HAL
open science

# On Effective Computation of Expectations in Large or Infinite Dimension

Nicolas Bouleau

## To cite this version:

Nicolas Bouleau. On Effective Computation of Expectations in Large or Infinite Dimension. Journal of Computational and Applied Mathematics, 1990, 31, pp.23-34. hal-00448696

**HAL Id: hal-00448696**

**https://hal.science/hal-00448696v1**

Submitted on 19 Jan 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 effective computation of expectations in large or infinite dimension

Nicolas BOULEAU *

published in *Jour. of Computational and Applied Mathematics*, 31, p23-34, (1990)

**Résumé**

Cette étude est une analyse des difficultés inhérentes à l'intégration par des méthodes de Monte Carlo ou de quasi- Monte Carlo. Contrairement à ce qui est écrit parfois ces méthodes ne sont praticables que dans certains cas. Pour le problème important du calcul de l'espérance de fonctionnelles de processus stochastiques, on présente les avantages d'une méthode de simulation en dimension grande ou infinie fondée sur une implémentation de l'opérateur de shift par des pointeurs.

**Abstract**

This study is an analysis of the natural difficulties of integration by Monte Carlo or quasi-Monte Carlo methods. In spite of what is sometimes written, these methods work only in some precise cases. For the important problem of the computation of expectations of functionals of stochastic processes, we present the advantages of a method based on the implementation of the Bernoulli shift operator by pointers.

**Key words**

Monte Carlo, quasi-Monte Carlo, Riemann integrable, discrepancy, effectivity, infinite dimension, Bernoulli shift, pointers.

*LAMM, Ecole Nationale des Ponts et Chaussées, Paris

On one hand simulation methods are often presented for solving stochastic problems, as the only ones which work when no direct computation is possible, but on the other hand they are known to be slow. Some reservations have to be brought to these two statements, first because these methods dont work in any cases but need precise assumptions which are often not satisfied, second because several improvements allow in some cases significant speeding up. The **first part** of this work is an attempt to make this analysis more accurate to clarify what succeeds and what fails when computing the expectation of a random variable by simulation, i.e. by Monte Carlo or Quasi-Monte Carlo sampling and averaging. For quasi-Monte Carlo methods this depends highly on the dimension of the space on which the random variable is defined. That leads to the concept of simulation morphisms. This is presented in paragraphs 1 and 2. But for both Monte Carlo and quasi-Monte Carlo cases a more fundamental difficulty comes from the fact that some random variables (often even very simple ones) are not Riemann integrable in the setting where they are naturally defined (see examples 2,3 and 4 of §3). Finding an other random variable with the same law and Riemann integrable, is in general a difficult problem which remains open in much cases. As exposed in §4 it is nevertheless possible to compute the expectation of non Riemann integrable random variables without proceeding to such a reduction to the Riemann integrable case, if the random variable can be shown to be effectively in $L^1$, that is, can be effectively approximated for the $L^1$-norm by bounded continuous functions. In this part many examples are given and some open questions are set.

The **second part** is devoted to the infinite dimensional case which is shown to be very usual when computing expectations of functionals of stochastic processes such as Markov chains or diffusions.

This is just for introducing the **third part** in which the **shift method** is exposed. Using Birkhoff's pointwise ergodic theorem instead of the law of large numbers is quite a natural idea, but what is more interesting is that, despite the fact that there is no standard speed of convergence for this theorem, this method runs surprisingly fast in usual cases, so much so that it is the only available one in some situations. This is due to the efficient use of pointers that this method allows. As the law of large numbers the shift method has its Monte Carlo and its quasi-Monte Carlo version, the latter being in connexion with the notion of rapidely normal numbers or sequences which replace here low discrepancy sequences. The shift method was exposed in a course of the author at Université Paris 6 in 1987-88, and in two lectures in European universities in 1988. At present time, this method asks several questions to fundamental research especially on inequalities for the speed of convergence and whether random sequences are the fastest or not in infinite dimension, a result of Flajolet, Kirschenhofer and Tichy [19] showing that, in a certain sense, they are not far to be the best ones.

The word effective in the title and along the paper is only a preformalized and naive version of recursive predicate theory.

# I Difficulties of simulation

## I.1 Fast integration in small dimension

Our study being mainly devoted to other cases (integration in large or infinite dimension) we recall this case just for completeness.

