On Buffon Machines and Numbers
Philippe Flajolet, Maryse Pelletier, Michèle Soria

To cite this version:

HAL Id: hal-00548904
https://hal.archives-ouvertes.fr/hal-00548904
Submitted on 20 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
On Buffon Machines and Numbers

Philippe Flajolet* Maryse Pelletier† Michèle Soria†

phillipe.flajolet@inria.fr, maryse.pelletier@lip6.fr, michele.soria@lip6.fr

Abstract

The well-known needle experiment of Buffon can be regarded as an analog (i.e., continuous) device that stochastically “computes” the number \(2/\pi \approx 0.63661\), which is the experiment’s probability of success. Generalizing the experiment and simplifying the computational framework, we consider probability distributions, which can be produced \emph{perfectly}, from a discrete source of unbiased coin flips. We describe and analyse a few simple Buffon machines that generate geometric, Poisson, and logarithmic-series distributions. We provide human-accessible Buffon machines, which require a dozen coin flips or less, on average, and produce experiments whose probabilities of success are expressible in terms of numbers such as \(\pi\), \(\exp(-1)\), \(\log 2\), \(\sqrt{3}\), \(\cos(\pi/4)\), \(\zeta(5)\). Generally, we develop a collection of constructions based on simple probabilistic mechanisms that enable one to design Buffon experiments involving compositions of exponentials and logarithms, polylogarithms, direct and inverse trigonometric functions, algebraic and hypergeometric functions, as well as functions defined by integrals, such as the Gaussian error function.

Introduction

Buffon’s experiment (published in 1777) is as follows \cite{1, 4}. Take a plane marked with parallel lines at unit distance from one another; throw a needle at random; finally, declare the experiment a success if the needle intersects one of the lines. Basic calculus implies that the probability of success is \(2/\pi\).

One can regard Buffon’s experiment as a simple analog device that takes as input real uniform \([0, 1]\)-random variables (giving the position of the centre and the angle of the needle) and outputs a discrete \([0, 1]\)-random variable, with 1 for success and 0 for failure. The process then involves exact arithmetic operations over real numbers. In the same vein, the classical problem of simulating random variables can be described as the construction of analog devices (algorithms) operating over the reals and equipped with a real source of randomness, which are both simple and computationally efficient. The encyclopedic treatise of Devroye \cite{11} provides many examples relative to the simulation of distributions, such as Gaussian, exponential, Cauchy, stable, and so on.

Buffon machines. Our objective is the perfect simulation of discrete random variables (i.e., variables supported by \(\mathbb{Z}\) or one of its subsets; see Fig. 1). In this context, it is natural to start with a discrete source of randomness that produces \emph{uniform random bits} (rather than uniform \([0, 1]\) real numbers), since we are interested in finite computation (rather than infinite-precision real numbers); cf Fig. 2

<table>
<thead>
<tr>
<th>Law</th>
<th>supp.</th>
<th>distribution</th>
<th>gen.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bernoulli</td>
<td>{0, 1}</td>
<td>(P(X = 1) = p)</td>
<td>GB(p)</td>
</tr>
<tr>
<td>geometric</td>
<td>(\mathbb{Z}_{\geq 0})</td>
<td>(P(X = r) = \lambda^r (1 - \lambda))</td>
<td>GF((\lambda))</td>
</tr>
<tr>
<td>Poisson</td>
<td>(\mathbb{Z}_{\geq 0})</td>
<td>(P(X = r) = e^{-\lambda} \frac{\lambda^r}{r!})</td>
<td>FP((\lambda))</td>
</tr>
<tr>
<td>logarithmic</td>
<td>(\mathbb{Z}_{\geq 0})</td>
<td>(P(X = r) = \frac{1}{\sqrt{L \pi}})</td>
<td>FL((\lambda))</td>
</tr>
</tbody>
</table>

Figure 1: Main discrete probability laws: support, expression of the probabilities, and naming convention for generators (\(L := \log(1 - \lambda)^{-1}\)).

Definition 1. A Buffon machine is a deterministic device belonging to a computationally universal class (Turing machines, equivalently, register machines), equipped with an external source of independent uniform random bits and input–output queues capable of storing integers (usually, \{0,1\}-bits), which is assumed to halt\(^\dagger\) with probability 1.

The fundamental question is then the following. How does one generate, exactly and efficiently, discrete distributions using only a discrete source of random bits and finitary computations? A pioneering work in this direction is that of Knuth and Yao \cite{20} who discuss the power of various restricted devices (e.g., finite-state machines), Knuth’s treatment \cite{18} §3.4 and the articles \cite{11 30} provide additional results along these lines.

\(^\dagger\)Machines that always halt can only produce Bernoulli distributions whose parameter is a dyadic rational \(s/2^t\); see \cite{20} and \cite{14}.
Our original motivation for studying discrete Buffon machines came from Boltzmann samplers for combinatorial structures, following the approach of Duchon, Flajolet, Louchard, and Schaeffer and Flajolet, Fusy, and Pivoteau. The current implementations rely on real number computations, and they require generating distributions such as geometric, Poisson, or logarithmic, with various ranges of parameters—since the objects ultimately produced are discrete, it is natural to try and produce them by purely discrete means.

**Numbers.** Here is an intriguing range of related issues. Let $M$ be an input-free Buffon machine that outputs a random variable $X$, whose value lies in $\{0, 1\}$. It can be seen from the definition that such a machine $M$, when called repeatedly, produces an independent sequence of Bernoulli random variables. We say that $M$ is a Buffon machine or Buffon computer for the number $p := P(X = 1)$. We seek simple mechanisms—Buffon machines—that produce, from uniform $\{0, 1\}$-bits, Bernoulli variables whose probabilities of success are numbers such as

$$1/\sqrt{2}, \ e^{-1}, \ \log 2, \ \frac{1}{\pi}, \ \pi - 3, \ \frac{1}{e - 1}. \tag{0.1}$$

This problem can be seen as a vast generalization of Buffon’s needle problem, adapted to the discrete world.

**Complexity and simplicity.** We will impose the loosely defined constraint that the Buffon machines we consider be short and conceptually simple, to the extent of being easily implemented by a human. Thus, emulating infinite-precision computation with multiprecision interval arithmetics or appealing to functions of high complexity as primitives is disallowed. Our Buffon programs only make use of simple integer counters, string registers (2) and stacks (3), as well as “bags” (4). The reader may try her hand at determining Buffon-ness in this sense of some of the numbers listed in (0.1).

We shall, for instance, produce a Buffon computer for the constant $\text{Li}_3(1/2)$ of Eq. (2.11) (which involves $\log 2$, $\pi$, and $\zeta(3)$), one that is human-compatible, that consumes on average less than 6 coin flips, and that requires at most 20 coin flips in 95% of the simulations. We shall also devise nine different ways of simulating Bernoulli distributions whose probability involves $\pi$, some requiring as little as five coin flips on average. Furthermore, the constructions of this paper can all be encapsulated into a universal interpreter, briefly discussed in Section 3 which has less than 60 lines of MAPLE code and produces all the constants of Eq. (0.1), as well as many more.

