On fractal distribution function estimation and applications

Stefano M. Iacus* Davide La Torre
Università degli Studi di Milano
Dipartimento di Economia Politica e Aziendale
Via Conservatorio 7 - I-20122 Milano

February 17, 2002

Abstract

In this paper we review some recent results concerning the approximations of distribution functions and measures on $[0,1]$ based on iterated function systems. The two different approaches available in the literature are considered and their relation are investigated in the statistical perspective. In the second part of the paper we propose a new class of estimators for the distribution function and the related characteristic and density functions estimators. Via Monte Carlo analysis we show that, for small sample sizes, the proposed estimator can be as efficient or even better than the empirical distribution function and the kernel density estimator respectively. This paper is to be considered as a first attempt in the construction of new class of estimators based on fractal objects.

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables, each having a common continuous cumulative distribution function F of a real random variable X with values on $[0,1]$, i.e. $F(x) = P(X \leq x)$, such that $F(x) = 0$, $x \leq 0$ and $F(x) = 1$ for $x \geq 1$. The empirical distribution function (e.d.f.)

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq x)$$

*Address for correspondance: Stefano M. Iacus, Dipartimento di Economia e Politica Aziendale, Via Conservatorio 7, I-20122 Milano, Italy. E-mail: *stefano.iacus@unimi.it*

is one commonly used estimator of the unknown distribution function F. Here $1(\cdot)$ is the indicator function. The e.d.f. has an impressive set of good statistical properties such as it is first order efficient in the minimax sense (see Dvoretsky *et al.* 1956, Kiefer and Wolfowitz 1959, Beran 1977, Levit 1978 and Millar 1979, Gill and Levit 1995). More or less recently, other second order efficient estimators have been proposed in the literature for special classes of distribution functions F. Golubev and Levit (1996a,b) and Efromovich(2001) are two of such examples.

It is rather curious that a step-wise function can be such a good estimator and, in fact, Efromovich (2001) shows that, for the class of analytic functions, for smal sample sizes, the e.d.f. is not the best estimator. Here we study the properties of a new class of continuous distribution function estimators based on iterated function systems (IFSs). IFSs have been introduced in Hutchinson (1981) and Barnsley and Demko (1985). These are particular fractal objects, hence the title of this note. The fractal nature of IFSs based estimators implies that they are nowhere differentiable and cannot be used directly in density estimation as in Efromovich (2001) but, to this end, we will show a Fourier analysis approach to bypass the problem.

The paper is organized as follows. Section 2 is dedicated to the theoretical background of two constructive methods of approximating measures and distribution functions respectively, with support on compact sets. The first method essentially consists in the minimization approach based on moment matching. This is rather common in the IFS literature. The second approach attacks directly the problem of approximating a distribution function with an IFS, by imposing conditions on the graph of the IFS. In practice, it is imposed to the IFS to pass through a fixed grid of points. IFS for measures are usually used not in a statistical context but mainly for image compression, here the main goal will be the problem of reconstructing a distribution function from sampled data. Even if we do not treat the problem here, the results are likely to hold for measures in any finite dimension $[0,1]^k$, $k \geq 1$. In particular, the case $k = 2$ is interesting for image reconstruction.

Section 3 recalls some results on the Fourier transform for affine IFS. These results are rather important in Section 4 where we propose two kinds of IFS estimators and a density estimator obtained as a Fourier series estimator. Last section is dedicated to the Monte Carlo analysis as it is rather difficult to study statistical properties of IFS-based estimators.

2 Theoretical background for affine IFSs

In this section we recall some of the results from Forte and Vrscay (1995) and Iacus and La Torre (2001) concerning the IFSs setup on the the space of distribution function. Let $\mathcal{M}(X)$ the set of probability measures on $\mathcal{B}(X)$, the σ-algebra of Borel subsets of X where (X, d) is a compact metric space (in our case will be $X = [0, 1]$ and d the Euclidean metric.)

In the IFS literature the following *Hutchinson* metric plays a crucial role

$$d_H(\mu,\nu) = \sup_{f\in \mathrm{Lip}(X)} \left\{ \int_X f \mathrm{d}\mu - \int_X f \mathrm{d}\nu \right\}, \quad \mu,\nu \in \mathcal{M}(X)\,,$$

where

$$\mathrm{Lip}(X) = \{f : X \to \mathbb{R}, |f(x) - f(y)| \le d(x,y), x, y \in X\}$$

thus $(\mathcal{M}(X), d_H)$ is a complete metric space (see Hutchinson, 1981).

As usual, we denote by $(\mathbf{w},\mathbf{p})$ an *N-map contractive IFS on X with probabilities* or simply an *N-maps IFS*, that is, a set of N affine contractions maps, $\mathbf{w} = (w_1, w_2, \dots, w_N)$,

$$w_i = s_i\, x + a_i, \quad \text{with } |s_i| < 1, \quad s_i, a_i \in \mathbb{R}, \quad i = 1, 2, \dots, N,$$

with associated probabilities $\mathbf{p} = (p_1, p_2, \dots, p_N)$, $p_i \ge 0$, and $\sum_{i=1}^{N} p_i = 1$. The IFS has a contractivity factor defined as

$$c = \max_{1\le i\le N} s_i < 1$$

Consider the following (usually called the *Markov*) operator $M : \mathcal{M}(X) \to \mathcal{M}(X)$ defined as

$$M\mu = \sum_{i=1}^{N} p_i \mu \circ w_i^{-1}, \quad \mu \in \mathcal{M}(X), \tag{1}$$

where w_i^{-1} is the inverse function of w_i and $\circ$ stands for the composition. In Hutchinson (1981) it was shown that M is a contraction mapping on $(\mathcal{M}(X), d_H)$: for all $\mu,\nu \in (X)$, $d_H(M\mu, M\nu) \le c d_H(\mu,\nu)$. Thus, there exists a unique measure $\bar\mu \in \mathcal{M}(X)$, the *invariant measure* of the IFS, such that $M\bar\mu = \bar\mu$ by Banach theorem.

2.1 Minimization approach

For affine IFS there exists a simple and useful relation between the moments of probabilty measures on $\mathcal{M}(X)$. Given an N-maps IFS$(\mathbf{w},\mathbf{p})$ with associated Markov operator M, and given a measure $\mu \in \mathcal{M}(X)$ then, for any continuous function $f : X \to \mathbb{R}$,

$$\int_X f(x)\mathrm{d}\nu(x) = \int_X f(x)\mathrm{d}(M\mu)(x) = \sum_{i=1}^{N} p_i \int_X (f\circ w_i)(x)\mathrm{d}\mu(x)\,, \tag{2}$$

where $\nu = M\mu$. In our case $X = [0,1] \subset \mathbb{R}$ so we readly have a relation involving the moments of μ and ν. Let

$$g_k = \int_X x^k \mathrm{d}\mu, \quad h_k = \int_X x^k \mathrm{d}\nu, \quad k = 0, 1, 2, \dots, \tag{3}$$

be the moments of the two measures μ and ν, with $g_0 = h_0 = 1$. Then, by (2), with $f(x) = x^k$, we have

$$h_k = \sum_{j=0}^{k} \binom{k}{j} \left\{ \sum_{i=1}^{N} p_i s_i^j a_i^{k-j} \right\} g_j, \quad k = 1, 2, \dots, .$$

Recursive relations for the moments and more details on polynomial IFSs can be found in Forte and Vrscay (1995). The following theorem is due to Vrscay and can be found in Forte and Vrscay (1995) as well.

Theorem 1. *Set $X = [0,1]$ and let μ and $\mu^{(j)} \in \mathcal{M}(X)$, $j = 1, 2, \dots$ with associated moments of any order g_k and*

$$g_k^{(j)} = \int_X x^k \mathrm{d}\mu^{(j)} .$$

Then, the following statements are equivalent (as $j \to \infty$ and $\forall k \geq 0$):

i) $g_k^{(j)} \to g_k$,

ii) $\forall f \in \mathbf{C}(X)$, $\int_X f \mathrm{d}\mu^{(j)} \to \int_X f \mathrm{d}\mu$, *(weak* convergence),*

iii) $d_H(\mu^{(j)}, \mu) \to 0$.

(here $\mathbf{C}(X)$ is the space of continuous functions on X). This theorem gives a way to find and appropriate set of maps and probabilities by solving the so called problem of moments matching. With the solution in hands, given the convergence of the moments, we also have the convergence of the measures and then the stationary measure of M approximates with given precision (in a sense specified by the collage theorem below) the target measure μ (see Barnsley and Demko, 1985).

