Complexity Questions in Non-Uniform Random Variate Generation

Luc Devroye

School of Computer Science
McGill University
Montreal, Canada H3A 2K6
lucdevroye@gmail.com

Abstract. In this short note, we recall the main developments in non-uniform random variate generation, and list some of the challenges ahead.

Keywords: random variate generation, Monte Carlo methods, simulation

1 The pioneers

World War II was a terrible event. But it can not be denied that it pushed science forward with a force never seen before. It was responsible for the quick development of the atomic bomb and led to the cold war, during which the United States and Russia set up many research labs and attracted the best and the brightest to run them. It was at Los Alamos and RAND that physicists and other scientists were involved in large-scale simulations. John von Neumann, Stan Ulam and Nick Metropolis developed the Monte Carlo Method in 1946: they suggested that we could compute and predict in ways never before considered. For example, the Metropolis chain method developed a few years later (Metropolis, Rosenbluth, Rosenbluth, Teller and Teller, 1953) can be used to simulate almost any distribution by setting up a Markov chain that has that distribution as a limit. At least asymptotically, that is. But it was feasible, because the computers were getting to be useful, with the creation of software and the FORTRAN compiler.

To drive the Markov chains and other processes, one would need large collections of uniform random numbers. That was a bit of a sore point, because no one knew where to get them. Still today, the discussion rages as to how one should secure a good source of uniform random numbers. The scientists eventually settled on something that a computer could generate, a sequence that looked random.

The early winner was the linear congruential generator, driven by $x_{n+1} = (ax_n + b) \mathrm{mod} m$, which had several well-understood properties. Unfortunately, it is just a deterministic sequence, and many of its flaws have been exposed in the last three decades. The built-in linear-congruential generator in the early FORTRAN package for IBM computers was RANDU. Consecutive pairs

Y. Lechevallier, G. Saporta (eds.), *Proceedings of COMPSTAT'2010*,
DOI 10.1007/978-3-7908-2604-3_1, © Springer-Verlag Berlin Heidelberg 2010

(x_n, x_{n+1}) produced by RANDU fall on just a few parallel lines, prompting Marsaglia (1968) to write a paper with the ominous title "Random numbers fall mainly in the plane". But bad linear congruential or related generators have persisted until today—the generator in Wolfram's Mathematica had a similar problem: their built-in generator Random uses the Marsaglia-Zaman subtract-with-borrow generator (1991), which has the amazing property that all consecutive triples (x_n, x_{n+1}, x_{n+2}) fall in only two hyperplanes of $[0,1]^3$, a fact pointed out to me by Pierre Lecuyer. Many thousands of simulations with Mathematica are thus suspect—I was made aware of this due an inconsistency between simulation and theory brought to my attention by Jim Fill in 2010. The company has never apologized or offered a refund to its customers, but it has quietly started using other methods, including one based on a cellular automaton (the default). Hoewever, they are still offering linear congruential generators as an option. The story is far from over, and physical methods may well come back in force.

Information theorists and computer scientists have approached randomness from another angle. For them, random variables uniformly distributed on $[0,1]$ do not and can not exist, because the binary expansions of such variables consist of infinitely many independent Bernoulli (1/2) random bits. Each random bit has binary entropy equal to one, which means that its value or cost is one. A bit can store one unit of information, and vice versa, a random bit costs one unit of resources to produce. Binary entropy for a more complex random object can be measured in terms of how many random bits one needs to describe it. The binary entropy of a random vector of n independent fair coin flips is n, because we can describe it by n individual fair coins.

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. In this model, one assumes that an i.i.d. sequence of independent perfect bits is available. This permits the development of an elegant information-theoretic theory. For example, Knuth and Yao (1976) showed that to generate a random integer X described by the probability distribution

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

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. 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.

For integer-valued random variables with $\mathbf{P}\{X = n\} = p_n, n \geq 0$, 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 $\mathbf{E}\{X+1\}$. 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. In any case, this value does not exceed $\log_2 N$, where N is the number of possible values X can take. The difference with the Knuth-Yao result is that one now needs to be able to store and add real numbers (the p_n's).