In this extended abstract, we focus the discussion on algorithmic design. Analytic estimates can be approached by means of (probability, counting) generating functions in the style of methods of analytic combinatorics; see the typical discussion in Subsection 2.1. The main results are Theorem 2.2 (Poisson and logarithmic generators), Theorem 2.3 (realizability of exps, logs, and trigs), Theorem 4.1 (general integrator), and Theorem 4.2 (inverse trig functions).

1 Framework and examples

Our approach consists in setting up a system based on the composition of simple probabilistic experiments, corresponding to simple computing devices. (For this reason, the Buffon machines of Definition 1 are allowed input/output registers.) The unbiased random-bit generator, with which Buffon machines are equipped will be named “flip”. The cost measure of a computation (simulation) is taken to be the number of flips. (For the restricted devices we shall consider, the overall simulation cost is essentially proportional to this measure.)

**Definition 2.** The function $\lambda \mapsto \phi(\lambda)$, defined for $\lambda \in (0, 1)$ and with values $\phi(\lambda) \in (0, 1)$, is said to be weakly realizable if there is a machine $M$, which,

\begin{itemize}
\item produces all the constants of Eq. (0.1), as well as many more.
\end{itemize}

\begin{itemize}
\item The informal requirement of “simplicity” can be captured by the formal notion of program size. All the programs we develop necessitate at most a few dozen register-machine instructions, see 5 and the Appendix, as opposed to programs based on arbitrary-precision arithmetics, which must be rather huge; cf 31. If program size is unbounded, the problem becomes trivial, since any Turing-computable number $a$ can be emulated by a Buffon machine with a number of coin flips that is $O(1)$ on average, e.g., by computing the sequence of digits of $a$ on demand; see 28 and Eq. (4.3) below.
\item This convention entails no loss of generality. Indeed, as first observed by von Neumann, suppose we only have a biased coin, where $P(1) = p, P(0) = 1 - p$, with $p \in (0, 1)$. Then, one should toss the coin twice: if 01 is observed, then output 0; if 10 is observed, then output 1; otherwise, repeat with new coin tosses. See, e.g., 26, 35 for more on this topic.
\end{itemize}
when provided on its input with a perfect generator of Bernoulli variables of (unknown) parameter \( \lambda \), outputs, with probability 1, a Bernoulli random variable of parameter \( \phi(\lambda) \). The function \( \phi \) is said to be realizable, resp., strongly realizable, if the (random) number \( C \) of coin flips has finite expectation, resp., exponential tail.

We shall also say that \( \phi \) has a [weak, strong] simulation if it is realizable by a machine [in the weak, strong sense]. Schematically:

\[
\begin{array}{c}
X_1, X_2, \ldots \in \text{Ber}(\lambda) \\
\rightarrow \mathcal{M} \\
Y \in \text{Ber}(\phi(\lambda))
\end{array}
\]

The devices are normally implemented as programs. Using \( \Gamma_B(\lambda) \) as a generic notation for a Bernoulli generator of parameter \( \lambda \), a Buffon machine that realizes the function \( \phi \) is then equivalent to a program that can call (possibly several times) a \( \Gamma_B(\lambda) \), as an external routine, and then outputs a random Bernoulli variable of parameters \( \phi(\lambda) \). It corresponds to a construction of type \( \Gamma_B(\lambda) \rightarrow \Gamma_B(\phi(\lambda)) \).

The definition is extended to machines with \( m \) inputs, in which case a function \( \phi(\lambda_1, \ldots, \lambda_m) \) of \( m \) arguments is realized; see below for several examples. A machine with no input register then computes a function \( \phi() \) of no argument, that is, a constant \( p \), and it does so based solely on its source of unbiased coin flips: this is what we called in the previous section a Buffon machine (or computer) for \( p \).

The fact that Buffon machines are allowed input registers makes it possible to compose them. For instance, if \( \mathcal{M} \) and \( \mathcal{N} \) realize the unary functions \( \phi \) and \( \psi \), connecting the output of \( \mathcal{M} \) to the input of \( \mathcal{N} \) realizes the composition \( \psi \circ \phi \). It is one of our goals to devise Buffon computers for special values of the success probability \( p \) by composition of simpler functions, eventually only applied to the basic flip primitive.

Note that there are obstacles to what can be done: Keane and O’Brien [16], for instance, showed that the “success doubling” function \( \min(2\lambda, 1) \) cannot be realized and discontinuous functions are impossible to simulate—this, for essentially measure-theoretic reasons. On the positive side, Nacu and Peres [27] show that every (computable) Lipschitz function is realizable and every (computable) real-analytic function has a

\(^{\text{strong simulation, but their constructions require computationally unrestricted devices; namely, sequences of approximation functions of increasing complexity. Instead, our purpose here is to show how sophisticated perfect simulation algorithms can be systematically and efficiently synthetized by composition of simple probabilistic processes, this without the need of hard-to-compute sequences of approximations.}}\)

To set the stage of the present study, we shall briefly review in sequence: (i) decision trees and polynomial functions; (ii) Markov chains (finite graphs) and rational functions.

**Decision trees and polynomials.** Given three machines \( P, Q, R \) with outputs in \( \{0, 1\} \) and with \( \Pr_P(1) = p, \Pr_Q(1) = q, \Pr_R(1) = r \), we can easily build machines, corresponding to loopless programs, whose probability of success is \( p \cdot q \cdot (1 - p) \), or any composition thereof: see Eq. (1.2) [top of next page] for the most important boolean primitives. We can then simulate a Bernoulli of parameter any dyadic rational \( s/2^k \), starting from the unbiased flip procedure, performing \( t \) draws, and declaring a success in \( s \) designated cases. Also, by calling a Bernoulli variate of (unknown) parameter \( \lambda \) a fixed number \( m \) of times, then observing the sequence of outcomes and its number \( k \) of 1s, we can realize any polynomial function \( a_{m,k} \lambda^k (1 - \lambda)^{m-k} \), with \( a_{m,k} \) any integer satisfying \( 0 \leq a_{m,k} \leq \binom{m}{k} \).

**Finite graphs (Markov chains) and rational functions.** We now consider programs that allow iteration and correspond to a finite control graph—these are equivalent to Markov chains, possibly parameterized by “boxes” representing calls to external Bernoulli generators. In this way, we can produce a Bernoulli generator of any rational parameter \( \lambda \in \mathbb{Q} \), by combining a simpler dyadic generator with iteration. (For instance, to get a Bernoulli generator of parameter \( 3/4 \), denoted by \( \Gamma_B(3/4) \), flip an unbiased coin twice; return success if 11 is observed, failure in case of 01 or 10, and repeat the experiment in case of 00.) Clearly, only rational numbers and functions can be realized by finite graphs.

A highly useful construction in this range of methods is the even-parity construction:

(1.3) \[ \text{[even-parity] do } \begin{cases} \text{if } P() = 0 \text{ then return(1); } \\
\text{if } P() = 0 \text{ then return(0). } \end{cases} \]

This realizes the function \( p \mapsto 1/(1 + p) \). Indeed, the probability of \( k + 1 \) calls to \( P() \) is \( (1 - p)^{k} \), which, when summed over \( k = 0, 2, 4, \ldots \), gives the probability of success as \( 1/(1 + p) \); thus, this function is realizable. Combining this function with complementation \( (x \mapsto 1 - x) \) leads, for instance, to a way of compiling any rational probability \( a/b \) into a generator \( \Gamma_B(\frac{a}{b}) \) whose size is proportional to the sum of digits of the continued
fraction representation of \(a/b\). (See also Eq. (1.5) below for an alternative based on binary representations and 
[20] for more on this topic.)