Next theorem, called the *collage* theorem is a standard result of IFS theory and is a consequence of the Banach theorem.

Theorem 2 (Collage theorem). *Let (Y, d_Y) be a complete metric space. Given an $y \in Y$, suppose that there exists a contractive map f on Y with contractivity factor $0 \leq c < 1$ such that $d_Y(y, f(y)) < \varepsilon$. If $\bar{y}$ is the fixed point of f, i.e. $f(\bar{y}) = \bar{y}$, then $d_Y(\bar{y}, y) < \varepsilon/(1-c)$.*

So if one wishes to approximate a function y with the fixed point $\bar{y}$ of an unknown contractive map f, it is only needed to solve the inverse problem of finding f which minimizes the collage distance $d_Y(y, f(y))$.

The main result in Forte and Vrscay (1995) that we will use to build on of the IFSs estimators is that the inverse problem can be reduce to minimize a suitable quadratic form in terms of the p_i given a set of affine maps w_i and the sequence of moments g_k of the target measure. Let

$$\Pi^N = \left\{ \mathbf{p} = (p_1, p_2, \ldots, p_N) : p_i \geq 0, \sum_{i=1}^{N} p_i = 1 \right\}$$

be the simplex of probabilities. Let $\mathbf{w} = (w_1, w_2, \ldots, w_N)$, $N = 1, 2, \ldots$ be subsets of $\mathcal{W} = \{w_1, w_2, \ldots\}$ the infinite set of affine contractive maps on $X = [0, 1]$ and let $\mathbf{g}$ be the vector of the moments of any order of $\mu \in \mathcal{M}(X)$. Denote by M the Markov operator of the N-maps IFS $(\mathbf{w}, \mathbf{p})$ and by $\nu_N = M\mu$ with associated moment vector of any order $\mathbf{h}_N$. The collage distance between the moment vector of μ and ν_N

$$\Delta(\mathbf{p}) = ||\mathbf{g} - \mathbf{h}_N||_{\bar{l}^2} : \Pi^N \to \mathbb{R}$$

is a continuous function and attains an absolute minimum $\Delta_{\min}$ on Π^N where

$$||\mathbf{x}||_{\bar{l}^2} = x_0^2 + \sum_{k=1}^{\infty} \frac{x_k^2}{k^2}\,.$$

Theorem 3 (Forte and Vrscay, 1995). $\Delta_{\min} \to 0$ *as* $N \to \infty$.

Thus, the collage distance can be made arbitrarily small by choosing a suitable number of maps and probabilities, N^*.

By the same authors, the inverse problem can be posed as a quadratic programming one in the following notation

$$S(\mathbf{p}) = (\Delta(\mathbf{p}))^2 = \sum_{k=1}^{\infty} \frac{(h_k - g_k)^2}{k^2}$$

$$D(X) = \{\mathbf{g} = (g_0, g_1, \ldots) : g_k = \int_X x^k \mathrm{d}\mu, k = 0, 1, \ldots, \mu \in \mathcal{M}(X)\}$$

Then by (2) there exists a linear operator $A : D(X) \to D(X)$ associated to M such that $\mathbf{h}_N = A\mathbf{g}$. In particular

$$h_k = \sum_{i=1}^{N} A_{ki} p_i, \quad k = 1, 2, \ldots,$$

where

$$A_{ki} = \sum_{j=0}^{n} s_i^j a_i^{k-j} g_j$$

Thus

$$(\mathbf{Q}) \qquad S(\mathbf{x}) = \mathbf{x}^t Q \mathbf{x} + \mathbf{b}^t \mathbf{x} + c, \quad \mathbf{x} \in \mathbb{R}^N,$$

where

$$q_{ij} = \sum_{k=1}^{\infty} \frac{A_{ki} A_{kj}}{k^2}, \quad i, j = 1, 2, \ldots, N,$$

$$b_i = -2\sum_{k=1}^{\infty} \frac{g_k}{k^2} A_{ki}, \quad i = 1, 2, \ldots, N, \quad \text{and} \quad c = \sum_{k=1}^{\infty} \frac{g_k^2}{k^2} \,.$$

The series above are convergent as $0 \leq A_{ni} \leq 1$ and the minimum can be found by minimizing the quadratic form on the simplex Π^N. This is the main result in Forte and Vrscay (1995) that can be used straight forwardly in statistical applications as we propose in Section 4.

On the other side Iacus and La Torre (2001) propose a different and direct approach to construction on IFSs on the space of distribution function on $[0, 1]$. Instead of constructing the IFS by matching the moments, the idea there is to have an IFS that has the same values of the target distribution function on a finite number of points.

2.2 Direct approach

We use directly the fractal nature of the IFS. Given a distribution function on $[0, 1]$, the idea is to rescale the whole function in abscissa and ordinate and copying it a number of times obtaining a function that is again a distribution function. Consider $\mathcal{F}([0, 1])$, the space of distribution functions on $[0, 1]$, then $(\mathcal{F}([0, 1]), d_\infty$ is a complete metric space, where $d_\infty(F, G) = \sup_{x\in[0,1]} |F(x) - G(x)|$. On $(\mathcal{F}([0, 1], d_\infty)$ we define the following operator (see Iacus and La Torre, 2001):

$$TF(x) = p_i F(w_i^{-1}(x)) + \sum_{j=1}^{i-1} p_j + \sum_{j=1}^{i-1} \delta_j \,, \quad x \in w_i([a_i, b_i)) \,, \quad i = 1, \ldots, N, \tag{4}$$

where $F \in \mathcal{F}([0, 1])$, $N \in \mathbb{N}$ is fixed and:

i) $w_i : [a_i, b_i) \to [c_i, d_i) = w_i([a_i, b_i))$, $i = 1, \ldots, N-1$, $w_N : [a_N, b_N] \to [c_N, d_N]$, with $a_1 = c_1 = 0$ and $b_N = d_N = 1$, $\bigcup_{i=1}^{N-1} [a_i, b_i) \cup [a_N, b_N] = \bigcup_{i=1}^{k-1} [c_i, d_i) \cup [c_N, d_N] = [0, 1]$;

ii) $w_i = s_i\, x + a_i$, $0 < s_i < 1$, $s_i, a_i \in \mathbb{R}$, $i = 1, \ldots, N$;

iii) $\bigcup_{i=1}^{N} w_i([a_i, b_i)) = [0, 1)$;

iv) $p_i \geq 0$, $i = 1, \ldots, N$, $\delta_i \geq 0$, $i = 1 \ldots N-1$, $\sum_{i=1}^{N} p_i + \sum_{i=1}^{N-1} \delta_i = 1$;

v) if $i \neq j$ then $w_i([a_i, b_i)) \cap w_j([a_j, b_j)) = \emptyset$;

In equation (4), when $i = 1$ the summation are to be intended to be equal to 0. We limit the treatise to affine maps w_i as in Forte and Vrscay (1995), but the general case of increasing and continuous maps can be treated as well (see cited reference of the authors). From now on, we consider the sets of maps w_i and parameters δ_i as given, thus the operator depends only on the probabilities p_i and we denote it by T_p.

Theorem 4 (Iacus and La Torre, 2001). *Under conditions i) to v):*

1. *T_p is an operator from $\mathcal{F}([0,1])$ to itself.*
2. *Suppose that $w_i(x) = x$, $p_i = p$, and $\delta_i \geq -p$, then $T_p : \mathcal{F}([0,1]) \to \mathcal{F}([0,1])$.*
3. *If $c = \max_{i=1,\dots,N} p_i < 1$, then T_p is a contraction on $(\mathcal{F}([0,1]), d_\infty)$ with contractivity constant c.*
4. *Let p, $p^* \in \mathbb{R}^k$ such that $T_p F_1 = F_1$ and $T_{p^*} F_2 = F_2$. Then*
$$d_\infty(F_1, F_2) \leq \frac{1}{1-c} \sum_{j=1}^{N} |p_j - p_j^*|$$
where c is the contractivity constant of T_p.

The theorem above assures the IFS nature of the operator T_p that can be denoted, as in the previous section, as a N-maps IFS($\mathbf{w}, \mathbf{p}$) with obvious notation.

