SIMULATING THE DICKMAN DISTRIBUTION

LUC DEVROYE AND OMAR FAWZI

Abstract. In this paper, we give a simple algorithm for sampling from the Dickman distribution. It is based on coupling from the past with a suitable dominating Markov chain.

1. Introduction

Following Embrechts and Goldie (1994), a perpetuity is a random variable

$$Y = W_1 + W_1W_2 + W_1W_2W_3 + \cdots + W_1W_2 \ldots W_n + \cdots$$

where the W_i's are i.i.d. random variables distributed as W, where $W \geq 0$ and $\mathbb{E}W = m < 1$. It occurs in various fields such as financial modelling, hydrology, analytic number theory, insurance mathematics, the analysis of Hoare's selection algorithm, weighted path lengths in increasing trees, the limit law for sums of powers of edge lengths in random minimal spanning trees, and in the study of component spectra in logarithmic combinatorial structures. Key references are de Bruijn (1951); Goldie and Grübel (1996); Grübel and Rösler (1996); Hwang and Tsai (2002); Kesten (1973); Kuba and Panholzer (2006); Nietlispach (2007); Penrose and Wade (2004); Takács (1955); Vervaat (1979). These are special cases of more general random variables that may be written as

$$A_0 + A_1W_1 + A_2W_1W_2 + A_3W_1W_2W_3 + \cdots$$

(Embrechts et al., 1997, section 8.4) and that occur as solutions of random recurrence relations or in financial mathematics.

The special case of a Vervaat perpetuity occurs when $W = U^{1/\beta}$, $\beta > 0$, where U is a uniform $[0, 1]$ random variable. When $W = U$, we obtain the well-known Dickman distribution Dickman (1930).

The question we ask here is how we can exactly generate such random variates given that we have an infinite source of i.i.d. uniform $[0, 1]$ random variates at our disposal. The literature on approximate algorithms for perpetuities is large; see e.g., Chamayou (1997) or Chamayou and Dunau (2003). As our perpetuities form a special subfamily of the infinitely divisible distributions, one might consider applying general approximation algorithms for infinitely divisible distributions with given Lévy measures. The earliest mention of this is in Bondesson (1982) and Bouleau (1988). More recent work by Damien et al. (1995) and Walker (1996) uses Gibbs sampling.

Devroye (2001) derived a rejection algorithm for the Vervaat perpetuities. Knape (2006) refined this work. Fill (2004) and Fill and Huber (2009) found an exact algorithm based on the perfect simulation paradigm proposed by Propp and Wilson (1996). They had to modify the Propp-Wilson method and introduced the notion of a dominating Markov chain from which one could sample backwards. Kendall and Thönnes (2004) provide a similar method that can also be applied to Vervaat perpetuities.

In this paper, we give our own version of this dominated coupling from the past method, which distinguishes itself from the algorithms mentioned above in the closeness of the dominating chain's stationary distribution to the Dickman distribution. For example, from Goldie and Grübel (1996); Grübel and Rösler (1996), we know that the tails of Vervaat perpetuities drop off like the tails

of a Poisson distribution. So, by selecting a dominating Poisson random variable, things will fall into place very naturally. In our method for the Dickman distribution, the expected number of iterations before coalescence is less than 2.32.

2. The procedure

Preliminaries. Our objective is to sample from the stationary distribution of the Markov chain (Y_t) defined by

$$Y_{t+1} = (1 + Y_t)U_{t+1}.$$

where $U_1, U_2 \dots$ are i.i.d. uniform $[0, 1]$. We use a Markov chain defined in a slightly different way. Define the function f as

$$f : (x, u, v) \mapsto \lfloor u(x+1) \rfloor + \begin{cases} v & \text{if } \lfloor u(x+1) \rfloor \leq \lfloor x \rfloor \\ v(x - \lfloor x \rfloor) & \text{else} \end{cases}, (x, u, v) \in [0, \infty) \times [0, 1]^2.$$

and define the sequence

$$X_{t+1} = f(X_t, U_{t+1}, V_{t+1})$$

where U_{t+1} and V_{t+1} are independent uniform $[0, 1]$. For a fixed x, $f(x, U_{t+1}, V_{t+1})$ is uniformly distributed on $[0, x + 1]$. Thus, (X_t) and (Y_t) have the same transition kernel, and thus the same stationary distribution. The update function f is used instead of the simpler function $(x, u) \mapsto (1+x)u$ in order to ensure coupling in finite time. In fact, for fixed U and V, $(1+x)U \neq (1+x')U$ when $x \neq x'$ whereas the range of $x \mapsto f(x, U, V)$ is discrete.