Here are furthermore two important characterizations based on works of Nacu–Peres [27] (see also Wästlund [35]) and Mossel–Peres [26]:

**THEOREM 1.1.** ([26, 27, 35]) (i) Any polynomial \(f(x)\) with rational coefficients that maps \((0, 1)\) into \((0, 1)\) is strongly realizable by a finite graph. (ii) Any rational function \(f(x)\) with rational coefficients that maps \((0, 1)\) into \((0, 1)\) is strongly realizable by a finite graph.

(\(p\) and \(Q\) are not assumed to be known.)

The even-parity construction implicitly makes use of this. The special \(\Gamma G(\frac{1}{2})\) then simply iterates on the basis of a flip.

The idea of generating certain probability distributions by way of their Taylor expansion seems to go back to von Neumann. An early application discussed by Knuth and Yao [20], followed by Flajolet and Saheb [11], is the exact simulation of an exponential variate by means of random \([0, 1]\)-uniform variates, this by a “continuation” process that altogether avoids multiprecision operations. We build upon von Neumann’s idea and introduce in Subsection 2.1 a general schema for random generation—the von Neumann schema. We then explain in Subsection 2.2 how this schema may be adapted to realize classical transcendental functions, such as \(e^{-\lambda},\) \(\cos(\lambda),\) only knowing a generator \(\Gamma B(\lambda)\).

### 2 The von Neumann schema

2.1 Von Neumann generators of discrete distributions. First a few notations from [12]. Start from a class \(\mathcal{P}\) of permutations, with \(\mathcal{P}_n\) the subset of permutations of size \(n\) and \(\mathcal{P}_n\) the cardinality of \(\mathcal{P}_n\). The (counting) exponential generating function, or EGF, is

\[
P(z) := \sum_{n \geq 0} P_n \frac{z^n}{n!}.
\]

For instance, the classes \(\mathcal{Q}, \mathcal{R}, \mathcal{S}\) of, respectively, all permutations, sorted permutations, and cyclic permutations have EGFs given by \(Q(z) = (1-z)^{-1}, R(z) = e^z, S(z) = \log(1-z)^{-1}\), since \(Q_n = n!, R_n = 1, S_n = (n-1)!\), for \(n \geq 1\). We observe that the class \(\mathcal{S}\) of cyclic permutations is isomorphic to the class of permutations such that the maximum occurs in the first position: \(U_1 > U_2, \ldots, U_N\). (It suffices to “break” a cycle at its maximum element.) We shall also denote by \(\mathcal{S}\) the latter class, which is easier to handle algorithmically.

Let \(U = (U_1, \ldots, U_n)\) be a vector of real numbers.

By replacing each \(U_j\) by its rank in \(U\), we obtain a permutation \(\sigma = (\sigma_1, \ldots, \sigma_n)\), which is called the (order) type of \(U\) and is written \(\text{type}(U)\). For instance: \(\text{type}(1.41, 0.57, 3.14, 2.71) = (2, 1, 4, 3)\). The von Neumann schema, relative to a class \(\mathcal{P}\) of permutations is described in Fig. 3 and denoted by \(\text{IVN}[\mathcal{P}](\lambda)\). Observe that it only needs a geometric generator \(\Gamma G(\lambda)\), hence, eventually, only a Bernoulli generator \(\Gamma B(\lambda)\), whose parameter \(\lambda\) is not assumed to be known.

**Table**

<table>
<thead>
<tr>
<th>Name</th>
<th>realization</th>
<th>function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conjunction ((P \land Q))</td>
<td>if (P() = 1) then return((Q())) else return((0))</td>
<td>(p \land q = p \cdot q)</td>
</tr>
<tr>
<td>Disjunction ((P \lor Q))</td>
<td>if (P() = 0) then return((Q())) else return((1))</td>
<td>(p \lor q = p + q - pq)</td>
</tr>
<tr>
<td>Complementation ((-P))</td>
<td>if (P() = 0) then return((1)) else return((0))</td>
<td>(1 - p)</td>
</tr>
<tr>
<td>Squaring ((P \land P))</td>
<td></td>
<td>(p^2)</td>
</tr>
<tr>
<td>Conditional ((R \rightarrow P</td>
<td>Q))</td>
<td>if (R() = 1) then return((P())) else return((Q()))</td>
</tr>
<tr>
<td>Mean</td>
<td>if flip() then return((P())) else return((Q()))</td>
<td>(\frac{1}{2}(p+q)).</td>
</tr>
</tbody>
</table>
\[ \Gamma VN[\mathcal{P}](\lambda) := \{ \text{do } \}
\begin{align*}
\text{let } U &:= (U_1, \ldots, U_N), \text{ vector of [0, 1]-uniform bits of the } U_j \text{ are produced on call-by-need basis} \\
\text{let } \tau &:= \text{trie}(U); \text{ let } \sigma := \text{type}(U); \\
\text{if } \sigma \in \mathcal{P}_N \text{ then return}(N) \}. 
\end{align*}
\]

Figure 3: The von Neumann schema \( \Gamma VN[\mathcal{P}](\lambda) \), in its basic version, relative to a class of permutations \( \mathcal{P} \) and a parameter \( \lambda \) (equivalently given by its Bernoulli generator \( \Gamma B(\lambda) \)).

First, by construction, a value \( N \) is, at each stage of the iteration, chosen with probability \((1 - \lambda)\lambda^n\). The procedure consists in a sequence of failed trials (when type(U) is not in \( \mathcal{P} \)), followed by eventual success. An iteration (trial) that succeeds then returns the value \( N = n \) with probability
\[
\frac{(1 - \lambda)P_n\lambda^n/n!}{(1 - \lambda)\sum_n P_n\lambda^n/n!} = \frac{P_n\lambda^n}{P(\lambda)/n!}.
\]

For \( \mathcal{P} \) one of the three classes \( Q, R, S \) described above this gives us three interesting distributions:

\[
\begin{array}{|c|c|c|}
\hline
\text{all (}Q\text{)} & \text{sorted (}R\text{)} & \text{cyclic (}S\text{)} \\
\hline
(1 - \lambda)\lambda^n & e^{-\lambda} \frac{\lambda^n}{n!} & \frac{1}{n} \frac{\lambda^n}{H_n} \quad \text{logarithmic.} \\
\hline
\end{array}
\]

The case of all permutations \( Q \) is trivial, since no order-type restriction is involved, so that the initial value of \( N \in \text{Geo}(\lambda) \) is returned. It is notable that, in the other two cases \( R, S \), one produces the Poisson and logarithmic distributions, by means of permutations obeying simple restrictions.