Here $X$ is a real random variable defined on the space $([0,1]^s,\ \mathcal{B}([0,1]^s),\ dx)$ and $s$ is a small natural integer. Several papers tends to made the word "small" more precise (cf for instance Warnock [9] or Sarkar and Prasad [10]), (roughly, $s \leq 15$ say). More over $X$ is supposed to be Riemann integrable i.e. can be inserted between two continuous functions whose integrals do not differ from much than $\epsilon$, for every $\epsilon > 0$.

In this case which is quite usual for applications for engineers there is better to do than to pick randomly independently points $U_n$ in $[0,1]^s$ and to compute $\mathbb{E}X$ by the law of large numbers:

$$\mathbb{E}X = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} X(U_n) \quad dx - \text{a.s.}$$

that is to say, in practice, by using pseudo-random sequences.

Indeed, if $\xi = (\xi_n)_{n\in\mathbb{N}^*}$ is a uniformly distributed sequence on $[0,1]^s$, i.e. if the measure $\mu_N = \frac{1}{N}\sum_{n=1}^{N} \delta_{\xi_n}$ converges narrowly to the Lebesgue measure $dx$, it is well known, and easy to check, that $X$ being Riemann integrable, it holds :

$$\int X\, d\mu_N \longrightarrow \int X\, dx.$$

But for some sequences this convergence is faster than with independent uniform random samples. The discrepancy being defined by

$$D_n^*(\xi) = \sup_{x\in[0,1]^s} |\mu_n([0,x_1]\times\ldots\times[0,x_s]) - x_1\ldots x_s|,$$

for independent samples it cannot be asymptotically better than $\sqrt{\frac{\log\log n}{2n}}$, by a form of the iterated logarithm law, while several sequences are known for which it is of the order $\frac{c(\log n)^s}{n}$.

**Examples**

**a) Halton sequences**

Let $r$ be an integer $\geq 2$. If the $r$-adic expansion of $n$ is

$$n = a_0 + a_1 r + \ldots + a_m r^m \quad a_i \in \{0,1,\ldots,r-1\}$$

then $\Phi_r(n)$ is defined by

$$\Phi_r(n) = \frac{a_0}{r} + \ldots + \frac{a_m}{r^{m+1}}$$

which belongs to $[0,1]$, and the Halton sequence is the following

$$\xi_n = (\Phi_{r_1}(n), \ldots, \Phi_{r_s}(n))$$

where $r_1, \ldots, r_s$ are the $s$ first prime numbers. Then it holds [1]

$$D_n^*(\xi) \leq \frac{1}{n} \prod_{i=1}^{s} \frac{r_i \log(r_i n)}{\log r_i} \quad \text{as soon as } n > r_s$$

**b) Hammersley sequences.** The second time a simulation is run, it is generally possible to improve it by using finite sequences, that is the case for Halton sequences. If $(\xi_m)$, $1 \leq m \leq n$ is defined by

$$\begin{aligned} \xi_m &= (\tfrac{m}{n}, \Phi_{r_1}(m), \ldots, \Phi_{rs-1}(m)) \\ \text{then } D_n^*(\xi) &\leq \tfrac{1}{n} \textstyle\prod_{n=1}^{s-1} \tfrac{r_i \log(r_i n)}{\log r_i} \quad \text{as soon as } n \geq r_{s-1}. \end{aligned}$$

cf Hammersley [2]

**c) Irrational translations of the torus.** Let $\alpha_i \in (0,1)$, $i = 1, \ldots, s$, be such that $1, \alpha_1, \ldots, \alpha_s$ be independent over $\mathbb{Q}$, then the sequence

$$\xi_n = (\{n\alpha_1\}, \ldots, \{n\alpha_s\})$$

(where $\{x\} = x \bmod 1$ ) satisfies

$$\forall \epsilon > 0,\ \exists c(\xi, \epsilon),\ \forall N \quad N^*(\xi) \leq c(\xi, \epsilon) \frac{1}{N^{1-\epsilon}}.$$

cf Schmidt [3]

With such sequences, moreover, the stopping criterions are deterministic when using the discrepancy with the Koksma-Hlawka inequality

$$|\mathbb{E}X - \int X \, d\mu_N| \leq V(X) D_N^*(\xi)$$

cf [4], [5] where $V(x)$ is the <u>variation</u> of $X$, in the sense of Hardy-Krauze (what is modulo some technical points related to the mass on the boundary of the cube, the total mass of the measure $\nu$ if it exists such that $\nu([0, x_1] \times \ldots \times [0, x_s]) = X(x) - \mathbb{E}X$ where $x = (x_1, \ldots, x_s) \in [0,1]^s$.