For the algorithm we describe, f acts as a multigamma coupler (Murdoch and Green, 1998) $f'(x, u, v) = u(x + 1)\mathbb{1}_{u(x+1)\geq 1} + v\mathbb{1}_{u(x+1)<1}$. The reason we mention f instead of f' is that f has more chances of coalescence than f'. Even though this is not explicitly used by the algorithm we present, it might be useful in other contexts that we mention later in the paper. Moreover, it will be shown that f can be dominated using a simple update function g.

Coupling from the past cannot be applied directly to this Markov chain because it is not uniformly ergodic, which basically means that coalescence cannot happen (Foss and Tweedie, 1998). In Kendall (1998), Dominated CFTP was introduced in order to handle unbounded state spaces. The idea is to introduce a Markov chain (D_t) coupled to our Markov chain (X_t), such that $X_t \leq D_t$ for all t, and (D_t) has a stationary distribution that we can simulate.

Defining the dominating chain. Observe that

$$f(x, u, v) \leq g(\lfloor x \rfloor, u) + v,$$

where $g : (n, u) \mapsto \lfloor u(n + 2) \rfloor, (n, u) \in \mathbb{N} \times [0, 1]$. Define (Z_t) taking integer values as

$$Z_{t+1} = g(Z_t, U_{t+1}) = \lfloor U_{t+1}(Z_t + 2) \rfloor.$$

The sequence (Z_t) dominates (X_t) in the sense that if $\lfloor X_t \rfloor \leq Z_t$, then

$$\lfloor X_{t+1} \rfloor \leq \lfloor g(\lfloor X_t \rfloor, U_{t+1}) + V_{t+1} \rfloor \leq g(Z_t, U_{t+1}) = Z_{t+1}$$

with probability 1. Now the important observation is that the limiting distribution of the Markov chain (Z_t) is Poisson with parameter 1. In fact, if Z follows the Poisson distribution with parameter

1,

$$\begin{aligned}\mathbb{P}\{g(Z,U)=i\} &= \sum_{k=i-1}^{\infty} \mathbb{P}\{Z=k\}\mathbb{P}\{i \le U(k+2) < i+1\} \\ &= \frac{1}{e}\sum_{k=i-1}^{\infty} \frac{1}{k!}\frac{1}{k+2} \\ &= \frac{1}{e}\sum_{k=i-1}^{\infty} \Big[\frac{1}{(k+1)!} - \frac{1}{(k+2)!}\Big] \\ &= \mathbb{P}\{Z=i\}.\end{aligned}$$

We can thus simulate the stationary distribution of the dominating chain (Z_t). It remains to compute the time-reversal of the transition kernel of (Z_t), so that we can simulate the dominating chain into the past.

For $t \ge 0$, given Z_{-t}, we need to compute Z_{-t-1}, and the uniform U_{-t} that will be used to calculate X_{-t}. By definition of U_{-t}, we should have

$$Z_{-t} = \lfloor U_{-t}(Z_{-t-1}+2)\rfloor.$$

So it is easy to compute Z_{-t-1},

$$\mathbb{P}\{Z_{-t-1}=i|Z_{-t}=k\} = \frac{\mathbb{P}\{Z_{-t}=k|Z_{-t-1}=i\}\mathbb{P}\{Z_{-t-1}=i\}}{\mathbb{P}\{Z_{-t}=k\}}.$$

