A chaotic non-uniform random variate generator

Alain BERLINET and Gérard BIAU*

Department of Probability and Statistics
University Montpellier II
Place Eugène Bataillon
34095 Montpellier cedex 5
France

Abstract

This paper deals with non-uniform random variable generation. Given an *a.e.* positive probability density f_* defined on an interval of the real line, it is always possible to construct a statistically stable transformation with f_* as invariant density. We use this result to design a new non-uniform random variable generator that generalizes the classical congruential algorithm. First, some important concepts of dynamical systems pertaining to ergodic theory are reviewed. Then, the construction of the random generator is given. Finally, results of simulations are presented and discussed.

Index Terms **— Dynamical system, Frobenius-Perron operator, invariant measure, exactness, statistical stability, random variable generator.**

1 Introduction

Uniform random numbers - and more generally sequences of numbers with a specified distribution - are basic ingredients for simulation, Monte Carlo methods, survey sampling, numerical analysis, computer programming, experimental physics and many other fields of applied sciences where random events play a major role. This, together with the increasing development in computer science, explains the constant interest of the statistical community in random variable generation. Background material in random generation can be found in the books by Knuth [14] (uniform generation) and Devroye [7] (non-uniform generation). For lists of references, see for example the surveys by Eichenauer-Herrmann [9], [10], L'Ecuyer [17], L'Ecuyer and Hellekalek [18] and Ripley [23].

Suppose we are given some probability distribution function F on the real line and let us set the problem of random number generation on computers as follows. Starting from some

*All correspondence to be addressed to this author. Email: biau@ensam.inra.fr.

initial values we would like to get, after some (quick) computation, a set of N numbers mimicking as much as possible the behavior of a set of N independent random variables with common distribution function F. Let us begin with $N = 1$ and a given initial random variable X_0. We are looking for a deterministic transformation S (a *discrete dynamical system*) that generates recursively a sequence

$$\begin{cases} X_n = S(X_{n-1}), \quad n \geq 1, \\ X_0 \end{cases}$$

of random variables whose distribution approaches the distribution F, as n increases. Denoting by F^{-1} the generalized inverse of F, we know that if X is a random variable with distribution function F, $F(X)$ is uniformly distributed on $]0,1[$ and that if U is uniform on $]0,1[$, $F^{-1}(U)$ has distribution function F. Now,

$$F(X_n) = F \circ S \circ F^{-1}\big(F(X_{n-1})\big).$$

So, if the distributions of $F(X_n)$ and $F(X_{n-1})$ tend to the uniform one, it is reasonable to think that good candidates S should be such that $T = F \circ S \circ F^{-1}$ preserves the uniform distribution. Thus we are led to look for "chaotic enough" transformations preserving the uniform distribution.

A lot of work has been devoted in recent years to understand and characterize properties of deterministic (*i.e.* non-random) dynamical systems. Very often, the approximation of the behavior of a dynamical system is done by direct simulations of the individual trajectories. This method is particularly useful in the situation where a specific trajectory has to be approximated for a finite period of time. However, if one is interested in the long term behavior and if the underlying system exhibits complicated dynamics, then the information derived from trajectories may prove inadequate. Rather in this case it seems more appropriate to determine a statistical description of the dynamical behavior, and this is typically achieved with the concepts of *invariant measure* and *Frobenius-Perron operator* (see Collet [6] and Lasota and Mackey [16] for general results). In this way, deterministic dynamical systems have recently drawn the attention of a large part of the statistical community, see for example the survey papers by Berliner [1], Chatterjee and Yilmaz [5], Tong [24], or the studies of Bosq and Guégan [4] and Guerre and Maës [12].

This paper mixes questions about random variable generation and dynamical systems theory. More precisely, we explore the possibility of using the statistical properties of some specific deterministic (discrete) systems, called *statistically stable*, in order to design a non-uniform random variable generator (preferably called random *variate* generator, see Devroye [7]). Roughly speaking, statistical stability of a dynamical system means that any distribution of initial states will be transformed by successive iterations of the system into an invariant distribution, independently of the initial one. Now, given an *a.e.* positive probability density f_* defined on an interval of the real line, it is always possible to construct a statistically stable transformation with f_* as invariant density. We use this result to design a non-uniform random variate generator that generalizes in some way the classical congruential algorithm.

The content of the paper is as follows. In Section 2, we review some essential concepts of deterministic (discrete) dynamical systems pertaining to ergodic theory (Frobenius-Perron operators, invariant measures, exactness and statistical stability) which will be needed later on. This first section is intended for readers not aware of dynamical systems theory. Our main reference is the work by Lasota and Mackey [16]. In Section 3, we first briefly survey the random variate generation field. We then construct and test the new non-uniform random generator.

2 Some results on deterministic dynamical systems

2.1 Transformations of densities

Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space and let $S : \mathcal{X} \to \mathcal{X}$ be a measurable transformation[1]. Consider a random variable X defined on some probability space $(\Omega, \zeta, \mathbf{P})$ with values in $\mathcal{X}$ and assume that the probability distribution P_X of X is absolutely continuous with respect to μ, with density

$$f = \frac{\mathrm{d}P_X}{\mathrm{d}\mu}. \tag{1}$$

Often, one is interested in finding a probability density for the random variable $S \circ X$, which is the transform of the random variable X by the measurable map S. Suppose, for example, that $(\mathcal{X}, \mathcal{A}, \mu) = (U, \mathcal{B}(U), \lambda)$, where U is an open set in $\mathbb{R}^d$ ($d \geq 1$), $\mathcal{B}(U)$ is the σ-algebra of Borel sets in U and λ is the Lebesgue measure. Then, provided the transformation S is continuously differentiable and invertible, it is well known (see Billingsley [3] for details) that the random variable $S \circ X$ is absolutely continuous with respect to λ, with probability density

$$f_1(x) = f\big(S^{-1}(x)\big)|J(x)|, \quad x \in U, \tag{2}$$

where J is the determinant of the Jacobian matrix of the inverse transformation S^{-1}.

Formula (2) can be extended to more general measurable spaces and transformations with the help of the so-called *Frobenius-Perron operator*. For this purpose, denote by P_{SX} the probability distribution of the random variable $S \circ X$ and assume for the moment that $S \circ X$ admits a probability density f_1 with respect to μ (rigor will come later on). Then, for every $A \in \mathcal{A}$,

$$P_{SX}(A) = \int_A f_1(x)\, \mu(\mathrm{d}x).$$

[1] Throughout this paper, a measure space will always be understood to be σ-finite.