Next, implementation details should be discussed. Once \( N \) has been drawn, we can imagine producing the \( U_j \) in sequence, by generating at each stage only the minimal number of bits needed to distinguish any \( U_j \) from the other ones. This corresponds to the construction of a digital tree, also known as “trie” \( [19, 23, 33] \) and is summarized by the command “let \( \tau := \text{trie}(U) \)” in the schema of Fig. 3. As the digital tree \( \tau \) is constructed, the \( U_j \) are gradually sorted, so that the order type \( \sigma \) can be determined—this involves no additional flips, just bookkeeping. The test \( \sigma \in \mathcal{P}_N \) is of this nature and it requires no flip at all.

The general properties of the von Neumann schema are summarized as follows.

**Theorem 2.1.**

(i) Given an arbitrary class \( \mathcal{P} \) of permutations and a parameter \( \lambda \in (0, 1) \), the von Neumann schema \( \Gamma VN[\mathcal{P}](\lambda) \) produces exactly a discrete random variable with probability distribution
\[
P(N = n) = \frac{1}{P(\lambda)} \frac{P_n\lambda^n}{n!}.
\]

(ii) The number \( K \) of iterations has expectation \( 1/s \), where \( s = (1 - \lambda)P(\lambda) \), and its distribution is \( 1 + \text{Geo}(1 - s) \).

(iii) The number \( C \) of flips consumed by the algorithm (not counting the ones in \( \Gamma \mathcal{P}(\lambda) \)) is a random variable with probability generating function
\[
\mathbb{E}(q^C) = \frac{H^+(\lambda, q)}{1 - H^-(\lambda, q)},
\]

where \( H^+, H^- \) are computable from the family of polynomials in \( [2.10] \) below by
\[
H^+(z, q) = (1 - z) \sum_{n=0}^{\infty} P_n / n! h_n(q) z^n
\]
\[
H^-(z, q) = (1 - z) \sum_{n=0}^{\infty} \left( 1 - \frac{P_n}{n!} \right) h_n(q) z^n.
\]

The distribution of \( C \) has exponential tails.

**Proof.** [Sketch] Items (i) and (ii) result from the discussion above. Regarding Item (iii), a crucial observation is that the digital tree created at each step of the iteration is only a function of the underlying set of values. But there is complete independence between this set of values and their order type. This justifies \( (2.8) \), where \( H^+, H^- \) are the probability generating functions associated with success and failure of one trial, respectively.

We next need to discuss the fine structure of costs. The cost of each iteration, as measured by the number of coin flips, is exactly that of generating the tree \( \tau \) of random size \( N \). The number of coin flips to build \( \tau \) coincides with the path length of \( \tau \), written \( \omega(\tau) \), which satisfies the inductive definition
\[
(2.9) \quad \omega(\tau) = |\tau| + \omega(\tau_0) + \omega(\tau_1), \quad |\tau| \geq 2,
\]
where \( \tau = (\tau_0, \tau_1) \) and \( |\tau| \) is the size of \( \tau \), that is, the number of elements contained in \( \tau \).

Path length is a much studied parameter, starting with the work of Knuth in the mid 1960s relative to the analysis of radix-exchange sort \([19\text{ pp. } 128-134]\); see also the books of Mahmoud \( [23] \) and Szpankowski \( [33] \) as well as Vallée et al.’s analyses under dynamical source models \( [5, 34] \). It is known from these works that the expectation of path length, for \( n \) uniform binary sequences, is finite, with exact value
\[
\mathbb{E}_n[\omega] = n \sum_{k=0}^{\infty} \left[ 1 - \left( 1 - \frac{1}{2^k} \right)^{n-1} \right],
\]
and asymptotic form (given by a Mellin transform analysis \([10, 23, 33]\)): \( \mathbb{E}_n[\omega] = n \log_2 n + O(n) \). A consequence of this last fact is that \( \mathbb{E}(C) \) is finite, i.e., the generator \( \Gamma VN[\mathcal{P}](\lambda) \) has the basic simulation (realisability) property.

The distribution of path length is known to be asymptotically Gaussian, as \( n \to \infty \), after Jacquet and Régnier \([19]\).
and Mahmoud et al. [25], see also [24, §11.2]. For our purposes, it suffices to note that the bivariate EGF
\[
H(z, q) := \sum_{n=0}^{\infty} \frac{E_n[z^n]}{n!} q^n
\]
satisfies the nonlinear functional equation
\[
H\left(\frac{z}{1 - z}, q\right) = H\left(\frac{z}{q}, q\right)^2 + z(1 - q),
\]
with \(H(0, q) = 1\). This equation fully determines \(H\), since it is equivalent to a recurrence on coefficients, \(h_n(q) := n! z^n[H(z, q)]\), for \(n \geq 2\):

\[
(2.10) \quad h_n(q) = \frac{1}{1 - q^n} \sum_{k=1}^{n-1} \frac{1}{2^n} \binom{n}{k} h_k(q) h_{n-k}(q).
\]

The computability of \(H^+\), \(H^-\) then results. In addition, large deviation estimates can be deduced from (2.10), which serve to establish exponential tails for \(C\), thereby ensuring a strong simulation property in the sense of Definition 2.

A notable consequence of Theorem 2.1 is the possibility of generating a Poisson or logarithmic variate by a simple device: as we saw in the discussion preceding the statement of the Theorem, only one branch of the tree needs to be maintained, in the case of the classes \(R\) and \(S\) of (2.7).

**Theorem 2.2.** The Poisson and logarithmic distributions of parameter \(\lambda \in (0, 1)\) have a strong simulation by a Buffon machine, \(\Gamma_{VN[R]}(\lambda)\) and \(\Gamma_{VN[S]}(\lambda)\), respectively, which only uses a simple string register.

Since the sum of two Poisson variates is Poisson (with rate the sum of the rates), the strong simulation result extends to any \(X \in \text{Poi}(\lambda)\), for any \(\lambda \in \mathbb{R}_{>0}\). This answers an open question of Keuth and Yao in [20, p. 426]. We may also stress here that the distributions of costs are easily computable: with the symbolic manipulation system MAPLE, the cost of generating a Poisson(1/2) variate is found to have probability generating function (Item (iii) of Theorem 2.1)

\[
\frac{3}{4} + \frac{7}{128} q^2 + \frac{119}{4096} q^4 + \frac{19}{1024} q^5 + \frac{2023}{131072} q^6 + \frac{179}{16384} q^7 + \cdots.
\]

Interestingly enough, the analysis of the logarithmic generators involves ideas similar to those relative to a classical leader election protocol [8, 30].

### 2.2 Buffon computers: logarithms, exponentials, and trig functions

We can also take any of the previous constructions and specialize it by declaring a success whenever a special value \(N = a\) is returned, for some predetermined \(a\) (usually \(a = 0, 1\)), declaring a failure, otherwise. For instance, the Poisson generator with \(a = 0\) gives us in this way a Bernoulli generator with parameter \(\lambda' = \exp(-\lambda)\). Since the von Neumann machine only requires a Bernoulli generator \(\Gamma_B(\lambda)\), we thus have a purely discrete construction \(\Gamma_B(\lambda) \rightarrow \Gamma_B(e^{-\lambda})\). Similarly, the logarithmic generator restricted to \(a = 1\) provides a construction \(\Gamma_B(\lambda) \rightarrow \Gamma_B(\log(1/e^\lambda)/\lambda)\). Naturally, these constructions can be enriched by the basic ones of Section 1 in particular, complementation.