As we know that Z_{-t-1} follows the Poisson distribution with parameter 1,

$$\mathbb{P}\{Z_{-t-1}=i|Z_{-t}=k\} = \mathbb{1}_{i\ge k-1}\frac{1}{i+2}\frac{k!}{i!}. \tag{1}$$

Thus,

$$\mathbb{P}\{Z_{-t-1}=i|Z_{-t}=k\} = \mathbb{1}_{i\ge k-1}(i+1)\frac{k!}{(i+2)!} = \mathbb{1}_{i\ge k-1}k!\Big(\frac{1}{(i+1)!} - \frac{1}{(i+2)!}\Big).$$

The update random variable U_{-t} follows:

$$U_{-t} = \frac{Z_{-t}+U^*_{-t}}{Z_{-t-1}+2} \tag{2}$$

with U^*_{-t} an independent uniform variable. We can check that U_{-t} has the desired distribution. In fact, we know by construction that conditional on Z_{-t-1}, Z_{-t} is uniformly distributed on $\{0,\dots,Z_{-t-1}+1\}$, so it follows immediately that given Z_{-t-1}, U_{-t} follows the uniform $[0,1]$ distribution.

The algorithm. Note that if the dominating process reaches $Z_{-T}=0$ at step $-T$, then we are assured that all chains using update f beginning at time $-t'$ before $-T$ ($t'>T$) will coalesce at time $-T$. This is because, for a fixed U and V, the update $f(.,U,V)$ is built in such a way that its image has only one point in the interval $[0,1)$. So the algorithm waits until $Z_{-T}=0$, and then simulates the chain (X_t) forwards, starting at $X_{-T}=V_{-T}$, a uniform $[0,1]$ random variable, and computing the transitions using uniforms U_{-t} defined as in equation (2) and fresh uniforms V_{-t}, for $t \le T$.

Note that $Z_{-t} = 0$ is only a sufficient condition for coalescence, and it is possible for coalescence to happen even if $Z_{-t} > 0$. An algorithm that simulates two chains starting from V_{-t} and $Z_{-t}+V_{-t}$ and checks for coalescence at time 0 for every t would definitely require fewer steps into the past, but it will slightly complicate the implementation and actually slow down the execution because of the additional computations needed for testing coalescence. However, if the objective is to minimize the amount of randomness used in the procedure, testing for coalescence at each step can be a good idea. In this case, our coupler f will coalesce faster than a multigamma coupler. As an example, if $Z_{-t} = 1$, given that $Z_{-t+1} > 0$, the probability of coalescence in one step is a little more than 7%.

Here is an outline of our procedure:

Dickman distribution simulation

Generate a Poisson random variable Z_0 with parameter 1. Set t=0.
While $Z_{-t} > 0$ do
  Let W be a uniform $[0,1]$, and define $Z_{-t-1} = \max\{i \geq Z_{-t} - 1; W \geq 1 - \frac{Z_{-t}!}{(i+2)!}\}$.
  Define and store $U_{-t} = \frac{Z_{-t}+U^*_{-t}}{Z_{-t-1}+2}$ where U^*_{-t} is a fresh uniform $[0,1]$.
  $t \leftarrow t + 1$.
Let $T = t$.
Generate new independent uniform $[0,1]$ random variates $V_{-T}, V_{-T+1}, \ldots, V_0$.
Define $X_{-T} = V_{-T}$.
Compute $X_{-t+1} = f(X_{-t}, U_{-t+1}, V_{-t+1})$ forwards for $t = T$ to 1.
Return X_0.

Remark. The easiest way to simulate a Poisson random variable is to generate i.i.d exponential random variables $E_1, E_2, \ldots$ and return $\max\{n; E_1 + \cdots + E_n \leq 1\}$. Other methods can be found in Devroye (1986).

Remark. With probability e^{-1}, $Z_0 = 0$, in which case the method returns a uniform directly, with 0 steps into the past.

Theorem 1. *The algorithm terminates with probability* 1*, with an expected number of steps to the past* T *equal to* $1 + \int_0^1 \frac{e^t-1}{t} dt < 2.32$*. Moreover, the random variable* X_0 *returned by the algorithm has a Dickman distribution.*

Proof. For the analysis of the running time, see the next section.