Other sequences are known (cf for instance [6]) satisfying the estimate

$$N D_N^*(\xi) \leq C_s (\log N)^s + O((\log N)^{s-1}) \quad N \geq 2 \tag{1}$$

some with constants $C_s$ as small as possible such as Faure sequences [7] themselves improved by Niederreiter [8]. But the practical difficulty comes from the fact that in (1) the righthand side increases rapidely with the dimension $s$ (despite the fact that $C_s$ itself could be decreasing). That is so true that for $s = 20$ several millions of iterations are necessary to reach similar results as with random sampling (see [9], [10]).

**Remark.** It is nevertheless interesting to note that equidistributed sequences allow the use of *rejection method.* More precisely, if $Y$ is a real random variable which can be simulated by rejection method using a dominating random variable simulated on $[0,1]^p$ and if the random variable $X$ whose expectation is to be computed is given by

$$X = F(Y, U_1, \ldots, U_q)$$

where the $U_i$'s are independent uniform variables independent with $Y$ and $F$ is supposed to be Riemann integrable, then $\mathbb{E}X$ can be computed by an equidistributed sequence on $[0,1]^{p+q}$, see for instance [11].

## I.2 Changes of dimension.

Let $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1)$ and $(\Omega_2, \mathcal{A}_2, \mathbb{P}_2)$ be two Polish spaces equipped with their Borel $\sigma$-field and probability measures $\mathbb{P}_1$, $\mathbb{P}_2$, then if $\Phi$ is a mapping from $\Omega_1$ into $\Omega_2$ such that

i) $\mathbb{P}_1\{x : \Phi \text{ is not continuous at } x\} = 0$

ii) $\Phi_* \mathbb{P}_1 = \mathbb{P}_2$

then if $f$ is Riemann integrable from $\Omega_2$ into $\mathbb{R}$, $f \circ \Phi$ is Riemann integrable over $\Omega_1$ and if $\xi = (\xi_n)$ is $\mathbb{P}_1$-uniformly distributed

$$\frac{1}{N} \sum_{n=1}^{N} f \circ \Phi(\xi_n) \to \int f \circ \Phi \, d\mathbb{P}_1 = \int f \, d\mathbb{P}_2$$

hence $\Phi(\xi_n)$ is $\mathbb{P}_2$-uniformly distributed over $\Omega_2$.

$\Phi$ can be called a *simulation morphism.*

**Example 1.** Let be

$$x = \sum_{n=1}^{\infty} \frac{a_n}{2^n} \quad a_n \in \{0,1\} \text{ with } \forall N \ \sum_{n=N}^{\infty} a_n \neq 0.$$

Then define a map $\Phi_s$ from $(0,1]$ into $[0,1]^s$ by

$$\Phi_s \ : \ x \in (0,1] \longrightarrow (y_1, \ldots, y_s) \in [0,1]^s$$

$$\text{with } y_p = \sum_{n=1}^{\infty} \frac{a_{sn+p-1}}{2^n}$$

then $\Phi_s$ is a simulation morphism.
**Questions** 1) Are there simulation morphisms which from good low discrepancy sequences on $[0,1]$ give good sequences on $[0,1]^s$ ?
2) Are there simulation morphisms $\Phi$ from $[0,1]$ into $[0,1]^s$ such that $f \circ \Phi$ be of bounded variation as soon $f$ is of bounded variation on $[0,1]^s$ ?

## I.3 Reduction of a non Riemann integrable random variable.

**a)** Clearly, if $X$ is a real random variable on $([0,1], \mathcal{B}([0,1]), dx)$ which is not Riemann integrable it is not possible to use directly $X$ to know the law of $X$ by simulation. If $(\xi_n)$ is uniformly distributed, the average

$$\frac{1}{N}\sum_{n=1}^{N} X(\xi_n)$$

does not converge to $\mathbb{E}X$ in general.
**Example 2.** Let be $x \in [0,1]$ given by

$$x = \sum_{n=1}^{\infty} \frac{a_n}{2^n} \quad a_n \in \{0,1\} \text{ with } \forall N \sum_{n=N}^{\infty} a_n \neq 0.$$