But

$$P_{SX}(A) = \mathbf{P}(S \circ X \in A) = \mathbf{P}\big(X \in S^{-1}(A)\big) = \int_{S^{-1}(A)} f(x)\,\mu(\mathrm{d}x),$$

so that f and f_1 are linked by the functional equation

$$\int_A f_1(x)\,\mu(\mathrm{d}x) = \int_{S^{-1}(A)} f(x)\,\mu(\mathrm{d}x), \quad \text{for } A \in \mathcal{A}. \tag{3}$$

Here and throughout, $S^{-1}(A)$ means the counterimage of the set A, that is

$$S^{-1}(A) = \{x \in \mathcal{X} : S(x) \in A\}.$$

Now, an interesting problem is to derive sufficient conditions for equation (3) to define without ambiguity a probability density for the random variable $S \circ X$. In order to give an answer, we need the following definition:

Definition 2.1 *Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space. A measurable transformation $S : \mathcal{X} \to \mathcal{X}$ is* nonsingular *if $\mu\big(S^{-1}(A)\big) = 0$ for all $A \in \mathcal{A}$ such that $\mu(A) = 0$.*

Assume now that we are given a nonsingular transformation $S : \mathcal{X} \to \mathcal{X}$ as well as a random variable with probability density given by (1). Then, some algebra and the Radon-Nikodym theorem (see Lasota and Mackey [16] for a rigorous proof) show that the random variable $S \circ X$ is absolutely continuous with respect to μ and that its density f_1 is μ-*a.e.* uniquely determined by the functional equation (3). Thus we are led to the following definition:

Definition 2.2 *Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space. If $S : \mathcal{X} \to \mathcal{X}$ is a nonsingular transformation, then the unique operator $P : L_1(\mathcal{X}) \to L_1(\mathcal{X})$ defined by*

$$\int_A Pf(x)\,\mu(\mathrm{d}x) = \int_{S^{-1}(A)} f(x)\,\mu(\mathrm{d}x), \quad \textit{for } f \in L_1(\mathcal{X}) \textit{ and } A \in \mathcal{A},$$

is called the Frobenius-Perron operator *corresponding to S.*

As usual, $L_1(\mathcal{X})$ means the space of classes of all possible real-valued integrable functions on $\mathcal{X}$ equipped with the norm

$$||f||_1 = \int_{\mathcal{X}} |f(x)|\,\mu(\mathrm{d}x).$$

Elements of $L_1(\mathcal{X})$ are represented by functions that can differ on a set of μ-measure zero. Thus, for any of these functions, properties such as $f \geq 0$ or $Pf \geq 0$ hold μ-*a.e.* When it is clear that we are dealing with elements of $L_1(\mathcal{X})$, we will drop the "μ-*a.e.*" mention. The following properties of P are more or less immediate.

1. P is a linear operator, *i.e.*

$$P(\alpha f + \beta g) = \alpha Pf + \beta Pg$$

for all $(\alpha, \beta) \in \mathbb{R}^2$ and $(f, g) \in L_1(\mathcal{X}) \times L_1(\mathcal{X})$.

2. $Pf \geq 0$ whenever $f \geq 0$.

3. For all $f \in L_1(\mathcal{X})$,

$$\int_{\mathcal{X}} Pf(x)\,\mu(\mathrm{d}x) = \int_{\mathcal{X}} f(x)\,\mu(\mathrm{d}x).$$

   In particular, P preserves the L_1 norm of a nonnegative function and transforms a probability density into a probability density.

4. For all $f \in L_1(\mathcal{X})$,

$$||Pf||_1 \leq ||f||_1,$$

   i.e. P is a *contraction*.

5. Define $S^0 = \mathrm{Id}_{\mathcal{X}}$ (the identity on $\mathcal{X}$), $S^1 = S$ and $S^n = S \circ S^{n-1}$ for $n \geq 2$. If P_n is the Frobenius-Perron operator corresponding to S^n, then $P_n = P^n$, where P is the Frobenius-Perron operator corresponding to S.

It is worth noting that the linear operator P tells us how the density f of a random variable will be transformed by a given map S into a new density Pf. As an example, it will be useful in this paper to have in mind the transformation

$$S_r : \big([0,1], \mathcal{B}\big([0,1]\big), \lambda\big) \to \big([0,1], \mathcal{B}\big([0,1]\big), \lambda\big)$$

given by

$$S_r(x) = rx \pmod 1, \tag{4}$$

where r is an integer greater than 1 (for asymptotic results when r is irrational, see Hlawka [13] and Kuipers, Niederreiter and Shiue [15]). The notation $rx \pmod 1$ means $rx - [rx]$, where $[rx]$ is the integer part of rx, *i.e.* the largest integer q such that $rx - q \geq 0$. The mapping (4) is usually called the *r-adic transformation* and is illustrated in Figure 1 for $r = 2$ (the *dyadic* transformation) and for $r = 5$. It is easily seen that the counterimage under S_r of any interval $[0, x]$ in $[0, 1]$ is given by

$$S_r^{-1}\big([0,x]\big) = \bigcup_{i=0}^{r-1} \Big[\frac{i}{r}, \frac{i}{r} + \frac{x}{r}\Big],$$

so that S_r is nonsingular and admits as associated Frobenius-Perron operator

$$P_r f(x) = \frac{\mathrm{d}}{\mathrm{d}x} \int_{S_r^{-1}([0,x])} f(t)\,\lambda(\mathrm{d}t) = \frac{\mathrm{d}}{\mathrm{d}x} \sum_{i=0}^{r-1} \int_{i/r}^{i/r+x/r} f(t)\,\lambda(\mathrm{d}t),$$

i.e.

$$P_r f(x) = \frac{1}{r} \sum_{i=0}^{r-1} f\Big(\frac{i}{r} + \frac{x}{r}\Big). \tag{5}$$

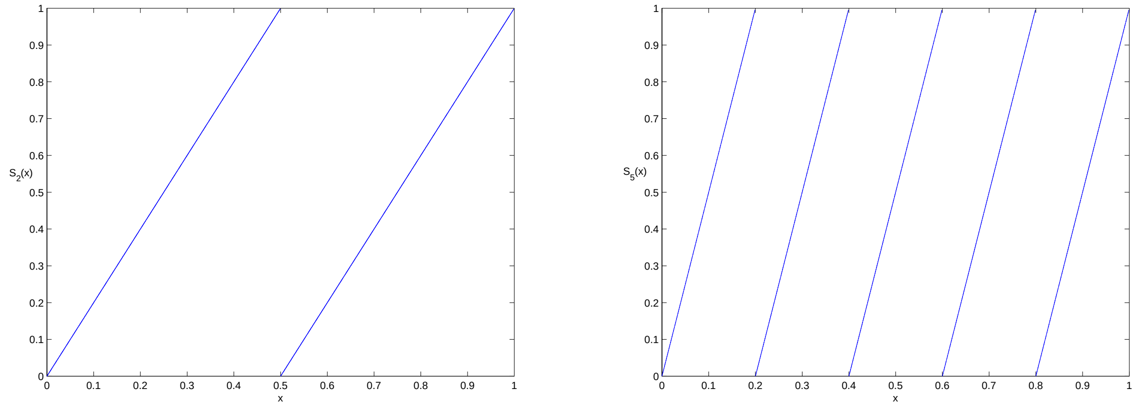

Figure 1: The **2**-adic (left) and the **5**-adic (right) transformations.

2.2 Dynamical systems and invariant measures

Assume that we are given a measurable space $(\mathcal{X}, \mathcal{A}, \mu)$ and a nonsingular transformation $S : \mathcal{X} \to \mathcal{X}$ associated with a Frobenius-Perron operator denoted by P. Assume further that X_0 is a random variable with values in $\mathcal{X}$. Successive applications of the map S lead to consider the sequence of random variables

$$X_0, \quad S(X_0), \quad S^2(X_0) = S \circ S(X_0), \ldots \tag{6}$$

If the initial random variable X_0 has density f_0 with respect to the measure μ, then Property 5 (page 5) of the Frobenius-Perron operator P gives us a sequence of probability densities associated with the sequence (6), namely

$$f_0, \quad P f_0, \quad P^2 f_0, \ldots$$

In this iterative framework, the measure space $(\mathcal{X}, \mathcal{A}, \mu)$ together with the application S is usually called a *deterministic discrete dynamical system.* The set $\mathcal{X}$, that represents the set of all possible states of the system, is the *state space.* Given an initial condition x_0 in $\mathcal{X}$ (*i.e.* given a *realization* of the random variable X_0), the mapping S generates recursively a sequence

$$x_0, \quad S(x_0), \quad S^2(x_0) = S \circ S(x_0), \ldots$$

called the *trajectory* or the *orbit* of the point x_0. The sequence $\big(S^n(x_0)\big)_{n \geq 0}$ can be considered as describing the state of a system at successive instants, and the map S performs the (discrete) time evolution.