Consider the sequence of random variables defined as follows. Y_0 is a uniform $[0,1]$, and Y_{t+1} is defined by $Y_{t+1} = W_{t+1}(Y_t + 1)$ with W_{t+1} independent uniform $[0,1]$ random variables. Let us define $(X_t^{-t_0})_{t \geq -t_0}$ by $X_{-t_0}^{-t_0} = V_{-t_0}$ and $X_{-t+1} = f(X_{-t}, U_{-t+1}, V_{-t+1})$, for $1 \leq t \leq t_0$. Then, clearly $X_0^{-t} \stackrel{\mathcal{L}}{=} Y_t$ for all t, so X_0^{-t} converges to a Dickman distribution because Y_t does so.

For T such that $Z_{-T} = 0$, consider $X_{-T}^{-t'}$ for $t' \geq T$, i.e., chains that start before time $-T$. Because the process $(Z_t)_{t \geq -t'}$ dominates $(X_t^{-t'})_{t \geq -t'}$, we have $X_{-T}^{-t'} = X_{-T}^{-T}$ for all $t' \geq T$. Hence, $X_0^{-t'} = X_0^{-T}$ because the same randomness is used. And because $T < \infty$ with probability 1, the sequence of random variables X_0^{-t} converges almost surely to X_0^{-T}. But as we know that X_0^{-t} converges weakly to the Dickman distribution, the returned random variable X_0^{-T} follows the Dickman distribution. □

Analysis of running time. The measure of the running time considered is the number of steps to the past needed to obtain coalescence. This time is the number of steps T taken by the time-reversal of the dominating chain to reach 0. To compute $\mathbb{E}\{T\}$, we consider T_{n0}, the number of steps to go from state n to state 0. It is worth observing that the time-reversed (Z_t) is downward skip-free, i.e., downward jumps are only of unit size. Thus, we start by calculating $\mathbb{E}\{T_{k(k-1)}\}$, the expected time to go from $Z = k$ to $Z = k - 1$.

Clearly,

$$\mathbb{E}\{T_{k(k-1)}\} = 1 + \mathbb{E}\{N_{\text{up}}\}\mathbb{E}\{T_{kk}^{\geq k}\},$$

where N_{up} is the number of excursions made from state k before going to state $k-1$, and $T_{kk}^{\geq k}$ is the length of one such excursion.

Starting at state k with $k \geq 1$, equation (1) implies that time-reversed (Z_t) goes to state $k-1$ with probability $\frac{k}{k+1}$. Thus,

$$\mathbb{E}\{N_{\text{up}} + 1\} = \frac{k+1}{k}.$$

Now we consider $\mathbb{E}\{T_{kk}^{\geq k}\}$. To evaluate this quantity, we consider the chain restricted to states $\geq k$. The quantity we are looking for is the mean recurrence time for state k in this chain. This time can be computed from the stationary distribution of this conditioned Markov chain. Recall the transition kernel of the time-reversal of the dominating Markov chain is defined by

$$\mathbb{P}\{Z_{-1} = i | Z_0 = j\} = \mathbb{1}_{i \geq j-1} \frac{1}{i+2} \frac{j!}{i!}.$$

Because of conditioning on $i \geq k$, the transition probabilities starting from state k should be divided by $\frac{1}{k+1}$. If $(p_i)_{i \geq k}$ is the stationary probability distribution, we have $p_{i+1} = \frac{1}{i+1} p_i$ for all $i \geq k$, as can be shown by induction. Indeed,

$$p_k = (k+1) \times \frac{1}{k+2} p_k + \frac{k+1}{k+2} p_{k+1},$$

which proves

$$p_{k+1} = \frac{1}{k+1} p_k.$$

And for $i \geq k$, by induction,

$$\begin{aligned} p_i &= (k+1) \times \frac{1}{i+2} \frac{k!}{i!} p_k + \sum_{j=k+1}^{i} \frac{1}{i+2} \frac{j!}{i!} p_j + \frac{i+1}{i+2} p_{i+1} \\ &= \frac{k+1}{i+2} \frac{k!}{i!} i \times \cdots \times (k+1) p_i + \sum_{j=k+1}^{i} \frac{1}{i+2} \frac{j!}{i!} i \times \cdots \times (j+1) p_i + \frac{i+1}{i+2} p_{i+1} \\ &= \frac{k+1}{i+2} p_i + \sum_{j=k+1}^{i} \frac{1}{i+2} p_i + \frac{i+1}{i+2} p_{i+1}. \end{aligned}$$

