draft Will be set by the publisher
URL: http://159.149.74.117/~web/ifs/

APPROXIMATING DISTRIBUTION FUNCTIONS BY ITERATED FUNCTION SYSTEMS

Stefano M. Iacus[1] and Davide La Torre[1]

Abstract. In this small note an iterated function system on the space of distribution functions is built. The inverse problem is introduced and studied by convex optimization problems. Applications of this method to approximation of distribution functions and estimation are presented.

Résumé. Dans cette petite note un système de fonction itéré sur l'espace de fonctions de repartition est construit. Le problème inverse est introduit et étudié par des problèmes d'optimisation convexes. Des applications de cette méthode à l'approximation de fonctions de repartition et à l'estimation est présenté.

1991 Mathematics Subject Classification. 62E17, 62H10, 37H.

.

1. Introduction

The Iterated Function Systems (IFSs) were born in mid eighties [2,7] as applications of the theory of discrete dynamical systems and as useful tools to build fractals and other similar sets. Some possible applications of IFSs can be found in image processing theory [6], in the theory of stochastic growth models [13] and in the theory of random dynamical systems [1,4,9]. The fundamental result [2] on which the IFS method is based is Banach theorem.

In practical applications a crucial problem is the so-called *inverse problem.* This can be formulated as follows: given f in some metric space (S,d), find a contraction $T : S \to S$ that admits a unique fixed point $\tilde{f} \in S$ such that $d(f, \tilde{f})$ is small enough. In fact if one is able to solve the inverse problem with arbitrary precision, it is possible to identify f with the operator T which has it as fixed point.

The paper is organized as follows: Section 2 is devoted to introduce a contractive operator T on the space of distribution functions while, in Section 3, the inverse problem for T is studied in details. Section 4 is divided into two parts: in the first some examples of inverse problems are analyzed and explicit solutions are given. In the second one, we introduce an estimator of the unknown distribution function based on IFSs.

Keywords and phrases: iterated function systems, optimization, distribution function estimation

[1] Dipartimento di Economia e Politica Aziendale, Via Conservatorio 7, I-20122 Milano, Italy; e-mail: stefano.iacus@unimi.it; e-mail: davide.latorre@unimi.it.

2. A contraction on the space of distribution functions

Let us denote by the space of distribution functions F on $[0,1]$ by $\mathcal{F}([0,1])$ and by $\mathcal{B}([0,1])$ the space of real bounded functions on $[0,1]$. Let us further define, for $F, G \in \mathcal{B}([0,1])$, $d_\infty(F,G) = \sup_{x\in[0,1]} |F(x) - G(x)|$. So that $(\mathcal{F}([0,1]), d_\infty)$ is a complete metric space.

Let $N \in \mathbb{N}$ be fixed and let:

i) $w_i : [a_i, b_i) \to [c_i, d_i) = w_i([a_i, b_i))$, $i = 1, \dots, N-1$, $w_N : [a_N, b_N] \to [c_N, d_N]$, with $a_1 = c_1 = 0$ and $b_N = d_N = 1$;

ii) w_i, $i = 1 \dots N$, are increasing and continuous;

iii) $\bigcup_{i=1}^{N-1} [c_i, d_i) \cup [c_N, d_N] = [0,1]$;

iv) if $i \neq j$ then $[c_i, d_i) \cap [c_j, d_j) = \emptyset$.

v) $p_i \geq 0$, $i = 1, \dots, N$, $\delta_i \geq 0$, $i = 1 \dots N-1$, $\sum_{i=1}^{N} p_i + \sum_{i=1}^{N-1} \delta_i = 1$.