Another possibility is to make use of the number \(K\) of iterations, which is a shifted geometric of rate \(s = (1 - \lambda)P(\lambda)\); see Theorem 2.1 Item (ii). If we declare a success when \(K = b\), for some predetermined \(b\), we then obtain yet another brand of generators. The Poisson generator used in this way with \(b = 1\) gives us \(\Gamma_B(\lambda) \rightarrow \Gamma_B(1 - \lambda)e^{\lambda} \), \(\Gamma_B(\lambda e^{1-\lambda})\), where the latter involves an additional complementation.

Trigonometric functions can also be brought into the game. A sequence \(U = (U_1, \ldots, U_n)\) is said to be alternating if \(U_1 < U_2 > U_3 < U_4 > \cdots\). It is well known that the EGFs of the classes \(A^+\) of even-sized and \(A^-\) of odd-sized permutations are respectively 
\[
A^+(z) = \sec(z) = 1/cos(z), \quad A^-(z) = \tan(z) = \sin(z)/\cos(z).
\]
This result, due to Désiré André around 1880, is in most books on combinatorial enumeration, e.g., [12, 13, 32]. Remember that the property of being alternating can once more be tested sequentially: only a partial expansion of the current value of \(U_j\) needs to be stored at any given instant. By making use of the properties \(A^+, A^-\), with, respectively \(N = 0, 1\), we then obtain new trigonometric constructions.

In summary:

**Theorem 2.3.** The following functions admit a strong simulation:

\[
\begin{align*}
\frac{e^{-x}}{x}, & \quad (1 - x)e^{x}, \quad xe^{1-x}, \\
\frac{\log(1-x)}{1-x} - 1, & \quad \frac{\log(1/x)}{1-x} - \log 1 - 1, \quad x \log 1 - x, \\
\cos(x), & \quad \frac{1-x}{\cos(x)}, \quad \frac{x}{\tan(x)}, \quad (1-x)\tan(x).
\end{align*}
\]

### 2.3 Polylogarithmic constants

The probability that a vector \(U\) is such that \(U_1 > U_2, \ldots, U_n\) (the first element is largest) equals \(1/n\), a property that underlies the logarithmic series generator. By sequentially drawing \(r\) several such vectors and requiring success on all \(r\) trials, we deduce constructions for families involving the polylogarithmic functions, \(\Li_r(z) := \sum_{n=1}^{\infty} z^n/n^r\), with \(r \in \mathbb{Z}_{\geq 1}\). Of course, \(\Li_1(1/2) = \log 2\). The few special evaluations known for polylogarithmic values (see the books by Berndt on Ramanujan [3 Ch. 9] and by Lewin [21, 22]) include \(\Li_2(1/2) = \frac{\pi^2}{12} - \frac{1}{2}\log^2 2\) and

\[
(2.11) \Li_3 \left( \frac{1}{2} \right) = \frac{\log^3 2}{6} - \frac{\pi^2 \log 2}{12} + \frac{7\zeta(3)}{8}, \quad \zeta(s) := \sum_{n \geq 1} \frac{1}{n^s}.
\]
By rational convex combinations, we obtain Buffon computers for $\frac{\pi^2}{6}$ and $\frac{\pi}{\sqrt{3}}(3)$. Similarly, the celebrated BBP (Bailey–Borwein–Plouffe) formulae \[2\] can be implemented as Buffon machines.

3 Square roots, algebraic, and hypergeometric functions

We now examine a new brand of generators based on properties of ballot sequences, which open the way to new constructions, including an important square-root mechanism. The probability that, in $2n$ tosses of a fair coin, there are as many heads as tails is $\bar{w}_n = \frac{1}{2^n} \binom{2n}{n}$. The property is easily testable with a single integer counter $R$ subject only to the basic operation $R := R + 1$ and to the basic test $R = 2$. From this, one can build a square-root computer and, by repeating the test, certain algebraic constants can be obtained.

3.1 Square-roots Let $N$ be a random variable with distribution $\text{Geo}(\lambda)$. Assume we flip $2N$ coins and return a success, if the score of heads and tails is balanced. The probability of success is

$$S(\lambda) := \sum_{n=0}^{\infty} (1 - \lambda)^n \bar{w}_n = \sqrt{1 - \lambda}.$$ 

(The final simplification is due to the binomial expansion of $(1 - x)^{-1/2}$.) This simple property gives rise to the square-root construction due to Wästlund \[35\] and Mossel–Peres \[26\]:

$$\Gamma B(\sqrt{1 - \lambda}) := \{ \text{let } N := \Gamma G(\lambda); \text{draw } X_1, \ldots, X_{2N} \text{ with } \mathbb{P}(X_j = +1) = \mathbb{P}(X_j = -1) = \frac{1}{2}; \text{set } \Delta := \sum_{j=1}^{2N} X_j; \text{if } \Delta = 0 \text{ then } \text{return}(1) \text{ else return}(0) \}.$$ 

The mean number of coin flips used is then simply obtained by differentiation of generating functions.

**Theorem 3.1.** \[\{26\} \[35\] \] The square-root construction yields a Bernoulli generator of parameter $\sqrt{1 - \lambda}$, given a $\Gamma B(\lambda)$. The mean number of coin flips required, not counting the ones involved in the calls to $\Gamma B(\lambda)$, is $\frac{\Delta}{\lambda}$. The function $\sqrt{1 - \lambda}$ is strongly realizable.

By complementation of the Bernoulli generator, we also have a construction $\Gamma B(\lambda) \rightarrow \Gamma B(1 - \lambda) \rightarrow \Gamma B(\sqrt{1 - \lambda})$, albeit one that is irregular at 0.

**Note 1.** Compatibility with a pushdown automaton. It can be seen that the number $N$ in the square-root generator never needs to be stored explicitly: an equivalent form is

$$\Gamma B(\sqrt{1 - \lambda}) := \{ \text{do } \{ \Delta := 0; \text{if } \Gamma B(\lambda) = 0 \text{ then break; \text{if } flip = 1 \text{ then } \Delta := \Delta + 1 \text{ else } \Delta := \Delta - 1; \text{if } flip = 1 \text{ then } \Delta := \Delta + 1 \text{ else } \Delta := \Delta - 1 \} \text{ if } \Delta = 0 \text{ then } \text{return}(1) \text{ else return}(0) \}.\$$

In this way, only a stack of unary symbols needs to be maintained: the stack keeps track of the absolute value of $|\Delta|$ stored in unary, the finite control can keep track of the sign of $\Delta$. We thus have realizability of the square-root construction by means of a pushdown (stack) automaton.

