, serious accelerations are possible. For example, if f is monotone decreasing on $[0,1]$, binary search may be used to reduce the search region rapidly. If f is strictly unimodal, then a slight adaptation of binary search may help. To illustrate the latter point for $d = 1$, assume that both f and f^{inv} are available in black box format, and that f is supported on the real line. Then the following algorithm always converges, where f_ℓ and f_r denote the left and right solutions of $f(x) = Y$:

```
Set $X \leftarrow x \in \mathbb{R}$.
Repeat forever:
    Generate $U$ uniformly on $[0,1]$.
    Set $Y \leftarrow Uf(X)$.
    Generate $X$ uniformly on $[f_\ell(Y), f_r(Y)]$.
```

Note that normalization or multiplicative constants in f and f^{inv} can be dropped altogether.

In any case, convergence theorems for the Gibbs sampler imply that the rejection Markov chain given above produces a sequence of vectors $\{(X_n, Y_n), n \geq 1\}$, with the property that (X_n, Y_n) tends in total variation to (X, Y), a random vector with the uniform distribution on A. Therefore, X_n tends in total variation to X, which has density f. It is thus a universal, but approximate, method.

There are stumbling blocks if one wishes to obtain an exact universal method. One which requires the knowledge of the tail function

$$T(t) = \int_{x:f(x)>t} f(x)\,dx, t > 0,$$

can easily be constructed. We consider the density in slabs of size one and generate a uniform random vector in A slab-wise. Define $p_n = T(n+1) - T(n)$, $n \geq 0$, and note that $\sum_n p_n = 1$. Selection of a random slab can be done by inversion on the fly—it does not require storage of the p_n's. Thus, a simple mixture and rejection algorithm would do this:

```
Generate $Z \geq 0$ with probability distribution $(p_0, p_1, p_2, \ldots)$.
Repeat forever:
    Generate $(X, Y)$ uniformly on $[0,1]^d \times [0,1]$.
    Until $Z + Y \leq f(X)$.
Return $X$.
```

6.9. Coupling from the past. Asmussen, Glynn and Thorisson (1992) and Propp and Wilson (1996) shocked the simulation world by announcing a way of simulating a finite discrete time ergodic Markov chain that produces an output with the stationary distribution. Their method is called perfect sampling, or CFTP, coupling from the past. We shall denote the unique stationary probability

vector by $(\pi_1, \ldots, \pi_n)$, where n is the number of states. We reserve t for the time index.

Let ω_t denote a random element at time t, governing transitions from all states. That is, X_{t+1}, the state at time $t+1$, is a function $\varphi(X_t, \omega_t)$, and indeed, we know and design φ and ω_t such that the transition probabilities

$$\mathbb{P}\{X_{t+1} = y | X_t = x\} = q(x, y),$$

for all x, y, as prescribed. The random elements ω_t form an independent sequence in t. Interestingly, the transitions at time t may be coupled. For example, if the Markov chain is a random walk on a cycle graph, then ω_t may just flip a fair coin. If it comes up heads, then all transitions are to the left, and if comes up tails, all transitions are to the right. The coupling is thus up to the designer. In the CFTP method, the sequence ω_t is fixed once and for all, but we need to generate only a finite number of them, luckily, to run the algorithm.

Pick a positive integer T. A basic run of length T consists of n Markov chains, each fired up from a different state for T time units, starting at time $-T$ until time 0, using the random elements $\omega_t, -T \le t < 0$. This yields n runs $X_t(i), 1 \le i \le n$ (i denotes the starting point), $-T \le t \le 0$, with $X_{-T}(i) = i$. Observe that if two runs collide at a given time t, then they will remain together forever. If $X_0(i)$ is the same for all i, then all runs have collided. In that case, Propp and Wilson showed that $X_0(.)$ has the stationary distribution. If there is no agreement among the $X_0(i)$'s, then replace T by $2T$ and repeat, always doubling the horizon until all $X_0(i)$'s agree. Return the first $X_0(.)$ for which agreement was observed.