On $(\mathcal{F}([0,1], d_\infty)$ we define an operator in the following way:

$$TF(x) = \begin{cases} p_1 F(w_1^{-1}(x)), & x \in [c_1, d_1) \\ p_i F(w_i^{-1}(x)) + \sum_{j=1}^{i-1} p_j + \sum_{j=1}^{i-1} \delta_j\,, & x \in [c_i, d_i)\,,\ i = 2, \dots, N-1 \\ p_N F(w_N^{-1}(x)) + \sum_{j=1}^{N-1} p_j + \sum_{j=1}^{N-1} \delta_j\,, & x \in [c_N, d_N] \end{cases} \tag{1}$$

where $F \in \mathcal{F}([0,1])$. In many pratical cases w_i are affine maps. The new distribution function TF is union of distorted copies of F; this is the fractal nature of the operator.

A similar approach has been discussed in [10] but here a more general operator is defined.

We stress here that in the following we will think to the maps w_i and to the parameters δ_j as fixed whistle the parameters p_i have to be chosen. To put in evidence the dependence of the operator T on the vector $p = (p_1, \dots, p_N)$ we will write T_p instead of T.

In Remark 2.2 the hypotheses ii) and v) will be weakened to allow more general functionals.

Theorem 2.1. *T_p is an operator from $\mathcal{F}([0,1])$ to itself.*

Proof. It is trivial that $T_pF(0) = 0$ and $T_pF(1) = 1$. Furthermore if $x_1 > x_2$, without loss of generality, we will consider the two cases:

i) $x_1, x_2 \in w_i([a_i, b_i))$;

ii) $x_1 \in w_{i+1}([a_{i+1}, b_{i+1}))$ and $x_2 \in w_i([a_i, b_i))$.

In case *i*), recalling that w_i are increasing maps, we have:

$$\begin{aligned} T_pF(x_1) &= p_i F(w_i^{-1}(x_1)) + \sum_{j=1}^{i-1} p_j + \sum_{j=1}^{i-1} \delta_j \\ &\geq p_i F(w_i^{-1}(x_2)) + \sum_{j=1}^{i-1} p_j + \sum_{j=1}^{i-1} \delta_j \\ &= T_pF(x_2) \end{aligned}$$

In case *ii*) we obtain:

$$\begin{aligned} T_pF(x_1) - T_pF(x_2) &= p_i + \delta_i + p_{i+1} F(w_{i+1}^{-1}(x_1)) - p_i F(w_i^{-1}(x_2)) \\ &= p_i(1 - F(w_i^{-1}(x_2))) + p_{i+1} F(w_{i+1}^{-1}(x_1)) + \delta_i \geq 0 \end{aligned}$$

since $p_i \geq 0$, $\delta_i \geq 0$ and $0 \leq F(y) \leq 1$. Finally, one can prove without difficulties the right continuity of $T_p f$. $\square$

The following remark will be useful for the applications in Section 4.

Remark 2.2. If hypotheses i), ii) and v) in the definition of T_p are replaced by the following

i'+ii') $w_i(x) = x$, $a_i = c_i$, $b_i = d_i$, $i = 1, \dots, N$,

v') $p_i = p$, $\delta_i \geq -p$, $Np + \sum_{i=1}^{N-1} \delta_i = 1$,

then $T_p : \mathcal{F}([0,1]) \to \mathcal{F}([0,1])$.

Theorem 2.3. *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.*

Proof. Let $F, G \in (\mathcal{F}([0,1]), d_\infty)$ and let it be $x \in w_i([a_i, b_i))$. We have

$$|T_p F(x) - T_p G(x)| = p_i \left| F(w_i^{-1}(x)) - G(w_i^{-1}(x)) \right| \leq c\, d_\infty(F, G)\,.$$

This implies $d_\infty(T_p F, T_p G) \leq c\, d_\infty(F, G)$. $\square$

The following theorem states that small perturbations of the parameters p_i produce small variations on the fixed point of the operator.

Theorem 2.4. *Let $\mathbf{p}$, $\mathbf{p}^* \in \mathbb{R}^N$ such that $T_p F_1 = F_1$ and $T_{p^*} F_2 = F_2$. Then*

$$d_\infty(F_1, F_2) \leq \frac{1}{1-c} \sum_{j=1}^{N} \left| p_j - p_j^* \right|$$

where c is the contractivity constant of T_p.

