The Double CFTP method

Luc Devroye and Lancelot F. James

January 20, 2010

Abstract. We consider the problem of the exact simulation of random variables Z that satisfy the distributional identity $Z \stackrel{\mathcal{L}}{=} VY + (1-V)Z$, where $V \in [0,1]$ and Y are independent, and $\stackrel{\mathcal{L}}{=}$ denotes equality in distribution. Equivalently, Z is the limit of a Markov chain driven by that map. We give an algorithm that can be automated under the condition that we have a source capable of generating independent copies of Y, and that V has a density that can be evaluated in a black box format. The method uses a doubling trick for inducing coalescence in coupling from the past. Applications include exact samplers for many Dirichlet means, some two-parameter Poisson–Dirichlet means, and a host of other distributions related to occupation times of Bessel bridges that can be described by stochastic fixed point equations.

Keywords and phrases. Random variate generation. Perpetuities. Coupling from the past. Random partitions. Stochastic recurrences. Stochastic fixed point equations. Distribution theory. Markov chain Monte Carlo. Simulation. Expected time analysis. Bessel bridge. Poisson-Dirichlet. Dirichlet means.

AMS 2000 Mathematics Subject Classification: Primary 11K45. Secondary 65C05, 65C10, 68U20.

ACM 1998 Computing Classification System: G. Mathematics of Computing G.3 PROBABILITY AND STATISTICS Subjects: Probabilistic algorithms (including Monte Carlo), Random number generation.

Corresponding authors' address: Luc Devroye, School of Computer Science, McGill University, 3480 University Street, Montreal, Canada H3A 2K6. Second author's address: Department of Information Systems, Business Statistics and Operations Management, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong. The first author's research was sponsored by NSERC Grant A3456. The second author was supported by RGC-HKUST grant 600907 of the Hong Kong SAR.

Introduction

Consider random variables Z that satisfy the recursive distributional identity

$$Z \stackrel{\mathcal{L}}{=} VY + (1-V)Z, \tag{1}$$

where V, Y, Z on the right-hand side are independent, and $V \in [0,1]$, and $\stackrel{\mathcal{L}}{=}$ denotes equality in distribution. Such random variables are of interest to us. If $V_t, t \geq 1$ is an i.i.d. sequence distributed as V, $Y_t, t \geq 1$ is an i.i.d. sequence distributed as Y, and

$$W_1 = V_1, \quad W_t = \prod_{s=1}^{t-1}(1-V_s) \times V_t, \quad t \geq 2,$$

then expanding (1) over and over again shows that

$$Z \stackrel{\mathcal{L}}{=} \sum_{t=1}^{\infty} W_t Y_t. \tag{2}$$

The sequence W_t is also called a stick-breaking sequence. Given infinite resources, one could sample Z using the above infinite sum. But truncating that sum anywhere makes the result inaccurate. So, our first goal is to provide an exact simulation technique that is as general as possible so that we can deal with random variables Z as in (1). In the paper, we develop the Double CFTP (coupling from the past) method which is a particular extension of coupling from the past that uses a trick so that one does not need to know or even compute the distribution of Y. In other words, our method avoids case-by-case analysis and algorithm design, as the same program works for all Z under a modest condition on V.

It is important to note that in most cases, the density or distribution function of the solution Z of (1) cannot be written down in a simple form, yet our method permits its simulation. All other methods we are aware of require at the very least analytic knowledge of the law of Y.

In what follows, we adapt the following notation for random variables: U is uniform $[0,1]$, $B_{a,b}$ is beta (a,b), G_a is gamma (a) (with density $x^{a-1}e^{-x}/\Gamma(a), x>0$), E is exponential, ξ_p is Bernoulli (p), $\xi_{n,p}$ is binomial (n,p), P_λ is Poisson (λ), C is Cauchy, S_α ($\alpha \in (0,1)$) is positive stable of parameter α (its Laplace transform is $\mathbb{E}\{e^{-\lambda S_\alpha}\} = e^{-\lambda^\alpha}$), and N is standard normal. Furthermore, $L_\alpha \stackrel{\text{def}}{=} S_\alpha / S'_\alpha$ is the ratio of two independent positive stable random variables. We use the symbol L_α for this law because the given ratio has Lamperti's distribution [Zolotarev [1957], Lamperti [1958]; see also Zolotarev [1986], Chaumont and Yor [2003, p. 116] and James [2006]]: more specifically,

$$L_\alpha^\alpha \stackrel{\mathcal{L}}{=} \frac{\sin(U\pi\alpha)}{\sin((1-U)\pi\alpha)}.$$

Finally, when we write a distributional identity such as (1), it is always understood that all random variables on the right are independent. Accented random variables are independent copies of their unaccented cousins. For example, $U - U$ is zero, but $U - U'$ has a triangular distribution.

In a few cases, not of interest here, solutions Z of (1) can be explicitly derived. For example, when Y is symmetric stable $\mathcal{S}_\alpha$ (i.e., having characteristic function $\exp(-|t|^\alpha)$) and $V \equiv v \in (0,1)$ is

monoatomic (i.e., it puts all its mass on a point), then, using a sequence of i.i.d. symmetric stable random variables $\mathcal{S}_\alpha(t)$, $t \geq 1$, we have

$$Z \stackrel{\mathcal{L}}{=} \sum_{t=0}^{\infty} v(1-v)^t \mathcal{S}_\alpha(t) \stackrel{\mathcal{L}}{=} \left(\sum_{t=0}^{\infty} v^\alpha (1-v)^{\alpha t} \right)^{\frac{1}{\alpha}} \mathcal{S}_\alpha \stackrel{\mathcal{L}}{=} \frac{v}{(1-(1-v)^\alpha)^{\frac{1}{\alpha}}} \mathcal{S}_\alpha.$$

The Double CFTP algorithm

Doeblin coupling [1937] has been used in simulation in various ways, see, e.g., Devroye [1990], Thorisson [1986] and Lindvall [1992]. Coupling from the past, or CFTP, first suggested by Propp and Wilson [1996] but also implicit in the work of Asmussen, Glynn and Thorisson [1992], can be used to generate the unique stationary solution of Markovian maps. We use it for the map $Z \mapsto VY + (1-V)Z$, where V has a density on $[0,1]$. Note that $VY + (1-V)Z$ is a convex combination of Y and Z and takes values between Y and Z. Hence if $Y \in [0,c]$, then any solution Z of (1) satisfies $Z \in [0,c]$. Note also that $\mathbb{E}\{Z\} = \mathbb{E}\{Y\}$. Our set-up is the following:

(i) Y and V can be sampled from.

(ii) The density h of V can be computed exactly (in black box format).

(iii) The density of V is bounded from below on $[0,1]$ by a known constant $\beta > 0$. [This condition will be weakened further on.]

(iv) $0 \leq Y \leq c$, with c known, and Y is not monoatomic.

An important example occurs for $V = 1 - U^{1/\theta} \stackrel{\mathcal{L}}{=} B_{1,\theta}$, $\theta \in (0,1)$, in which case

$$h(x) = \theta(1-x)^{\theta-1} \mathbb{1}_{[0,1]}(x) \geq \theta \mathbb{1}_{[0,1]}(x).$$

Condition (iii) holds with $\beta = \theta$. Under the conditions above, applying the Markovian map $Z \mapsto VY + (1-V)Z$ starting with any $Z \in [0,c]$ gives a sequence of random variables that converges in distribution to the unique solution Z of

$$Z \stackrel{\mathcal{L}}{=} VY + (1-V)Z.$$

However, forward simulation, when stopped at a finite time, yields only approximations. To apply CFTP, one needs to be able to create coupled Markov chains that coalesce in (random) finite time with probability one.

We stress the fact that the algorithm below can basically be automated, something that differentiates it from most other attempts at applying CFTP to absolutely continuous distributions. Closest in spirit is the work of Murdoch and Green [1998, 1999] who look at the Markov map, and based upon explicit knowledge of the distribution of the random element, coalescence is forced by coupling using the minimal common density. However, that approach cannot be automated—the minimizing density needs to be computed by hand, which, in our case implies tedious computations involving the distributions of V and Y.

Automation can be achieved by the doubling trick of this paper: the distribution of Y does not have to be known or restricted. Coalescence occurs with a given positive probability at every step of our Markov chain by making use of the fact that if we randomize the transition by introducing an additional random dichotomy, then the conditional densities in the transition can be shown to have a fixed amount of overlap regardless of the current state. This is exploited in the method—thus, the doubling, or dichotomy, has two benefits, automation and coalescence.