It is interesting that coupling in the future does not work. We cannot simply start the n chains $X_0(i) = i$ and let T be the first time when they coalesce, because there are no guarantees that $X_T(.)$ has the stationary distribution. To see this, consider a directed cycle with n vertices, in which state 1 is the only state with a loop (occurring with probability $\epsilon > 0$). Coalescence can only take place at this state, so $X_T(.) = 1$, no matter what. Yet, the limiting distribution is roughly uniform over the states when $\epsilon > 0$ is small.

The effort needed may be prohibitive, as the time until agreement in a well designed coupling is roughly related to the mixing time of a Markov chain, so that n times the mixing time provides a rough bound of the computational complexity. Luckily, we have the freedom to design the random elements in such a manner that coalescence can be checked without doing n runs. For example, for a random walk on the chain graph with n vertices, where a loop is added (with probability 1/2) at each state to make the Markov chain aperiodic, we design the ω_t such that no two paths ever cross. Then coalescence can be checked by following the chains started at states 1 and n only. The coupling in ω_t is as follows: draw a uniform integer from $\{-2, -1, 1, 2\}$. If it is -2, then all transitions are to the left, except the transition from state 1, which stays. If it is 2, then all transitions are to the right, except the transition from state n, which stays. If it is -1, then all states stay, except state n, which takes a step to the left. If it is 1, then all states stay, except state 1, which takes a step to the right. As the number of steps until the two extreme runs meet is of the order of n^2, the expected complexity is also of that order. For more on the basics of random walks on graphs and mixing times for Markov chains, see Aldous and Fill (2004).

If the state space is countably infinite or uncountable (say, $\mathbb{R}^d$), then special constructions are needed to make sure that we only have to run finitely many chains, see, e.g., Green and Murdoch (2000), Wilson (2000), Murdoch (2000), or Mira, Moller and Roberts (2001). Most solutions assume the existence of a known dominating density g, with $f \le cg$ for a constant c, with random samples from g needed to move ahead. The resulting algorithms are still inferior to the corresponding rejection method, so further research is needed in this direction.

6.10. Further research. If one accepts approximations with density f_n of the density f, where n is the computational complexity of the approximation, then a study must be made that compares $\int |f_n - f|$ among various methods. For example, f_n could be the n-th iterate in the slice sampler, while a competitor could be g_n, the density in the approximate rejection method, which (incorrectly) uses rejection from n (so that g_n is proportional to $\min(f, n)$, and $\int |g_n - f| = 2 \int (f - n)_+$). Both methods have similar complexities (linear in n), but a broad comparison of the total variation errors is still lacking. In fact, the rejection method could be made adaptive: whenever an X is generated for which $f(X) > n$, then replace n by $f(X)$ and continue.

Simple exact simulation methods that are based upon iterative non-Markovian schemes could be of interest. Can coupling from the past be extended and simplified?

8. References

J. H. Ahrens, "Sampling from general distributions by suboptimal division of domains," vol. 19, Grazer Mathematische Berichte, 1993.

J. H. Ahrens, "A one-table method for sampling from continuous and discrete distributions," *Computing*, vol. 54, pp. 127–146, 1995.

J. H. Ahrens and U. Dieter, "Computer methods for sampling from gamma, beta, Poisson and binomial distributions," *Computing*, vol. 12, pp. 223–246, 1974.

J. H. Ahrens and U. Dieter, "Sampling from binomial and Poisson distributions: a method with bounded computation times," *Computing*, vol. 25, pp. 193–208, 1980.

J. H. Ahrens and U. Dieter, "Generating gamma variates by a modified rejection technique," *Communications of the ACM*, vol. 25, pp. 47–54, 1982.

J. H. Ahrens and U. Dieter, "An alias method for sampling from the normal distribution," *Computing*, vol. 42, pp. 159–17159-170, 1989.