Proof. In fact, recalling that w_i and δ_i are fixed, we have

$$\begin{aligned} d_\infty(F_1, F_2) &= d_\infty(T_p F_1, T_p F_2) \\ &= \max_{i=1,\dots,N} \sup_{x \in [c_i, d_i)} \left\{ \left| p_i F_1(w_i^{-1}(x)) + \sum_{j=1}^{i-1} p_j - p_i^* F_2(w_i^{-1}(x)) - \sum_{j=1}^{i-1} p_j^* \right| \right\} \\ &\leq \sum_{i=1}^{N} |p_i - p_i^*| + c\, d_\infty(F_1, F_2)\,, \end{aligned}$$

since

$$\begin{aligned} &\left| p_i F_1(w_i^{-1}(x)) + \sum_{j=1}^{i-1} p_j - p_i^* F_2(w_i^{-1}(x)) - \sum_{j=1}^{i-1} p_j^* \right| \\ &\leq \sum_{j=1}^{i-1} |p_j - p_j^*| + |p_i F_1(w_i^{-1}(x)) - p_i F_2(w_i^{-1}(x))| + |p_i F_2(w_i^{-1}(x)) - p_i^* F_2(w_i^{-1}(x))| \\ &\leq \sum_{j=1}^{i-1} |p_j - p_j^*| + p_i d_\infty(F_1, F_2) + |p_i - p_i^*| \\ &\leq c\, d_\infty(F_1, F_2) + \sum_{j=1}^{N} |p_j - p_j^*|\,. \end{aligned}$$

$\square$

3. The inverse problem as a convex constrained optimization problem

Choose $F \in (\mathcal{F}([0,1]), d_\infty)$. The aim of solving the inverse problem is to find a contractive map $T : \mathcal{F}([0,1]) \to \mathcal{F}([0,1])$ which has a fixed point "near" to F. In fact if it is possible to solve the inverse problem with an arbitrary precision one can identify the operator T with its fixed point. With a fixed system of maps w_i and parameters δ_j, the inverse problem can be solved, if it is possible, by using the parameters p_i. These have to be choosen in the following convex set:

$$C = \left\{ \mathbf{p} \in \mathbb{R}^N : p_i \geq 0, i = 1, \ldots, N, \sum_{i=1}^{N} p_i = 1 - \sum_{i=1}^{N-1} \delta_i \right\},$$

We have the following result that is trivial to prove.

Proposition 3.1. *Choose $\epsilon > 0$ and $\mathbf{p} \in C$ 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 $c = \max_{i=1,\ldots,N} p_i$ is the contractivity constant of T_p.

If we wish to find an approximated solution of the inverse problem, we have to solve the following constrained optimization problem:

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

It is clear that the ideal solution of $(\mathbf{P})$ consists of finding a $\mathbf{p}^* \in C$ such that $d_\infty(T_{p^*} F, F) = 0$. In fact this means that, given a distribution function F, we have found a contractive map T_p which has exactly F as fixed point. Indeed the use of Theorem 3.1 gives us only an approximation of F. This can be improved, once fixed the maps w_i, by increasing the number of parameters p_i.

The following result proves the convexity of the function $D(\mathbf{p}) = d_\infty(T_p F, F)$, $\mathbf{p} \in \mathbb{R}^N$.

Theorem 3.2. *The function $D(\mathbf{p}) : \mathbb{R}^N \to \mathbb{R}$ is convex.*

Proof. If we choose $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{R}^N$ and $\lambda \in [0,1]$ then:

$$D(\lambda \mathbf{p}_1 + (1-\lambda)\mathbf{p}_2) = \sup_{x \in [0,1]} |T_{\lambda p_1 + (1-\lambda)p_2} F(x) - F(x)| \leq$$

$$\lambda \sup_{x \in [0,1]} |T_{p_1} F(x) - F(x)| + (1-\lambda) \sup_{x \in [0,1]} |T_{p_2} F(x) - F(x)| = \lambda D(p_1) + (1-\lambda) D(p_2).$$

□

Hence for solving problem $(\mathbf{P})$ one can recall classical results about convex programming problems (see for instance [14]). A necessary and sufficient condition for $\mathbf{p}^* \in C$ to be a solution of $(\mathbf{P})$ can be given by Kuhn-Tucker conditions.