As an illustration, let us turn back to the example (4) of the r-adic transformation. Formula (5) leads by induction to

$$P_r^n f_0(x) = \frac{1}{r^n} \sum_{i=0}^{r^n - 1} f_0\Big(\frac{i}{r^n} + \frac{x}{r^n}\Big). \tag{7}$$

Thus, $P_r^n f_0(x)$ appears as the mean of the values of f_0 at points

$$\frac{x}{r^n}, \frac{x}{r^n} + \frac{1}{r^n}, \ldots, \frac{x}{r^n} + \frac{r^n - 1}{r^n}.$$

Observe that when f_0 is a continuous element of $L_1\big([0,1]\big)$, the right-hand side of equation (7) approaches the Riemann integral of f_0 on $[0,1]$, so that

$$\lim_{n\to\infty} P_r^n f_0(x) = \int_0^1 f_0(t)\,\lambda(\mathrm{d}t), \quad \text{uniformly in } x,$$

and thus

$$\lim_{n\to\infty} \left|\left| P_r^n f_0 - \int_0^1 f_0(t)\,\lambda(\mathrm{d}t) \right|\right|_1 = 0.$$

Now, since the set of continuous functions is dense in $L_1\big([0,1]\big)$ and since the operator P is contractive (Property 4, page 5), this limit holds for any f_0 in $L_1\big([0,1]\big)$. In particular, for any initial probability density f_0,

$$\lim_{n\to\infty} ||P_r^n f_0 - \mathbf{1}||_1 = 0,$$

where we denote by $\mathbf{1}$ the constant function of $L_1\big([0,1]\big)$ defined by $\mathbf{1}(x) = 1$ for all $x \in [0,1]$. Plainly, $P_r \mathbf{1} = \mathbf{1}$. This means that the uniform density is invariant under the action of P_r, or, equivalently, that any uniform random variable on $[0,1]$ will be transformed by S_r into a random variable with the same uniform distribution. Consequently, if we start the sequence (6) with a uniformly distributed random variable X_0, then all the elements of the sequence $\big(S_r^n(X_0)\big)_{n\geq 0}$ will have the same uniform distribution. This example leads us directly to the central concept of measure-preserving transformation.

Definition 2.3 *Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space and $S : \mathcal{X} \to \mathcal{X}$ a measurable transformation. Then S is said to be* measure-preserving *if*

$$\mu\big(S^{-1}(A)\big) = \mu(A) \quad \textit{for all } A \in \mathcal{A}.$$

Observe that any measure-preserving transformation is necessarily nonsingular (Definition 2.1). Since the property of measure preservation is dependent on S as well as μ, it is alternatively said that the measure μ is *invariant* under S. For a very complete study of measure-preserving maps and examples, the reader is referred to Billingsley [3]. The following result, whose proof is straightforward, establishes the connection between invariant measures and fixed points of the Frobenius-Perron operator.

Theorem 2.1 *Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space, $S : \mathcal{X} \to \mathcal{X}$ a nonsingular transformation, and P the Frobenius-Perron operator associated with S. Let f be a nonnegative element of $L_1(\mathcal{X})$. Then the measure μ_f given by*

$$\mu_f(A) = \int_A f(x)\,\mu(\mathrm{d}x), \quad \textit{for } A \in \mathcal{A},$$

is invariant under S if, and only if, f is a fixed point of P, that is $Pf = f$.

Basically, Theorem 2.1 means that a random variable with density f will be transformed by S into a random variable with the same density if, and only if, f is a fixed point of P. When the measure space is finite, *i.e.* $\mu(\mathcal{X}) < \infty$, we denote as above by $\mathbf{1}$ the constant function of $L_1(\mathcal{X})$ defined by $\mathbf{1}(x) = 1$ for all $x \in \mathcal{X}$. In this case, we observe that the original measure μ is invariant if, and only if, $P\mathbf{1} = \mathbf{1}$. This is for example the case for the r-adic transformation and the Lebesgue measure.

2.3 Exact transformations and statistical stability

We start with a definition.

Definition 2.4 *Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a probability space (i.e. $\mu(\mathcal{X}) = 1$) and $S : \mathcal{X} \to \mathcal{X}$ a measure-preserving transformation such that $S(A) \in \mathcal{A}$ for each $A \in \mathcal{A}$. If*

$$\lim_{n\to\infty} \mu\big(S^n(A)\big) = 1 \quad \text{for every } A \in \mathcal{A},\ \mu(A) > 0, \tag{8}$$

then S is called exact.

As pointed out in Lasota and Mackey [16], condition (8) has a very simple interpretation: *if we start with a set A of initial conditions of nonzero measure, then after a large number of iterations of an exact transformation S the points will have spread and completely filled the space $\mathcal{X}$.* In regard to this property, exact transformations are probably among the most chaotic one can think of. The next theorem deals with the invariant probability density of the Frobenius-Perron operator associated with an exact transformation.

Theorem 2.2 *Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a probability space, $S : \mathcal{X} \to \mathcal{X}$ a measure-preserving transformation, and P the Frobenius-Perron operator corresponding to S. Then*

(i) *If S is exact, $\mathbf{1}$ is the* unique *stationary probability density of P, i.e. $P\mathbf{1} = \mathbf{1}$.*

(ii) *S is exact if, and only if,*
$$\lim_{n\to\infty} ||P^n f_0 - \mathbf{1}||_1 = 0$$
for every initial probability density f_0 on $\mathcal{X}$.

Proof

Since the original measure is invariant under S, $P\mathbf{1} = \mathbf{1}$ (Theorem 2.1). Uniqueness of this invariant probability density under P stems from the ergodicity of the exact transformation. See Lasota and Mackey [16] for details and proof of point (ii).

■

Provided the transformation S is exact, point (i) of Theorem 2.2 means firstly that any uniformly distributed random variable will be transformed by S into a new uniformly distributed random variable and secondly, that the uniform random variable is the only absolutely continuous random variable with this property. Further, point (ii) of the theorem tells us that the probability distribution of any absolutely continuous initial random variable X_0 is strongly convergent towards the uniform distribution on $\mathcal{X}$ if, and only if, the mapping S is exact. As a matter of fact, Theorem 2.2 is a powerful tool to determine whether a given transformation S with an invariant measure is exact. Consider for example the r-adic transformation introduced in the previous paragraphs. We already know, from equation (5), that $P_r\mathbf{1} = \mathbf{1}$ and, from equation (7) that

$$\lim_{n\to\infty} ||P_r^n f_0 - \mathbf{1}||_1 = 0$$

holds for any initial probability density f_0. Thus, thanks to point (ii) of Theorem 2.2, we conclude that S_r is an exact transformation.

If P is the Frobenius-Perron operator associated with an exact transformation, Theorem 2.2 states that any sequence $(P^n f_0)_{n\geq 0}$ is consistent in the L_1 sense towards the uniform invariant density, whatever may be the choice of the initial probability density f_0. This is actually an amazing statistical regularity property that should be compared with the very irregular behavior of the individual trajectories induced by Definition 8. With this respect, an exact transformation is said to be *statistically stable*. The rigorous definition is:

Definition 2.5 *Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space, $S : \mathcal{X} \to \mathcal{X}$ a nonsingular transformation, and P the Frobenius-Perron operator associated with S. The transformation S is said to be* statistically stable *if there exists a unique stationary probability density f_* and if*

$$\lim_{n\to\infty} ||P^n f_0 - f_*||_1 = 0$$

for every initial probability density f_0.