We also mention for the sake of completeness other versions of CFTP, such as one shot CFTP [Roberts and Rosenthal, 2002, Beskos and Roberts, 2005], read once CFTP [Wilson, 2000a], CFTP with a dominating chain [Fill and Huber, 2004; Kendall and Thönnes, 2004], and Fill's method [Fill, 1997, 1998; Fill, Machida, Murdoch and Rosenthal, 2000]. More early work can be found in Fismen [1998], Kendall [1998], Wilson [2000b] and Foss and Tweedie [1998]. Some of these modifications of CFTP were created specifically to deal with unbounded spaces [Murdoch, 2000], and non-uniformly ergodic chains. Our state space is compact, so we won't need to consider any of those more advanced methods.

The main contribution of this paper is an algorithm, called Double CFTP, that achieves the sought automation. To grasp the idea of CFTP, imagine that the Markov chain Z_t runs as follows:

$$Z_{t+1} = \varphi(Z_t, R_t),$$

where φ is a certain function, and $R_t, t = \ldots, -2, -1, 0, 1, 2, \ldots$ are i.i.d. random elements (typically random vectors). Assume that this vector of R_t's is stored once and for all, for all t. This would force a simulation started at $Z_t = z_t$ for negative integer time t, to assume a value $\psi(z_t, R_t, R_{t+1}, \ldots, R_{-1})$ at time zero, where ψ is some map. We declare coalescence if all z_t yield the same value at time zero. In that case, as shown by Propp and Wilson [1996], the common value at time zero has the limit distribution. To detect coalescence, one can try running simulations first starting at $t = -1$, then from $t = -2$, until one is successful, say, starting at T. This necessitates the storage of a finite number of values, $R_{-T}, \ldots, R_{-1}$. But of course, there are uncountably many possible start values z_t, and one can't possibly run that many simulations. We develop a new method to deal with this.

If we were to define the Markov chain by

$$Z_{t+1} = V_t Y_t + (1 - V_t) Z_t,$$

with the Y_t's i.i.d. and distributed as Y, and the V_t's i.i.d. and distributed as V, we would have our set-up, with $R_t = (Y_t, V_t)$. However, coalescence is difficult to achieve in this manner. The reason is the following: the density of $X = Z_{t+1}$ given $Y_t = y$ and $Z_t = z$ is

$$f(x) = \frac{1}{|y - z|} h\left(\frac{x - z}{y - z}\right).$$

Note that f is zero unless either $z \le x \le y$ or $y \le x \le z$. Now, for every fixed x,

$$\inf_{z \in [0,c]} f(x) = 0.$$

Thus, there is no minorizing function of nonzero integral that works for all z, and so there is no hope for coalescence.

We are allowed to define φ and the random element in any equivalent manner, and thus, we can think about doubling the number of Y_t's. Assume that we have an additional i.i.d. sequence Y'_t, and a further i.i.d. sequence of Bernoulli (1/2) random variables B_t, and that we set

$$Z_{t+1} = V_t \left(Y_t B_t + Y'_t (1 - B_t) \right) + (1 - V_t) Z_t.$$

The random element is $R_t = (Y_t, Y'_t, V_t, B_t)$. It is clear that this is an equivalent Markov chain. But we still cannot hope to get coalescence, as this is a strictly monotone map in Z_t. The idea is to represent the density of Z_{t+1} in a special manner.

We drop the index t for now. Conditional on Y, Y', Z, the density of $V \left(YB + Y'(1 - B) \right) + (1 - V)Z$ is the mixture density

$$f(x) = \frac{1}{2} \left(h \left(\frac{x - Z}{Y - Z} \right) \frac{1}{|Y - Z|} + h \left(\frac{x - Z}{Y' - Z} \right) \frac{1}{|Y' - Z|} \right).$$

We clearly have

$$\inf_{x \in [\min(Y,Y'), \max(Y,Y')]} \inf_{Z \in [0,c]} f(x) \geq \frac{\beta}{2c},$$

where β is as in (iii) and c as in (iv). Thus, an equivalent decomposition is, using the notation $g_{a,b}$ for the uniform density on $[\min(a, b), \max(a, b)]$,

$$f(x) = f_1(x) + f_2(x),$$

with

$$f_1(x) = \frac{|Y - Y'|\beta}{2c} g_{Y,Y'}(x),$$
$$f_2(x) = \frac{1}{2} \left(h \left(\frac{x - Z}{Y - Z} \right) \frac{1}{|Y - Z|} + h \left(\frac{x - Z}{Y' - Z} \right) \frac{1}{|Y' - Z|} \right) - \frac{|Y - Y'|\beta}{2c} g_{Y,Y'}(x).$$

This decomposition is useful because f_1 does not depend upon Z, and will permit us to create a coalescent step. The probability of coalescence, given Y, Y', is $\int f_1 = |Y - Y'|\beta/(2c)$.

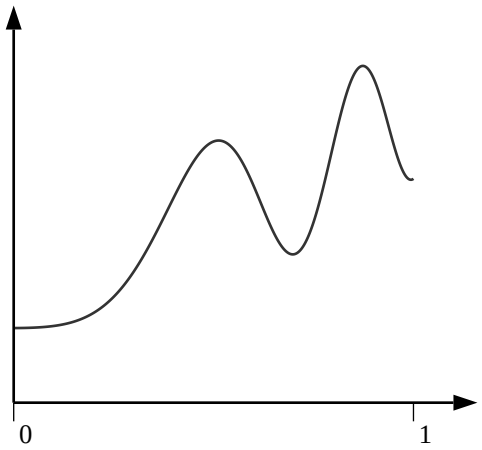

Fig. 1. The density h of V is shown. It is bounded away from zero on $[0, 1]$. This is the density used in Figure 2.

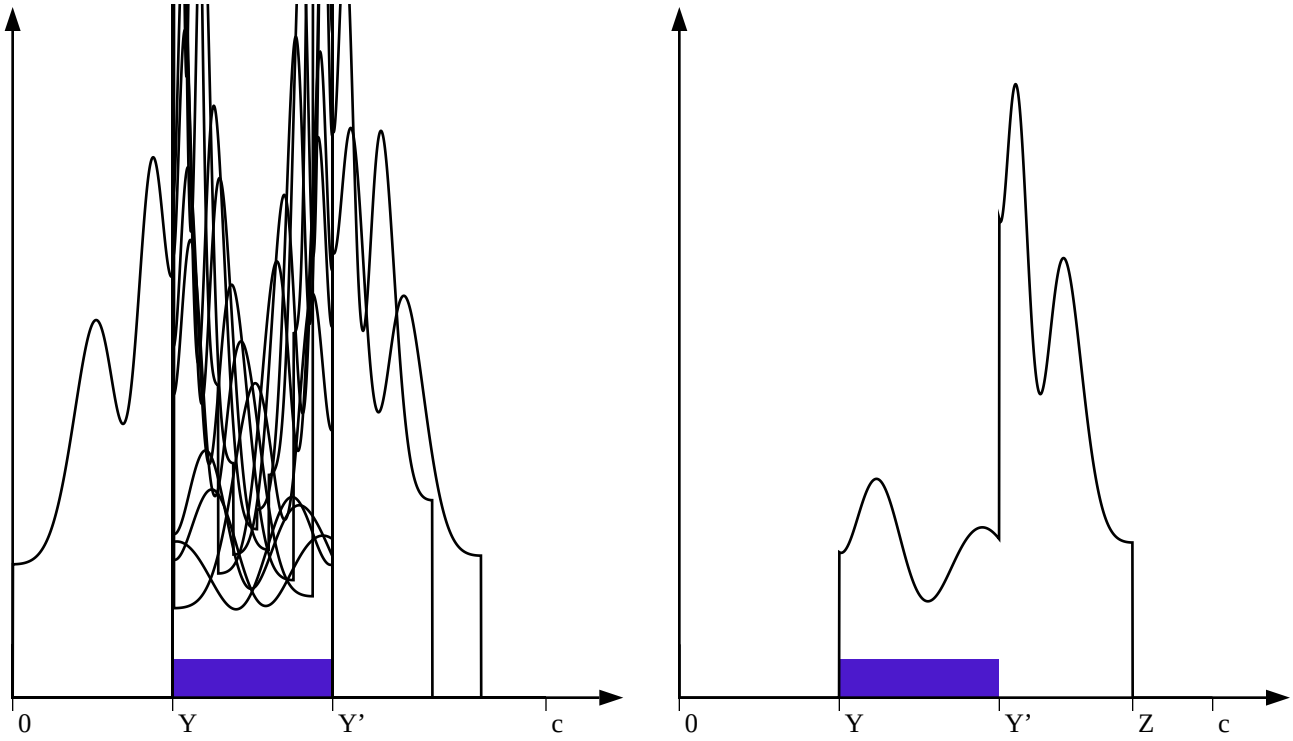

Fig. 2. On the right, the density f is shown for fixed values of Y, Y', Z when V has density h given in Figure 1. The blacked out box represents the lower bound f_1, valid no matter how Z is chosen on the interval. On the left, ten different densities f are shown for fixed Y and Y' but varying Z. Note that all are strictly above f_1.