This suggests a number of variants of the square-root construction, which are also computable by a pushdown automaton. For instance, assume that, upon the condition “flip=1”, one does $\Delta := \Delta + 2$ (and still does $\Delta := \Delta - 1$ otherwise). The sequences of $H$ (heads) and $T$ (tails) that lead to eventual success (i.e., the value 1 is returned) correspond to lattice paths that are bridges with vertical steps in $\{+2, -1\}$; see \[12\] §VII.8.1. The corresponding counting generating function is then $S(z) = \sum_{n\geq 0} \binom{2n}{n} z^{2n}$, and the probability of success is $(1 - \lambda)S\left(\frac{1}{\lambda}\right)$. As it is well known (via Lagrange inversion), the function $S(z)$ is a variant of an algebraic function of degree 3; namely, $S(z) = (1 - 3zY(z)^2)^{-1}$, where $Y(z) = z + zY(z)^3$, and $Y(z) = z + z^4 + 3z^7 + \cdots$ is a generating function of ternary trees. One can synthesize in the same way the family of algebraic functions

$$S(z) \equiv S^t(z) = \sum_{n\geq 0} \binom{tn}{n} z^{tn},$$

by updating $\Delta$ with $\Delta := \Delta + (t - 1)$.

As a consequence of Theorems 2.3 and 3.1, the function $\sin(\lambda)$ is strongly realizable, since $\sin(\lambda) = \sqrt{1 - \cos(\lambda)^2}$.

3.2 Algebraic functions and stochastic grammars. It is well known that unambiguous context-free grammars are associated with generating functions that are algebraic: see \[12\] for background (the Chomsky–Schützenberger Theorem).

**Definition 3.** A binary stochastic grammar (a “bistoach”) is a context-free grammar whose terminal alphabet is binary, conventionally $\{H, T\}$, where each production is of the form

$$X \rightarrow Hm + Tn,$$

with $m, n$ that are monomials in the non-terminal symbols. It is assumed that each non-terminal is the left hand side of at most one production.

Let $G$, with axiom $S$, be a bistoch. We let $L[G; S]$ be the language associated with $S$. By the classical theory of formal languages and automata, this language can be recognized by a pushdown (stack) automata. The constraint that there is a single production for each non-terminal on the left means that the automaton corresponding to a bistoch is deterministic. (It is then a simple matter to come up with a recursive procedure that parses a word in $\{H, T\}^*$ according to a non-terminal symbols $S$.) In order to avoid trivialities, we assume that all non-terminals are “useful”, meaning...
{let $N := FG(\lambda)$; 
draw $w := X_1 X_2 \cdots X_N$ with $P(X_j = H) = P(X_j = T) = \frac{1}{2}$; 
if $w \in L(G; S)$ then return(1) else return(0) \}.

Figure 4: The algebraic construction associated to the pushdown automaton arising from a bistoch grammar.

that each produces at least one word of $\{H, T\}^*$. For instance, the one-terminal grammar $Y = HY \cup Y + T$ generates all Lukasiewicz codes of ternary trees [12, §I.5.3] and is closely related to the construction of Note 1.

Next, we introduce the ordinary generating function (or ogf) of $G$ and $S$,

$$S(z) := \sum_{w \in L(G; S)} z^{|w|} = \sum_{n \geq 0} S_n z^n,$$

with $S_n$ the number of words of length $n$ in $L(G; S)$. The deterministic character of a bistoch grammar implies that the ogfs are bound by a system of equations (one for each nonterminal): from (3.12), we have $X = z\bar{m} + z\bar{n}$, where $\bar{m}, \bar{n}$ are monomials in the ogfs corresponding to the non-terminals of $m, n$; see [12, §I.5.4]. For instance, in the ternary tree case: $Y = z + zY^3$.

Thus, any ogf $y$ arising from a bistoch is a component of a system of polynomial equations, hence, an algebraic function. By elimination, the system reduces to a single equation $P(z, y) = 0$. We obtain, with a simple proof, a result of Mossel-Peres [26 Th. 1.2]:

Theorem 3.2. (26) To each bistoch grammar $G$ and non-terminal $S$, there corresponds a construction (Fig. 4), which can be implemented by a deterministic pushdown automaton and calls to a $\Gamma B(\lambda)$ and is of type $\Gamma B(\lambda) \rightarrow \Gamma B(S)\left(\frac{1}{2}\right)$, where $S(z)$ is the algebraic function canonically associated with the grammar $G$ and non-terminal $S$.

Note 2. Stochastic grammars and positive algebraic functions. First, we observe that another way to describe the process is by means of a stochastic grammar with production rules $X' \rightarrow \frac{1}{2}m + \frac{1}{2}n$, where each possibility is weighted by its probability (1/2). Then fixing $N = n$ amounts to conditioning on the size $n$ of the resulting object. This bears a superficial resemblance to branching processes, upon conditioning on the size of the total progeny, itself assumed to be finite. (The branching process may well be supercritical, as in the ternary tree case.)

The algebraic generating functions that may arise from such grammars and positive systems of equations have been widely studied. Regarding coefficients and singularities, we refer to the discussion of the Drmota–Lalley–Woods Theorem in [12] pp. 482–493]. Regarding the values of the generating functions, we mention the studies by Kiefer, Luttonberger, and Esparza [17] and by Pivotau, Salvy, and Soria [29]. The former is motivated by the probabilistic verification of recursive Markov processes, the latter by the efficient implementation of Boltzmann samplers.

It is not (yet) known whether a function as simple as $(1 - \lambda)^{-1/3}$ is realizable by a stochastic context-free grammar or, equivalently, a deterministic pushdown automaton. (We conjecture a negative answer, as it seems that only square-root and iterated square-root singularities are possible.)

3.3 Ramanujan, hypergeometrics, and a Buffon computer for $1/\pi$. A subtitle might be: What to do if you want to perform Buffon’s experiment but don’t have needles, just coins? The identity

$$\frac{1}{\pi} = \sum_{n=0}^{\infty} \frac{(2n)!!}{n!} \frac{3n + 1}{2^{2n+4}},$$

due to Ramanujan (see [14] for related material), lends itself somewhat miraculously to evaluation by a simple Buffon computer. The following simple experiment (the probabilistic procedure Rama) succeeds (returns the value 1) with probability exactly $1/\pi$. It thus constitutes a discrete analogue of Buffon’s original, one with only three counters ($T$, a copy of $T$, and $\Delta$).

procedure Rama(); {returns 1 with probability $1/\pi$} 
S1. let $T := X_1 + X_2$, where $X_1, X_2 \in \text{Geom}(\frac{1}{2})$; 
S2. with probability $\frac{1}{2}$ do $T := T + 1$; 
S3. for $j = 1, 2, 3$ do 
S4. draw a sequence of 2$\Delta$ coin flips; 
if ($\Delta = \# \text{ Heads} - \# \text{Tails} \neq 0$ then return(0); 
S5. return(1).

4 A Buffon integrator

Our purpose here is to develop, given a construction of type $\Gamma B(\lambda) \rightarrow \Gamma B(\phi(\lambda))$, a generator for the function

$$\Phi(\lambda) = \frac{1}{\lambda} \int_{0}^{\lambda} \phi(w) \, dw.$$

An immediate consequence will be a generator for $\lambda \Phi(\lambda)$; that is, an “integrator”.