J. H. Ahrens and U. Dieter, "A convenient sampling method with bounded computation times for Poisson distributions," in: *The Frontiers of Statistical Computation, Simulation and Modeling*, (edited by P. R. Nelson, E. J. Dudewicz, A. Öztürk, and E. C. van der Meulen), pp. 137–149, American Sciences Press, Columbus, OH, 1991.

J. H. Ahrens, K. D. Kohrt, and U. Dieter, "Algorithm 599. Sampling from gamma and Poisson distributions," *ACM Transactions on Mathematical Software*, vol. 9, pp. 255–257, 1983.

N. I. Akhiezer, *The Classical Moment Problem*, Hafner, New York, 1965.

D. Aldous and J. Fill, "Reversible Markov Chains and Random Walks on Graphs," Manuscript aavailable at `http://128.32.135.2/users/aldous/RWG/book.html`, 2004.

S. Asmussen, P. Glynn, and H. Thorisson, "Stationary detection in the initial transient problem," *ACM Transactions on Modeling and Computer Simulation*, vol. 2, pp. 130–157, 1992.

R. W. Bailey, "Polar generation of random variates with the t distribution," *Mathematics of Computation*, vol. 62, pp. 779–781, 1994.

J. Besag, "On the statistical analysis of dirty pictures (with discussions)," *Journal of the Royal Statistical Society, Series B*, pp. 259–302, 1986.

J. Besag and P. J. Green, "Spatial statistics and Bayesian computation," *Journal of the Royal Statistical Society B*, vol. 55, pp. 25–37, 1993.

D. J. Best, "A simple algorithm for the computer generation of random samples from a Student's t or symmetric beta distribution," in: *COMPSTAT 1978: Proceedings in Computational Statistics*, (edited by L. C. A. Corsten and J. Hermans), pp. 341–347, Physica Verlag, Wien, Austria, 1978.

D. J. Best, "Letter to the editor," *Applied Statistics*, vol. 27, p. 181, 1978.

D. J. Best, "A note on gamma variate generators with shape parameter less than unity," *Computing*, vol. 30, pp. 185–188, 1983.

L. Bondesson, "On simulation from infinitely divisible distributions," *Advances in Applied Probability*, vol. 14, pp. 855–869, 1982.

G. E. P. Box and M. E. Müller, "A note on the generation of random normal deviates," *Annals of Mathematical Statistics*, vol. 29, pp. 610–611, 1958.

R. E. Caflisch, W. J. Morokoff, and A. B. Owen, "Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension," *Journal of Computational Finance*, vol. 1, pp. 27–46, 1997.

J. M. Chambers, C. L. Mallows, and B. W. Stuck, "A method for simulating stable random variables," *Journal of the American Statistical Association*, vol. 71, pp. 340–344, 1976.

R. C. H. Cheng, "The generation of gamma variables with non-integral shape parameter," *Applied Statistics*, vol. 26, pp. 71–75, 1977.

R. C. H. Cheng and G. M. Feast, "Some simple gamma variate generators," *Applied Statistics*, vol. 28, pp. 290–295, 1979.

R. C. H. Cheng and G. M. Feast, "Gamma variate generators with increased shape parameter range," *Communications of the ACM*, vol. 23, pp. 389–393, 1980.

H. C. Chen and Y. Asau, "On generating random variates from an empirical distribution," *AIIE Transactions*, vol. 6, pp. 163–166, 1974.

L. Devroye, "The series method in random variate generation and its application to the Kolmogorov-Smirnov distribution," *American Journal of Mathematical and Management Sciences*, vol. 1, pp. 359–379, 1981a.

L. Devroye, "The computer generation of random variables with a given characteristic function," *Computers and Mathematics with Applications*, vol. 7, pp. 547–552, 1981b.