Note that f_1 and f_2 are nonnegative functions, and that $f = f_1 + f_2$ is again a mixture decomposition. Let us make the Markovian map a bit more complicated by enlarging the random element with a random uniform $[0,1]$ random variable U and replace V_t by a yet undescribed additional independent random element R'_t. With subscripts t, we would thus have $R_t = (Y_t, Y'_t, U_t, R'_t)$. The map that uses $f = f_1 + f_2$, and the fact that $\int f_1 = |Y - Y'|\beta/2c$ would be as follows:

$$Z_{t+1} = \begin{cases} \min(Y_t, Y'_t) + \frac{2cU_t}{\beta} & \text{if } U_t \le \frac{|Y_t - Y'_t|\beta}{2c} \\ \psi(Z_t, Y_t, Y'_t, R'_t) & \text{otherwise,} \end{cases},$$

where $\psi(Z, Y, Y, R)$ is a random variable with density proportional to f_2. Note that if $U_t \le \frac{|Y_t - Y'_t|\beta}{2c}$, then $2cU_t/\beta$ is uniform on $[0, |Y_t - Y'_t|]$, as required. Note that if Z_{t+1} is generated according to f_1 (first step above), then its value does not depend upon Z_t. Thus, coalescence occurs at that time!

Before we even discuss the simple problem of the structure of ψ for f_2, it suffices to note that we can simply determine the first negative time at which coalescence occurs: just keep generating triples (Y_t, Y'_t, U_t), $t = -1, -2, \ldots$ until for the first time

$$U_t \le \frac{|Y_t - Y'_t|\beta}{2c}.$$

That random time will be called T. We have as starting point a random variable uniformly distributed on $[\min(Y_T, Y'_T), \max(Y_T, Y'_T)]$:

$$Z_{T+1} = \min(Y_T, Y'_T) + \frac{2cU_T}{\beta}.$$

We need to apply the transitions from there forward, all according to f_2, using the random elements $R_{T+1}, \ldots, R_{-1}$. Now, fortunately, we do not need to store the full random elements, as Y_t, Y'_t suffice.

Let us describe ψ, or the generator for f_2. We are given Y, Y' and Z. One can clearly apply the rejection method with bounding density f. A simple algorithm to generate a random variate X with density f_2 (given Y, Y', Z) is as follows:

```
[Forward phase:  Generator for f_2 given Y, Y', Z.]
Repeat Generate U' uniform [0,1], B Bernoulli (1/2)
       Generate V with density h
       set X = (1 − V)Z + VYB + VY'(1 − B)
Until U'f(X) > β/(2c) [this is U'f(X) > f_1(X)] or X < min(Y, Y') or X > max(Y, Y')
Return X
```

We summarize the entire algorithm. A "stack" is a computer science term for a structure that accepts items, one on top of another, with addition and removal always done from the top (as with a stack of pancakes). To "pop" a stack means to remove the top element. The popped element is the element removed in that operation.

```
Define an empty stack S
[Backward phase until coalescence is detected.]
Repeat
       Generate U uniform [0,1]
       and generate i.i.d. random variates Y, Y'
       Put (Y, Y') on the stack S
Until U ≤ |Y − Y'|β/(2c).
Pop S
[Set starting point.]
Set Z = min(Y, Y') + 2cU/β (uniform on [min(Y, Y'), max(Y, Y')])
[Forward phase.]
While S is not empty:
       Pop S and let the popped element be (Y, Y')
       Generate X using the generator for f_2 for given Y, Y', Z
       Z ← X
Return Z
```

The stack size is proportional to the time taken by the algorithm. The probability of ending the piling of triples on the stack at each iteration is

$$p = \mathbb{P}\left\{U \le \frac{|Y - Y'|\beta}{2c}\right\} = \frac{\beta\mathbb{E}\left\{|Y - Y'|\right\}}{2c}.$$

Therefore, the expected stack size is exactly

$$\frac{2c}{\beta\mathbb{E}\{|Y - Y'|\}}.$$

The performance deteriorates with a worsening ratio of the spread of Y over the variability of Y. We would never use this algorithm if Y were a constant with probability one: in that case, $Z \equiv Y$, and nothing is to be done.

It is noteworthy that no computation of the density of Y is required in this procedure.

For the complexity, we count the total number of rejection steps over all iterations. This can be looked at backwards, from time 0 down. Let N_t be the number of rejection steps in the algorithm for f_2 at time t. As noted earlier, the total number of steps, and indeed the total stack size, T, is geometric, and

$$\mathbb{E}\{T\} = \frac{2c}{\beta\mathbb{E}\{|Y - Y'|\}}.$$

Furthermore, by Wald's identity,

$$\mathbb{E}\left\{\sum_{t=1}^{T} N_t\right\} = \mathbb{E}\{T\}\mathbb{E}\{N_1\}.$$

Given Y, Y', the expected number of steps for the rejection method for f_2 is

$$\begin{cases} \frac{1}{1-\frac{|Y-Y'|\beta}{2c}} & \text{with probability } 1 - \frac{|Y-Y'|\beta}{2c}, \\ 1 & \text{with probability } \frac{|Y-Y'|\beta}{2c}. \end{cases}$$

Thus, $\mathbb{E}\{N_1\} = 1 + \mathbb{E}\{|Y - Y'|\}\beta/(2c)$. We conclude that

$$\mathbb{E}\left\{\sum_{i=t}^{T} N_t\right\} = 1 + \frac{2c}{\beta\mathbb{E}\{|Y - Y'|\}}.$$

The rejection algorithm for f_2 can be avoided in special cases, provided that one can, for example, compute the distribution function for f_2 explicitly, and apply inversion to generate random variates from it. When V is uniform $[0, 1]$, say, this is easy to do. It results in small computational savings.

In some cases, programs may run out of storage when the stack size exceeds available space. The stack can be avoided altogether at a serious cost in terms of number of Y and U variates. Without a stack, we must reconstruct the (non-)coalescence decision. At the start of the forward phase, we need to generate a triple Y, Y', U that satisfies

$$U \leq |Y - Y'|\beta/(2c).$$

The last triple generated in the backward phase can be employed for that. Using Y, Y', a random variate with density f_1 can be used to start up the forward chain. But every other forward step requires a number of attempts: we must generate triples Y, Y', U until for the first time

$$U > |Y - Y'|\beta/(2c).$$

Then Y and Y' can be used as in the stack algorithm.

Double CFTP when the mixing density is not bounded away from zero

Finally, we can relax the boundedness condition on V given in (iii). It suffices that the density h of V can be bounded from below by a nonincreasing function g, such that $g(x) > 0$ for all $x \in (0, 1)$. This enlarges the field of application of Double CFTP considerably. We replace condition (iii) above by

(iii*) The density of V is bounded from below on $[0, 1]$ by a known function g that is nonincreasing and strictly positive for all $x \in (0, 1)$. Let $G(x) = \int_0^x g(u)\,du$ denote the integral, and let G^{inv} be the inverse of G.

In Figure 3, we show a possible density for V, together with the corresponding density f for fixed values of Z, Y and Y' (in the notation of the previous section). By the monotone nature of g, it takes just a moment to verify that, given $Y' > Y$, and uniformly over all values of Z,

$$f(x) \geq f_1(x) \stackrel{\text{def}}{=} \frac{1}{2c} \min\left(g\left(\frac{x}{Y'}\right), g\left(\frac{c-x}{c-Y}\right)\right) \mathbb{1}_{[Y,Y']}(x).$$

Examples of densities that are relevant to Dirichlet means and their generalizations include that of $B_{a,b}$ with $a \in (0, 1]$ and $b > 0$.