Even when taking bits at unit cost, one needs to be careful about the computational model. For example, is one allowed to store real numbers, or should we work with a model in which storage and computation time is also measured in terms of bits? We feel that the information-theoretic boundaries and lower bounds should be studied in more detail, and that results like those of Knuth and Yao should be extended to cover non-discrete random variables as well, if one can formulate the models correctly.

2 The assumptions and the limitations

Assume that we can indeed store and work with real numbers and that an infinite source of independent identically distributed uniform $[0,1]$ random variables, $U_1, U_2, \ldots$ is available at unit cost per random variable used. The random source excepted, the computer science community has embraced the so-called RAM (random access memory) model. While it unrealistic, designing random variate generators in this model 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 idealistic model. The subject of non-uniform random generation is to generate random variables with a given distribution—we call these random variates—, in (possibly random) finite time. We also assume that computations can be carried out with infinite precision, and we require that the results be theoretically exact.

For a given collection of operations (a computer language), one can define the collection of all distributions of random variables that can be generated in finite time using these operations. Classes of achievable distributions defined in this manner will be denoted by $\mathcal{D}$. For example, if we only allow addition and subtraction, besides the standard move, store and copy operations, then one can only generate the sums

$$c + \sum_{i=1}^{N} k_i U_i,$$

where $c \in \mathbf{R}$, and $N, k_1, \ldots, k_N$ are finite integers. This is hardly interesting. An explosion occurs when one allows multiplication and division, and introduces comparisons and loops as operators. The achievable class becomes quite large. We will call it the algebraic class.

The need for non-uniform random variates in Monte Carlo simulations prompted the post-World War II teams to seriously think about the problem. All probabilists understand the inversion method: a random variate with distribution function F can be obtained as

$$X = F^{\mathrm{inv}}(U),$$

where U is uniform $[0,1]$. This inversion method is useful when the inverse is readily computable. For example, a standard exponential random variable (which has density $e^{-x}, x > 0$), can be generated as $\log(1/U)$. Table 1 gives some further examples.

Table 1. Table 1: Some densities with distribution functions that are explicitly invertible. Random variates can be generated simply by appropriate transormations of a uniform $[0,1]$ random variable U.

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), $a > 0$	$\frac{a}{x^{a+1}}$, $x > 1$	$1 - 1/x^a$	$1/U^{1/a}$

However, note that only the Pareto distribution for values of a that are inverses of an integer is in the algebraic class. One can attempt to create functions of a finite number of uniforms, and in this way, one notes that the Cauchy too is in the algebraic class. We leave it as a simple exercise to show that the following method works. Keep generating independent random pairs of independent uniforms, (U, U'), until for the first time $U^2 + U'^2 \leq 1$ (now (U, U') is uniformly distributed in the positive quarter of the unit circle). Then set $X = SU/U'$, where $S \in \{-1, +1\}$ is a random sign. One can ask if the normal distribution is in the algebraic class for example. In fact, a good description of the algebraic class is sorely needed.

Assume now a much more powerful class, one that is based upon all operations for the algebraic class, plus the standard mathematical functions, exp, log, sin (and thus cos and tan). Call it the standard class. All inversion method examples in Table 1 describe distributions in the standard class.

Since we did not add the inverse of the normal distribution function to the allowed operations, it would appear at first that the normal distribution is not in the standard class. For future reference, the standard normal density is given by $\exp(-x^2/2)/\sqrt{2\pi}$. This was of great concern to the early simulationists because they knew how to calculate certain standard functions very well, but had to make do with approximation formulas for functions like the inverse gaussian distribution function. Such formulas became very popular, with researchers outcompeting each other for the best and the latest approximation.

Amazingly, it was not until 1958 that Box and Müller showed the world that the gaussian distribution was in the standard class. Until that year, all normal simulations were done either by summing a number of uniforms and rescaling in the hope that the central limit theorem would yield something good enough, or by using algebraic approximations of the inverse of the gaussian distribution function, as given, e.g., in the book of Hastings (1955).