A very interesting feature of the theory of deterministic dynamical systems is the following one: given a λ-*a.e.* positive probability density, say f_*, *defined on an interval $\mathcal{X}$ of the real line*, it is always possible to construct a statistically stable transformation $S : \mathcal{X} \to \mathcal{X}$ with f_* as invariant density. This construction is achieved via an exact transformation T. Clearly, the use of a different exact transformation T_1 will yield a different statistically stable transformation S_1, but one that has the same stationary probability density as S. We give the key theorem.

Theorem 2.3 *Let $T :]0,1[\to]0,1[$ be a nonsingular transformation and let f_* be a given* λ-a.e. *positive probability density on (a, b), with a and b finite or not. Let a second transformation $S : (a, b) \to (a, b)$ be given by $S = F_*^{-1} \circ T \circ F_*$, where F_* stands for the distribution function of f_*, that is*

$$F_*(x) = \int_a^x f_*(t)\,\lambda(\mathrm{d}t), \quad a \leq x \leq b.$$

Then T is exact if, and only if, S is statistically stable. f_* is then the probability density invariant with respect to S.

Proof

See Lasota and Mackey [16]. Observe however that since the probability density f_* is assumed to be λ-*a.e.* positive, the increasing function F_* is invertible on (a, b).

■

The conjugation mechanism of Theorem 2.3 is summarized by a diagram in Figure 2.

$$\begin{tikzcd}
]0,1[\arrow[d, "T"'] & (a,b) \arrow[l, "F_*"'] \arrow[d, "S"] \\
]0,1[\arrow[r, "F_*^{-1}"'] & (a,b)
\end{tikzcd}$$

Figure 2: Construction of a statistically stable transformation **S** from an exact transformation **T** (see Theorem 2.3).

As an example, let us go back to the r-adic transformation (4) with $r = 2$, denoted here by T and whose Frobenius-Perron operator is given in (5). Consider the exponential density $f_*(x) = \lambda \exp(-\lambda x)$ defined on $[0, +\infty[$ with $\lambda > 0$. A straightforward calculation shows that $S = F_*^{-1} \circ T \circ F_*$ is given by

$$S(x) = \begin{cases} -\ln\left(2\exp\left(-\lambda x\right) - 1\right)/\lambda & \text{if } x \in [0, \ln 2/\lambda[\\ x - \ln 2/\lambda & \text{if } x \in [\ln 2/\lambda, +\infty[. \end{cases}$$

Thanks to Theorem 2.3, we know that the transformation S is statistically stable with invariant density the exponential (λ) one. Hence, if we start the sequence (6) with a random variable X_0 distributed according to some initial probability density f_0 on $[0, +\infty[$, then, asymptotically, the random variable $S^n(X_0)$ should be distributed according to the exponential (λ) density. Assume now that we are given a set of N initial realizations $x_0^1, \ldots, x_0^N$ of the random variable X_0. To each of these states, we apply recursively the map

S, thereby obtaining N new states after each iteration, namely $S^n(x_0^1), \ldots, S^n(x_0^N)$. Thus, if n is sufficiently large, the data $S^n(x_0^1), \ldots, S^n(x_0^N)$ should be approximately distributed according to the exponential (λ) density on $[0, +\infty[$. By this way, we are now in a position to generate realizations of any given λ-*a.e.* positive probability density from an initial set of points. The properties of this random variate generator are studied in the next section.

3 The random variate generator

3.1 Random variate generation

In this paragraph, we briefly summarize the random variate generation problem. For clarity sake, we first review the uniform random variate generation framework and then turn to the non-uniform one.

3.1.1 Uniform random variate generation

The goal of a uniform random variate generator is to produce a sequence of numbers that mimics the realization of a sequence of *independent and identically distributed* (*i.i.d.*) random variables uniformly distributed over some set U, which is here (and most often) the unit interval $[0, 1]$. Since 60 years, a lot of effort has been spent on the design of (necessarily deterministic) computer programs that produce sequences of numbers appearing random. General background material on the so-called *pseudo-random numbers* and their generation can be found in the book by Knuth [14] and, for example, in the survey articles by Eichenauer-Herrmann [9], [10], L'Ecuyer [17], L'Ecuyer and Hellekalek [18] and Ripley [23]. By far the most popular random number generators in use today are special cases of the following scheme, called *linear congruential method*, and defined by

$$u_{n+1} = ru_n + s \pmod{M}, \quad n \geq 0, \tag{9}$$

where the integer u_0 $(0 \leq u_0 < M)$ is the *starting value*, the integer r $(1 \leq r < M)$ is the *multiplier*, the integer s $(0 \leq s < M)$ the *increment* and the integer $M > 0$ is the *modulus*. For example, the sequence obtained when $M = 32$, $r = 5$, $s = 0$ and $u_0 = 1$ is

$$1, 5, 25, 29, 17, 21, 9, 13, 1, \ldots \tag{10}$$

A sequence $(v_n)_{n \geq 0}$ of *linear congruential pseudo-random numbers* in the interval $[0, 1]$ is then obtained by the normalization $v_n = u_n/M$. Clearly, any sequence $(v_n)_{n \geq 0}$ generated by the process (9) can take at most M values of the form j/M for $j \in \{0, 1, \ldots, M-1\}$. A congruential sequence always gets trapped into a loop, *i.e.* there is ultimately a cycle of numbers that is repeated endlessly. The length of the cycle is called the *period*. If v_0 is the starting value, and if the sequence returns back to v_0, the period p is $\min\{i : v_i = v_0\}$ and it may of course depend on v_0. Thus $p \leq M$ and if equality occurs the generator is said to have *full period*. For example, sequence (10) has period 8. If M is large and if the coefficients r and s are carefully chosen, the sequence $(v_n)_{n \geq 0}$ will take most of the representable values in $[0, 1]$. We refer to Knuth [14] for conditions on the coefficients r, s and M leading to maximal period.

Uniformity and independence of the generated numbers $(v_n)_{n\geq 0}$ can be checked by statistical tests. Two kinds of tests are distinguished: the *empirical tests*, such as the chi-squared and the Kolmogorov-Smirnov tests, for which one manipulates groups of numbers of the sequence and evaluates certain statistics; and *theoretical tests*, which check the intrinsic quality of generators (see Knuth [14], L'Ecuyer and Hellekalek [18] and Niederreiter [20] for details). The last category of tests arises from the fact that linear congruential sequences exhibit a lattice structure which implies several regularities. This inherent deficiency stems from the simple nature of the underlying linear recursion (9) and cannot be overcome even by the most judicious choice of parameters. This feature can render linear congruential pseudo-random number generators useless in simulations where random irregularities are required. A typical example of regularities in the distribution of pairs of linearly generated numbers is shown in Figure 3.