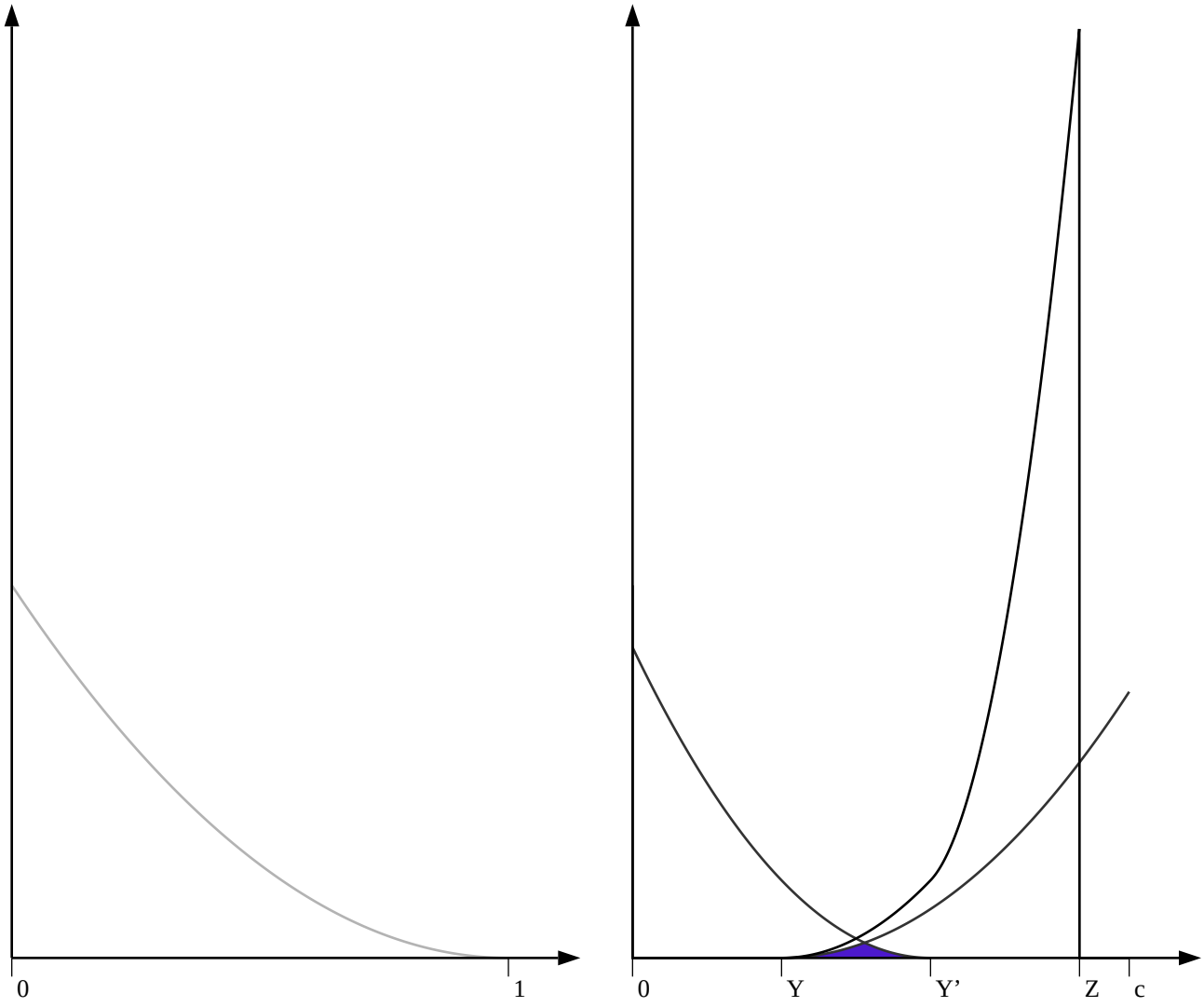

Fig. 3. On the left, the density h of V is given, and we can use $g = f$ because of monotonicity. On the right, the top curve with the large peak at Z is the density f for fixed values of Y, Y', Z. The shaded area represents the surface under the lower bound f_1, valid no matter how Z is chosen on the interval (but with Y and Y' fixed). The two curves hugging the shaded area just above it are the two functions g in the definition of f_1.

A few remarks regarding f_1 will be helpful here. Define the cut-off point $Y'' = Y'c/(c+\Delta)$, where $\Delta = Y' - Y$. This is the place where the two functions in the definition of the minimum in f_1 are equal. We have

$$\int_Y^{Y''} g\left(\frac{c-x}{c-Y}\right)\,dx = (c-Y)\int_{\frac{c}{c+\Delta}}^1 g(x)\,dx = (c-Y)\left(G(1) - G\left(\frac{c}{c+\Delta}\right)\right),$$

and

$$\int_{Y''}^{Y'} g\left(\frac{x}{Y'}\right)\,dx = Y'\int_{\frac{c}{c+\Delta}}^1 g(x)\,dx = Y'\left(G(1) - G\left(\frac{c}{c+\Delta}\right)\right),$$

Therefore,

$$p \stackrel{\text{def}}{=} \int f_1(x)\,dx = \frac{c+\Delta}{2c}\left(G(1) - G\left(\frac{c}{c+\Delta}\right)\right).$$

With probability p, we can cause coalescence, as f_1 does not depend upon Z. If $g \equiv \beta$, then verify that $p = \beta\Delta/(2c)$, just as in the original Double CFTP algorithm.

We also need a method for generating a random variable with density proportional to f_1. The part of f_1 corresponding to $[Y, Y'']$ has weight $(c-Y)/(c+\Delta)$, and the interval $[Y'', Y']$ has weight $Y'/(c+\Delta)$. Let Q be a random variable with density proportional to g on $[c/(c+\Delta), 1]$. It can be generated by inversion as

$$G^{\text{inv}}\left(G\left(\frac{c}{c+\Delta}\right) + U\left(G(1) - G\left(\frac{c}{c+\Delta}\right)\right)\right),$$

where U is uniform $[0,1]$. Then it is easy to see that a generator for f_1 is given by:

```
[Generator for f_1, given Y < Y'.]
Set Δ = Y' − Y
Generate a pair of i.i.d. uniform [0,1] random variables U, U'
Set Q ← G^inv( G(c/(c+Δ)) + U( G(1) − G(c/(c+Δ)) ) )
If U' < (c − Y)/(c + Δ) then X ← c − Q(c − Y)
                        else X ← QY'
Return X
```

The algorithm above suggests that we should be stacking (Y, Y') in the backwards phase. Note that f_1 is used to start the forward phase, just when coalescence occurs. In every other forward step, we need to generate a random variable from $f_2 = f - f_1$ using the current value of Z, and the appropriate pair Y, Y'. Modifying the earlier algorithm to replace the test $U'f(X) > \beta/(2c)$ by the test $U'f(X) > f_1(X)$, we obtain:

```
[Forward phase:  Generator for f_2 given Y, Y', Z.  Recall Y < Y'.]
Repeat Generate U' uniform [0,1], B Bernoulli (1/2)
       Generate V with density h
       Set X = (1 − V)Z + VYB + VY'(1 − B)
Until U'f(X) > f_1(X) or X < Y or X > Y
Return X
```

We repeat the Double CFTP algorithm with the necessary modifications.

```
Define an empty stack S
[Backward phase until coalescence is detected.]
Repeat
       Generate an i.i.d. pair Y,Y', and swap them if needed to make Y < Y'.  Define Δ = Y' − Y
       Generate U uniform [0,1]
       Put (Y,Y') on the stack S
Until U ≤ p ≝ (c+Δ)/(2c) (G(1) − G(c/(c+Δ)))
Pop S
[Set starting point.]
Generater Z with density proportional to f1, given the last pair Y,Y'
[Forward phase.]
While S is not empty:
       Pop S and let the top element be (Y,Y')
       Generate X using the generator for f2 for given Y,Y',Z
       Z ← X
Return Z
```

The analysis of this algorithm is not different from that of the original Double CFTP algorithm. The expected stack size (length of the backward phase) is $1/\mathbb{E}\{p\}$: if Y, Y' are i.i.d., and have the common distribution required for the recurrence, this is

$$\alpha \stackrel{\text{def}}{=} \frac{1}{\mathbb{E}\{p\}} = \frac{2c}{\mathbb{E}\left\{(c+|Y'-Y|)\left(G(1)-G\left(\frac{c}{c+|Y'-Y|}\right)\right)\right\}}.$$

Each forward step has a hidden second layer of iterations for the rejection generator for f_2. Arguing as earlier, the expected number of inner iterations is precisely $1+\alpha$. Observe also that conditions (iii*) and (iv) imply $\alpha < \infty$.

In the majority of the cases of interest in this paper, one can take $h \equiv g$, so that, among other things, $G(1) = 1$ in all formulas above.

Dirichlet means

Let $\theta > 0$. The class of random variables Z defined by

$$Z \stackrel{\mathcal{L}}{=} VY + (1-V)Z \tag{3}$$

where $V = 1 - U^{1/\theta} \stackrel{\mathcal{L}}{=} B_{1,\theta}$ are known in the literature as *Dirichlet means*. We say that Z is generated by Y, or that Y is the *generator* of Z. We will write Z_θ for such random variables without showing the dependence upon the law of Y in the notation. Thus, specifically,

$$Z_\theta \stackrel{\mathcal{L}}{=} B_{1,\theta}Y + (1-B_{1,\theta})Z_\theta \stackrel{\mathcal{L}}{=} B_{\theta,1}Z_\theta + (1-B_{\theta,1})Y. \tag{4}$$

Observe that for bounded positive Y, our algorithm can be used to perfectly sample Z_θ. In some cases, Dirichlet means have simple solutions:

(i) It is well known [see, e.g., Ferguson, 1973], that if $Y = \xi_q$ and $V = B_{1,\theta}$, then $Z_\theta \stackrel{\mathcal{L}}{=} B_{\theta q, \theta(1-q)}$.