The goal is again the solution of the inverse problem in terms of $\mathbf{p}$. Consider the following convex set:
$$C^N = \left\{ p \in \mathbb{R}^N : p_i \geq 0, i = 1, \dots, N, \sum_{i=1}^{N} p_i = 1 - \sum_{i=1}^{N-1} \delta_i \right\},$$
then we have the following results:

Theorem 5 (Iacus and La Torre, 2001). *Choose $\epsilon > 0$ and $p \in C^N$ such that $p_i \cdot p_j > 0$ for some $i \neq j$. If $d_\infty(T_p F, F) \leq \epsilon$, then:*
$$d_\infty(F, \tilde{F}) \leq \frac{\epsilon}{1-c},$$
where $\tilde{F}$ is the fixed point of T_p on $\mathcal{F}([0,1])$ and c is the contractivity constant of T_p. Moreover, the function $D(p) = d_\infty(T_p F, F)$, $p \in \mathbb{R}^N$ is convex.

Thus, the following constrained optimization problem:

$$(\mathbf{P}) \qquad \min_{p \in C^N} d_\infty(T_p F, F)$$

can be solved at least numerically.

Another way of choosing the form of T_p is the direct approach, that is the following. Choose $n = N+1$ points on $[0,1]$, $(x_1, \dots, x_n)$, and assume that $0 = x_1 < x_2 < \cdots < x_{n-1} < x_n = 1$. The proposed functional is the following
$$T_F u(x) = (F(x_{i+1}) - F(x_i)) u\left(\frac{x - x_i}{x_{i+1} - x_i}\right) + F(x_i), \quad x \in [x_i, x_{i+1}),$$

$i = 1, \ldots, n-1$, where u is any member in the space $\mathcal{F}([0,1])$. Notice that T_F is a particular case of T_p where $p_i = F(x_{i+1}) - F(x_i)$, $\delta_i = 0$ and $w_i(x) : [0,1) \to [x_i, x_{i+1}) = (x_{i+1} - x_i)x + x_i$. This is a contraction and, at each iteration, T_F passes exactly through the points $F(x_i)$. It is almost evident that, when n increases the fixed point of the above functional will be "close" to F.

For n small, the choice of a good grid of points is critical. So one question arises: how to choose the n points ? One can proceed case by case but as F is a distribution function so one can use its properties. We propose the following solution: take n points ($u_1 = 0, u_2, \ldots, u_n = 1$) equally spaced $[0,1]$ and define $q_i = F^{-1}(u_i)$, $i = 1, \ldots, n$. The points q_i are just the quantiles of F. In this way, it is assured that the profile of F is followed as smooth as possible. In fact, if two quantiles q_i and q_{i+1} are relatively distant each other, than F is slowly increasing in the interval (q_i, q_{i+1}) and viceversa. This method is more efficient than simply taking equally spaced points on $[0,1]$. With this assumption the functional T_F reads as

$$T_N u(x) = T_F u(x) = \frac{1}{N} u\left(\frac{x - q_i}{q_{i+1} - q_i}\right) + \frac{i-1}{N}, \quad x \in [q_i, q_{i+1}),\ i = 1, \ldots, N\,.$$

This form of the estimator proposes an intuitive (possibily) good candidate for distribution function estimation. Note that we overcome the problem of moment matching as we don't even need the existence of the moments.

Corollary 6. *As a corollary of the collage Theorem 4 we can anwser to the question: how many quantiles are needed to approximate a distribution function with a given precision, say ϵ ? The answer is: take the first integer N such that $N > 1/\epsilon$. This value of N is in fact the one that guarantees that the sup-norm distance between the true F and the fixed point $\tilde{F}$ of T_F is less than ϵ. In general, this distance could be considerably smaller as shown in Table 1.*

Proof. The proof is as follows: assume that we set $\epsilon < 1$ (the contrary will be absurd), then it holds true that

$$d_\infty(F, \tilde{F}) \le d_\infty(T_p F, F) \le \epsilon$$

as, by Theorem 4, $d_\infty(u, \tilde{F}) < \epsilon/(1-c)$ where $c = \max_i p_i = 1/N$. And this is in particular true for F. To estimate $d_\infty(T_p F, F)$ we can slipt the interval [0,1] as $[0 = q_0, q_1) \cup [q_1, q_2) \cup \cdots \cup [q_N, q_{N+1} = 1]$. In each of the intervals $[q_i, q_{i+1})$ the distance between $T_p F(x)$ and $F(x)$ is at most $1/N$, thus $\epsilon < 1/N$. It is worth to note that this is really a bad case because, all of the distributions we have tested has an estimated sup-norm less than $1/N$. □

2.3 Links between the two approaches

To investigate the asymptotic behaviour of T_p it is worth to show the relation between this functional on the space of distribution function and the one proposed by Forte and Vrscay (1995) on the space of measures.

Assuming that T_p as well as T_F is such that, for any $G \in \mathcal{F}([0,1])$,

$$T_pG(x) = 0, \quad \forall\, x \leq 0 \quad \text{and} \quad T_pG(x) = 1, \quad \forall\, x \geq 1\,,$$

Then T_p can be rewritten as

$$T_pF(x) = \begin{cases} 0, & x \leq 0, \\ \sum_{i=1}^{N} p_i F(w_i^{-1}(x)), & x \in (0,1), \\ 1, & x \geq 1\,. \end{cases}$$

Theorem 7 (Bridge theorem). *Given a set of N maps and probabilities $(\mathbf{w}, \mathbf{p})$, then the fixed points of $M \in \mathcal{M}([0,1])$ and $T_p \in \mathcal{F}([0,1])$ relates as follows*

$$\bar{\mu}((0,x]) = M\bar{\mu}((0,x]) = T_p\tilde{F}(x) = \tilde{F}(x), \quad \forall x \in [0,1]\,.$$

Proof. For each fixed $\mu \in \mathcal{M}([0,1])$ there corresponds a distribution function $F \in \mathcal{F}([0,1])$ defined by the relation $\mu((0,x]) = F(x)$, $\forall x \in [0,1]$. Thus, fixed a set of N-maps IFS $(\mathbf{w}, \mathbf{p})$

$$M\mu((0,x]) = T_pF(x)\,.$$

The second iteration is then

$$M(\,M\mu((0,x])\,) = T_p(\,T_pF(x)\,)$$

so, the fixed points in the relative spaces are

$$\bar{\mu}((0,x]) = M(\,M(\,\ldots M(\,\bar{\mu}((0,x])\,)\ldots)\,)$$

and

$$T_p\tilde{F}(x) = T_p(\,T_p(\,\ldots T_p(\,\tilde{F}(x)\,)\ldots)\,)$$

from which the statement of the theorem. $\square$

The previous theorem allows to reuse the results of Forte and Vrscay (1995) and in particular gives another way of finding the solution of (**P**) in terms of (**Q**) at least on the simplex Π^N by letting $\delta_i = 0$ in C^N. This is true in particular if we choose the maps as in T_F.

2.4 The choice of the affine maps

As we are mostly concerned with estimation, we briefly discuss the problem of choosing the maps. In Forte and Vrscay (1995) the following two sets of wavelet-type maps are proposed. Fixed and index $i^* \in \mathbb{N}$, define

$$WL_1 : \omega_{ij} = \frac{x+j-1}{2^i}, \quad i = 1, 2, \ldots, i^* \quad j = 1, 2, \ldots, 2^i$$

and

$$WL_2 : \omega_{ij} = \frac{x+j-1}{i}, \quad i = 2, \dots, i^* \quad j = 2, \dots, i\,.$$

Then set $N = \sum_{i=1}^{i^*} 2^i$ or $N = i^*(i^*-1)/2$ respectively. To choose the maps, consider the natural ordering of the maps ω_{ij} and operate as follows

$$\mathcal{W}_1 = \{w_1 = \omega_{11}, w_2 = \omega_{12}, w_3 = \omega_{21}, \dots, w_6 = \omega_{24}, w_7 = \omega_{31}, \dots, w_N = \omega_{i^* 2^{i^*}}\}$$

and

$$\mathcal{W}_2 = \{w_1 = \omega_{22}, w_2 = \omega_{32}, w_3 = \omega_{33}, w_4 = \omega_{42}, \dots, w_6 = \omega_{44}, \dots, w_N = \omega_{i^* i^*}\}$$

respectively. Our quantile based maps are of the following type $\mathcal{W}_q = \{w_i(x) = (q_{i+1} - q_i)x + q_i, i = 1, 2, \dots, N\}$ where $q_i = F^{-1}(u_i)$, and $0 = u_1 < u_2 < \dots < u_N < u_{N+1} = 1$ are $N+1$ equally spaced points on $[0,1]$.