Hence,

$$p_{i+1} = \frac{1}{i+1} p_i.$$

Then, the fact that the probabilities sum to 1 gives

$$\sum_{i=k}^{\infty} \frac{1}{(k+1) \times \cdots \times i} p_k = 1.$$

So the mean recurrence time of state k is

$$\begin{aligned}\frac{1}{p_k} &= k! \sum_{i=k}^{\infty} \frac{1}{i!} \\ &\leq \begin{cases} e-1 & \text{if } k=1 \\ 1+\sum_{i=1}^{\infty} \frac{1}{(k+1)^i} & k \geq 2 \end{cases} \\ &\leq e-1.\end{aligned} \tag{3}$$

We can thus bound $\mathbb{E}\{T_{k(k-1)}\}$,

$$\mathbb{E}\{T_{k(k-1)}\} \leq 1 + \frac{1}{k}(e-1) \leq e,$$

and as a consequence

$$\mathbb{E}\{T_{n0}\} = \sum_{k=1}^{n} \mathbb{E}\{T_{k(k-1)}\} \leq en.$$

Finally, we get the expected running time,

$$\begin{aligned}\mathbb{E}\{T\} &\leq \frac{1}{e} \sum_{n=1}^{\infty} \frac{1}{n!} en \\ &= e.\end{aligned}$$

Remark. Note that by applying Taylor's formula to the exact expression of $\frac{1}{p_k}$ in equation (3), it can be shown that

$$\mathbb{E}\{T\} = 1 + \int_0^1 \frac{e^t - 1}{t} dt < 2.32.$$

Comparison with other methods. Kendall and Thönnes (2004) and Fill and Huber (2009) also give procedures to perfectly simulate the Dickman distribution using Dominated CFTP. In both methods, the dominating chain is a reflected random walk whose stationary distribution is geometric. Kendall and Thönnes (2004) use a multishift coupler whereas Fill and Huber (2009) use a simpler multigamma coupler. Although these methods are more general than ours in that they work for all Vervaat perpetuities, our method is more efficient for the Dickman distribution. In fact, the expected number of steps taken by Fill and Huber's algorithm is proven to be at most 17 and experiments seem to show it to be a little more than 6. No running time analysis is provided in Kendall and Thönnes (2004). It is also worth observing that Kendall and Thönnes' algorithm needs to check for coalescence by simulating an upper and lower chain, whereas both Fill and Huber's and ours can test coalescence at time $-t$ only using values of the random variables associated to time $-t$.

An implementation of our algorithm in C generated ten million samples in 5.52 seconds. On average, the number of steps into the past was 2.32 as predicted, and on 99% of the samples it was less than 14. Moreover, the maximum number of steps in these trials was 58. In order to compare execution times, we also implemented Fill and Huber's algorithm under the same circumstances. Generating ten million samples took 11.4 seconds. The number of steps was 6.57 on average, at least 29 for around 1% of the samples, and reached a maximum of 123 steps. The code used in these experiments can be found at `http://www.cs.mcgill.ca/~ofawzi/dickman/`.

Remark. This algorithm can be easily adapted to generate Vervaat perpetuities with $\beta \leq 1$. In fact, as $U^{1/\beta} \leq U$, the same dominating process (Z_t) can be used. The exact same procedure described for the Dickman distribution can be applied by letting $X_{-T} = V_{-T}^{1/\beta}$ and replacing the update function f by $f_\beta(x, u, v) = u^{1/\beta}(x+1)\mathbb{1}_{u^{1/\beta}(x+1)\geq 1} + v^{1/\beta}\mathbb{1}_{u^{1/\beta}(x+1)<1}$.