(ii) Diaconis and Kemperman [1996] (see also Diaconis and Freedman [1999]) have shown that if U, U' (both uniform $[0,1]$) and C (Cauchy) are independent, then

$$(1-U')U + U'\frac{e^C}{e^C+1} \stackrel{\mathcal{L}}{=} U.$$

(iii) The Dirichlet mean with $Y \equiv U$ is trivially sampled from by Double CFTP because U is supported on $[0,1]$. When $\theta = 1$, Diaconis and Kemperman [1996] have shown that the density of Z is

$$\frac{e}{\pi}\sin(\pi z) z^{-z}(1-z)^{-(1-z)}\,\mathbb{1}_{[0,1]}(z).$$

This is the only case in which the density for this Dirichlet subfamily is explicitly known. Sampling from Z in this example is best achieved by rejection, as the Diaconis-Kemperman density is supported on $[0,1]$ and unimodal with mode at $1/2$ given by $2e/\pi$. This suggests a simple rejection algorithm. However, for $\theta \neq 1$, Double CFTP seems like the simplest option. In fact, we do not know any other exact simulation approach.

Guglielmi, Holmes and Walker [2002] considered the same problem we considered. For bounded $Y \in [0,c]$, they also suggest simulating Z by CFTP using the Markov chain proposed in this paper. However, they run a classical monotone chain, with simulations starting at 0 and c, and wait until approximate coalescence occurs, i.e., until the upper and lower chains are virtually indistinguishable up to machine precision. They ran simulations, which we will compare with ours further on. However, their approach is in fact an approximation.

The general study of the distributional properties of these random variables can be traced back to the work of Cifarelli and Regazzini [1979, 1990]. There are of course examples where the density of Z_θ can either be found by inversion formulae as above or by other means. See for instance Cifarelli and Regazzini [1979, 1990], Diaconis and Kemperman [1996], Cifarelli and Melilli [2001] and more recently James [2007, 2008], James Lijoi and Prünster [2008] and Bertoin, Fujita, Roynette and Yor [2006]. Some of these examples along with additional reference are in the survey paper by James, Roynette and Yor [2008]. However in general it is a difficult task to obtain explicit densities for these quantities, especially in the case where $0 < \theta < 1$.

Let us take the example $V = 1 - U^{1/\theta} \stackrel{\mathcal{L}}{=} B_{1,\theta}$, $Y \equiv U$, discussed above, and analyze the expected time performance with respect to θ. Simulations are reported in a later section. For $\theta \leq 1$, we can use the first algorithm with $\beta = \theta$, $c = 1$. Note that $|Y - Y'| \stackrel{\mathcal{L}}{=} B_{1,2}$. The expected stack size, which is proportional to the overall time taken by the algorithm, is

$$\frac{2c}{\beta \mathbb{E}\{|Y-Y'|\}} = \frac{6}{\theta},$$

and deteriorates polynomially as $\theta \downarrow 0$. For $\theta \geq 1$, only the second algorithm will be applicable. Let us choose $g \equiv h$, as h is nonincreasing. Noting that

$$G(x) = 1 - (1-x)^\theta,$$

we see that the expected stack size is

$$\frac{2}{\mathbb{E}\left\{(1+|Y'-Y|)\left(\left(1-\frac{1}{1+|Y'-Y|}\right)^\theta\right)\right\}} = \frac{2}{\mathbb{E}\left\{(1+|Y'-Y|)^{1-\theta}|Y'-Y|^\theta\right\}}.$$

The expected value in this expression is

$$\int_0^1 2(1-x^2)\left(\frac{x}{1+x}\right)^\theta dx.$$

By the dominated convergence theorem, one can show that as $\theta \uparrow \infty$, this is asymptotic to $(4/3)2^{-\theta}$. Thus, the expected stack size grows as

$$(3/2)2^\theta.$$

Our simulation results will confirm this behaviour.

Generalized gamma convolutions

We say that X is a generalized gamma convolution (or GGC) with parameters (θ, Y) and write GGC(θ, Y), with Y being a random variable and $\theta > 0$ if $X \stackrel{\mathcal{L}}{=} G_\theta Z_\theta$ and Z_θ is defined by the recursive identity

$$Z_\theta \stackrel{\mathcal{L}}{=} B_{\theta,1} Z_\theta + (1 - B_{\theta,1})Y \tag{5}$$

[Bondesson [1992]; see also Cifarelli and Regazzini, 1990]. This definition follows along the lines of James, Roynette and Yor [2008]. The Double CFTP method thus permits the exact simulation of all GGC(θ, Y) whenever Y is of compact support. We believe that no method has ever been published for the exact simulation of GGC random variables.

We recall that G_θ can be generated in expected time uniformly bounded over the parameter range, see, e.g., Ahrens and Dieter [1974, 1982], Ahrens, Kohrt and Dieter [1983], Best [1978a, 1978b, 1983], Cheng [1977], Cheng and Feast [1979, 1980], Devroye [1986], LeMinh [1988], Marsaglia [1977], and Marsaglia and Tsang [2001].

Two parameter Poisson–Dirichlet means

We now deal with some situations in which the V_t's are no longer i.i.d. Interestingly, we can say quite a bit for random variables related to mean functionals of the two parameter Poisson–Dirichlet random probabilty measure, otherwise known as the Pitman-Yor process [Pitman and Yor, 1997a] as named in Ishwaran and James [2001], and parametrized by $0 \leq \alpha < 1$ and $\theta > -\alpha$. When $\alpha = 0$ this model reduces to the Dirichlet process discussed earlier in this paper. Hereafter we assume $0 < \alpha < 1$.

The processes arose originally in the work of Perman, Pitman and Yor [1992] and Pitman and Yor [1997a] in connection with excursions of Bessel processes. Ishwaran and James [2001, 2003] suggested

these as natural priors for Bayesian applications [see also Pitman [1996]]. Recently, Goldwater, Griffiths and Johnson [2006] and Teh [2006] have argued that the power law behavior, in terms of numbers of distinct words, induced by general Pitman-Yor processes is appropriate when applied to various natural language models. James, Lijoi and Prünster [2008] describe sums in which V is a general beta, in the context of Poisson-Dirichlet processes.

Similar to the Dirichlet means case, we define a family of random variables, which we shall call Poisson–Dirichlet means and denote by $Z_{\alpha,\theta}$ [James, Lijoi and Prünster, 2008]. This family satisfies the distributional identity

$$Z_{\alpha,\theta} \stackrel{\mathcal{L}}{=} B_{\theta+\alpha,1-\alpha} Z_{\alpha,\theta+\alpha} + (1 - B_{\theta+\alpha,1-\alpha})Y, \tag{6}$$

where $\theta > -\alpha$, and Y is a random variable said to be the "generator" of $Z_{\alpha,\theta}$. Repeated expansions yields

$$Z_{\alpha,\theta} \stackrel{\mathcal{L}}{=} \sum_{t=1}^{\infty} Y_t V_t \prod_{s=1}^{t-1} (1 - V_s),$$

where all variables on the right are independent, the Y_t are i.i.d. and distributed as Y, and

$$V_t = B_{1-\alpha,\theta+t\alpha}, t \geq 1.$$

There is no known technique to perfectly sample $Z_{\alpha,\theta}$. The difficulty stems from the fact that the V_t in the infinite sum are not identically distributed. However, Theorem 2.1 of James, Lijoi and Prünster [2008] shows that for $\theta > 0$, $Z_{\alpha,\theta}$ is in fact a Dirichlet mean of order θ generated by $Z_{\alpha,0}$. In particular, Theorem 6.1 of that paper shows that

$$Z_{\alpha,\theta} \stackrel{\mathcal{L}}{=} B_{\theta,1} Z_{\alpha,\theta} + (1 - B_{\theta,1}) Z_{\alpha,0}. \tag{7}$$

Thus, we can perfectly sample $Z_{\alpha,\theta}$ by Double CFTP whenever $Z_{\alpha,0}$ is of bounded support and can be perfectly sampled! Note that

$$Z_{\alpha,0} \stackrel{\mathcal{L}}{=} \sum_{t=1}^{\infty} Y_t V_t \prod_{s=1}^{t-1} (1 - V_s)$$

where $V_t \stackrel{\mathcal{L}}{=} B_{1-\alpha,t\alpha}$ are independent. There is no indication that $Z_{\alpha,0}$ is a Dirichlet mean functional. However, an explicit expression for its density is given in Theorem 4.1 of James, Lijoi and Prünster [2008], which in many cases is amenable to rejection sampling.