For each given sets of maps $\mathbf{w}$ above different $\mathbf{p}$'s will be solution of ($\mathbf{Q}$) (or ($\mathbf{P}$)). Wether the corresponding fixed point is closer a given F in the three cases is not always clear. As an example, in Table 1 we show the relative performance of the approximation based on the quantity

$$\Delta_m(\mathbf{p}) = \sum_{k=1}^{m} \frac{1}{k^2} \left(\sum_{i=1}^{N} A_{ki} p_i - g_k \right)^2$$

(that is an approximation of the collage distance) and on the sup-norm d_∞ and the average mean square error, AMSE. We also report the contractivity constant in both the space $\mathcal{M}([0,1])$ and the space $\mathcal{F}([0,1])$. Recall that the collage theorem for the moments establishes that, if $\mathbf{g}$ is the vector of moments of a the target measure μ (of a distribution function F) and $\bar{\mathbf{g}}$ is the moment vector of the invariant measure $\bar{\mu}$ of the IFS $(\mathbf{w}, \mathbf{p})$ then

$$||\mathbf{g} - \bar{\mathbf{g}}||_{\bar{l}^2} < \frac{\Delta}{1-c}\,.$$

Table 1 shows that, at least in this classical example of the IFS literature, for a fixed number of maps N, T_N is a better approximator than M relatively to the sup-norm and the AMSE while the contrary is true in terms of the approximate collage distance $\Delta_m(\mathbf{p})$. As noted in Forte and Vrscay (1995), M uses not all the maps, in the sense that N', the number of non null probabilities, is usually smaller than N. It is evident that, two alternatives seem promising in the perspective of distribution function estimation: M with $\mathcal{W}_1$ and T_N (i.e. M with $\mathcal{W}_q$ and $p_i = 1/N$). Note that it is apparently simpler to use T_N because there is no need to calculate moments.

3 Fourier analysis results

The results presented in this section, taken from Forte and Vrscay (1998) Sec. 6, are rather straight forward to prove but it is essential to recall them since we will use these in density estimation later on.

IFS	$\mathbf{w}$	N	N'	$\Delta_m(\mathbf{p})$	d_∞	AMSE	$\max \mathbf{p}$	$c = \max \mathbf{s}$
M	$\mathcal{W}_1$	6	5	7.79e-08	0.06253	5.31e-4	0.255	0.500
M	$\mathcal{W}_2$	6	3	3.40e-05	0.25024	9.62e-3	0.483	0.500
M	$\mathcal{W}_q$	6	6	8.32e-08	0.12718	2.51e-3	0.165	0.259
T_N	$\mathcal{W}_q$	6	6	9.90e-05	0.04550	4.54e-4	$0.16\bar{6}$	0.259
M	$\mathcal{W}_1$	10	10	2.45e-07	0.03948	3.45e-4	0.291	0.500
M	$\mathcal{W}_2$	10	6	1.54e-06	0.17870	5.66e-3	0.678	0.500
M	$\mathcal{W}_q$	10	10	5.00e-08	0.04060	3.55e-4	0.351	0.195
T_N	$\mathcal{W}_q$	10	10	3.34e-05	0.02778	1.46e-4	0.100	0.195
M	$\mathcal{W}_1$	14	11	5.38e-7	0.02983	1.93e-4	0.266	0.500
M	$\mathcal{W}_2$	14	12	9.56e-7	0.09822	2.17e-3	0.218	0.500
M	$\mathcal{W}_q$	14	14	2.66e-8	0.02546	1.43e-4	0.106	0.163
T_N	$\mathcal{W}_q$	14	14	1.57e-5	0.01973	6.93e-5	0.100	0.163

Table 1: Approximation results for the different N-maps IFS $(\mathbf{w}, \mathbf{p})$ for the targe tdistribution function $F(x) = x^2(3-2x)$ as in Forte and Vrscay (1995). N = number of maps used, AMSE = average MSE, max $\mathbf{p}$ is the contractivity constant of T_p in $\mathcal{F}([0,1])$, $\mathbf{s}$ is the contractivity constant of M in $\mathcal{M}([0,1])$. N' the number of non null probabilities. For the rest of the notation see text.

Given a measure $\mu \in \mathcal{M}(X)$, the Fourier transform (FT) $\phi : \mathbb{R} \to \mathbb{C}$, where $\mathbb{C}$ is the complex space, is defined by the relation

$$\phi(t) = \int_X e^{-itx} \mathrm{d}\mu(x), \quad t \in \mathbb{R}\,.$$

with the well known properties $\phi(0) = 1$ and $|\phi(t)| \leq 1$, $\forall\, t \in \mathbb{R}$. We denote by $FT(X)$ the set of all FT's associated to the measures in $\mathcal{M}(X)$. Given two elements ϕ and ψ in $FT(X)$, the following metric can be defined

$$d_{FT}(\phi, \psi) = \left(\int_{\mathbb{R}} |\phi(t) - \psi(t)|^2 t^{-2} \mathrm{d}t \right)^{\frac{1}{2}}$$

and the above integral is always finite (see cited paper). With this metric $(FT(X), d_{FT})$ is a complete metric space. Given an N-maps (affine) IFS$(\mathbf{w}, \mathbf{p})$ and its Markov operator M it is possibile to define a new linear operator $B : FT(X) \to FT(X)$ as follows

$$\psi(t) = \sum_{k=1}^{N} p_k e^{-ita_k} \phi(s_k t), \quad t \in \mathbb{R}\,,$$

where ϕ is the FT of a measure μ and ψ is the FT of $\nu = M\mu$.

Theorem 8 (Forte and Vrscay, 1998). *The operator B is contractive in $(FT(X), d_{FT})$ and has a unique fixed point. In particular, if $\bar{\phi}$ is the FT of the invariant measure of the*

Markov operator M, then the fixed point is

$$\bar{\phi}(t) = \sum_{k=1}^{N} p_k e^{-ita_k} \bar{\phi}(s_k t), \quad t \in \mathbb{R}\,.$$

The following final results holds true.

Theorem 9 (Collage Theorem for FT, (Forte and Vrscay, 1998)). *Let (X, d) be a compact metric space and $\mu \in \mathcal{M}(X)$ with FT ϕ, $\phi \in FT(X)$. Let $(\mathbf{w}, \mathbf{p})$ be an N-maps IFS with contractivity factor $c \in [0, 1)$ such that $d_{FT}(\phi, \psi) < \epsilon$, where $\psi = B(\phi)$ is the FT of $\nu = M\mu$. Then*

$$d_{FT}(\phi, \bar{\phi}) < \frac{\epsilon}{c}\,,$$

where $\bar{\phi}$, $\bar{\phi} \in FT(X)$, is the FT of the invariant measure $\bar{\mu}$ of M, i.e. $\bar{\mu} = M\bar{\mu}$.

4 Statistical applications

It is rather natural to propose two estimators for a distribution function, the Markov operator M with wavelets maps $\mathcal{W}_1$ and the T_N IFS. By Corollary 6 one can easily note that using the sample quantiles, it is not possible in general to achieve a precision $\epsilon = 1/N$ if the sample size n is less than N. But when $n = N$ than, in the most defavorable case $\epsilon = 1/N$, we just have the empirical distribution function for which we have the identity $T_N(x) = \hat{F}_n(x)$ for $x = x_i$, $i = 1, \ldots, N + 1$ and a linear interpolant between $\hat{F}_n(x_i)$ and $\hat{F}_n(x_{i+1})$. Thus apparently, the worst thing one can do with the estimator T_N is to estimate the unknown distribution function with an interpolated version of $\hat{F}_n$. The target of having $\epsilon = 1/100$ means that at least 100 quantiles are needed and, non asymptotically, this is a to severe condition because, even having $n = 100$ observations, the empirical quantiles are not good estimates of the true quantiles. As we have seen in the previous section (see Table 1) for having an error of order $\epsilon = 1/50$ only 14 quantiles are needed: around 1/3 of ϵ. So, as a rule of thumb we suggest to use a number of quantiles between $n/2$ and $n/3$. In our Monte Carlo analysis of Section 5 we convain to use $n/2$. This strategy is computationally heavy when n is large as the time to calculate the estimator increases too much, thus from a certain sample size n it is better to use a fixed amount of quantiles. Our experience shows that $N = 50$ for large sample sizes is big enough, but for very large sample sizes we suggest to use the empirical distribution function instead. Moreover, it has to be reminded that for $N = 50$ one can attend, in the worst case an error in sup-norm of 2%.