Let us define

$$X(x) = \sup_n \frac{1}{n}\sum_{k=1}^{n} a_k$$

then $X$ is not Riemann integrable. Indeed, let be

$$A = \{x : \lim \frac{1}{n}\sum_{k=1}^{n} a_k = \frac{1}{2} \text{ and } a_1 = 0\},$$

clearly for Lebesgue measure $\mathbb{P}(A) = \frac{1}{2}$ and if $x \in A$ then $X(x) < 1$. But in every neighbourhood of $x$ there is a $y$ with $X(y) = 1$. □
**Example 3.** Here is a more probabilistic example. The $X_n$'s being i.i.d. with law $\frac{dx}{2}$ on $[-1,1]$, let

$$S_n = X_1 + \cdots + X_n$$

be the symmetrical random walk. Let us put

$$X = \left(\left(\sup_{n\geq 2} \frac{S_n}{\frac{1}{3}\sqrt{2n\log\log n}}\right) \wedge 2\right) \vee 0$$

then $X$ is bounded but not Riemann integrable. □
**b)** To see where the difficulty lies, let us recall what can be called the fundamental theorem of simulation

**Theorem 1** *For every probability measure $\mu$ on $\mathbb{R}^d$ with compact support, there is a Riemann integrable random variable $H$ with value in $\mathbb{R}^d$ defined on $([0,1], dx)$ with law $\mu$.*

This theorem is effective, its proof gives $h$ explicitely when $\mu$ is known ,see [11]. A more general theorem exists for stochastic processes ,see [12].
**c)** These facts lead us to introduce the following notion:

**Definition 2** *A non Riemann integrable random variable $X$ is said to be* reduced *if a Riemann integrable random variable $Y$ can be exhibited with the same law as $X$.*

Most of almost sure convergence theorems of probability theory give examples of random variables whose reduction is known. The reduction is in general a less difficult operation than to find an analytical expression of the law even if analytical is understood in a large sense (allowing the use of non holomorphic usual functions).
**Example 4.** Let $(U_n)$ be the coordinate mappings of $([0,1]^{\mathbb{N}}, dx^{\otimes \mathbb{N}})$. Let us consider the Markov chain

$$\begin{cases} x_n = \frac{1}{n+2}((n+1)x_{n-1} + 1_{\{U_n < x_{n-1}\}}) \\ x_0 \in (0,1) \end{cases}$$

which is a manner of writing Polya's urn. If $\mathcal{F}_n = \sigma(U_i, i \leq n)$, $(x_n)$ is an $\mathcal{F}_n$-martingale in $[0,1]$ and if $x_\infty = \limsup_n x_n$, then $x_\infty$ is not Riemann integrable. Nevertheless if $x_0 = \frac{1}{2}$, the law of $x_\infty$ is the Lebesgue measure on $[0,1]$. Hence $x_\infty$ is reduced.

## I.4 Computation of the expectations of non bounded or non Riemann integrable random variables: random variables effectively in $L^1$.

**a)** Non bounded random variables are very commonly used in simulation. For instance the log function on $[0,1]$ is used for simulating the exponential law by

$$X = -\frac{1}{\lambda} \log U.$$

But the use of non bounded random variable is impossible in simulation if the position of the poles are unknown. Even when this position is known, the result will depend on the speed with which the uniformly distributed sequence nears the poles.
**Example 5.** To compute $\int_0^1 \frac{dx}{\sqrt{x}} = 2$ by simulation, let us use a uniformly distributed sequence $\xi = (\xi_n)$ over $(0,1)$ and let us consider

$$\frac{1}{N} \sum_{n=1}^{N} \frac{1}{\sqrt{\xi_n}}.$$

If the sequence $\xi_n$ is changed on a nul set $A \subset \mathbb{N}$ (i.e. such that $\frac{1}{N}\sum_{n=1}^{N} 1_A(n) \to 0$) the new sequence is still uniformly distributed. Let us take $\xi_{p^2} = \frac{1}{p^8}$ then

$$\limsup_n \frac{1}{N}\sum_{n=1}^{N} \frac{1}{\sqrt{\xi_n}} = +\infty.$$

**b)** Nevertheless, it is often possible to compute the expectation of non bounded or even of non Riemann integrable functions (without to proceed to a reduction to the Riemann integrable case).