The distribution of $Z_{\alpha,0}$ looms large for the simulation of two parameter Poisson–Dirichlet means generated by some general Y. James, Lijoi and Prünster [2008] have shown some distributional results that link the law of Y with that of $Z_{\alpha,0}$. These can be rephrased as follows. Consider the equation

$$\frac{\mathbb{E}\{(Y - x)_+^{\alpha}\}}{\mathbb{E}\{(x - Y)_+^{\alpha}\}} = L_{\alpha}^{\alpha} \stackrel{\text{def}}{=} \frac{\sin(U\pi\alpha)}{\sin((1-U)\pi\alpha)}.$$

This is to be taken as the definition of the Lamperti random variable L_α. Its solution $X = x$ (note that the left-hand-side is strictly monotonically decreasing in x) is distributed as $Z_{\alpha,0}$. This is to be interpreted as some sort of inversion method. If it happens that the distribution of Y has sufficient structure for that ratio to be explicitly known, then a simple solution of this equation permits random variate generation. In general, one can apply the series method of Devroye [1986] based on approximations of the expected values to develop exact generators for $Z_{\alpha,0}$.

The expected complexity (stack size) grows as $1/\theta$ times $1/\mathbb{E}\{|Z_{\alpha,0} - Z'_{\alpha,0}|\}$ as $\theta \downarrow 0$, where $Z_{\alpha,0}, Z'_{\alpha,0}$ are i.i.d.

In the next two sections, we present the details for two simple Poisson-Dirichlet mean families, one generated by $Y \equiv U$ and one generated by $Y \equiv \xi_{1/2}$ in (6). For other, more general choices of generator, we are not aware of any exact simulation methods in the literature.

A first example: Poisson-Dirichlet means generated by uniform random variables

When $Y \stackrel{\mathcal{L}}{=} U$, the left-hand-side of the equation defining $Z_{\alpha,0}$ is $((1-x)/x)^{\alpha+1}$, and thus,

$$Z_{\alpha,0} \stackrel{\mathcal{L}}{=} \frac{1}{1 + L_\alpha^{\frac{\alpha}{\alpha+1}}} \stackrel{\mathcal{L}}{=} \frac{\sin^{\frac{1}{\alpha+1}}(U\pi\alpha)}{\sin^{\frac{1}{\alpha+1}}(U\pi\alpha) + \sin^{\frac{1}{\alpha+1}}((1-U)\pi\alpha)}.$$

This implies that for the Poisson-Dirichlet mean generated by U, (7) reduces to

$$Z_{\alpha,\theta} \stackrel{\mathcal{L}}{=} B_{\theta,1} Z_{\alpha,\theta} + (1 - B_{\theta,1}) \frac{\sin^{\frac{1}{\alpha+1}}(U\pi\alpha)}{\sin^{\frac{1}{\alpha+1}}(U\pi\alpha) + \sin^{\frac{1}{\alpha+1}}((1-U)\pi\alpha)}.$$

The "Y" variable in this distributional identity takes values on $[0,1]$. Therefore, Double CFTP can be used.

A second example: Poisson-Dirichlet means generated by Bernoulli random variables

Consider the Poisson-Dirichlet means with Bernoulli generator:

$$Z_{\alpha,\theta} \stackrel{\mathcal{L}}{=} B_{\theta+\alpha,1-\alpha} Z_{\alpha,\theta+\alpha} + (1 - B_{\theta+\alpha,1-\alpha})\xi_{1/2},$$

where $\theta > -\alpha$, $\alpha \in [0,1]$. To distinguish this family from others, we will use the notation $O_{\alpha,\theta}$. The Dirichlet mean representation (7) then reads

$$O_{\alpha,\theta} \stackrel{\mathcal{L}}{=} B_{\theta,1} O_{\alpha,\theta} + (1 - B_{\theta,1}) O_{\alpha,0}.$$

Remarkably, $O_{\alpha,0}$ has a simple repesentation, as shown by Barlow, Pitman and Yor [1989]:

$$O_{\alpha,0} \stackrel{\mathcal{L}}{=} \frac{L_\alpha}{L_\alpha + 1},$$

where L_α is Lamperti's law. As pointed out above, L_α, and thus $O_{\alpha,0}$, can be simulated in one line of code. Since $O_{\alpha,0} \in [0,1]$, we have a representation for $O_{\alpha,\theta}$ as a Dirichlet mean with Y variable equal to $L_\alpha/(1+L_\alpha)$. The Double CFTP method can thus be used to generate $O_{\alpha,\theta}$. For $\theta \leq 1$, its expected time is proportional to

$$\frac{2}{\theta \mathbb{E}\{|O_{\alpha,0} - O'_{\alpha,0}|\}},$$

where $O_{\alpha,0}, O'_{\alpha,0}$ are i.i.d. One can verify that $L_\alpha^\alpha \stackrel{\mathcal{L}}{\to} U/(1-U)$ as $\alpha \downarrow 0$, so that $L_\alpha/(1+L_\alpha) \stackrel{\mathcal{L}}{\to} \xi_{1/2}$. Also, as $\alpha \uparrow 1$, $L_\alpha/(1+L_\alpha) \to 1/2$ in probability. The crucial parameter range is thus for α near one. Calculations show that $\mathbb{E}\{|O_{\alpha,0} - O'_{\alpha,0}|\} = \Theta(1-\alpha)$ as $\alpha \uparrow 1$, and this shows that the expected complexity of Double CFTP is bounded from above by a constant times $1/(\theta(1-\alpha))$.

Several special cases stand out.

Occupation time of a Bessel bridge. Let $\theta = \alpha$. Barlow, Pitman and Yor [1989] showed that the occupation time of a Bessel bridge of fractional dimension $2 - 2\alpha$ up to time one is given by the Poisson-Dirichlet mean

$$O_{\alpha,\alpha} \stackrel{\mathcal{L}}{=} (1 - B_{1,\alpha})O_{\alpha,\alpha} + B_{1,\alpha}O_{\alpha,0}. \tag{8}$$

The occupation time is the time spent on the positive side, and Bessel bridges generalize the Brownian bridge (which occurs at $\alpha = 1/2$). They are important processes in the modeling of volatility models in financial mathematics [Campolietti and Makarov, 2007]. While they can be simulated exactly by using randomized gamma laws [Makarov and Glew, 2009], various parameters of these processes, such as occupation times, offer simulation challenges. Pitman and Yor [1997b] [see also Yano and Yano, 2008] obtain the generalized Stieltjes transform of $O_{\alpha,\alpha}$. For $\alpha = 1/2$, we have $O_{\alpha,\alpha} \stackrel{\mathcal{L}}{=} U$ [Lévy's theorem, see, e.g., Ito and McKean, 1974, p. 58]. This transform does not lend itself easily to random variate generation. Our Double cftp method, applied to (8), runs in expected time bounded from above by a constant times $(\alpha(1-\alpha))^{-1}$. Further references on $O_{\alpha,\alpha}$ include Pitman and Yor [1997a,b] and Watanabe [1993].

The uniform case. Let $\theta = 1$. We then have the distributional identity

$$O_{\alpha,1} \stackrel{\mathcal{L}}{=} (1 - U)O_{\alpha,1} + UO_{\alpha,0}.$$

This is trivially simulated by Double cftp in expected time $O(1/(1 - \alpha))$. The results of James, Lijoi and Prünster [2008] show that $O_{\alpha,1}$ can also easily be sampled by rejection sampling.

Phylogenetic tree models. The case $\theta = 1-\alpha$ is also of interest: $O_{\alpha,1-\alpha}$ was shown to be a limit law for a parameter in a phylogenetic tree model by Haas, Miermont, Pitman and Winkel [2008, proposition 20]:

$$O_{\alpha,1-\alpha} \stackrel{\mathcal{L}}{=} (1 - B_{1,1-\alpha})O_{\alpha,1-\alpha} + B_{1,1-\alpha}O_{\alpha,0}. \tag{9}$$

Double cftp can be applied and runs in expected time (expected stack size) $O((1 - \alpha)^{-2})$. Thus, it becomes inefficient as $\alpha \uparrow 1$. However, there is a small algebra of relations between variables in the $O_{\alpha,\theta}$ family. For example, we have

$$O_{\alpha,1-\alpha} \stackrel{\mathcal{L}}{=} \xi_{1/2} + B_{1,1-\alpha}O_{\alpha,1}(1 - 2\xi_{1/2}) \tag{10}$$

[James, Lijoi and Prünster, 2008, corollary 6.1]. Using this, and the previous remark about $O_{\alpha,1}$, we see that $O_{\alpha,1-\alpha}$ can be generated in expected time $O(1/(1-\alpha))$, which is better than if (9) was used directly.