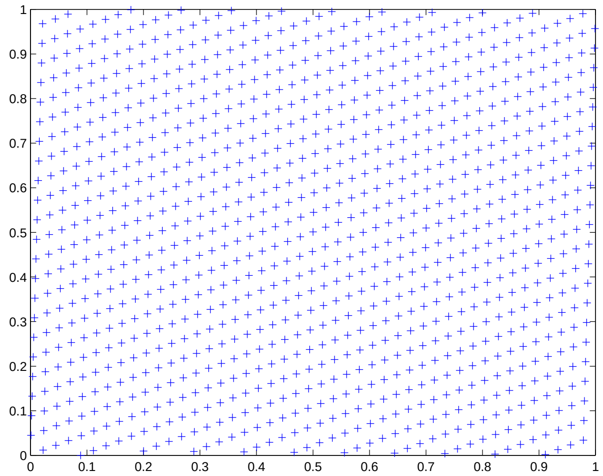

Figure 3: Pairs $(\mathbf{v_n}, \mathbf{v_{n+1}})$ of the linear congruential generator (9) with $\mathbf{r} = \mathbf{45}$, $\mathbf{s} = \mathbf{1}$ and $\mathbf{M} = \mathbf{1024}$. Observe the very high regularity of the set of points resulting of this random generator.

One can establish a connection between the r-adic transformation described in example (4) and the linear congruential pseudo-random number generator (9). For that purpose, assume that the linear congruential generator (9) is *multiplicative*, *i.e.* that $s = 0$ or

$$u_{n+1} = r u_n \pmod{M}, \quad n \geq 0.$$

The sequence $(v_n)_{n\geq 0}$ of pseudo-random numbers in the interval $[0, 1]$ is thus defined by

$$v_{n+1} = \frac{r u_n \pmod{M}}{M}, \quad n \geq 0,$$

which is easily seen to be the same as

$$v_{n+1} = rv_n \pmod 1, \quad n \geq 0.$$

Consequently, denoting by S_r the r-adic mapping, the sequence $(v_n)_{n\geq 0}$ is given by

$$(v_n)_{n\geq 0} = \left(S_r^n(v_0)\right)_{n\geq 0}, \quad \text{with } v_0 = \frac{u_0}{M}.$$

The pseudo-random number sequence produced by a multiplicative linear congruential generator is thus nothing else than the sequence of states $\left(S_r^n(v_0)\right)_{n\geq 0}$ generated by the r-adic map S_r from the initial state $v_0 = u_0/M$, $0 \leq u_0 < M$.

3.1.2 Non-uniform random variate generation

In this framework, an unlimited uniform $[0,1]$ *i.i.d.* sequence $(v_n)_{n\geq 0}$ is assumed to be available. The aim is to develop algorithms to convert this sequence to independent realizations of a *known* univariate distribution. Several general algorithms have been designed and tested so far. See Devroye [8] for a comprehensive review on the topic. See also Niederreiter [21] and Niederreiter and Shiue [22]. One of the oldest and simplest method to generate a sample with a given distribution function F is the *inversion method* based on the following lemma:

Lemma 3.1 *Let F be a distribution function on $\mathbb{R}$ with generalized inverse F^{-1} defined on $]0,1[$ by*

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}.$$

If U is a uniform random variable on $]0,1[$, then $F^{-1}(U)$ has distribution function F.

In the context of Lemma 3.1, the distribution function F need not be invertible in the usual sense (compare with Theorem 2.3). This lemma is in general used for random variables whose (generalized) inverse distribution function F^{-1} is explicitly known. This is for example the case for the exponential, Cauchy, Rayleigh, triangular, Pareto or Gumbel distributions. As pointed out by Devroye [8]: *The inversion method is the only truly universal method*: *if all we can do is compute $F(x)$ for all x, and we have enough* (i.e., *infinite*) *time on our hands, then we can generate random variates with distribution function F.* Many other non-uniform generation algorithms are available, for continuous as well as for discrete distributions (acceptance-rejection, specific methods as *e.g.* Box-Müller method...). For more information on the topic, the reader is referred to the above-mentioned literature.

3.2 Construction

Let us now turn to the very construction of the non-uniform random variate generator. We assume given a λ-*a.e.* positive probability density f_*, distinct from the uniform one and defined on an interval (a,b) of the real line, a and b finite or not. Throughout, we denote by F_* (*resp.* F_*^{-1}) the distribution function (*resp.* the inverse distribution function) associated with f_*. Precisely:

$$F_*(x) = \int_a^x f_*(t)\,\lambda(\mathrm{d}t), \quad a \leq x \leq b.$$

Our goal is to generate N independent realizations $x_1, \dots, x_N$ of a random variable distributed according to F_*. As pointed out in the previous section, the simulation algorithm is based upon Theorem 2.3 and can be summarized as follows:

1. Denote by S_r the exact r-adic ($r > 1$) transformation defined in (4).

2. Generate N independent realizations $x_0^1, \dots, x_0^N$ of a random variable distributed according to any probability density on (a, b) with respect to the Lebesgue measure.

3. Repeat, from $j = 0$ to $n - 1$ ($n \geq 1$ is fixed and is the number of iterations)

$$x_{j+1}^i = F_*^{-1} \circ S_r \circ F_*(x_j^i), \quad 1 \leq i \leq N,$$

   or perform, more efficiently,

$$x_n^i = F_*^{-1} \circ S_r^n \circ F_*(x_0^i), \quad 1 \leq i \leq N.$$

If the probability density f_* is assumed to be the uniform density on $[0, 1]$, then $F_*^{-1} \circ S_r \circ F_*$ reduces to the r-adic transformation S_r, which has been seen to be closely related to the linear congruential generator (9). It is in this sense that we say that the algorithm under study generalizes the usual uniform $[0, 1]$ congruential method. For a related approach, see Li and Yorke [19]. See also Góra, Byers and Boyarsky [11].

According to Theorem 2.3, for large enough n, $x_n^1, \dots, x_n^N$ should be independent (since $x_0^1, \dots, x_0^N$ are) realizations of a random variable distributed according to F_*. Let us now go back to the proof of Theorem 2.3. If Q_r (*resp.* P_r) denotes the Frobenius-Perron operator associated with the transformation $F_*^{-1} \circ S_r \circ F_*$ (*resp.* with the transformation S_r), it can be shown (Lasota and Mackey [16]) that

$$||Q_r^n f_0 - f_*||_1 = ||P_r^n \varphi(f_0) - \mathbf{1}||_1,$$

where f_0 is any initial probability density on (a, b) and $\varphi(f_0)$ is a probability density on $]0, 1[$, depending upon f_0. Now, from equation (7),

$$||P_r^n \varphi(f_0) - \mathbf{1}||_1 = \mathrm{O}\Big(\frac{1}{r^n}\Big),$$

so that

$$||Q_r^n f_0 - f_*||_1 = \mathrm{O}\Big(\frac{1}{r^n}\Big). \tag{11}$$

Thus, by an adequate choice of the parameter r, the quantity $||Q_r^n f_0 - f_*||_1$ can be made arbitrarily small after only one iteration. For instance, to achieve $||Q_r^1 f_0 - f_*||_1 = \mathrm{O}(10^{-\alpha})$ $(\alpha \geq 0)$ after the first iteration $\big(n = 1$ in equation (11)$\big)$, the choice $r = 10^\alpha$ will do.

Remark The limit of double-precision computer packages like *Matlab* is about 16 decimal places. Consequently, the result of the operation $10^\alpha x$ (mod 1) will be accurate up to about $16 - \alpha$ decimals. The work-around is easy. Code the state x as long strings of $0's$ or $1's$, and do binary arithmetic. Another solution is to use formal computation on softwares (like *Maple* or *Mathematica*) that allow to work with virtually unlimited length numbers.