4. Inverse problem for distribution functions and applications

In this section we consider different problems. We show that for a particular class of distribution functions the inverse problem can be solved exactly without solving any optimization problem. Then we discuss two ways of construct IFS to approximate a distribution function F with a finite number of parameters p_i and maps w_i.

As is usual in statistical applications, given a sample of n independent and identically distributed observations, $(x_1, x_2, \ldots, x_n)$, drawn from an unknown distribution function F, one can easily contruct the empirical distribution function (e.d.f.) $\hat{F}_n$ that reads

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} \chi_{(-\infty,x]}(x_i), \quad x \in \mathbb{R}\,,$$

where χ_A is the indicator function of the set A. Asymptotic properties of optimality of $\hat{F}_n$ as estimator of the unknown F when n goes to infinity are well known and studied [11,12]. This function has an IFS representation that is exact and can be found without solving any optimization problem. We assume that the x_i in the sample are all different (this assumption is natural if F is a continuous distribution function). Let $w_i(x) : [x_{i-1}, x_i) \to [x_{i-1}, x_i)$, when $i = 1 \ldots n$ and $w_1(x) : [0, x_1) \to [0, x_1)$, $w_{n+1}(x) : [x_n, x_{n+1}] \to [x_n, x_{n+1}]$,with $x_0 = 0$ and $x_{n+1} = 1$. Assume also that every map is of the form $w_i(x) = x$. If we choose $p_i = \frac{1}{n}$, $i = 2 \ldots n+1$, $p_1 = 0$ and

$$\delta_1 = \frac{n-1}{n^2}, \quad \delta_i = -\frac{1}{n^2}$$

then the following representation holds:

$$T_p\hat{F}_n(x) = \begin{cases} 0, & i = 1 \\ \frac{1}{n}\hat{F}_n(x) + \frac{n-1}{n^2}, & i = 2 \\ \frac{1}{n}\,\hat{F}_n(x) + \frac{i-1}{n} + \frac{n-i+1}{n^2}, & i = 3, \ldots, n+1. \end{cases}$$

when $x \in [x_{i-1}, x_i)$. Furthermore, from Remark 2.2 we are guaranteed that

$$\lim_{s\to\infty} d_\infty(T_p^{(s)}u, \hat{F}_n) \to 0, \quad \forall\, u \in \mathcal{F}[0,1].$$

Note that from the point of view of applications, constructing the e.d.f. or iterate the IFS with the given maps is exactly equivalent if one start, for example, with a uniform distribution on [0,1] in the first iteration. So this is just a case when we can present an IFS that gives exact result for this particular class of distribution functions.

What follows, on the contrary, is more attractive from the point of view of applications. Suppose that one knows the distribution function F and wants to construct the IFS which has F as fixed point. In general one has to provide an infinite set of affine maps $\{w_i, i \in \mathbb{N}\}$ and solve an extremal problem to find the corresponding sequence of weigths p_i, $i \in \mathbb{N}$. This problem has not a general solution but at the same time the solution in terms of a finite, possibily few, number of maps and weigths is crucial in applications like image compression and trasmission.

The idea is the following: one can think at n points $(x_1, x_2, \ldots, x_n)$ as they were drawn from the distribution function F, and use the same maps w_i of the e.d.f. $\hat{F}_n$, then instead of using the p_i equal to $1/n$ one solve the extremal problem as it is usual in IFS application. The corresponding IFS should have a fixed point that is a "good" approximation on F. So it is sufficient to store the simulated data and the weights instead of F itself.