The two estimators are of the following types:

a) The Markov-Wavelets IFS

$$\hat{M}_{\mathcal{W}_1} u(x) = \sum_{i=1}^{N} \hat{p}_i u(w_i^{-1}(x))$$

where $w_i \in \mathcal{W}_1$ and the $\hat{p}_i$ are the solutions of the quadratic problem (**Q**) with vector of empirical moments $\hat{\mathbf{g}}$ instead of $\mathbf{g}$. The number of empirical moments ($m = N + 1$) used is linked to the number of wavelet maps $N = \sum_{i=1}^{i^*} 2^i$, for $i^* = 1, 2, 3, 4$.

b) The quantile-based IFS

$$\hat{T}_N u(x) = \sum_{i=1}^{N} \frac{1}{N} u(\hat{w}_i^{-1}(x))$$

where $\hat{w}_i^{-1}(x) = (x - \hat{q}_i)/(\hat{q}_{i+1} - \hat{q}_i)$ with $\hat{q}_i$ the empirical quantiles, being $\hat{q}_1 = 0$ and $\hat{q}_{N+1} = 1$.

In both cases u is any distribution function on $[0, 1]$, for example the Uniform distribution, that is to be considered at the starting point in $\mathcal{F}([0, 1])$ form which the iteration of the functionals start.

Asymptotic properties of the fixed point of both $\hat{M}_{\mathcal{W}_1}$ and $\hat{T}_N$ derive as a natural consequency from the properties of moments estimators and empirical quantiles. So, one can expect that for a fixed number N of maps $\hat{M}$ converges in probability to M as the sample size increases and that $\hat{T}_N$ converges to the fixed point of T_N as $n \to \infty$.

4.1 Characteristic function and Fourier density estimation

Using the results of Section 3 is now feasible to propose a Fourier expansion estimator of the density function of F assuming tha it exists. We assume that all the minimal conditions to proceed in the Fourier analysis of this section are fulfilled. Thus, given and N-map IFS($\mathbf{w}, \mathbf{p}$), we have seen that the IFS estimator is the fixed point of the operator

$$T_p u(x) = \sum_{i=1}^{N} p_i u(w_i^{-1}(x)), \quad x \in [0, 1],$$

for any $u \in \mathcal{F}([0, 1])$ or, equivalently, in the space of measure $\mathcal{M}([0, 1])$

$$M\mu(A) = \sum_{i=1}^{N} p_i \mu(w_i^{-1}(A)), \quad A \subset [0, 1]$$

with maps and coefficients eventually estimated. Now, let $\bar{\phi}$ be the fixed point of the operator B in Section 3, i.e.

$$\bar{\phi}(t) = \sum_{k=1}^{N} p_k e^{-ia_k t} \bar{\phi}(s_k t), \quad t \in \mathbb{R}\,.$$

Then $\bar{\phi}$ is nothing else that the characteristic function of $f(\cdot)$ where $f(\cdot)$ is the density function of the underlying unknown distribution function $F(\cdot)$ that generates the sample data $X_1, X_2, \ldots, X_n$. Now (see e.g. Tarter and Lock, 1993) it is possible to derive a Fourier expansion density estimator in this way.

$$\bar{\phi}(t) = \int_0^1 f(x)e^{-itx}\mathrm{d}x = \int_0^1 e^{-itx}\mathrm{d}F(x)$$

and, given $\bar{\phi}(t)$ the density function $f(\cdot)$ can be rewritten as

$$f(x) = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} B_k e^{ikx} = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{i=1}^{\infty}\Big(\mathrm{Re}(B_k)\,\cos(kx) - \mathrm{Im}(B_k)\,\sin(kx)\Big)$$

where

$$B_k = \int_0^1 f(x)e^{-ikx}\mathrm{d}x = \phi(k)$$

Denoting by $\hat{\phi}$ (the fixed point of) the characteristic function estimator based on quantiles

$$\hat{\phi}(t) = \sum_{k=1}^{N}\frac{1}{N}e^{-i\hat{a}_k t}\psi(\hat{s}_k t) \quad \hat{a}_k = \hat{q}_k, \quad \hat{s}_k = \hat{q}_{k+1} - \hat{q}_k\,, \tag{5}$$

with $\hat{q}_0 = 0$ and $\hat{q}_{N+1} = 1$, a density estimator is the following

$$\hat{f}(x) = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} c_k\hat{B}_k e^{ikx} \tag{6}$$

where $\{c_k, k = 0, \pm1, \pm2, \ldots\}$ is a sequence of suitable multipliers not to be estimated and $\hat{B}_k = \hat{\phi}(k)$. One choice for the multipliers is $c_k = 1$ for $|k| \le m$ and $c_k = 0$ if $|k| > m$. In such a case the estimator reduces to the raw Fourier expansion estimator

$$\hat{f}_{FT}(x) = \frac{1}{2\pi}\sum_{k=-m}^{+m}\hat{B}_k e^{ikx} = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{i=1}^{m}\Big(\mathrm{Re}(\hat{B}_k)\,\cos(kx) - \mathrm{Im}(\hat{B}_k)\,\sin(kx)\Big)\,.$$

A detailed discussion on which family of multipliers is to be choosen can be found in Tarter and Lock (1993) and can be applied to this case as well. As it is well known, the fact that the Fourier expansion is a convergent series it is always possible to differentiate or integrate it in order to obtain an estimator for the first derivative of the density

$$\hat{f}'(x) = \frac{1}{2\pi}\frac{\mathrm{d}}{\mathrm{d}x}\hat{f}(x) = \frac{1}{2\pi}\sum_{k=-m}^{+m}\frac{\mathrm{d}}{\mathrm{d}x}\hat{B}_k e^{ikx} = \sum_{k=-m,\,k\neq0}^{+m}\frac{ik}{2\pi}\hat{B}_k e^{ikx} \tag{7}$$

which is a particular case of (6) with $c_k = ik$, $|k| \le m$, $k \neq 0$ and $c_k = 0$ for $k = 0$ or $|k| > m$. We can also propose another distribution function estimator

$$\begin{aligned}\hat{F}_{FT}(x) &= \frac{1}{2\pi}\left(x + \sum_{k=-m,\,k\neq0}^{+m}\frac{\hat{B}_k}{ik}\left(e^{ikx} - 1\right)\right)\\ &= \frac{1}{2\pi}\left(x + 2\sum_{k=1}^{m}\Big(\mathrm{Re}(\hat{B}_k)\sin(kx) + \mathrm{Im}(\hat{B}_k)(\cos(kx) - 1)\Big)\right)\end{aligned} \tag{8}$$

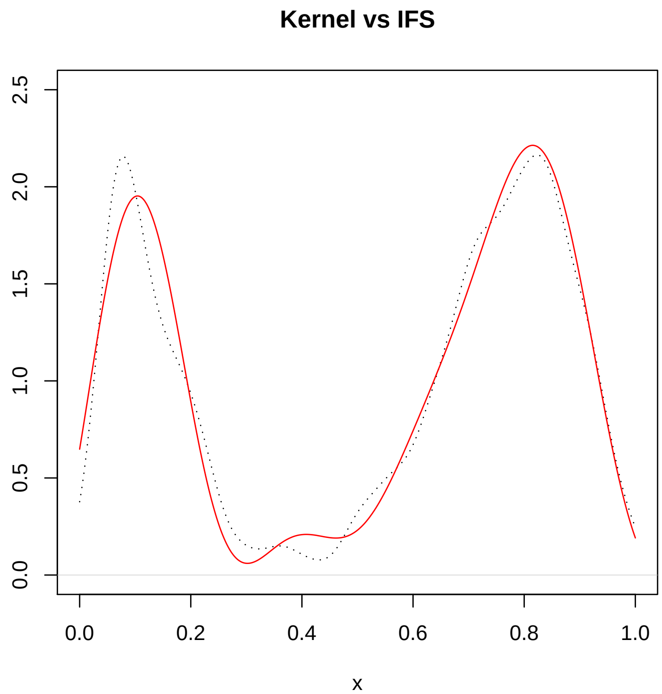

Figure 1: Old Faithful gayser data rescaled on [0,1]. Dotted line is the kernel estimator (bw=0.03, kernel=Gaussian), solid line is the IFS-Fourier expansion estimator (iterated 2 times, 26 Fourier coefficients).

that can be used as a smooth estimator derived from IFS techinques instead of applying direclty the fractal $\hat{M}$ or $\hat{T}_N$ estimator.