As in our Cauchy example, Box and Müller noted that one should only look at simple transformations of 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 (1958), a pair of independent standard normal random variates is obtained by setting

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

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, 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

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

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 $a \geq 1/2$. 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 $a > 0$:

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

Until Bailey's paper, only rather inconvenient rejection methods were available for Student's 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

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

is said to be stable with parameters $\alpha \in (0,2]$ and $|\beta| \leq 1$. 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

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

We end this section with a few questions about the size and nature of the standard class. Let us say that a distribution is k-standard (for fixed integer k) if it is in the standard class and there exists a generator algorithm that uses only a fixed number k of uniforms. The standard class is thus the union of all k-standard classes. Even more restrictive is the loopless k-standard class, one in which looping operations are not allowed. These include distributions for which we can write the generator in one line of code. The gaussian and indeed all stable laws are loopless 2-standard. We do not know if the gamma density

$$\frac{x^{a-1}e^{-x}}{\Gamma(a)}, x > 0,$$

is loopless k-standard for any finite k not depending upon the gamma parameter $a > 0$. Similarly, this is also unknown for the general beta family. Luckily, the gamma law is in the standard class, thanks to the rejection method, which was invented by von Neumann and is discussed in the next section.

It would be a fine research project to characterize the standard class and the (loopless) k-standard classes in several novel ways. Note in this respect that all discrete laws with the property that p_n can be computed in finite time using standard operations are 1-standard. Note that we can in fact use the individual bits (as many as necessary) to make all the necesary comparisons of U with a threshold. Only a random but finite number of these bits are needed for each variate generated. Let us define the class of distributions with the property that only a (random) finite number of bits of U suffice 0-standard. The full use of all bits in a uniform is only needed to create an absolutely continuous law.

Are absolutely continuous laws that are describable by standard operations k-standard for a given universal finite k?

Finally, it seems that even the simplest singular continuous laws on the real line are not in the standard class, but a proof of this fact would be nice to have. Take as an example a random variable $X \in [0,1]$ whose binary expansion has independent Bernoulli (p) bits. If $p = 1/2$, X is clearly uniform on $[0,1]$. But when $p \notin \{0, 1/2, 1\}$, then X is singular continuous. It is difficult to see how standard functions can be used to recreate such infinite expansions. If this is indeed the case, then the singular continuous laws, and indeed many fractal laws in higher dimensions, have the property that no finite amount of resources suffices to generate even one of them exactly. Approximations on the real line that are based on uniforms and standard functions are necessarily atomic or absolutely continuous in nature, and thus undesirable.

3 The rejection method

The Cauchy method described above uses a trick called rejection. The rejection method in its general form is due to von Neumann (1951). Let X have density f on $\mathbf{R}^d$. Let g be another density with the property that for some finite constant $c \geq 1$, called the rejection constant,

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

For any nonnegative integrable function h on $\mathbf{R}^d$, define the body of h as $B_h = \{(x,y) : x \in \mathbf{R}^d, 0 \leq y \leq h(x)\}$. 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 $Ucg(X) \leq f(X)$
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.