**Example 6.** Let $U_n$ be the coordinates of $([0,1]^{\mathbb{N}}, dx^{\otimes \mathbb{N}})$. Let us put

$$X_n = \sum_{i=1}^{n} \frac{2U_i - 1}{i}$$

$$\text{and } Y = \limsup_n |X_n|$$

It can be proved that, $Y$ is not bounded and that $(Y \wedge 1) \vee (-1)$ is not Riemann integrable. But the martingale $X_n$ converges almost surely and in $L^2$ and

$$\|Y - |X_n|\|_{L^2}^2 \leq \sum_{i=n+1}^{\infty} \frac{1}{i^2}.$$

Hence to compute $\mathbb{E}Y$ with the precision $\epsilon > 0$, it is sufficient to chose $n$ large enough so that $\sum_{n+1}^{\infty} \leq \frac{\epsilon}{2}$ and then to compute by simulation $\mathbb{E}|X_n|$ with the accuracy $\frac{\epsilon}{2}$. □

On the classical probability spaces used in simulation which are compact, $\mathbb{E}Y$ can be computed as soon as a sequence of continuous functions $f_n$ is effectively known such that $\|Y - f_n\|_{L^1} \to 0$.

In practice on $[0,1]^s$, this property of being *effectively in* $L^1$ splits into two properties :

i) An effective bound decreasing to zero is known showing that

$$\lim_{k \uparrow \infty} \int_{|Y|>k} |Y| \, d\mathbb{P} = 0$$

ii) For each $k$ a sequence $f_{n,k}$ of polynomials with rational coefficients is known such that the limit

$$\lim_n \|f_{n,k} - (Y \wedge k) \vee (-k)\|_{L^1} = 0$$

is proved with effective bound.

**Remark** : If the belonging of $Y$ to $L^1$ is known only by Fatou's lemma, the variable $Y$ does not need to be effectively in $L^1$. In general the almost sure limits of non uniformly integrable positive supermartingales are not effectively in $L^1$ unless they can be obtained also by an other way. Even the dominated convergence theorem does not yield the $f_{n,k}$'s.

# II Integration of functionals of stochastics processes

In situations related to stochastic processes, one has often to do with integration in infinite dimension. When computing (expectations of functions of) entrance times or entrance points of a stochastic process in a set or in a event, or more general functionals of the sample paths, one is dealing with random variables which can be bounded and Riemann integrable but defined on $([0,1]^{\mathbb{N}},\ \mathcal{B}([0,1]^{\mathbb{N}}),\ dx^{\otimes\mathbb{N}})$.

Let us consider as a generic example the case of Markov chains (to which reduces also the case of diffusions by discretization of SDE's) and let us look how the classical simulation method works.

Let $X_n$ be a Markov chain defined by

$$X_{n+1} = F(X_n, n, U_{n+1}) \quad , X_0 = x, \tag{2}$$

where $F : \mathbb{R}^d \times \mathbb{N} \times [0,1]^k \longrightarrow \mathbb{R}^d$ is given and where the $U_n$'s are i.i.d. with law $dx$ on $[0,1]^k$.

If the following quantities have to be computed

- $\mathbb{E}(T)$, $\mathbb{E}(X_T)$, $\mathbb{E}[G(X_T,T)]$ for $T$ stopping time such as $T = \inf\{n : (X_n, n) \in A\}$,
- $\mathbb{E}\sum_{k=1}^{T} C(X_k, X_{k+1})$ for $C : \mathbb{R}^2 \to \mathbb{R}$ representing costs,
- or more generally $\mathbb{E}[H(X_\bullet)]$ where $H : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$,

by (2) $X_n$ can actually be defined on the probability space $([0,1]^k)^{\mathbb{N}}$ with $U_i$ being the coordinates, and the problem is to find good double sequences

$$\begin{array}{rcll} \mathbf{u}_1 & = & (u_{01}, \ldots, u_{i1}, \ldots) & \in \ ([0,1]^k)^{\mathbb{N}} \\ & \vdots & & \vdots \\ \mathbf{u}_j & = & (u_{0j}, \ldots, u_{ij}, \ldots) & \in \ ([0,1]^k)^{\mathbb{N}} \\ & \vdots & & \vdots \end{array}$$

in such a way that

$$\frac{1}{N}\sum_{j=1}^{N} H(X_\bullet(\mathbf{u}_j))$$

converges rapidely to $\mathbb{E}[H(X_\bullet)]$.

**a)** Let us take $k = 1$ for simplicity. Then let us remark at first that if we take

$$u_{ij} = V_i + \xi_j \mod 1$$

where $(V_i)_{i\in\mathbb{N}}$ is i.i.d. on $[0,1]$ and $(\xi_j)_{j\geq 1}$ is a (low-discrepancy) uniformly distributed sequence on $[0,1]$, one gets a sequence $(\mathbf{u}_j)$ which *is not* uniformly distributed over $([0,1])^{\mathbb{N}}$ for $dx^{\otimes\mathbb{N}}$.