To conclude this section, we have to say that it is still possible to build IFSs in the space of density functions but direct application to estimation is less straightforward and this will be the object of another paper as it requires a different class of IFS systems, namely the local-IFS approach.

5 Monte Carlo analysis

Before going into details with simulations results, we want to remark that the IFS estimators are fractal objects, this means that they are nowhere differentiable and they are self-similar. In Figure 2 we have represented the distribution function estimator $\hat{T}_N$ of an underlying truncated normal distribution. It is evident that the curve simply replicates of itself. To put in evidence this fractal nature, we have "zoomed" the curve 4 times. As it is possible to see the curve is the same at any scale. Figure 1 shows an application of the density estimator to real data. In particular, we have choosen the classical textbook Old Faithful gayser data rescaled on [0,1]. It is evident that $\hat{f}_{FT}$ is capable of discriminate the two curves as the kernel estimator does.

As seen in the previous sections, it is rather difficult to establish statistical properties of the estimators based on the IFS as it is not yet clear to us, how to characterize the fixed points of the IFSs. So in this section we will show some numerical results both for distribution function and density estimation. We have choosen the Beta family of random variable as they allow compact support, moments existance, different shapes and well tested pseudo random number generators. As criterior for evaluating the performance of the estimators we consider the average mean square error (AMSE) and the sup-norm distance. We also consider small sample sizes $n = 10, 20, 30, 50, 75, 100$ as asymptotically the IFS converges to the e.d.f. based estimators. Four estimators are considered for the distribution function: $\hat{M}$ with $\mathcal{W}_1$, $\hat{T}_N$, $\hat{F}_n$, $\hat{F}_{FT}$. For $\hat{T}_N$ we have choose $n/2$ quantiles. For the density estimator, we compare a standard kernel estimator and $\hat{f}_{FT}$, the Fourier transform estimator based on the IFS. It is well known that kernel estimators are particular Fourier expansion estimators by a proper choice of the multiplier c_k when the e.d.f is used in the Fourier transform. The number of terms used in the Fourier series estimators of the distribution function, is choosen accordingly to the following rule

$$\text{if } \left|\hat{B}_{m+1}\right|^2 \text{ and } \left|\hat{B}_{m+2}\right|^2 < \frac{2}{n+1} \quad \text{then use the first } m \text{ coefficents}$$

as suggested in Tarter and Lock (1993). The rule of thumb we use cannot be considered optimal in any sense but its principle is to minimize the integrated MSE. The software used is R (Ihaka and Gentleman, 1996) with a beta 'ifs' package available soon at `http://159.149.74.117/~web/R/ifs/`. Kernel density estimation is as in Silverman (1986) and implemented in R with the `density()` function (see also Venables and Ripley, 2002) in the R implementation. All the estimates are evaluated in 512 points in order to calculates AMSE and the sup-norm. For density estimation we calculate the average of the absolute error (MAE) instead of the sup-norm as this index is influenced by the bad performance of density estimators in the endpoints (0 and 1) of the support of the distributions.

Tables 2 and 3 are organized as follows: there are five main columns, one for the distribution investigated, two for the distribution function estimators and the last two are for density estimation. Under column AMSE, the $\hat{M}_{\mathcal{W}_1}$ column reports the ratio, in percentage, between the AMSE of $\hat{M}_{\mathcal{W}_1}$ and the AMSE of the $\hat{F}_n$ and similarly for the entire row. This means that we indicate the relative efficiency of the three estimators $\hat{M}_{\mathcal{W}_1}$, $\hat{T}_N$ anf $\hat{F}_{FT}$ with respect to $\hat{F}_n$. Under the column marked SUP-NORM the same scheme as been applied but considering the sup-norm distance.

The last two columns are for density estimation. This time the columns represents the ratio in percentage, of the distance for the Fourier estimator $\hat{f}_{FT}$ and the kernel estimator.

The tables shows that, in the average the $\hat{T}_N$ estimator is equivalent to to the e.d.f. for ugly distributions like the beta(.1,.9) or beta(.1,.1), while it is somewhat better in the other case

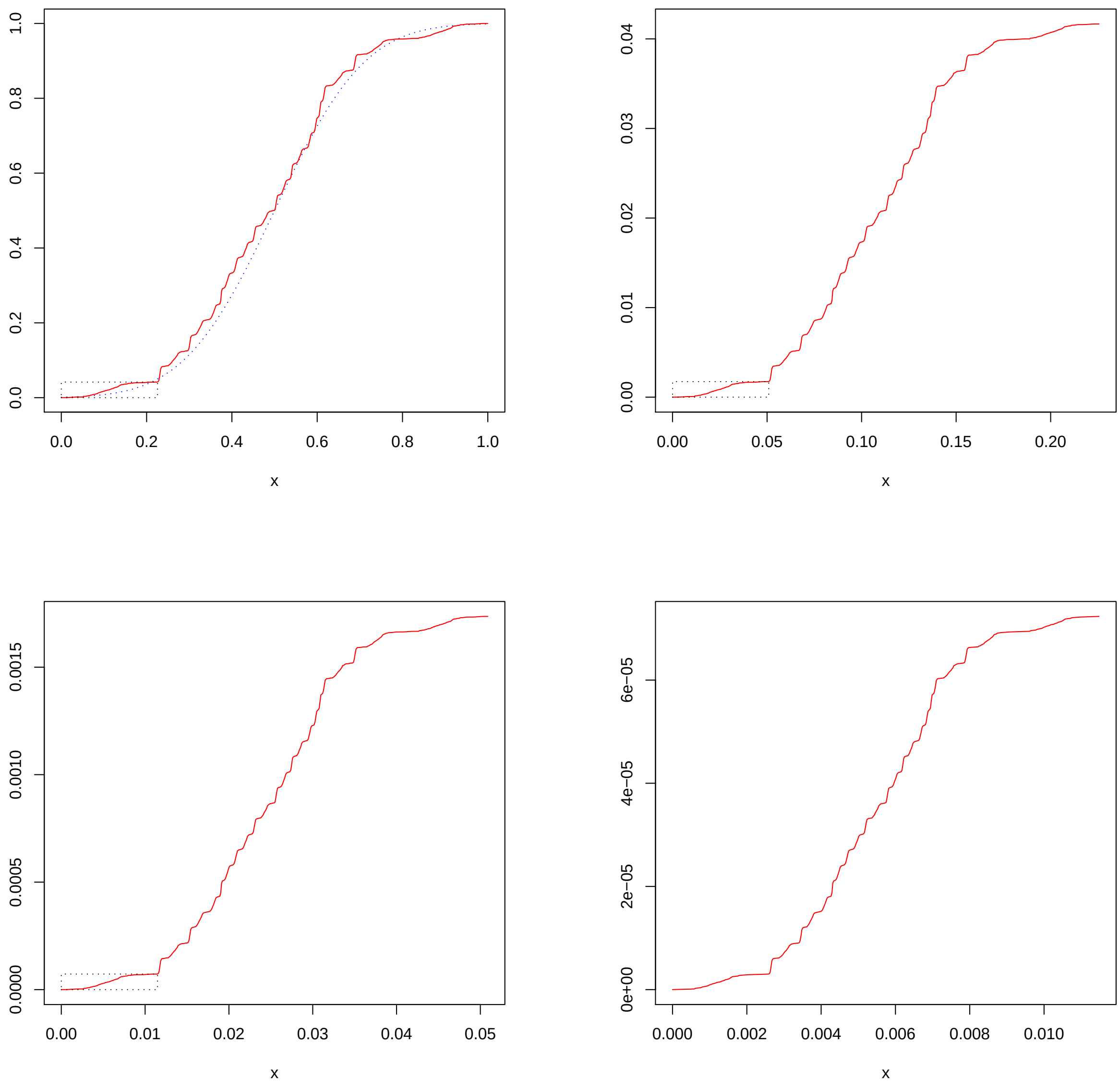

Figure 2: The fractal nature of the IFS distribution function estimator $\hat{T}_N$. The dotted line is the underlying truncated Gaussian distribution. The dotted rectangle is to represent the area zoomed in the next plot (left to right, up to down). The dotted boxes are in the order: $[0, \hat{q}_2] \times [0, \hat{q}_2]$, $[0, \hat{q}_2^2] \times [0, \hat{q}_2^2]$ and $[0, \hat{q}_2^3] \times [0, \hat{q}_2^3]$.