We take the functional T_pF with the particular choice of $\delta_i = 0$. This choice is in principle not necessary but simplifies the solution of the problem. We simulated n i.i.d. observations from the distribution function F and we use the maps of the e.d.f. above.

We now try to solve the extremal problem

$$\min d_\infty(T_pF, F)$$

under the constrain $\sum_{i=1}^{n} p_i = 1$, $p_i \geq 0$, $i = 1, \ldots, n$ (with some $p_i > 0$). The optimal solution will be $\{\hat{p}_i, i = 1, \ldots, n\}$ such that $d_\infty(T_{\hat{p}}F, F) = 0$ that it is true in at least one case: if F equals $\hat{F}_n$ and $p_i = 1/n$.

Otherwise we will obtain some positive number. That means that, in principle, in the worst case we can approximate F with its empirical distribution function $\hat{F}_n$. But we can generally do better. So let us solve the problem: let us fix $x_0 = 0$ and $x_{n+1} = 1$, then

$$\begin{aligned}
d_\infty(T_pF, F) &= \sup_{x\in[0,1]} |T_pF(x) - F(x)| \\
&= \max_{i=1,\dots,n+1} \left\{ \sup_{[x_{i-1},x_i)} |T_pF(x) - F(x)| \right\} \\
&= \max_{i=1,\dots,n+1} \left\{ \sup_{[x_{i-1},x_i)} \left| \sum_{j=1}^{i-1} p_j - (1-p_i)F(x) \right| \right\} \\
&= \max_{i=1,\dots,n+1} \left\{ \left| \sum_{j=1}^{i-1} p_j - (1-p_i)F(x_{i-1}) \right|, \left| \sum_{j=1}^{i-1} p_j - (1-p_i)F(x_i^-) \right| \right\}
\end{aligned}$$

and the last line is due to the non-decreasing property of F.

Example 4.1. Suppose that F is the distribution function of a uniform distribution on [0,1] and suppose that we can only draw one observation from F (or choose a point) x_1. The empirical distribution function $\hat{F}_1(x) = \chi_{\{x_1,1\}}(x)$ is usless if we have in mind to approximate F. Let us use the second technique: fix $w_1 : [0, x_1) \to [0, x_1)$ and $w_2 : [x_1, 1] \to [x_1, 1]$. We try to solve the above extremal problem.

$$\begin{aligned}
d_\infty(T_pF, F) &= \max_{i=1,\dots,n+1} \left\{ \left| \sum_{j=1}^{i-1} p_j - (1-p_i)F(x_i) \right|, \left| \sum_{j=1}^{i-1} p_j - (1-p_i)F(x_{i+1}) \right| \right\} \\
&= \max\left\{ \Big| -(1-p_1)\cdot 0 \Big|, \Big| -(1-p_1)x_1 \Big|, \Big| p_1 - (1-p_2)x_1 \Big|, \Big| p_1 - (1-p_2) \Big| \right\} \\
&= \max\{0, x_1(1-p_1), p_1(1-x_1), 0\},
\end{aligned}$$

$x_1 \in (0, 1)$, $p_1 + p_2 = 1$. Now minimazing with respect to p_1 and p_2 under the constrain $p_1 + p_2 = 1$ one obtains simply $p_1 = x_1$. The resulting functional will be

$$T_{x_1}u(x) = \begin{cases} x_1\, u(x), & x \in [0, x_1) \\ (1-x_1)\, u(x) + x_1, & x \in [x_1, 1] \end{cases}$$

and it is clear that $T_{x_1}u(x) = T_{x_1}x$, for one iteration only, is closer than $\hat{F}_1$ to $F(x) = x$ and the approximation is better and better as $x_1 \to 0$ or $x_1 \to 1$.