There is only one other avenue we are aware of for generating $O_{\alpha,\theta}$, and it involves the tilted stable law. Let $S_{\alpha,\theta}$, $\theta > -\alpha$, denote a polynomially tilted stable, i.e., if $f_\alpha(x)$ is the density of S_α, then the density of $S_{\alpha,\theta}$ is

$$\frac{\Gamma(1+\theta)}{\Gamma(1+\theta/\alpha)}\, x^{-\theta} f_\alpha(x), x > 0.$$

The generalized Lamperti law of James [2007] is $L_{\alpha,\theta} = S_\alpha/S_{\alpha,\theta}$. Random variates from the tilted stable law can be generated by the method of Devroye [2009] in uniformly bounded time. Stable variates can be

generated in one line of code by Kanter's method [Kanter, 1975; Chambers, Mallows and Stuck, 1976]. Proposition 5.1 of James [2007] establishes the following relationship between densities,

$$\mathbb{P}\{0_{\alpha,\theta} \in dy\} = 2^{\theta/\alpha}(1-y)^{\theta}\mathbb{P}\left\{\frac{L_{\alpha,\theta}}{L_{\alpha,\theta}+1} \in dy\right\}$$

for all $\theta > -\alpha$. This equation shows that rejection can be applied, but it also shows that rejection will be inefficient for some values of the parameters, e.g., when θ/α is large.

A limited simulation

We wrote computer programs for a benchmark family of distributions, the Dirichlet means with $Y \stackrel{\mathcal{L}}{=} U$ and $V \stackrel{\mathcal{L}}{=} 1 - U^{1/\theta}$, $\theta > 0$. By varying θ, we can clearly illustrate the merits, shortcomings and pitfalls of our method. It should be noted that except for $\theta = 1$, no explicit distributional expression is known for Z. At $\theta = 1$, we obtain the Diaconis-Kemperman law which was described in the text.

Four algorithms were compared.

(i) For $\theta = 1$, we implemented a rejection algorithm for the Diaconis-Kemperman law, as suggested in the text.

(ii) For $\theta \leq 1$, the original Double CFTP method was used with lower bound $\beta = \theta$. This method is not valid for $\theta > 1$.

(iii) For $\theta \geq 1$, the modified Double CFTP method was used with lower bound $g \equiv h$ (as h is nonincreasing). This method cannot be extended to $\theta < 1$.

(iv) We can simulate the infinite sum to machine precision. That is, we compute $Z'_t = \sum_{s=1}^{t} W_s Y_s$ (with W_s and Y_s as in the introduction), which is a lower bound on $Z = \sum_{s=1}^{\infty} W_s Y_s$, and $Z''_t = \sum_{s=1}^{t} W_s Y_s + cW_{t+1}/V_{t+1}$ (an upper bound on Z), until for the first time $Z'_t = Z''_t$ where $=$ means "up to machine precision".

Clearly, (iv) is an approximate method, and it is also vulnerable to numerical errors because sums should not be taken starting with the largest terms first.

We gathered satistical information on numbers of uniform random variates needed, number of Y_t's needed, number of V_t's needed, and so forth. These are independent of particular implemenations. Furthermore, we ran a simple check of the means of the random variables returned, which were all in the acceptable range. Finally, timing experiments were carried out. So as not to have to worry about compiler implementations, especially since a stack is used, we settled on an interpreted language that comes closest to machine language, PostScript. In it, we have full control of the details of the stack and all operations. Our UNIX machine ran FreeBSD 6.1-RELEASE-p7. The interpreter was GhostScript (GNU Ghostscript 7.07 (2003-05-17) Copyright (C) 2003 artofcode LLC, Benicia, CA).

Table (i). Timings in Milliseconds per Random Variate Generated, Averaged over 10,000 Runs

θ:	10	5	2	1	0.5	0.2	0.1
Rejection (Diaconis-Kemperman)				0.0047			
Double CFTP, original				0.0414	0.0891	0.2508	0.5555
Double CFTP, extension		4.5390	0.1880	0.0540			
Approximation	0.4610	0.2344	0.0766	0.0344	0.0219	0.0156	0.0101

The timings above pinpoint the main problem with both Double CFTP algorithms—that the expected times are not uniformly bounded even over the limited family of Dirichlet means. It is especially bad for our second algorithm when θ is large—our bounds show that the expected time grows exponentially quickly as $\theta \uparrow \infty$. On the other hand, we know of no exact method for this family besides Double CFTP. For the only explicitly known case, $\theta = 1$, the specially designed rejection method is nine times faster than Double CFTP. On the other hand, approximations by evaluating infinite sums to machine precision, seem to perform acceptably well. Since our sums are roughly exponentially decreasing, expect the timings to increase linearly in the number of bits of machine precision. For example, simply doubling the present precision would make Double CFTP outperform the simple approximate algorithm.

All CFTP algorithms of a certain sophistication face the stack storage problem. Stack sizes are geometrically distributed and thus relatively well-behaved. Bounds were given in the text. For the range $\theta \in [0.1, 1]$, averaged over a thousand runs, we observed average sizes of 59.436 ($\theta = 0.1$), 29.222 ($\theta = 0.2$), 11.960 ($\theta = 0.5$) and just 5.930 at $\theta = 1$ [recall that the theoretical value was $6/\theta = 6$]. The maximal stack sizes, were, respectively, 325, 207, 79 and 43. The growth in stack size is logarithmic in the number of runs, as the maximum of a geometric collection of random variables grows logarithmically.

At this point, we can compare with a simulation reported for an approximate CFTP algorithm by Guglielmi, Holmes and Walker [2002]. They too took $Y \equiv U$ and considered the same Dirichlet mean. Their coalescent times (our stack size) was about 20 for $\theta = 1$. Our value, 5.930, compares favorably.

Consumption of random variates per generated random variable is also of interest. Clearly, they all run in step with the stack size, showing the same growth patterns. At $\theta = 1$, the average number of uniforms generated, the average number of Y random variables generated, and the average number of V random variables generated in vector notation, averaged over one thousand runs, was $(15.253, 11.86, 5.969)$. Note that in all cases, Y and V are just transformation of uniforms, but we do not double count these, to keep the comparison clean. For the approximation algorithm, the corresponding vector is $(0, 19.737, 19.737)$. For the rejection method, it is just $(3.461, 0, 0)$.

Conclusions

Universal algorithms like the ones presented here can be used as "black boxes" for plug and play. The disadvantage, as shown in the simulation and our theoretical analysis, is that the expected time and expected storage are not uniformly bounded over all input distributions (of Y and V). In fact, there is no hope of a finite universal bound on the performance. Nevertheless, it is of interest to look for variations on the CFTP theme that halt quickly.

One referee mentioned distributional identities of the form

$$Z \stackrel{\mathcal{L}}{=} \varphi(Z, Z', Y),$$

where Z and Z' on the right are independent and identically distributed, Y is a general random element, and φ is a given function. Devroye and Neininger [2002] dealt with the simulation of such Z variables under some conditions on φ. Their solution was not elegant, but good enough for the exact simulation of the limit distribution of Quicksort [Devroye, Fill and Neininger, 2000], for example. However, there is no Markov chain that corresponds to such distributional identities—it is rather an infinite tree. We do not know how to generalize the ideas in CFTP to such tree-like recurrences, but it is certainly worth pursuing.

Finally, our distributional identity was motivated by Dirichlet means and Poisson-Dirichlet means. There will be many other distributional identities waiting to be given the same treatment to cause coalescence in continuous settings.

Acknowledgment

The first author thanks the second author for his hospitality during his visit at HKUST in Hong Kong. Both thank the referees who convincingly argued for a clearer presentation, a modification of the second algorithm, and additional material.

References

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

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

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

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

Barbu, Gh., 1987. A new fast method for computer generation of gamma and beta random variables by transformations of uniform variables. *Statistics 18*, 453–464.

M. Barlow, M., Pitman, J. and Yor, M., 1989. Une extension multidimensionnelle de la loi de l'arc sinus. in: *Seminaire de Probabilites, XXIII*, *vol. 1372*, 294–314. Lecture Notes in Mathematics, Springer-Verlag, Berlin.

Bertoin, J., Fujita, T., Roynette, B. and Yor, M., 2006. On a particular class of self-decomposable random variables: the durations of Bessel excursions straddling independent exponential times. *Probability and Mathematical Statistics 26*, 315–366.

Beskos, A. and Roberts, G.O., 2005. One-Shot CFTP: application to a class of truncated Gaussian densities. *Methodology and Computing in Applied Probability 7*, 407–437.

Best, D.J., 1978a. 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, 341–347. Physica Verlag, Wien, Austria.

Best, D.J., 1978b. Letter to the editor. *Applied Statistics 27*, 181.

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

Bondesson, L., 1992. *Generalized Gamma Convolutions and Related Classes of Distributions and Densities*, Springer-Verlag, New York.

Campolieti, G. and Makarov, R., 2007. Pricing path-dependent options on state dependent volatility models with a Bessel bridge. *International Journal of Theoretical and Applied Finance 10*, 1–38.

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

Chaumont, L. and Yor, M., 2003. *Exercises in Probability. A Guided Tour from Measure Theory to Random Processes, via Conditioning*, Cambridge University Press, Cambridge.