L. Devroye, "The computer generation of Poisson random variables," *Computing*, vol. 26, pp. 197–207, 1981c.

L. Devroye, "Methods for generating random variates with Pólya characteristic functions," *Statistics and Probability Letters*, vol. 2, pp. 257–261, 1984a.

L. Devroye, "Random variate generation for unimodal and monotone densities," *Computing*, vol. 32, pp. 43–68, 1984b.

L. Devroye, "A simple algorithm for generating random variates with a log-concave density," *Computing*, vol. 33, pp. 247–257, 1984c.

L. Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986a.

L. Devroye, "An automatic method for generating random variables with a given characteristic function," *SIAM Journal of Applied Mathematics*, vol. 46, pp. 698–719, 1986b.

L. Devroye, "The analysis of some algorithms for generating random variates with a given hazard rate," *Naval Research Logistics Quarterly*, vol. 33, pp. 281–292, 1986c.

L. Devroye, "A simple generator for discrete log-concave distributions," *Computing*, vol. 39, pp. 87–91, 1987a.

L. Devroye, *A Course in Density Estimation*, Birkhäuser, Boston, 1987b.

L. Devroye, "Generating sums in constant average time," in: *Proceedings of the 1988 Winter Simulation Conference*, (edited by M. A. Abrams, P. L. Haigh and J. C. Comfort), pp. 425–431, IEEE, San Diego, CA., 1988.

L. Devroye, "On random variate generation when only moments or Fourier coefficients are known," *Mathematics and Computers in Simulation*, vol. 31, pp. 71–89, 1989.

L. Devroye, "A note on Linnik's distribution," *Statistics and Probability Letters*, vol. 9, pp. 305–306, 1990.

L. Devroye, "Algorithms for generating discrete random variables with a given generating function or a given moment sequence," *SIAM Journal on Scientific and Statistical Computing*, vol. 12, pp. 107–126, 1991.

L. Devroye, "Random variate generation in one line of code," in: *1996 Winter Simulation Conference Proceedings*, (edited by J. M. Charnes, D. J. Morrice, D. T. Brunner and J. J. Swain), pp. 265–272, ACM, San Diego, CA, 1996.

L. Devroye, "Simulating theta random variates," *Statistics and Probability Letters*, vol. 31, pp. 2785–2791, 1997.

L. Devroye, J. Fill, and R. Neininger, "Perfect simulation from the quicksort limit distribution," *Electronic Communications in Probability*, vol. 5, pp. 95–99, 2000.

L. Devroye, "Simulating perpetuities," *Methodologies and Computing in Applied Probability*, vol. 3, pp. 97–115, 2001.

L. Devroye and R. Neininger, "Density approximation and exact simulation of random variables that are solutions of fixed-point equations," *Advances of Applied Probability*, vol. 34, pp. 441–468, 2002.

K. Dickman, “On the frequency of numbers containing prime factors of a certain magnitude,” *Arkiv for Mathematik, Astronomi oc Fysik*, vol. 22A, p. 1014, 1930.

D. Dugué and M. Girault, “Fonctions convexes de Pólya,” *Publications de l'Institut de Statistique des Universités de Paris*, vol. 4, pp. 3–10, 1955.

M. Evans and T. Swartz, “Random variable generation using concavity properties of transformed densities,” *Journal of Computational and Graphical Statistics*, vol. 7, pp. 514–528, 1996.

M. Evans and T. Swartz, *Approximating Integrals Via Monte Carlo and Deterministic Methods*, Oxford University Press, Oxford, UK, 2000.

W. Feller, “On the Kolmogorov-Smirnov limit theorems for empirical distributions,” *Annals of Mathematical Statistics*, vol. 19, pp. 177–189, 1948.

J. Fill, “An interruptible algorithm for perfect sampling via Markov chains,” *The Annals of Applied Probability*, vol. 8, pp. 131–162, 1998.

P. Flajolet and N. Saheb, “The complexity of generating an exponentially distributed variate,” *Journal of Algorithms*, vol. 7, pp. 463–488, 1986.