Acknowledgments

The authors would like to thank Jim Fill and an anonymous reviewer for many useful comments.

References

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

N. Bouleau. *Processus Stochastiques et Applications*. Hermann, Paris, 1988.

J.F. Chamayou. Pseudo random numbers for the Landau and Vavilov distributions. *Journal of Computational Physics*, 1997.

J.F. Chamayou and J.L. Dunau. Random difference equations: an asymptotic result. *Journal of Computational and Applied Mathematics*, 154:183–193, 2003.

P. Damien, P.W. Laud, and A.F.M. Smith. Approximate random variate generation from infinitely divisible distributions with applications to Bayesian inference. *Journal of the Royal Statistical Society*, B57:547–563, 1995.

N.G. de Bruijn. The asymptotic behaviour of a function occurring in the theory of primes. *Journal of the Indian Mathematical Society*, 15:25–32, 1951.

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

Luc Devroye. Simulating perpetuities. *Methodologies and Computing in Applied Probability*, 3: 97–115, 2001.

K. Dickman. On the frequency of numbers containing prime factors of certain relative magnitude. *Arkiv für Mathematik, Astronomi och Fysik*, 22:1–14, 1930.

P. Embrechts and C.M. Goldie. Perpetuities and random equations. In *Asymptotic Statistics: Proceedings of the Fifth Prague Symposium*, pages 75–86, Heidelberg, 1994. Physica-Verlag.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events*. Springer-Verlag, Berlin, 1997.

James Allen Fill. Personal communication, 2004.

James Allen Fill and Mark Huber. Perfect simulation of perpetuities. preprint, 2009. URL `http://arxiv.org/abs/0908.1733`.

S.G. Foss and R.L. Tweedie. Perfect simulation and backward coupling. *Stochastic Models*, 14: 187–203, 1998.

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

R. Grübel and U. Rösler. Asymptotic distribution theory for Hoare's selection algorithm. *Advances of Applied Probability*, 28:252–269, 1996.

Hsien-Kuei Hwang and Tsung-Hsi Tsai. Quickselect and the Dickman function. *Combinatorics, Probability and Computing*, 11:353–371, 2002.

W. Kendall. Perfect simulation for the area-interaction point process. In L. Accardi and C.C. Heyde, editors, *Probability Towards 2000*, pages 218–234. Springer-Verlag, 1998.

Wilfrid Kendall and Elke Thönnes. Random Walk CFTP. Technical report, Department of Statistics, University of Warwick, 2004.

H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131:207–248, 1973.

M. Knape. Approximating perpetuities. Diploma Thesis, Department of Mathematics, University of Frankfurt, 2006.

M. Kuba and A. Panholzer. On weighted path lengths and distances in increasing trees. Technical report, Institut für Diskrete Mathematik und Geometrie, Technical University of Vienna, 2006.

D.J. Murdoch and P.J. Green. Exact sampling from a continous space. *Scandinavian Journal of Statistics*, 25:483–502, 1998.

B. Nietlispach. *Quasi-logarithmic structures*. PhD thesis, Zurich, 2007.

M.D. Penrose and A.R. Wade. Random minimal directed spanning trees and Dickman-type distributions. *Advances in Applied Probability*, 36:691–714, 2004.

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

L. Takács. On stochastic processes connected with certain physical recording apparatuses. *Acta Mathematica Academiae Scientificarum Hungarica*, 6:363–379, 1955.

W. Vervaat. On a stochastic difference equation and a representation of nonnegative infinitely divisible random variables. *Advances in Applied Probability*, 11:750–783, 1979.

S.G. Walker. Random variate generation from an infinitely divisible distribution via Gibbs sampling. Technical report, Imperial College, London, 1996.

School of Computer Science, McGill University, 3480 University Street, Montreal H3A 2K6, Canada

E-mail address: lucdevroye@gmail.com

School of Computer Science, McGill University, 3480 University Street, Montreal H3A 2K6, Canada

E-mail address: ofawzi@cs.mcgill.ca