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

Cheng, R.C.H., 1978. Generating beta variates with nonintegral shape parameters. *Communications of the ACM 21*, 317–322.

Cheng, R.C.H. and Feast, G.M., 1978. Some simple gamma variate generators. *Applied Statistics 28*, 290–295.

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

Cifarelli, D.M. and Regazzini, E., 1979. Considerazioni generali sull'impostazione bayesiana di problemi non parametrici. Le medie associative nel contesto del processo aleatorio di Dirichlet I, II. *Riv. Mat. Sci. Econom. Social 2*, 39–52.

Cifarelli, D.M. and Regazzini, E., 1990. Distribution functions of means of a Dirichlet process. *Annals of Statistics 18*, 429–442. Correction in Annals of Statistics (1994) 22, 1633–1634.

Cifarelli, D.M. and Melilli, E., 2000. Some new results for Dirichlet priors. *Annals of Statistics 28*, 1390–1413.

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

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

Devroye, L., 1988. 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, 425–431. IEEE, Winter Simulation Conference, San Diego, CA.

Devroye, L., 1990. Coupled samples in simulation. *Operations Research 38*, 115–126.

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

Devroye, L., 2001. Simulating perpetuities. *Methodologies and Computing in Applied Probability 3*, 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 34*, 441–468.

Devroye, L., 2009. Random variate generation for exponentially and polynomially tilted stable distributions. *ACM Transactions on Modeling and Computer Simulation 19*, Article 18.

Diaconis, P. and Kemperman, J., 1996. Some new tools for Dirichlet priors. in: *Bayesian Statistics 5*, edited by J. M. Bernardo, J. O. Berger, A. P. , Dawid, and A. F. M. Smith, 97–106. Oxford University Press, New York.

Diaconis, P. and Freedman, D.A., 1999. Iterated random functions. *SIAM Review 41*, 45–67.

Doeblin, W., 1937. Exposé de la théorie des chaines simples constantes de Markov à un nombre fini d'états. *Revue Mathématique de l'Union Interbalkanique 2*, 77–105.

Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. *Annals of Statistics 1*, 209–230.

Fill, J.A., 1997. An interruptible algorithm for perfect sampling via Markov chains. in: *Proceedings of the Twenty-ninth Annual ACM symposium on Theory of Computing, El Paso, TX*, 688–695.

Fill, J.A., 1998. An interruptible algorithm for perfect sampling via Markov chains. *Annals of Applied Probability 81*, 131–162.

Fill, J.A., Machida, M., Murdoch, D.J. and Rosenthal, J.S., 2000. Extension of Fill's perfect rejection sampling algorithm to general chains. *Random Structures and Algorithms 17*, 290–316.

Fill, J.A. and Huber, M., 2004. Personal communication.

Fismen, M., 1998. Exact simulation using Markov chains. Technical Report 6/98, Department for Mathematical Sciences, The Norwegian University of Technology and Science, 7034 Trondheim,

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

Goldwater, S., Griffiths, T.L. and Johnson, M., 2006. Interpolating between types and tokens by estimating power-law generators. in: *Advances in Neural Information Processing Systems 18*,

Green, P.J. and Murdoch, D.J., 1999. Exact sampling for Bayesian inference: towards general purpose algorithms (with discussion). in: *Bayesian Statistics 6*, edited by J. Bernardo, J. Berger, A. Dawid, and A. Smith eds., 301–321, The Clarendon Press Oxford University Press.

Guglielmi, A., Holmes, C.C. and Walker, S.G., 2002. Perfect simulation involving functionals of a Dirichlet process. *Journal of Computational and Graphical Statistics 11*, 306–310.

Haas, B., Miermont, G., Pitman, J. and Winkel, M., 2008. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *Annals of Probability 36*, 1790–1837.

Ishwaran, I., and James, L.F., 2001. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association 96*, 161–173.

Ito K. and McKean, H.P., 1974. *Diffusion processes and their sample paths*, Springer-Verlag, Berlin.

James, L.F., 2007. Lamperti type laws: positive Linnik, Bessel bridge occupation and Mittag-Leffler functions. *arXiv:0708, 0618v1.*

James, L.F., Lijoi, A. and Prünster, I., 2008. Distributions of linear functionals of two parameter Poisson-Dirichlet random measures. *Annals of Aplied Probability 18*, 521–551.

James, L.F., 2005. Functionals of Dirichlet processes, the Cifarelli-Regazzini identity and Beta-Gamma processes. *Annals of Statistics 33*, 647–660.

James, L.F., 2008. Dirichlet mean identities and laws of a class of subordinators. Technical Report, Hong Kong University of Science and Technology.

James, L.F., Roynette, B. and Yor, M., 2008. Generalized Gamma Convolutions, Dirichlet means, Thorin measures, with explicit examples. *Probability Surveys 5*, 346–415.

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

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

Kendall, W. and Thönnes, E., 2004. Random Walk CFTP. Technical Report, Department of Statistics, University of Warwick.

Lamperti, J., 1958. An occupation time theorem for a class of stochastic processes. *Transactions of the American Mathematical Society 88*, 380–387.

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

Lindvall, T., 1992. *Lectures on the Coupling Method*, John Wiley, New York.

Makarov R.N. and Glew, D., 2009. Exact simulation of Bessel diffusions. *arXiv, 0910.4177.*

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

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

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

Murdoch, D.J., 2000. Exact sampling for Bayesian inference: Unbounded state spaces. in: *Proceedings of the Workshop on MCMC Methods, Fields Institute, October 1998*, 111–121. Fields Institute Communications.

Perman, M., Pitman, J. and Yor, M., 1992. Size-biased sampling of Poisson point processes and excursions. *Probability Theory and Related Fields 92*, 21–39.

Pitman, J. and Yor, M., 1992. Arcsine laws and interval partitions derived from a stable subordinator. *Proc. London Math. Soc. 65*, 326–356.

Pitman, J., 1996. Some developments of the Blackwell-MacQueen urn scheme. in: *Statistics, probability and game theory 30*, 245–267. IMS Lecture Notes Monograph Series, IMS, Hayward, CA.

Pitman, J. and Yor, M., 1997a. The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Annals of Probability 25*, 855–900.

Pitman, J. and Yor, M., 1997b. On the relative lengths of excursions derived from a stable subordinator. in: *Séminaire de Probabilités I*, edited by J. Azema, M. Emery, and M. Yor, *vol. 1655*, 287–305. Lecture Notes in Mathematics, Springer, Berlin.

Pitman, J., 1999. Brownian motion, bridge, excursion, and meander characterized by sampling at independent uniform times. Technical Report No. 545, Department of Statistics, University of California at Berkeley,

Pitman, J., 2003. Poisson Kingman partitions. in: *Science and Statistics: A Festschrift for Terry Speed*, edited by D. R. Goldstein, 1–34. Lecture Notes Monograph Series Vol. 30, Institute of Mathematical Statistics, Hayward, CA.

Pitman, J., 2006. Combinatorial stochastic processes. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002. Lecture Notes in Mathematics, *vol. 1875*, Springer-Verlag, Berlin.

Propp, J.G. and Wilson, D.B., 1998. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms 9*, 223–252.

Propp, J.G. and Wilson, D.B., 1998. How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph. *Journal of Algorithms 27*, 170–217.

Roberts, G.O. and Rosenthal, J.S., 2002. One-shot coupling for certain stochastic recursive sequences. *Stochastic Processes and Applications 99*, 195–208.

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

Teh, Y.W., 2006. *A Hierarchical Bayesian Language Model based on Pitman-Yor Processes*, Coling/ACL.

Thorisson, H., 1986. On maximal and distributional coupling. *Annals of Probability 14*, 873–876.

von Neumann, J., 1951. Monte Carlo method. *National Bureau of Standards Series 12*, 36–38.

von Neumann, J., 1963. Various techniques used in connection with random digits. *Collected Works 5*, 768–770. Pergamon Press.

Watanabe, S., 1995. Generalized arc-sine laws for one-dimensional diffusion processes and random walks. in: *Proceedings of the Symposium on Pure MathematicsRandom number generation 57*, 157–172. Stochastic Analysis (Ithaca, NY, 1993), American Mathematical Society, Providence, RI, 1993.

Wilson, D.B., 2000a. How to couple form the past using a read-once source of randomness. *Random Structures and Algorithms 16*, 85–113.

Wilson, D.B., 2000b. Layered multishift coupling for use in perfect sampling algorithms. in: *Monte Carlo Methods 26*, edited by N. Madras, 143–179. Fields Institute Communications, American Mathematical Society.

Yano, K. and Yano, Y., 2008. Remarks on the density of the law of the occupation time for Bessel bridges and stable excursions. *Statistics and Probability Letters 78*, 2175–2180.