B. L. Fox, *Strategies for Quasi-Monte Carlo*, Kluwer Academic Publishers, Boston, 1999.

A. E. Gelfand and A. F. M. Smith, “Sampling-based approaches to calculating marginal densities,” *Journal of the American Statistical Association*, vol. 85, pp. 398–409, 1990.

S. Geman and D. Geman, “Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.

S. Geman and D. McClure, “Bayesian image analysis: An application to single photon emission tomography,” in: *Proceedings of the Statistical Computing Section*, pp. 12–18, Washington, DC, 1985.

W. R. Gilks and P. Wild, “Adaptive rejection for Gibbs sampling,” *Applied Statistics*, vol. 41, pp. 337–348, 1992.

W. R. Gilks, N. G. Best, and K. K. C. Tan, “Adaptive rejection Metropolis sampling,” *Applied Statistics*, vol. 41, pp. 337–348, 1995.

C. M. Goldie and R. Grübel, “Perpetuities with thin tails,” *Advances in Applied Probability*, vol. 28, pp. 463–480, 1996.

P. J. Green and D. J. Murdoch, “Exact sampling for Bayesian inference: towards general purpose algorithms (with discussion),” in: *Monte Carlo Methods*, (edited by J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), pp. 301–321, Bayesian Statistics, vol. 6, Oxford university Press, Oxford, 2000.

O. Häggström, *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press, Cambridge, UK, 2002.

W. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, vol. 57, pp. 97-110, 1970.

L. Holden, "Geometric convergence of the Metropolis-Hastings simulation algorithm," *Statistics and Probability Letters*, vol. 39, pp. 371–377, 1998.

W. Hörmann, "The generation of binomial random variates," *Journal of Statistical Computation and Simulation*, vol. 46, pp. 101–110, 1993a.

W. Hörmann, "The transformed rejection method for generating Poisson random variables," *Insurance: Mathematics and Economics*, vol. 12, pp. 39–45, 1993b.

W. Hörmann, "A universal generator for discrete log-concave distributions," *Computing*, vol. 52, pp. 89–96, 1994.

W. Hörmann and G. Derflinger, "Universal generators for correlation induction," in: *COMPSTAT 94*, (edited by R. Dutter and W. Grossmann), pp. 52–57, Physica-Verlag, Heidelberg, 1994.

W. Hörmann, "A rejection technique for sampling from T-concave distributions," *ACM Transactions on Mathematical Software*, vol. 21, pp. 182–193, 1995.

W. Hörmann, J. Leydold, and G. Derflinger, *Automatic Nonuniform Random Variate Generation*, Springer-Verlag, Berlin, 2004.

D. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the IRE*, vol. 40, pp. 1098–1101, 1952.

D. Jackson, "The Theory of Approximation," Colloquia Publications, vol. 11, American Mathematical Society, Providence, RI, 1930.

S. F. Jarner and E. Hansen, "Geometric ergodicity of Metropolis algorithms," Department of Statistics, University of Copenhagen, 1998.

N. L. Johnson and S. Kotz, *Distributions in Statistics: Discrete Distributions*, John Wiley, New York, N.Y., 1969.

V. Kachitvichyanukul and B. W. Schmeiser, "Binomial random variate generation," *Communications of the ACM*, vol. 31, pp. 216–222, 1988.

V. Kachitvichyanukul and B. W. Schmeiser, "Algorithm 678: BTPEC: sampling from the binomial distribution," *ACM Transactions on Mathematical Software*, vol. 15, pp. 394–397, 1989.

M. Kanter, "Stable densities under change of scale and total variation inequalities," *Annals of Probability*, vol. 3, pp. 697–707, 1975.

M. S. Keane and G. L. O'Brien, "A Bernoulli factory," *ACM Transactions on Modeling and Computer Simulation*, vol. 4, pp. 213–219, 1994.