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 $Ue^{1/2-|X|} \leq e^{-X^2/2}$
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 $X \leftarrow SE_1$
```

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

The laws statisticians care about have one by one fallen to the rejection method. As early as 1974, Ahrens and Dieter showed how to generate beta, gamma, Poisson and binomial random variables efficiently. All these distributions are in the standard class. However, if the density f or the probability p_n is not computable in finite time using standard functions, then the distribution is not obviously in the standard class.

4 The alternating series method

To apply the rejection method, we do not really need to know the ratio $f(x)/(cg(x))$ exactly. Assume that 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 $n \to \infty$. In that case, we let n increase until for the first time, either

$$U \leq \xi(X)$$

(in which case we accept X), or

$$U \geq \psi_n(X)$$

(in which case we reject X). 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 $n \leftarrow n + 1$ until $U \le \xi_n(X)$ or $U \ge \psi_n(X)$
until $U \le \xi_n(X)$
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 is finite almost surely. Also, $N > t$ implies that $U \in [\xi_t(X), \psi_t(X)]$, and thus,

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

and

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

We cannot stress strongly enough how important the alternating series method is, as it frees us from having to compute f exactly. When ξ_n and ψ_n are computable in finite time with standard functions, and g is in the standard class, then f is in the standard class.

It is indeed the key to the solution of a host of difficult non-uniform random variate generation problems. For example, since the exponential, logarithmic and trigonometric functions have simple Taylor series expansions, one can approximate densities that use a finite number of these standard functions from above and below by using only addition, multiplication and division, and with some work, one can see that if a law is (k-)standard, then it is (k-)algebraic. Both gamma and gaussian are algebraic if one invokes the alternating series method using Taylor series expansions. To the programmer, this must seem like' a masochistic approach—if we have the exponential function, why should we not use it? But for the information theorist and computer scientist, the model of computation matters, and lower bound theory is perhaps easier to develop using more restricted classes.

But one can do better. Assume that a given density is Riemann integrable. Then it can be approximated from below by histograms. It takes only a moment to verify that such densities can be written as infinite mixtures of uniforms on given intervals. The mixture weights define a discrete law, which we know is 0-standard. A random variate can be written as

$$a_Z + b_Z U,$$

where Z is a discrete random variable, and $[a_i, b_i]$, $i \geq 1$, denote the intervals in the mixture decomposition. So, given one uniform random variable, first use a random number of bits from its expansion to generate Z, and then note that the unused bits, when shifted, are again uniformly distributed. This shows that Riemann integrable densities are 1-standard if we can compute the density at each point using only standard functions. In particular, the gamma and normal laws are 1-standard. This procedure can be automated, and indeed, several so-called table methods are based on such mixture decompositions. See, e.g., Devroye (1986a), or Hörmann, Leydold and Derflinger (2004).

5 Oracles

Oracles are a convenient way of approaching algorithms. Engineers call them "black boxes". One can imagine that one has an oracle for computing the value of the density f at x. Armed with one or more oracles, and our infinite source of uniforms, one can again ask for the existence of generators for certain dustributions.

For example, given a density oracle, is there an exact finite time method for generating a random variate with that density? Is there such a method that is universal, i.e., that works for all densities? The answer to this question is not known. In contrast, when given an oracle for the inverse of a distribution function, a universal method exists, the inversion method.

Given that we do not know the answer for the density oracle, it is perhaps futile at this point to ask for universal generators for characteristic function, Laplace transform or other oracles. It is perhaps possible to achieve success in the presence of two or more oracles. In the author's 1986 book, one can find partial success stories, such as a density oracle method for all log-concave densities on the line, or a combined density / distribution function (not the inverse though) moracle method for all monotone densities.

Complexity is now calculated in terms of the numbers of uniforms consumed and as a function of the number of consultations of the oracle. This should allow one to derive a number of negative results and lower bounds as well.

6 Open questions