To start with, we discuss a purely discrete implementation of $\Gamma B(\lambda) \rightarrow \Gamma B(U(\lambda))$, with $U \in [0, 1]$, uniformly, where multiple invocations of $\Gamma B(\lambda)$ must involve the same value of $U$. Conceptually, it suffices to draw $U$ as an infinite sequence of flips, then make use of this $U$ to get a $\Gamma B(U)$ and then appeal to the conjunction (product) construction to get a $\Gamma B(U') \cdot \Gamma B(\lambda)$. To implement this idea, it suffices to resort to lazy evaluation. One may think of $U$ as a potentially infinite vector $(v_1, v_2, \ldots)$, where $v_j$ represents the $j$th bit of $U$. Only a finite section of the vector is used at any given time and the $v_j$ are initially undefined (or unspecified). Remember that a $\Gamma B(U)$ is simply obtained by fetching the bit of $U$ that is of order $J$, where $J \in 1 + \text{Geo}(\frac{1}{2})$: cf. Eq. (1.5). In our relaxed lazy context, whenever such a bit $v_j$ is fetched, we first examine
whether it has already been assigned a \{0,1\}-value; if so, we return this value; if not, we first perform a flip, assign it to \(v_j\), and return the corresponding value: see Fig. 5. (The implementation is obvious: one can maintain an association list of the already “known” indices and values, and update it, as the need arises; or keep a boolean vector of not yet assigned values; or encode a yet unassigned position by a special symbol, such as ‘?’, or ‘−1’; See the Appendix for a simple implementation.)

Assume that \(\phi(\lambda)\) is realized by a Bernoulli machine that calls a Bernoulli generator \(\Gamma B(\lambda)\). If we replace \(\Gamma B(\lambda)\) by \(\Gamma B(\lambda U)\), as described in the previous paragraph, we obtain a Bernoulli generator whose parameter is \(\phi(\lambda U)\), where \(U\) is uniform over \([0,1]\). This is equivalent to a Bernoulli generator whose parameter is \(\int_0^1 \phi(\lambda u) \, du = \Phi(\lambda)\), with \(\Phi(\lambda)\) as in (4.13).

**Theorem 4.1.** Let \(\phi(\lambda)\) be realizable by a Buffon machine \(\mathcal{M}\). Then the function \(\Phi(\lambda) = \frac{1}{\lambda} \int_{\lambda}^{1} \phi(w) \, dw\) is realizable by addition of a geometric bag to \(\mathcal{M}\). In particular, if \(\phi(\lambda)\) is realizable, then its integral taken starting from 0 is also realizable.

This result paves the way to a large number of derived constructions. For instance, starting from the even-parity construction of \(\mathcal{M}\), we obtain \(\Phi_0(\lambda) := \frac{1}{\lambda} \int_0^\lambda \frac{1}{1+w} \, dw = \frac{1}{\lambda} \log(1+\lambda)\), hence, by product, a construction for \(\Phi(1+\lambda)\). When we now combine the parity construction with “squaring”, where a \(\Gamma B(p)\) is replaced by the product \(\Gamma B(p) \cdot \Gamma B(p)\), we obtain \(\Phi_1(\lambda) := \frac{1}{\lambda} \int_0^\lambda \frac{1}{1+w^2} \, dw = \frac{1}{\lambda} \arctan(\lambda)\), hence also \(\arctan(\lambda)\). When use is made of the exponential (Poisson) construction \(\lambda \mapsto e^{-\lambda}\), one obtains (by squaring and after multiplications) a construction for \(\Phi_2(\lambda) := \int_0^\lambda e^{-w^2/2} \, dw\), so that the error function (“erf”) is also realizable. Finally, the square-root construction combined with parity and integration provides \(\Phi_3(\lambda) := \int_0^\lambda \frac{\sqrt{1-w^2}}{1+w} \, dw = -1 + \sqrt{1-\lambda^2} + \arcsin(\lambda)\), out of which we can construct \(\frac{1}{2} \arcsin(\lambda)\). In summary:

\[
\Phi(\lambda) = \frac{1}{\lambda} \log(1+\lambda) = \Phi(1+\lambda) = \frac{1}{\lambda} \arctan(\lambda) = \frac{1}{\lambda} \text{erf}(\lambda) = \frac{1}{\lambda} \arcsin(\lambda) + \text{const.}
\]

**Theorem 4.2.** The following functions are strongly realizable (0 ≤ \(x < 1\)):

\[
\log(1+x), \quad \arctan(x), \quad \frac{1}{2} \arcsin(x), \quad \int_0^x e^{-w^2/2} \, dw.
\]

The first two only require one bag; the third requires a bag and a stack; the fourth can be implemented with a string register and bag.

**Buffon machines for \(\pi\).** The fact that \(\Phi_1(1) = \arctan(1) = \frac{\pi}{4}\) yields a Buffon computer for \(\pi/4\). There are further simplifications due to the fact that \(\Gamma B(1)\) is trivial: this computer then only makes use of the \(U\) vector. Given its extreme simplicity, we can even list the complete code of this Madhava–Gregory–Leibniz (MGL) generator for \(\pi/4\):

```plaintext
MGL := proc() do if bag(U)=0 then return(1) fi; if bag(U)=0 then return(0) fi; if bag(U)=0 then return(0) fi; od; end;
```

The Buffon computer based on \(\arctan(1)\) works fine for small simulations. For instance, based on 10,000 experiments, we infer the approximate value \(\pi/4 \approx 0.7876\), whereas \(\pi/4 \approx 0.78539\), with a mean number of flips per experiment about 76. However, values of \(U\) very close to 1 are occasionally generated (the more so, as the number of simulation increases). Accordingly, the expected number of flips is infinite, a fact to be attributed to slow convergence in the Madhava–Gregory–Leibniz series,

\[
\frac{\pi}{4} = \frac{1}{4} - \frac{1}{6} + \frac{1}{8} - \frac{1}{10} + \cdots.
\]

The next idea is to consider formulae of a kind made well-known by Machin, who appealed to arctangent addition formulae in order to reach the record computation of 100 digits of \(\pi\) in 1706. For our purposes, a formula without negative signs is needed, the simplest of which,

\[
\frac{\pi}{4} = \arctan\left(\frac{1}{2}\right) + \arctan\left(\frac{1}{3}\right),
\]
specified as a composition of basic constructions, such as $f \mapsto \exp(-f) \cdot \text{expn}$, $f \mapsto \sqrt{f \cdot \text{sqrt0}}$, $f \mapsto \int f \cdot \text{intl}$, and so on. An interpreter using as source of randomness the built-in function `random` then takes this description and produces a $\{0, 1\}$ result; this interpreter, which is comprised of barely 60 lines, contains from one to about a dozen instructions for each construction. In accordance with the conditions of our contract, only simple register manipulations are ever used, so that a transcription in C or Java would be of a roughly comparable size.

We see here that even a complicated constant such as the “value” of the probability associated with $Z_4$, $$e^{-\frac{1}{6} + \frac{2}{13} \zeta(3) - \frac{1}{4} \sqrt{\pi} \int_0^1 \frac{\frac{2}{3}}{\ln 2} + \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{3}{4} \cdot \text{arctan} \left( \frac{1}{3} \right) \cdot \text{d}Z},$$
is effectively simulated with an error of $4 \times 10^{-3}$, which is once more consistent with normal statistical fluctuations. For such a complicated constant, the observed mean number of flips is a little above 100. Note that the quantity $\zeta(3)$ is produced (as well as retrieved automatically by MAPLE’s symbolic engine!) from Beuker’s triple integral: $\frac{7}{8} \zeta(3) = \int_0^1 \int_0^1 \int_0^1 \frac{1}{1 + x y z} \cdot dx \cdot dy \cdot dz$. (On batches of $10^5$ experiments, that quantity alone consumed an average of 6.5 coin flips, whereas the analogous $\frac{1}{11} \zeta(5)$ required barely 6 coin flips on average.)