**b)** On the contrary if $(u_{ij})$ is chosen in such a way that for every $q \in \mathbb{N}$

$$u_j^q = (u_{0j}, \ldots, u_{qj})$$

be uniformly distributed over $[0,1]^q$ then $(\mathbf{u}_j)$ is uniformly distributed over $[0,1]^{\mathbb{N}}$ and

$$\frac{1}{N}\sum_{j=1}^{N} \delta_{\mathbf{u}_j} \longrightarrow dx^{\otimes \mathbb{N}}$$

in the narrow sense.

With Halton sequences, we get infinite dimensional Halton sequences which where studied by Sobol' [13]. For such sequences $(\mathbf{u}_j)$ the average

$$\frac{1}{N}\sum_{n=1}^{N} g(\mathbf{u}_j)$$

converges to $\mathbb{E}g$ for smooth $g$ i.e. for "almost cylindrical" functions $g$. Nevertheless these sequences are unusable in practice because, even for cylindrical functions, the convergence is two slow (or more exactly two late) in high dimension as we have already seen in part I.

# III The Bernoulli shift method.

The infinite or high dimensional case turns out to have its specific features which allow to avoid picking completely different sample paths and doing the average on them.

## III.1 An efficient method.

Although the pointwise ergodic theorem has no typical speed of convergence (such as the iterated logarithm law) the method, we shall explain now, seems to be the most efficient one at present for computing high dimensional functionals of stochastic processes: The implementation of this algorithm takes great advantage of the notion of pointers (or equivalent device of piles management).

Let us consider the preceding Markov chain

$$X_{n+1} = F(X_n, n, U_{n+1}) \quad , X_0 = x$$

where the $(U_n)$ are i.i.d. random variables with law $dx$ on $[0,1]$. As we saw the process $(X_n)_{n\geq 1}$ can be represented on $([0,1]^{\mathbb{N}},\ dx^{\otimes \mathbb{N}})$. For the notations to be simple, it is actually convenient here to represent it on $([0,1]^{\mathbb{N}^*},\ dx^{\otimes \mathbb{N}^*})$. The $U_n$'s being the coordinate mappings, $n \geq 1$.

Let us take, for instance, an entrance time $T$, $T = \inf\{n \geq 1 : \ X_n \in A\}$ and a functional $G(X_T, T)$. The expectation $\mathbb{E}G(X_t, T)$ can be computed by Birkhoff's theorem

$$\mathbb{E}G(X_T, T) = \lim_{N \uparrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} G(X_t, T) \circ \theta^n \quad \text{a.s.} \tag{3}$$

as soon as $G(X_T, T) \in L^1$, where $\theta$ is the shift operator satisfying

$$\forall \omega \in [0,1]^{\mathbb{N}^*} \quad U_n \circ \theta(\omega) = U_{n+1}(\omega) \quad \forall n \geq 1$$

The use of pointers can be explained as follows. For applying formula (3) we have to compute $G(X_T, T)$ on successive points $\omega$, $\theta(\omega)$, $\theta^2(\omega), \ldots$ of $[0,1]^{\mathbb{N}^*}$, these points are sequences of points in $[0,1]$

$$\begin{array}{rcl} \omega & = & (U_1(\omega), U_2(\omega), \ldots, U_k(\omega), \ldots) \\ \theta(\omega) & = & (U_2(\omega), U_3(\omega), \ldots, U_{k+1}(\omega), \ldots) \\ & \vdots & \end{array}$$

Then if the stopping time $T$ is finite , what we suppose here, only a finite length of each of these sequences have to be computed : the sequence $\omega$ being computed until $T(\omega)$ and this is told by the test

$$\text{¡test¿} \begin{cases} X_k(\omega) \text{ still outside } A & \Rightarrow \quad k < T(\omega) \\ X_k(\omega) \in A \text{ for the first time} & \Rightarrow \quad k = T(\omega) \end{cases}$$

Suppose the sequence $\omega = (U_1, U_2, \ldots)$ has been picked out, and the pseudo-random numbers $U_1$, $U_2, \ldots$ have been put in pointers as indicated below :