D. E. Knuth and A. C. Yao, "The complexity of nonuniform random number generation," in: *Algorithms and Complexity*, (edited by J. E. Traub), pp. 357–428, Academic Press, New York, N.Y., 1976.

S. Kotz and I. V. Ostrovskii, "A mixture representation of the Linnik distribution," *Statistics and Probability Letters*, vol. 26, pp. 61–64, 1996.

M. G. Krein, "On one extrapolation problem of A.N. Kolmogorov," *Dokl. Akad. Nauk SSSR*, vol. 46, pp. 339–342, 1944.

R. G. Laha, "On a class of unimodal distributions," *Proceedings of the American Mathematical Society*, vol. 12, pp. 181–184, 1961.

D. Le Minh, "Generating gamma variates," *ACM Transactions on Mathematical Software*, vol. 14, pp. 261–266, 1988.

G. Letac, "On building random variables of a given distribution," *Annals of Probability*, vol. 3, pp. 298–306, 1975.

P. A. W. Lewis and G. S. Shedler, "Simulation of nonhomogeneous Poisson processes by thinning," *Naval Research Logistics Quarterly*, vol. 26, pp. 403–413, 1979.

J. Leydold, "Automatic sampling with the ratio-of-uniforms method," *ACM Transactions on Mathematical Software*, vol. 26, pp. 78–98, 2000a.

J. Leydold, "A note on transformed density rejection," *Computing*, vol. 65, pp. 187–192, 2000b.

J. Leydold, "A simple universal generator for continuous and discrete univariate T-concave distributions," *ACM Transactions on Mathematical Software*, vol. 27, pp. 66–82, 2001.

G. D. Lin, "On the moment problem," *Statistics and Probability Letters*, vol. 35, pp. 85–90, 1997.

Yu. V. Linnik, "Linear forms and statistical criteria,I,II," *Selected Translations in Mathematical Statistics and Probability*, vol. 3, pp. 1–90, 1962.

J. S. Liu, "Metropolized independent sampling with comparisons to rejection sampling and importance sampling," *Statistics and Computing*, vol. 6, pp. 113–119, 1996.

G. Marsaglia and W. W. Tsang, "A simple method for generating gamma variables," *ACM Transactions on Mathematical Software* , vol. 24, pp. 341–350, 2001.

G. Marsaglia, M. D. Maclaren, and T. A. Bray, "A fast procedure for generating normal random variables," *Communications of the ACM*, vol. 7, pp. 4–10, 1964.

G. Marsaglia, "The squeeze method for generating gamma variates," *Computers and Mathematics with Applications*, vol. 3, pp. 321–325, 1977.

G. Marsaglia and W. W. Tsang, "A fast, easily implemented method for sampling from decreasing or symmetric unimodal density functions," *SIAM Journal on Scientific and Statistical Computing*, vol. 5, pp. 349–359, 1984.

N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemical. Physics*, vol. 21, p. 10871091, 1953.

S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, Liondon, 1993.

A. Mira, J. Moller, and G. O. Roberts, "Perfect slice samplers," *Journal of the Royal Statistical Society B*, vol. 63, pp. 593–606, 2001.

D. J. Murdoch, "Exact sampling for Bayesian inference: unbounded state spaces," in: *Monte Carlo Methods*, (edited by N. Madras), pp. 111–121, Fields Institute Communications, vol. 6, American Mathematical Society, 2000.

R. B. Nelsen, *An Introduction to Copulas* , Springer-Verlag, Berlin, 1999.

J. G. Propp and D. B. Wilson, "Exact sampling with coupled Markov chains and applications to statistical mechanics," *Random Structures and Algorithms*, vol. 9, pp. 223–252, 1996.

S. T. Rachev, *Probability Metrics and the Stability of Stochastic Models*, John Wiley, New York, 1991.

A. Rényi and G. Szekeres, "On the height of trees," *Journal of the Australian Mathematical Society*, vol. 7, pp. 497–507, 1967.