We discussed the need for descriptions of operator-dependent classes, and the creation of models that can deal with singular continuity. The rejection and alternating series methods enable us to generate random variates with any distribution provided two conditions hold: we have an explicitly known finite dominating measure of finite, and we can approximate the value of the density or discrete probability locally by convergent and explicitly known upper and lower bounds. This has been used by the author, for example, to deal with distributions that are given by infinite series (Devroye, 1981a, 1997, 2009), distributions specified by a characteristic function (Devroye, 1981b, 1986b), Fourier coefficients (Devroye, 1989), a sequence of moments (Devroye, 1991), or their Laplace transforms. It should also be possible to extend this to laws whose Mellin transforms are known, or infinitely divisible laws that are specified in terms of Lévy or Khinchin measures (see Sato for definitions; Bondesson (1982) offers some approximative solutions). In all these examples, if a density exists, there are indeed inversion formulae that suggest convergent and explicitly known upper and lower bounds of the density.

It is hopeless to try to remove the requirement that a dominating measure be known—a characteristic function of a singular continuous distribution is a particularly unwieldy beast, for example. Some distributions have asymptotic distributional limits. As an example, consider

$$X = \sum_{i=0}^{\infty} \theta^i \xi_i,$$

where the ξ_i are independent Bernoulli (p), and $\theta \in (-1, 1)$. When $p = 1/2, \theta = 1/2$, X is uniform $[0, 1]$, while for $p \notin \{0, 1/2, 1\}$, $\theta = 1/2$, X is singular continuous. Using $\overset{\mathcal{L}}{=}$ for distributional identity, we see that

$$X \overset{\mathcal{L}}{=} \xi_0 + \theta X.$$

It seems unlikely that the distribution of X is in the standard class for all parameter values.

This leads to the question of determining which X, given by simple distributional identities of the form

$$X \overset{\mathcal{L}}{=} \phi(X, U)$$

are in the standard class. Note that the map $X \leftarrow \phi(X, U)$ defines in some cases a Markov chain with a limit. Using CFTP (coupling from the past; see Propp and Wilson (1996), Asmussen, Glynn and Thönnes (1992), Wilson (1998), Fill (2000), Murdoch and Green (1998)) or related methods, some progress has been made on such distributional identities if one assumes a particular form, such as

$$X \overset{\mathcal{L}}{=} U^{\alpha}(X + 1)$$

(its solutions are known as Vervaat perpetuities, Vervaat (1979). We refer to Kendall and Thönnes (2004), Fill and Huber (2009), Devroye (2001), and Devroye and Fawzi (2010) for worked out examples.

Identities like

$$X \overset{\mathcal{L}}{=} AX + B$$

occur in time series, random partitions, fragmentation processes, and as indirect descriptions of limit laws. Solutions are in the form of general perpetuities

$$X \overset{\mathcal{L}}{=} B_0 + \sum_{i=1}^{\infty} B_i \prod_{j=0}^{i-1} A_j,$$

where (A_i, B_i) are i.i.d. pairs distributed as (A, B). Necessary and sufficient conditions for the existence of solutions are known (Goldie and Maller, 2000; see also Alsmeyer and Iksanov, 2009, for further discussion). It suffices, for example, that

$$\mathbf{E}\{\log |A|\} \in (-\infty, 0), \mathbf{E}\{\log^+ |B|\} < \infty.$$

Yet one needs to describe those perpetuities that are in the standard class, and give algorithms for their generation.

Even more challenging are identities of the form

$$X \overset{\mathcal{L}}{=} \psi(X, X', U),$$

where X and X' on the right-hand-side are independent copies of X. Such identities do not lead to Markov chains. Instead, the repeated application of the map ψ produces an infinite binary tree. One should explore methods of random variate generation and constructively determine for which maps ψ, there is a solution that is in the standard class. A timid attempt for linear maps ψ was made by Devroye and Neininger (2002).

References

AHRENS, J.H. and DIETER, U. (1974): Computer methods for sampling from gamma, beta, Poisson and binomial distributions. *Computing, vol. 12, pp. 223–246.*

AKHIEZER, N.I. (1965): *The Classical Moment Problem*, Hafner, New York.

ALSMEYER, G. and IKSANOV, A. (2009): A log-type moment result for perpetuities and its application to martingales in supercritical branching random walks.' *Electronic Journal of Probability, vol. 14, pp. 289–313.*

ASMUSSEN, S., GLYNN, P. and THORISSON, H. (1992): Stationary detection in the initial transient problem. *ACM Transactions on Modeling and Computer Simulation, vol. 2, pp. 130–157.*

BAILEY, R.W. (1994): Polar generation of random variates with the t distribution (1994): *Mathematics of Computation, vol. 62, pp. 779–781.*

BONDESSON, L. (1982): On simulation from infinitely divisible distributions. *Advances in Applied Probability, vol. 14, pp. 855–869.*

BOX, G.E.P. and MÜLLER, M.E. (1958): A note on the generation of random normal deviates. *Annals of Mathematical Statistics, vol. 29, pp. 610–611.*