Thus, for large enough r, point 3 of the generation algorithm can be restated as follows:

$$x_i = F_*^{-1} \circ S_r \circ F_*(x_0^i), \quad 1 \leq i \leq N,$$

where we have changed the notation x_1^i into x_i for clarity sake. Save the choice of the constant r, the random variate generator defined in this paragraph is quite similar to the inversion random variate generator defined in Lemma 3.1, the main difference being that the initial probability density need not be uniform any more. But there is more. Choose a^* and b^* in the interval (a, b) so that $a < a^* < b^* < b$. Assume now that the starting values $x_0^1, \dots, x_0^N$ form a regular grid on $[a^*, b^*]$, *i.e.* that

$$x_0^1 = a^*, \quad x_0^2 = a^* + \frac{b^* - a^*}{N - 1}, \quad \dots, \quad x_0^N = b^*.$$

The values $x_0^1, \dots, x_0^N$ can be thought of as *dependent* realizations of a random variable distributed according to the uniform density on $[a^*, b^*]$. If the constant r associated with the r-adic transformation is large enough it has been observed in the simulations (see the next paragraph for details) that the linear structure of the initial sample is completely broken and disappears after the first iteration. Thus, the generated realizations $x_1, \dots, x_N$ will appear independent, or at least uncorrelated, although the initial ones $x_0^1, \dots, x_0^N$ were not. This interesting feature is due to the exactness property of the r-adic transformation and is enlightened in the next paragraph with the use of a rigorous statistical test.

3.3 Simulations

The notations are those adopted in the previous paragraph. We have chosen to simulate N independent realizations $x_1, \dots, x_N$ of a random variable distributed according to the Gumbel density f_* on the real line $\mathbb{R}$:

- **Density**:

$$f_*(x) = \exp\big(-x - \exp(-x)\big) \quad \text{on } (a, b) =]-\infty, +\infty[.$$

- **Distribution function**:

$$F_*(x) = \exp\big(-\exp(-x)\big), \quad x \in \mathbb{R}.$$

- **Inverse distribution function**:

$$F_*^{-1}(x) = -\ln\big(-\ln(x)\big), \quad x \in]0,1[.$$

In accordance with the results of the previous paragraph, the value of the constant r has been fixed once and for all to $r = 10^7$. Two initial starting data were considered:

(i) $N = 1000$ independent realizations of a random variable distributed according to the standard normal density $\mathcal{N}(0,1)$.

(ii) The regular grid $-1, -0.9980, -0.9960, \ldots, 0.9960, 0.9980, 1$ (size $N = 1001$).

Figure 4 **(b)** and **(c)** shows the kernel estimates $\hat{f}_*$ (see Devroye [8]) of the simulated probability density based upon the data $x_1, \ldots, x_N$ in both cases (i) and (ii). As a benchmark, a kernel estimate obtained with Gumbel distributed variates generated by the inversion method from a uniform sample of size $N = 1000$ is also shown in **(a)**. In **(a)**, **(b)** and **(c)** the kernel is Gaussian and the smoothing parameter is optimal according to the L_2 criterion error (Berlinet and Devroye [2]). Of course, the use of a kernel estimate in this framework is purely didactic as we know the theoretical form of the underlying probability density. Our goal is simply to *illustrate* the behavior of the algorithm under study.

Figure 5 depicts the pairs (x_i, x_{i+1}) obtained from the regular grid described in (ii). This figure clearly shows that the initial linear structure of the input data has been broken. Following Li and Yorke [19], we tested the independence of x_i from x_{i+1} as follows. We split the real line into 20 intervals, say $A_1, \ldots, A_{20}$, of equal probability with respect to the Gumbel density, and we let $\psi(j,k,N)$ be the number of pairs (x_i, x_{i+1}), $i = 1, \ldots, N$, for which $x_i \in A_j$ and $x_{i+1} \in A_k$ (as suggested by Li and Yorke [19] we define the final pair to be (x_N, x_1) for simplicity). Independence of the pairs (x_i, x_{i+1}) implies an expected value $\psi^*(j,k,N)$ for $\psi(j,k,N)$ to be $\psi^*(j,k,N) = 1/20 * 1/20 * N$. The goodness of the test is based on the statistic

$$S_N = \sum_{j,k=1}^{20} \frac{\big[\psi(j,k,N) - \psi^*(j,k,N)\big]^2}{\psi^*(j,k,N)}.$$

Independence implies for this statistic an asymptotic χ^2 distribution with $d = 20*(20-1)-1 = 379$ degrees of freedom since $\sum_{j=1}^{20}\psi(j,k,N) = \sum_{k=1}^{20}\psi(j,k,N)$, or, almost equivalently, $Z_N = (2S_N)^{1/2} - (2d-1)^{1/2}$ should be normally distributed with unit variance and mean 0. In order to strengthen the value of the test, the calculation of the statistics S_N and Z_N was conducted for 101 distinct Gumbel data sets. Each of these data sets was obtained by the algorithm under study from a regular grid between -1 and 1: the first grid had 950 equally spaced points, the second had 951 points, ... and the last had 1050 points. Table 1, **A** displays the mean results. Since $Z_N < 2$ we may accept the null hypothesis that the pairs (x_i, x_{i+1}) are independent at level 0.05. This "mixing" property could be used in some cases to circumvent the regularity problem of pseudo-random numbers generated by