parameters		AMSE			SUP-NORM			AMSE	MAE
n	law	$\hat{M}_{\mathcal{W}_1}$	$\hat{T}_N$	$\hat{F}_{FT}$	$\hat{M}_{\mathcal{W}_1}$	$\hat{T}_N$	$\hat{F}_{FT}$	$\hat{f}_{FT}$	$\hat{f}_{FT}$
10	beta(.9,.1)	105260.1	94.86	186.81	2179.91	98.10	193.42	—	—
10	beta(.1,.9)	827.42	99.79	2097.27	608.94	100.33	262.18	—	—
10	beta(.1,.1)	34067.84	99.26	560.99	5879.58	100.05	190.82	—	—
10	beta(2,2)	153.01	80.99	133.33	102.51	82.18	68.48	171.16	123.05
10	beta(5,5)	114.13	89.91	210.05	99.60	89.74	81.67	162.46	123.63
10	beta(5,3)	76.14	98.57	163.17	79.46	92.86	71.24	185.88	119.36
10	beta(3,5)	142.96	90.70	194.36	99.41	91.34	79.51	154.86	125.25
10	beta(1,1)	58094.14	81.23	88.16	2741.62	80.04	80.07	148.71	117.11

parameters		AMSE			SUP-NORM			AMSE	MAE
n	law	$\hat{M}_{\mathcal{W}_1}$	$\hat{T}_N$	$\hat{F}_{FT}$	$\hat{M}_{\mathcal{W}_1}$	$\hat{T}_N$	$\hat{F}_{FT}$	$\hat{f}_{FT}$	$\hat{f}_{FT}$
20	beta(.9,.1)	80402.60	101.56	216.58	676.97	99.10	251.70	—	—
20	beta(.1,.9)	3156.71	113.31	4028.99	1265.92	114.50	371.70	—	—
20	beta(.1,.1)	30695.03	99.77	999.28	6611.92	98.84	231.88	—	—
20	beta(2,2)	91.73	86.05	100.44	82.46	83.34	61.55	121.19	109.07
20	beta(5,5)	154.93	89.27	120.92	114.32	87.49	65.56	120.54	109.30
20	beta(5,3)	83.89	92.89	133.44	87.12	89.31	67.28	123.48	104.81
20	beta(3,5)	198.84	88.27	116.82	114.82	87.29	67.64	132.69	117.18
20	beta(1,1)	8601.04	84.89	79.26	751.59	79.27	71.67	153.64	118.69

parameters		AMSE			SUP-NORM			AMSE	MAE
n	law	$\hat{M}_{\mathcal{W}_1}$	$\hat{T}_N$	$\hat{F}_{FT}$	$\hat{M}_{\mathcal{W}_1}$	$\hat{T}_N$	$\hat{F}_{FT}$	$\hat{f}_{FT}$	$\hat{f}_{FT}$
30	beta(.9,.1)	143440.9	97.17	245.11	1357.53	92.91	283.57	—	—
30	beta(.1,.9)	3740.29	118.80	6842.92	1328.19	115.38	486.58	—	—
30	beta(.1,.1)	39983.05	99.62	1352.06	8086.76	98.57	284.39	—	—
30	beta(2,2)	92.86	86.23	86.09	82.66	84.01	55.83	103.84	99.35
30	beta(5,5)	191.88	89.54	114.96	132.21	89.21	67.27	118.48	106.07
30	beta(5,3)	98.55	91.67	135.96	97.77	87.90	66.60	113.78	105.41
30	beta(3,5)	241.12	88.31	108.49	124.99	88.34	66.76	130.77	115.17
30	beta(1,1)	483.23	87.16	87.77	181.99	81.00	75.80	168.50	129.67

Table 2: Relative efficiency of IFS-based estimator with respect to the empirical distribution function and the kernel density estimator. Small sample sizes.

(10 to 20% better). The Fourier series estimator based on IFS, $\hat{F}_{FT}$ is preferable to the e.d.f only for bell shaped distributions and seems unbeatable for simmetric shaped laws. This is somewhat expected by a Fourier expansion estimator. The same argument applies to the density estimator: for bell shaped symmetric distributions, it seems as good as the kernel estimator and in some cases even better.

For the beta(.1,.9) or the beta(.1,.1) the density estimators (both kernel and our Fourier) are of no use, we have omitted the corresponding ratios.

6 Applications to survival analysis

Let T denote a random lifetime (or time until failure) with distribution function F. On the basis of a sample of n independent replications of T the object of inference are usually quantities derived from the so-called survival function $S(t) = 1 - F(t) = P(T < t)$. If F has a density f then it is possible to define the hazard function $h(t) = \lim_{\Delta t \to 0} P(t \leq T < t + \Delta t | T \geq T)/\Delta t = f(t)/S(t)$ and in particular the cumulative hazard function

parameters		AMSE			SUP-NORM			AMSE	MAE
n	law	$\hat{M}_{\mathcal{W}_1}$	$\hat{T}_N$	$\hat{F}_{FT}$	$\hat{M}_{\mathcal{W}_1}$	$\hat{T}_N$	$\hat{F}_{FT}$	$\hat{f}_{FT}$	$\hat{f}_{FT}$
50	beta(.9,.1)	4462.59	98.59	318.90	524.57	95.87	361.67	—	—
50	beta(.1,.9)	7577.62	107.85	11017.55	1875.88	113.85	619.15	—	—
50	beta(.1,.1)	41958.58	97.77	1740.82	8419.93	99.87	318.48	—	—
50	beta(2, 2)	99.40	91.55	75.24	88.29	87.40	54.08	93.06	92.63
50	beta(5, 5)	279.55	91.57	91.36	158.18	89.64	56.17	88.25	91.67
50	beta(5, 3)	124.40	97.80	96.04	112.71	92.73	59.33	89.63	96.36
50	beta(3, 5)	327.86	91.90	84.84	145.24	91.40	60.54	115.95	105.87
50	beta(1, 1)	548.05	92.84	104.46	144.12	86.24	80.56	173.79	132.25

parameters		AMSE			SUP-NORM			AMSE	MAE
n	law	$\hat{M}_{\mathcal{W}_1}$	$\hat{T}_N$	$\hat{F}_{FT}$	$\hat{M}_{\mathcal{W}_1}$	$\hat{T}_N$	$\hat{F}_{FT}$	$\hat{f}_{FT}$	$\hat{f}_{FT}$
75	beta(.9,.1)	49097.24	98.85	409.18	853.06	97.66	427.17	—	—
75	beta(.1,.9)	2518.57	122.62	15348.49	1206.46	122.78	721.83	—	—
75	beta(.1,.1)	52407.20	100.37	2460.06	10475.81	106.85	381.39	—	—
75	beta(2, 2)	109.62	94.77	66.97	90.00	90.42	52.11	89.97	91.61
75	beta(5, 5)	338.18	97.41	72.95	177.11	92.39	53.67	72.79	86.68
75	beta(5, 3)	139.85	104.31	111.65	120.93	96.29	63.52	93.45	97.32
75	beta(3, 5)	407.03	95.14	87.93	163.53	91.80	61.76	107.14	104.07
75	beta(1, 1)	107.36	98.39	130.28	90.23	89.05	83.62	158.01	131.13

parameters		AMSE			SUP-NORM			AMSE	MAE
n	law	$\hat{M}_{\mathcal{W}_1}$	$\hat{T}_N$	$\hat{F}_{FT}$	$\hat{M}_{\mathcal{W}_1}$	$\hat{T}_N$	$\hat{F}_{FT}$	$\hat{f}_{FT}$	$\hat{f}_{FT}$
100	beta(.9,.1)	322.52	98.22	575.82	425.48	96.65	457.07	—	—
100	beta(.1,.9)	1191.05	114.91	22660.02	895.01	152.41	890.91	—	—
100	beta(.1,.1)	65060.00	102.02	3901.37	9778.29	110.96	455.34	—	—
100	beta(2, 2)	113.49	97.03	74.57	97.12	91.68	58.62	96.45	95.48
100	beta(5, 5)	425.20	95.08	69.37	198.27	93.37	50.80	67.01	80.75
100	beta(5, 3)	179.28	95.52	100.85	137.34	93.96	60.62	82.91	92.25
100	beta(3, 5)	489.18	97.59	76.83	176.56	94.23	58.68	92.38	98.10
100	beta(1, 1)	108.74	101.58	141.98	91.94	92.26	84.65	146.78	125.03

Table 3: Relative efficiency of IFS-based estimator with respect to the empirical distribution function and the kernel density estimator. Moderate sample sizes.