CHAMBERS J.M., MALLOWS, C.L. and STUCK, B.W. (1976): A method for simulating stable random variables. *Journal of the American Statistical Association, vol. 71, pp. 340–344.*

DEVROYE, L. (1981a): 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.*

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

DEVROYE, L. (1986a): *Non-Uniform Random Variate Generation*, Springer-Verlag, New York.

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

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

DEVROYE, L. (1991): 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.*

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

DEVROYE, L. (1997): Simulating theta random variates. *Statistics and Probability Letters, vol. 31, pp. 2785–2791.*

DEVROYE, L., FILL, J., and NEININGER, R. (2000): Perfect simulation from the quicksort limit distribution. *Electronic Communications in Probability, vol. 5, pp. 95–99.*

DEVROYE, L. (2001): Simulating perpetuities. *Methodologies and Computing in Applied Probability, vol. 3, pp. 97–115.*

DEVROYE, L. and NEININGER, R. (2002): Density approximation and exact simulation of random variables that are solutions of fixed-point equations. *Advances of Applied Probability*, vol. 34, pp. 441–468.

DEVROYE, L. (2009): On exact simulation algorithms for some distributions related to Jacobi theta functions. *Statistics and Probability Letters*, vol. 21, pp. 2251–2259.

DEVROYE, L. and FAWZI, O. (2010): Simulating the Dickman distribution. *Statistics and Probability Letters, vol. 80, pp. 242–247.*

FILL, J. (1998): An interruptible algorithm for perfect sampling via Markov chains. *The Annals of Applied Probability*, vol. 8, pp. 131–162.

FILL, J.A. and HUBER, M (2009): *Perfect simulation of perpetuities*, To appear.

FLAJOLET, P. and SAHEB, N. (1986): The complexity of generating an exponentially distributed variate. *Journal of Algorithms, vol. 7, pp. 463–488.*

GOLDIE, C.M. and MALLER, R.A. (2000): Stability of perpetuities. *Annals of Probability, vol. 28, pp. 1195–1218.*

GREEN, P.J. and MURDOCH, D.J. (2000): Exact sampling for Bayesian inference: towards general purpose algorithms (with discussion). In: *Monte Carlo Methods*, Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. (eds.), pp. 301–321, Bayesian Statistics, vol. 6, Oxford university Press, Oxford.

HASTINGS, C. (1955): *Approximations for Digital Computers*, Princeton University Press, Princeton, New Jersey.

HÖRMANN, W., LEYDOLD, J., and DERFLINGER, G. (2004): *Automatic Nonuniform Random Variate Generation*, Springer-Verlag, Berlin.

HUFFMAN, D. (1952): A method for the construction of minimum-redundancy codes. *Proceedings of the IRE, vol. 40, pp. 1098–1101.*

KANTER, M. (1975): Stable densities under change of scale and total variation inequalities. *Annals of Probability, vol. 3, pp. 697–707.*

KEANE, M.S., and O'BRIEN, G.L. (1994): A Bernoulli factory. *ACM Transactions on Modeling and Computer Simulation, vol. 4, pp. 213–219.*

KENDALL, W. (2004): Random walk CFTP. Thönnes ed., Department of Statistics, University of Warwick.

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

MARSAGLIA, G. (1968): Random numbers fall mainly in the planes. *Proceedings of the National Academy of Sciences, vol. 60, pp. 25–28.*

MARSAGLIA, G. and ZAMAN, A. (1991): A new class of random number generators. *Annals of Applied Probability, vol. 1, pp. 462–480.*

METROPOLIS, N., ROSENBLUTH, A., ROSENBLUTH, M., TELLER, A., and TELLER, E. (1953): Equations of state calculations by fast computing machines. *Journal of Chemical Physics, vol. 21, p. 1087–1091.*

MURDOCH, D.J. and GREEN, P.J. (1998): Exact sampling from a continous space. *Scandinavian Journal of Statistics, vol. 25, pp. 483–502.*

PROPP, G.J. and WILSON, D.B. (1996): Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms, vol. 9, pp. 223–252.*

RÖSLER, U. and RÜSHENDORF, L. (2001): The contraction method for recursive algorithms. *Algorithmica, vol. 29, pp. 3–33.*

K. SATO (2000): *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge.

ULRICH, U. (1984): Computer generation of distributions on the m-sphere. *Applied Statistics, vol. 33, pp. 158–163.*

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

VON NEUMANN, J. (1963): Various techniques used in connection with random digits. *Collected Works*, vol. 5, pp. 768–770, Pergamon Press. Also in (1951): Monte Carlo Method. *National Bureau of Standards Series, Vol. 12, pp. 36-38.*

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

ZOLOTAREV, V. M. (1959): On analytic properties of stable distribution laws. *Selected Translations in Mathematical Statistics and Probability, vol. 1, pp. 207–211.*

ZOLOTAREV, V. M. (1966): On the representation of stable laws by integrals. *Selected Translations in Mathematical Statistics and Probability, vol. 6, pp. 84–88.*

ZOLOTAREV, V. M. (1981): Integral transformations of distributions and estimates of parameters of multidimensional spherically symmetric stable laws. In: *Contributions to Probability*, pp. 283–305, Academic Press.

ZOLOTAREV, V. M. (1986): *One-Dimensional Stable Distributions*, American Mathematical Society, Providence, R.I..