Note that the implementation is, by nature, freely extendible. Thus, given integration and our basic primitives (e.g., even($f$) $\equiv \frac{1}{1 + f}$), we readily program an arc-tangent as a one-liner,

$$(5.16) \quad \text{arctan}(f) = f \cdot \left[ \frac{1}{\pi} \int_0^f \frac{\text{d}x}{1 + x^2} \right]$$

and similarly for sine, arc-sine, erf, etc, with the symbolic engine automatically providing the symbolic and numerical values, as a validation.

Here is finally a table recapitulating nine ways of building Buffon machines for $\pi$-related constants, with, for each of the methods, the value, and empirical average of the number of coin flips, as observed over $10^4$ simulations:

<table>
<thead>
<tr>
<th>Li$_4$($\frac{1}{2}$)</th>
<th>Rama</th>
<th>arcsin [1; $\frac{1}{2}$; $\frac{1}{2}$]</th>
<th>$\text{arctan} \left[ \frac{1}{2} + \frac{1}{3} \right]$;</th>
<th>$\zeta(4)$; $\zeta(2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{\pi^2}{12}$</td>
<td>$\frac{\pi}{4}$</td>
<td>$\frac{\pi}{4}$</td>
<td>$\frac{\pi}{4}$</td>
<td>$\frac{\pi}{4}$</td>
</tr>
<tr>
<td>$7.9$</td>
<td>$10.8$</td>
<td>$76.5$ (1000)</td>
<td>$16.2$</td>
<td>$4.9$</td>
</tr>
</tbody>
</table>

(The tag “∞” means that the expected cost of the simulation is infinite—a weak realization.)

6 Conclusion

As we pointed out in the introduction, every computable number can be simulated by a machine, but one that,
in general, will violate our charter of simplicity (as measured, typically, by program size). Numbers accessible to our framework seem not to include Euler’s constant \( \gamma \approx 0.57721 \), and we must leave it as an open problem to come up with a “natural” experiment, whose probability of success is \( \gamma \). Perhaps the difficulty of the problem lies in the absence of a simple “positive” expression that could be compiled into a correspondingly simple Buffon generator. By contrast, exotic numbers, such as \( \pi^{-1/\pi} \) or \( e^{-\sin(1/\sqrt{7})} \) are easily simulated...

On another note, we have not considered the generation of continuous random variables \( X \), specified by a distribution function \( F(x) = P(X \leq x) \). Von Neumann’s original algorithm for an exponential variate belongs to this paradigm. In this case, the bits of \( X \) are obtained by a short computation of \( O(1) \) initial bits, continued by the production of an infinite flow of uniform random bits. This theme is thoroughly explored by Knuth and Yao in [20]. It would be of obvious interest to be able to hybridize the von Neumann-Knuth-Yao generators of continuous distributions with our Buffon computers for discrete distributions. Interestingly, the fact that the Gaussian error function, albeit restricted to the interval \((0, 1)\), is realizable by Buffon machines suggests the possibility of a totally discrete generator for a (standard) normal variate.

The present work is excellently summarized by Keane and O’Brien’s vivid expression of “Bernoulli factory” [19]. It was initially approached with a purely theoretical goal. It then came as a surprise that a priori stringent theoretical requirements—those of perfect generation, discreteness of the random source, and simplicity of the mechanisms—could lead to computationally efficient algorithms. We have indeed seen many cases, from Bernoulli to logarithmic and Poisson generators, where short programs and the execution of just a few dozen instructions suffice!

Acknowledgements. This work was supported by the French ANR Projects GAMMA, BOOLE, and MAGNUM. The authors are grateful to the referees of SODA-2011 for their perceptive and encouraging comments.

References

Appendix: A complete Buffon machine for $\frac{\pi}{8}$

Here is the complete pseudo-code (in fact an executable MAPLE code), cf procedure Pi8 below, for a $\pi/8$ experiment, based on Eq. (4.14) and not using any high-level primitive. It exemplifies many constructions seen in the text. The translation to various low level languages, such as C, should be immediate, possibly up to inlining of code or macro expansions, in case procedures cannot be passed as arguments of other procedures. The expected number of coin flips per experiment is about 4.92.

The flip procedure (returns a pseudo-random bit):

```
flip:=proc() if rand()/10.*2<1/2 then return(1) else return(0) fi end:
```

A $\Gamma B(\frac{1}{2})$ returns a geometric of parameter $\frac{1}{2}$; cf Eq. (1.4):

1. geo_half:=proc () local j; j := 0; while flip() = 0 do j:=j+1 od; return j end:

2. A $\Gamma B(\frac{1}{3})$ returns a Bernoulli of parameter $\frac{1}{3}$; cf Eq. (1.5):

```
bern_third:=proc() local a,b;
    a:=flip(); b:=flip();
    if not ((a=1) and (b=1)) then break fi;
    if (a=0) and (b=0) then return(1)
    else return(0) fi;
end:
```

Bags. Initialization and result of a comparison with a random $U \in [0,1]$; cf Fig. 5 and 4

```
INFINITY:=50:
8. init_bagU:=proc() local j; global U;
    for j from 1 to INFINITY do U[j]:=1/od;
end:
```

10. bagU:=proc() local k; global U; k:=1+geo_half();
11. if U[k]=-1 then U[k]:=flip() fi; return(U[k]);
end:

```
Pi8:=proc() if flip()=0 then ATAN(flip)
    else ATAN(bern_third) fi end:
```

18. if bagU()=1 then if f()=1 then return(f()) fi; fi;
19. return(0); end;
20. return(f()); end;
end:
```

The EVENP construction takes $f \in \Gamma B(\lambda)$ and produces a $\Gamma B(\frac{1}{2\lambda})$; cf Eq. (1.3):

```
EVENP:=proc(f) do if f()=0 then return(1); fi;
13. if f()=0 then return(0) fi; od; end:
```

The main atan construction, based on bags implementing integration, takes an $f \in \Gamma B(\lambda)$ and produces a $\Gamma B(\arctan(\lambda))$; the auxiliary procedure $g()$ builds a $\Gamma B(\lambda^2 U^2)$, with $U \in [0,1]$ random; cf Eq. (16) and Eq. (5.16):

```
ATAN:=proc(f) local g;
14. if f()=0 then return(0) fi;
15. g:=proc() if bagU()=1 then if f()=1 then return(f()) fi; fi;
16. init_bagU();
17. g:=proc() if bagU()=1 then if f()=1 then return(f()) fi; fi;
18. if bagU()=1 then return(f()) fi; fi;
19. return(0); end;
20. EVENP(g); end;
end:
```

```
Pi8:=proc() if flip()=0 then ATAN(flip)
22. else ATAN(bern_third) fi end:
```