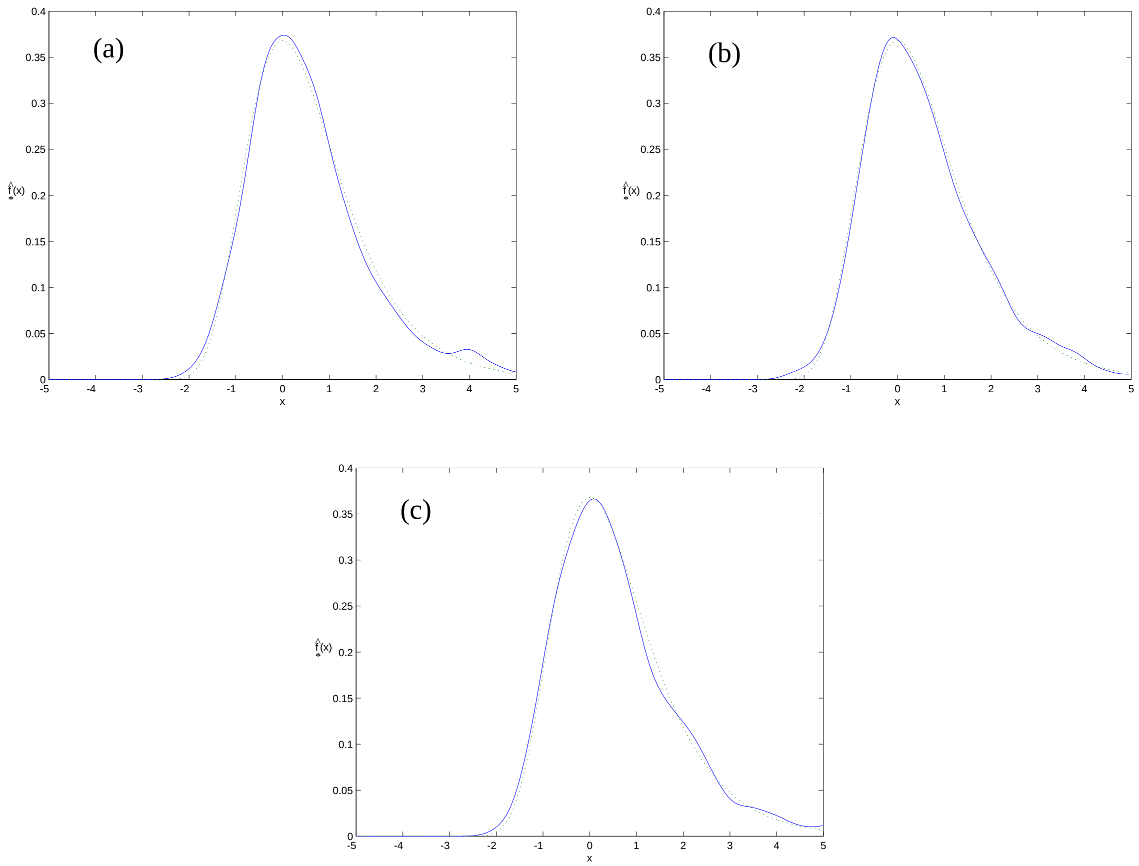

Figure 4: **(a)** Kernel estimate obtained with Gumbel data generated from a uniform sample by the inversion method ($\mathbf{N = 1000}$). **(b)** Kernel estimate obtained with Gumbel data generated from a Gaussian sample by the algorithm under study ($\mathbf{N = 1000}$). **(c)** Kernel estimate obtained with Gumbel data generated from the regular grid described in **(ii)** by the algorithm under study ($\mathbf{N = 1001}$). The Gumbel density is also shown (dotted lines).

Situation	S_N	Z_N
A	406.0439	0.9590
B	73.6563	-15.3764
C	398.6563	0.7231

Table 1: Values of the $\mathbf{S_N}$ and $\mathbf{Z_N}$ statistics. **(A)** Gumbel data generated by the algorithm under study from **101** regular grids (see text). **(B)** Gumbel data generated by the inversion method from the uniform data described in Figure 3. **(C)** Gumbel data generated by the algorithm under study from the uniform data described in Figure 3.

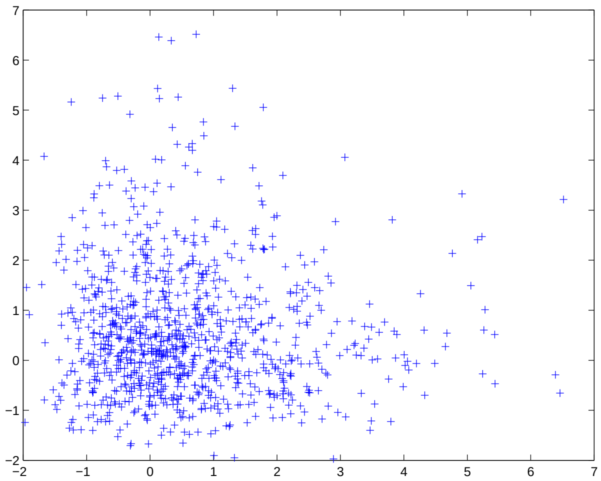

Figure 5: Pairs (x_i, x_{i+1}) of the Gumbel data generated by the algorithm under study from the regular grid described in **(ii)**.

the linear congruential method (9). Indeed, we saw in the first paragraph of this section that the numbers generated by a linear congruential algorithm present a lattice structure inherent to the nature of the recursion (9) (see Figure 3). When such pseudo-random numbers are used as input of the inversion method described in Lemma 3.1, then the non-uniform data generated by this method will exhibit an undesirable regular structure (see Figure 6, **a**). Further, Table 1, **B** clearly shows that the independence hypothesis of the generated Gumbel data is not viable in this case. On the other hand, the regular lattice structure of the linear recursive pseudo-random numbers disappears under the action of the non-uniform generator presented in this paper (Figure 6, **b**). This is confirmed by Table 1, **C** that tells us that we may accept the null hypothesis that the generated pairs (x_i, x_{i+1}) are independent.

Lastly, we illustrate in Figure 7 the three stages occuring in the random variate generator S under study (remember that $S = F_*^{-1} \circ S_r \circ F_*$, where S_r is the r-adic map). Part **(a)** of the figure recalls the regular repartition in the unit square of the uniform congruential data pairs (same as Figure 3). Part **(b)** depicts the pairs $\big(F_*(x_i), F_*(x_{i+1})\big)$. Part **(c)** shows the pairs $\big(S_r \circ F_*(x_i), S_r \circ F_*(x_{i+1})\big)$ and part **(d)** presents the final Gumbel-like data pairs $\big(F_*^{-1} \circ S_r \circ F_*(x_i), F_*^{-1} \circ S_r \circ F_*(x_{i+1})\big)$ (same as Figure 6, **b**). These figures clearly emphasize the major role played by the r-adic transformation in the simulation process.

4 Conclusion

According to Devroye [7], a good non-uniform random variate generator has to be *exact*. In this context, exactness means that the random variates obtained via the generator in use have the exact required distribution. If we assume that:

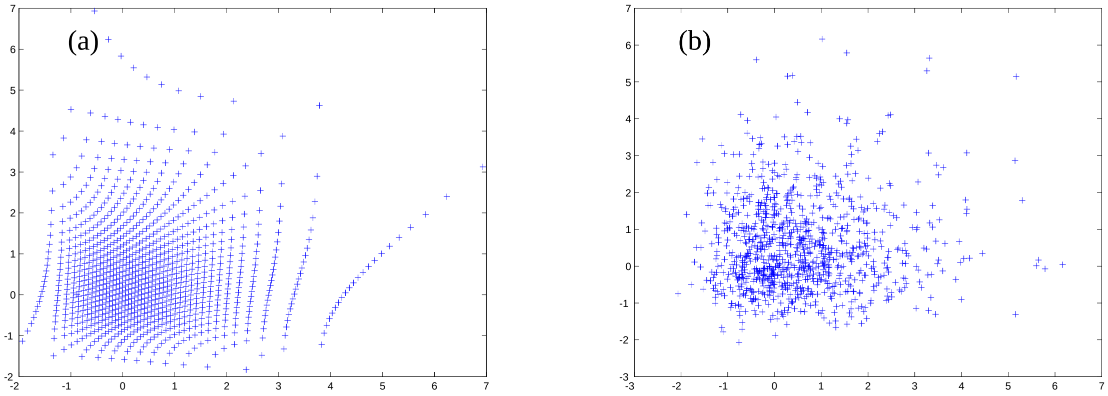

Figure 6: **(a)** Pairs $(\mathbf{x_i}, \mathbf{x_{i+1}})$ of Gumbel data generated by the inversion method from the pseudo-uniform data described in Figure 3. **(b)** Pairs $(\mathbf{x_i}, \mathbf{x_{i+1}})$ of Gumbel data generated by the algorithm under study from the same data.