$H(t) = \int_0^t h(s)\mathrm{d}s = -\log S(t)$. Usually T is thought to take values in $[0, \infty)$, but we can think to consider the estimation conditionally to the last sample failure, say τ, and rescale the interval $[0, \tau]$ to $[0, 1]$. So we will assume, from now on, all the failure times occur in [0,1], being 1 the instant of the last failure when the experiment stops. In this scheme of observation $\hat{S}(t) = 1 - \hat{F}(t)$ is a natural estimator of S, with $\hat{F}$ any estimator of F and, in particular, the IFS estimator. A more realistic situation is when some censoring occurs, in the sense that, as time pass by, some of the initial n observations are removed at random times C not because of failure (or death) but for some other reasons. In this case, a simple distribution function estimator is obviously not good. Let us denote by $t_0 = 0 < t_1 < \cdots < t_{d-1} < t_d = 1$ the observed instants of failure (or death). A well known estimator of S is the Kaplan-Meyer estimator

$$\hat{S}(t) = \prod_{t_i < t} \frac{r(t_i) - d_i}{r(t_i)}$$

where $r(t_i)$ are the subject exposed to risk of death at time t_i and d_i are the dead in the time interval $[t_i, t_{i+1})$ (see the original paper of Kaplan and Meyer, 1958 or for a modern account Fleming and Harrington, 1991). In our case d_i is one as t_i are the instants when failures occur. Subjects exposed to risk are those still present in the experiment and not yet dead or censored. This estimator has good properties whenever T and C are independent. Related to the quantities $r(t_i)$ and d_i it is also available the Nelson estimator for the function H that is defined as $\hat{H}(t) = \sum_{t_i < t} d_i / r(t_i)$. We assume for simplicity that there are not ties, in the sense that in each instant t_i only one failure occurs. The function $\hat{H}(t)$ is a increasing step-function. Now let $\hat{\mathcal{H}}(t) = \hat{H}(t)/\hat{H}(1)$. $\hat{\mathcal{H}}(t)$ can be thought as an empirical estimates of a distribution function $\mathcal{H}$ on [0,1]. To derive and IFS estimator for the cumulative hazard function H we construct the sample quantiles by simply taking the inverse of $\hat{\mathcal{H}}$. Suppose we want to deal with $N+1$ quantiles, being $\hat{q}_1 = 0$ and $\hat{q}_{N+1} = 1$. One possible definition of the empirical quantile of order k/N is obtained by the formula

$$\hat{q}_{k+1} = t_i + \frac{t_{i+1} - t_i}{\hat{\mathcal{H}}(t_{i+1}) - \hat{\mathcal{H}}(t_i)} \cdot \left(\frac{k}{N} - \hat{\mathcal{H}}(t_i) \right), \quad \text{if} \quad \hat{\mathcal{H}}(t_i) \leq \frac{k}{N} < \hat{\mathcal{H}}(t_{i+1}), \tag{9}$$

for $i = 0, 1, \ldots, d-1$ and $k = 1, 2, \ldots, N-1$. Now set $p_i = 1/N$, $i = 1, 2, \ldots, N$ and $\hat{q}_i$, $i = 1, 2, \ldots, N+1$ as in (9). An IFS estimator of H is $\hat{H}(1) \cdot \tilde{H}(t)$ where $\tilde{H}(t)$ is the following IFS:

$$\tilde{H}(t) = \tilde{H}u(t) = \sum_{i=1}^{N} \frac{1}{N} u(\hat{w}_i^{-1}(x))$$

where $\hat{w}_i^{-1}(x) = (x - \hat{q}_i)/(\hat{q}_{i+1} - \hat{q}_i)$ and u is any member of the space of distribution function on $[0, 1]$. In (9) we have assumed that $\mathcal{H}$ is the distribution function of a continuous random variable, with $\mathcal{H}$ varying linearly between t_i and t_{i+1}, but of course any other assumption than linearity can be made as well (for example an exponential behaviour). A Fleming-Harrington (or Altshuler) IFS-estimator of S is then

$$\tilde{S}(t) = \exp\{-\hat{H}(1) \cdot \tilde{H}(t)\}, \quad t \in [0, 1] .$$

7 Final remarks

We have shown how it is relatively powerful to adopt IFS technique in distribution function estimation and related quantities (density and Fourier transform). There are several open issues about the estimators themself. The main open problem is about a better characterization of the fixed points of the IFS in order to establish non asymptotic properties for the estimators. The second, and commonly not discussed in the IFS literature, is the problem of choosing the maps $\mathbf{w}$. There recently appeared some papers that discuss the relationship of some class of IFS and wavelets analysis as well as some papers on local IFS (possible candidates to density function approximators) but the results there in are not directly useful to statistics.

8 References

Barnsley, M.F., Demko, S. (1985), Iterated function systems and the global construction of fractals, *Proc. Roy. Soc. London, Ser A*, **399**, 243-275.

Beran, R. (1977), Estimating a distribution function, *Ann. Statist.*, 5, 400-404.

Dvoretsky, A., Kiefer, J. and Wolfowitz, J. (1956), Asymptotic minimax character of the sample distribution function and of the classical multinomial estimators, *Ann. Math. Statist.*, 27, 642-669.

Efromovich, S. (2001), Second order efficient estimating a smooth distribution function and its applications, *Meth. Comp. App. Probab.*, 3, 179-198.

Kaplan, E. and Meyer,P. (1958) Nonparametric estimator from incomplete observations, J. Amer. Statist. Ass, **53**, 457-81.

Forte, B., Vrscay, E.R. (1995), Solving the inverse problem for function/image approximation using iterated function systems, I. Theoretical basis, *Fractal*, **2**, 3, 325-334.

Forte, B., Vrscay, E.R. (1998), Inverse problem methods for generalized fractal transforms, in *Fractal Image Encoding and Analysis*, NATO ASI Series F, Vol. 159, ed. Y. Fisher, Springer Verlag, Heidelberg.

Fleming, T.R. and Harrington, D.P. (1991), Counting processes and survival analysis, Wiley, New York.

Kiefer, J. and Wolfowitz, J. (1959), Aymptotic minimax character of the sample distribution function for vector chance variables, *Ann. Math. Stat.*, 30, 463-489.

Gill, R. D. and Levit, B. Y. (1995), Applications of the van Trees inequality: A Bayesian Cramér-Rao bound, *Bernoulli*, 1, 59-79.

Golubev, G. K. and Levit, B. Y. (1996a), On the second order minimax estimation of distribution functions, *Math. Methods. Statist.*, 5, 1-31.

Golubev, G. K. and Levit, B. Y. (1996b), Asymptotic efficient estimation for analytic distributions, *Math. Methods. Statist.*, 5, 357-368.

Hutchinson, J., (1981), Fractals and self-similarity, *Indiana Univ. J. Math.*, **30**, 5, 713-747.

Iacus, S.M. and La Torre, D. (2001), Approximating distribution functions by iterated function systems, *submitted*, available as Acrobat PDF file at `http://159.149.74.117/~web/R/ifs/ifs.pdf`

Ihaka, R. and Gentleman, R. (1996), R: A Language for Data Analysis and Graphics, *Journal of Computational and Graphical Statistics*, 5, 299-314.

Levit, B.Y. (1978), Infinite-dimensional information inequalities, *Theory Probab. Applic.*, 23, 371-377.

Millar, P.W. (1979), Aymptotic minimax theorems for sample distribution functions, *Z. Warsch. Verb. Geb.*, 48, 233-252.

Silverman, B. W. (1986) Density Estimation, London, Chapman and Hall.

Tarter, M.E. and Lock, M.D, Model free curve estimation, Chapman & Hall, New York.

Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S-PLUS, New York, Springer, forthcoming.