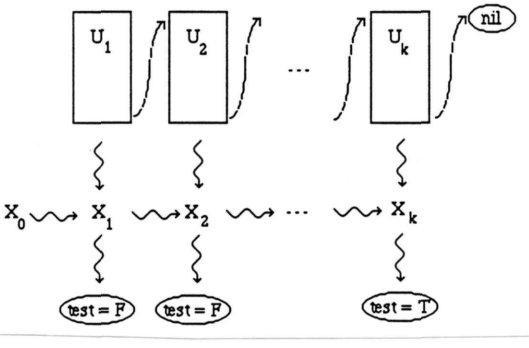


Figure 1

The following sequence $\theta(\omega) = (U_2, U_3, \ldots)$ is partially already chosen, either it

is long enough or it must be lengthened in what case we have the new scheme :

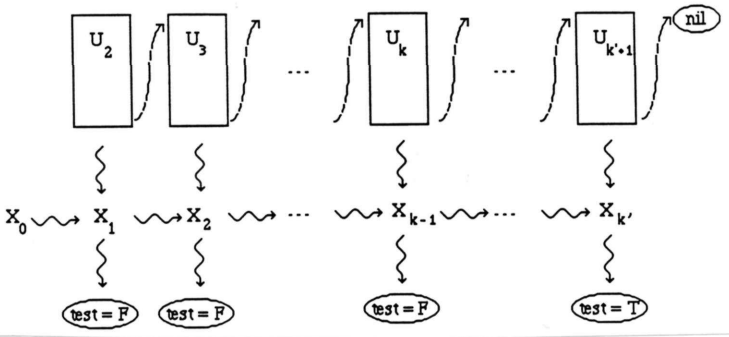


Figure 2

In these pictures the arrows $\rightsquigarrow$ represent computations to do, that is, mainly, to compute

$$X_{n+1} = F(X_n, n, U_{n+1})$$

but in this computation of $F(x, n, y)$ only the variable $x$ (and $n$) is new and partial computations depending only on $y$ can be stored in the box of $U_{n+1}$ in such a way that only easy computations remain to be done. This storage has of course to be made during the lengthening of the sequence when it occurs as explained above.

## III.2 Speed of convergence

Let us quote the following two results on speed of convergence for the Bernoulli shift $([0,1]^{\mathbb{N}}, dx^{\otimes \mathbb{N}})$, cf Krengel [14].

**Proposition 3** *For every sequence $(\alpha_n)_{n\in\mathbb{N}}$, $\alpha_n > 0$, $\alpha_n \to 0$, there exists a continuous function $f$ from $[0,1]^{\mathbb{N}}$ into $\mathbb{R}$ such that*

$$\frac{\left|\frac{1}{n}\sum_{i=0}^{n-1} f \circ \theta^i - \mathbb{E}f\right|}{\alpha_n} \to +\infty \quad a.s.$$

**Proposition 4** *For every sequence $(c_n)$, $c_n > 0$, increasing to infinity, with $c_1 \geq 2$, there exists a Borel set $A$ with $\mathbb{P}(A) = \frac{1}{2}$ on which*

$$\left|\frac{1}{n}\sum_{i=0}^{n-1} 1_A \circ \theta^i - \frac{1}{2}\right| \leq \frac{c_n}{n} \quad \forall n$$

They show that the speed of convergence can be arbitrarily slow or arbitrarily near of $O(\frac{1}{n})$.

Nevertheless, a speed of convergence can be obtained if the function whose expectation is to be computed fulfills additional assumptions. Some of these results will be published elsewhere.

## III.3 Connexion with normal numbers

Let us replace $[0,1]^{\mathbb{N}}$ by $\{0,1\}^{\mathbb{N}}$ for simplicity of the discussion.

The best points for integration by the shift method are fast normal numbers. Let us recall that a real number $x$ in $[0,1]$, or a sequence of binary digits, is called normal if the sequence $x \circ \theta^n$ is uniformly distributed over $\{0,1\}^{\mathbb{N}}$ equipped with the measure $(\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1)^{\otimes \mathbb{N}}$. Explicit normal numbers are known, the Champernowne sequence for instance. There are other known normal sequences which are generalisations of Champernowne's one (see for instance Rauzy [15] [16] or Dumont [18]). Some of these constructions make use of $(s,m)$-de Bruijn sequences which are periodic mappings $u$ from $\mathbb{N}$ into $\{0,\ldots,m-1\}$ of period $m^s$ such that the $m^s$ strings of length $s$

$$(u(n),\ldots,u(n+s-1)) \quad n = 0,1,\ldots,m^s-1$$

are all different. Algorithms are known for generating infinite de Bruijn sequences (see Ivanyi [17] and its references), that is to say infinite sequences whose initial sections of length $m^s$, considered as loops, are $(s,m)$-de Bruijn sequences for every $s$.