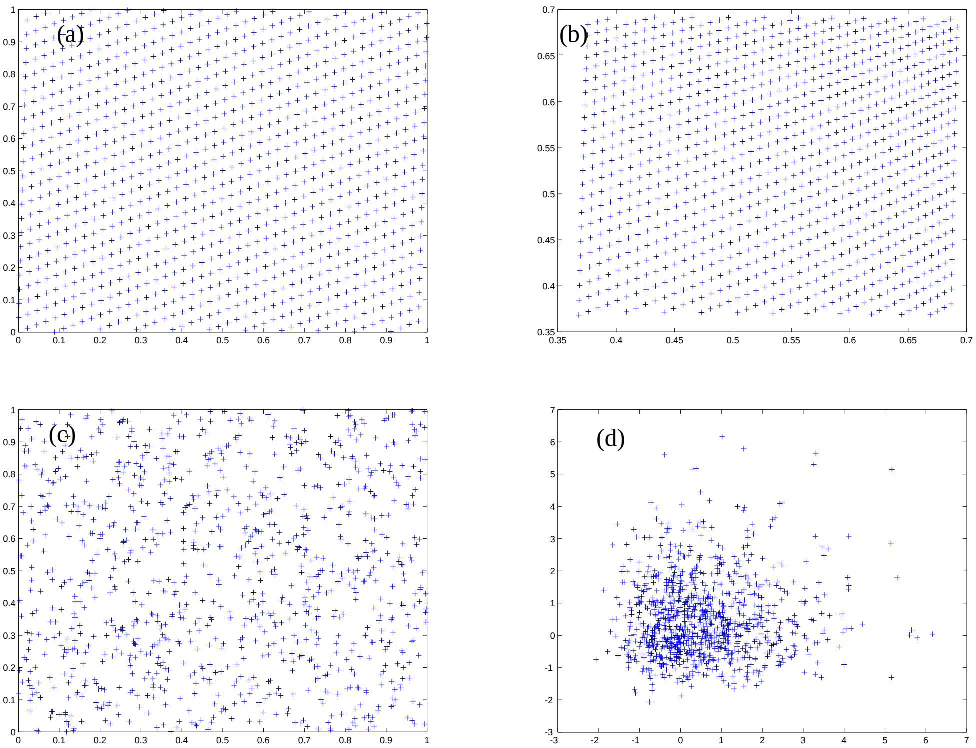

Figure 7: Pairs of points obtained at the different stages of the random variate generator $\mathbf{S} = \mathbf{F}_*^{-1} \circ \mathbf{S_r} \circ \mathbf{F}_*$.

1. Our computer can store and manipulate real numbers,

2. A perfect uniform $]0, 1[$ random generator is available,

then, provided the (generalized) inverse distribution function F^{-1} is explicitly known, the inversion method (Lemma 3.1) is exact. In other cases, the equation $F(X) = U$ has to be solved numerically, and this necessarily leads to an inexact algorithm (see Devroye [7] for discussion). Devroye [7] writes: *inexact algorithms, which are usually algorithms that are based upon a mathematical approximation of sorts, are forever excluded, because neither extended precision arithmetic nor improvements in the basic random number generator make them more exact.* Even if we assume that an explicit form for F_*^{-1} is available, the algorithm described in this paper is definitely *not* exact because of the (arbitrary) choice of the constant r that controls the accuracy of the approximation $||Q_r^n f_0 - f_*||_1$. Further, the user should be aware that the very high sensitivity of the r-adic transformation may cause some numerical problems with softwares that do not allow to work with virtually unlimited length numbers. However, the approach developed in this paper seems interesting because of two points:

- Firstly, it generalizes the classical congruential algorithm to a more general class of λ-*a.e.* positive probability densities defined on an interval (a, b) of the real line.

- Secondly, it is insensitive to both the distribution and independence of the initial sample. No doubt that there is here room for future research.

Acknowledgments. We would like to thank Professor Luc Devroye for valuable comments on a first version of this paper.

References

[1] Berliner, L.M. (1992) Statistics, probability and chaos (with discussion), *Statistical Science*, **Vol. 7**, pp. 69–122.

[2] Berlinet, A. and Devroye, L. (1994) A comparison of kernel density estimates, *Publications de l'Institut de Statistique de l'Université de Paris*, **Vol. 38**, pp. 3–59.

[3] Billingsley, P. (1995) *Probability and Measure*, 3rd edition, John Wiley and Sons, New York.

[4] Bosq, D. and Guégan, D. (1995) Nonparametric estimation of the chaotic function and the invariant measure of a dynamical system, *Statistics and Probability Letters*, **Vol. 25**, pp. 201–212.

[5] Chatterjee, S. and Yilmaz, M.R. (1992) Chaos, fractals and statistics (with discussion), *Statistical Science*, **Vol. 7**, pp. 49–122.

[6] Collet, P. (1996) Some ergodic properties of maps of the interval, in *Dynamical Systems (Temuco, 1991/1992)*, Travaux en Cours, **Vol. 52**, pp. 55–91, Hermann, Paris.

[7] Devroye, L. (1986) *Non-Uniform Random Variate Generation*, Springer–Verlag, New York.

[8] Devroye, L. (1987) *A Course in Density Estimation*, Birkhäuser, Boston.

[9] Eichenauer-Herrmann, J. (1992) Inversive congruential pseudorandom numbers: a tutorial, *International Statistical Review*, **Vol. 60**, pp. 167–176.

[10] Eichenauer-Herrmann, J. (1995) Pseudorandom number generation by nonlinear methods, *International Statistical Review*, **Vol. 63**, pp. 247–255.

[11] Góra, P., Byers, W. and Boyarsky, A. (1991) Measures on periodic orbits for continuous transformations on the interval, *Stochastic Analysis and Applications*, **Vol. 9**, pp. 263–270.

[12] Guerre, E. and Maës, J. (1999) Optimal rate for nonparametric estimation in deterministic dynamical systems, *Statistical Inference for Stochastic Processes*, **Vol. 1**, pp. 157–173.

[13] Hlawka, E. (1984) *The Theory of the Uniform Distribution*, Academic, New York.

[14] Knuth, D.E. (1981) *The Art of Computer Programming,* **Vol. 2**, *Seminumerical Algorithms*, 2nd edition, Addison–Wesley, Reading.

[15] Kuipers, L., Niederreiter, H. and Shiue, J.S. (1975) Uniform distribution of sequences in the ring of Gaussian integers, *Bulletin of the Institute of Mathematics of Academia Sinica*, **Vol. 3**, pp. 311–326.

[16] Lasota, A. and Mackey, M.C. (1994) *Chaos, Fractals, and Noise – Stochastic Aspects of Dynamics*, Springer–Verlag, New York.

[17] L'Ecuyer, P. (1994) Uniform random number generation, *Annals of Operations Research*, **Vol. 53**, pp. 77–120.

[18] L'Ecuyer, P. and Hellekalek, P. (1998) Random number generators: selection criteria and testing, in *Random and Quasi-Random Point Sets*, Lecture Notes in Statistics, **Vol. 138**, pp. 223–266, Springer–Verlag, New York.

[19] Li, T.Y. and Yorke, J.A. (1978) Ergodic maps on $[0,1]$ and nonlinear pseudo-random number generators, *Nonlinear Analysis Theory, Methods and Applications*, **Vol. 2**, pp. 473–481.

[20] Niederreiter, H. (1978) Statistical tests for linear congruential pseudo-random numbers, in *COMPSTAT, Third Symposium on Computational Statistics, Leiden*, pp. 398–404, Physica–Verlag, Wien.

[21] Niederreiter, H. (1992) *Random Number Generation and Quasi-Monte Carlo Methods*, Journal of the Society for Industrial and Applied Mathematics, Philadelphia.

[22] Niederreiter, H. and Shiue, J.S. (1995) Monte Carlo and quasi-Monte Carlo methods in scientific computing, in *Proceedings of a Conference at the University of Nevada, Las Vegas*, Lecture Notes in Statistics, **Vol. 106**, Springer–Verlag, New York.

[23] Ripley, B.D. (1983) Computer generation of random variables: a tutorial, *International Statistical Review*, **Vol. 51**, pp. 301–319.

[24] Tong, H. (1995) A personal overview of non-linear time series analysis from a chaos perspective (with discussion), *Scandinavian Journal of Statistics*, **Vol. 22**, pp. 399–445.