U. Rösler and L. Rüschendorf, "The contraction method for recursive algorithms," *Algorithmica*, vol. 29, pp. 3–33, 2001.

K. Sato, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge, 2000.

H. Scheffé, "A useful convergence theorem for probability distributions," *Annals of Mathematical Statistics*, vol. 18, pp. 434–458, 1947.

B. W. Schmeiser and A. J. G. Babu, "Beta variate generation via exponential majorizing functions," *Operations Research*, vol. 28, pp. 917–926, 1980.

B. W. Schmeiser and R. Lal, "Squeeze methods for generating gamma variates," *Journal of the American Statistical Association*, vol. 75, pp. 679–682, 1980.

E. Stadlober, "Sampling from Poisson, binomial and hypergeometric distributions: ratio of uniforms as a simple fast alternative," Habilitationsschrift, Institute of Statistics, Technical University of Graz, Austria, 1988.

E. Stadlober, "Binomial random variate generation: a method based on ratio of uniforms," *American Journal of Mathematical and Management Sciences*, vol. 9, pp. 1–20, 1989.

E. Stadlober, "The ratio of uniforms approach for generating discrete random variates," *Journal of Computational and Applied Mathematics*, vol. 31, pp. 181–189, 1990.

J. Stoyanov, "Krein condition in probabilistic moment problems," *Bernoulli*, vol. 6, pp. 939–949, 2000.

R. Swendsen and J. Wang, "Non-universal critical dynamics in Monte Carlo simulation," *Physical review Letters*, vol. 58, pp. 86–88, 1987.

L. Tierney, "Markov chains for exploring posterior distributions," *Annals of Statistics*, vol. 22, pp. 1701–1762, 1994.

G. Ulrich, "Computer generation of distributions on the m-sphere," *Applied Statistics*, vol. 33, pp. 158–163, 1984.

W. Vervaat, "On a stochastic difference equation and a representation of non-negative infinitely divisible random variables," *Advances in Applied Probability*, vol. 11, pp. 750–783, 1979.

J. von Neumann, “Various techniques used in connection with random digits,” *Collected Works*, vol. 5, pp. 768–770, Pergamon Press, 1963. also: in Monte Carlo Method, National Bureau of Standards Series, Vol. 12, pp. 36-38, 1951.

A. J. Walker, “New fast method for generating discrete random numbers with arbitrary frequency distributions,” *Electronics Letters*, vol. 10, pp. 127–128, 1974.

A. J. Walker, “An efficient method for generating discrete random variables with general distributions,” *ACM Transactions on Mathematical Software*, vol. 3, pp. 253–256, 1977.

R. L. Wheeden and A. Zygmund, *Measure and Integral*, Marcel Dekker, New York, N.Y., 1977.

D. B. Wilson, “Layered multishift coupling for use in perfect sampling algorithms (with a primer on CFTP),” in: *Monte Carlo Methods*, (edited by N. Madras), pp. 141–176, Fields Institute Communications, vol. 6, American Mathematical Society, 2000.

H. Zechner and E. Stadlober, “Generating beta variates via patchwork rejection,” *Computing*, vol. 50, pp. 1–18, 1993.

V. M. Zolotarev, “On analytic properties of stable distribution laws,” *Selected Translations in Mathematical Statistics and Probability*, vol. 1, pp. 207–211, 1959.

V. M. Zolotarev, “On the representation of stable laws by integrals,” *Selected Translations in Mathematical Statistics and Probability*, vol. 6, pp. 84–88, 1966.

V. M. Zolotarev, “Integral transformations of distributions and estimates of parameters of multidimensional spherically symmetric stable laws,” in: *Contributions to Probability*, pp. 283–305, Academic Press, 1981.

V. M. Zolotarev, *One-Dimensional Stable Distributions*, American Mathematical Society, Providence, R.I., 1986.