The efficiency of these sequences is difficult to evaluate, and is largely an open subject. One of the main difficulty is to summarize the performance of a sequence by a single scale while there is naturally an infinity of discrepancies one on each finite product space.

On the space $\{0,1\}^{\mathbb{N}}$ Flajolet et al. [19] give an asymptotic result which shows that in a certain sense, random sequences are not far to be the best ones.

**Ending remarks.**

**a)** What was explained for a Markov chain is of course valid for a large class of stochastic processes or stochastic fields which, in usual cases, can be defined in a natural way on $([0,1]^{\mathbb{N}}, \mathcal{B}([0,1]^{\mathbb{N}}), dx^{\otimes \mathbb{N}})$ after discretization if necessary.

**b)** It can happen that the process, whose functional is studied, be itself mixing or simply ergodic. In this case the shift method can be applied to the process itself instead of to the underlying coordinates. This works as soon as a sample path of arbitrary length of the process can be simulated.

**c)** It is well known that pointers are particularly convenient to represent tree structures, for instance binary trees. Then the space which takes the place of the product $[0,1]^{\mathbb{N}}$ is the space $[0,1]^B$ or a space of the form $A^B$ where $A$ is some probability space and $B = \{l,r\}^{(\mathbb{N})}$ (finite sequences of $l$'s and $r$'s). Such a space with a suitable extension of Birkhoff's theorem is an interesting device for dealing with branching processes which are of great practical interest. Nevertheless the "linear" above explained shift method works actually for branching processes as well, and after our experiments this "linear" implementation runs even faster.

**d)** The shift method can be slightly improved by using the two-sided Bernoulli shift, which allows to never throw a box away. The idea is to simulate two independent

copies of the process, one with the classical right to left shift the other with a left to right shift beginning with the box left by the first one. But the benefit is negligible in high dimension.

# References

[1] J.H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multidimensional integrals.
Numer. Math. 2,1960,84-90

[2] J.M. Hammersley. MonteCarlo methods for solving multivariate problems.
Ann. New York Acad. Sci. 86,1960,844-874

[3] W.M. Schmidt. Simultaneous approximation to agebraic numbers by rationals.
Acta Math. 125,1970,189-201

[4] I.F. Koksma. Een algemeene stelling uit de theorie der gelikmatige Verdeeling modulo 1.
Math. B (Zutphen), 11,1942-43,7-11

[5] E. Hlawka. Funktionen von beschränkter variation in der theorie Gleichverteilung.
Ann. Math. pure Appl., 54,1961,325-333

[6] Hua Loo Keng, Wang Yuan. Application of number theory to numerical analysis.
Springer 1981

[7] H. Faure. Discrépance de suites associées à un système de numération (en dimension $s$)
Acta Arithm.,41,1982,337-351

[8] H. Niederreiter. Low-discrepancy and low-dispersion sequences.
J. of number theory, vol 30,1,51-70

[9] T. Warnock. in "Application of number theory to numerical analysis" S.K. Zaremba ed.
Acad. Press N.Y. 1972

[10] P.K. Sarkar,M.A. Prasad. A comparative study of pseudo and quasi random sequences for the solution of integral equations.
J. of computational Physics,68,1987,66-88

[11] N. Bouleau. Probabilités de l'ingénieur, variables aléatoires et simulation.
Hermann,Paris,1986

[12] A.V. Skorohod. Studies in the theory of random processes.
Addison Wesley,Reading,1965

[13] J.M. Sobol'.Onthe convergence of infinite dimensional cubature and simulation of Markov chains.
Voprosy Vyčisl. i Prikl. Mat. (Taškent),32,1975,161-167

[14] U. Krengel. Ergodic theorems
de Gruyter,1985

[15] G. Rauzy. Propriétés statistiques des suites arithmétiques.
Presses Univ. de France, 1985

[16] G. Rauzy. Dicrépance d'une suite complètement équirépartie.
Ann. Fac. Sci Toulouse, vol III,1981,105-112

[17] A. Ivanyi. Construction of infinite de Bruijn arrays.
Discrete App. Math. 22,1988-89,289-293

[18] J.M. Dumont, A. Thomas. Une modification multiplicative des nombres g-normaux.
Ann. Fac. Sci. Toulouse, vol III, 1986-87,367-373

[19] P. Flajolet, P. Kirchenhofer,R.F. Tichy. Deviations from uniformity in random strings.
Probab. Th. Rel. fields , 80,1988,139-150