We propose now a more efficient method to approximate F when F is not to be estimated. We have already mentioned that the e.d.f. is the better estimator of an unknown distribution function F, so one can think to sample n points from F and use their values to approximate F by $\hat{F}_n$. As $n \to \infty$, the statistical literature assures almost sure convergence of $\hat{F}_n(x)$ to $F(x)$ for every x. We also have shown the exact IFS representation of $\hat{F}_n$. But this method is not efficient. On the contrary, suppose that F is a continuous distribution function. As we know F, we can think to approximate it by means of continuous functions instead of simple functions like $\hat{F}_n$. Choose n points $(x_1, \dots, x_n)$ and assume that $x_0 = 0$ and $x_{n+1} = 1$. One can costruct the following functional

$$T_F u(x) = (F(x_i) - F(x_{i-1}))u\left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right) + F(x_{i-1}), \quad x \in [x_{i-1}, x_i),$$

$i = 1, \dots, n+1$. Notice that T_F is a particular case of (1) where $p_i = F(x_i) - F(x_{i-1})$, $\delta_i = 0$ and $w_i(x) : [0, 1) \to [x_{i-1}, x_i)$. 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. So again, instead of sending an infinite set of weigths and maps, one can send n points and the values of F evaluated at these points. All in summary, only $2 \cdot n$ informations should be sent to reconstruct F.

For n small, the choice of a good grid of point is critical. So one question arises: how to choose them ? One can proceed case by case but as F is a distribution function one can use its properties. We propose the following solution: take n points $(u_1, u_2, \dots, u_n)$ equally spaced $[0,1]$ and define $x_i = F^{-1}(u_i)$, $i = 1, \dots, n$. The points x_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 x_i and x_{i+1} are relatively distant each other, then F is slowly increasing in the interval (x_i, x_{i+1}) and viceversa. This method is more efficient than simply taking equally spaced points on $[0,1]$. If this method of choosing the points is used, then the functional simply reads

$$T_F u(x) = \frac{1}{n} u\left(\frac{x - x_{i-1}}{x_i - x_{i-1}}\right) + \frac{i-1}{n}, \quad x \in [x_{i-1}, x_i),\, i = 1, \dots, n+1\,.$$

And this suggests an empirical estimator of F. If $\hat{q}_i$, $i = 1, 2, \dots, k$, $k < n$, are the empirical quantiles of the sample $(x_1, x_2, \dots, x_n)$ of order $\frac{i}{k}$, then an estimator of the unknown distribution function F should be written as

$$\tilde{F}_{(k)} u(x) = \frac{1}{k} u\left(\frac{x - \hat{q}_i}{\hat{q}_{i+1} - \hat{q}_i}\right) + \frac{i-1}{k}, \quad x \in [\hat{q}_i, \hat{q}_{i+1}),$$

$i = 1, \dots, k$, with $\hat{q}_0 = 0$ and $\hat{q}_{k+1} = 1$ As n and $k = k(n)$ go to infinity $\hat{F}_{(k)}$ converges to F. Relative efficiency of $\tilde{F}_{(k)}$ with respect to $\hat{F}_n$ is investigated via simulations. The results are reported in Table 1 for differently shaped distribution functions and sample sizes. What emerges is that $\tilde{F}_{(k)}$ is equivalent to the e.d.f. in the sense of the sup-norm. If we choose $k = n$ a natural choice of the $\hat{q}_i$ are the order statistics $x_{(i)}$. Since, every $u \in \mathcal{F}([0,1])$ is such that $u(x) = 0$, $x \leq 0$ and $u(x) = 1$, $x \geq 1$, the operator $\tilde{F}_{(k)}$ can be rewritten as

$$\tilde{F}_{(n)} u(x) = \frac{1}{n} \sum_{i=1}^{n} u\left(\frac{x - x_{(i)}}{x_{(i+1)} - x_{(i)}}\right), \quad x \in \mathbb{R}\,.$$

The fixed point of the above operator, $\tilde{F}^*_{(n)}(x)$, satisfies

$$\tilde{F}^*_{(n)}(x) = \frac{1}{n} \sum_{i=1}^{n} \tilde{F}^*_{(n)}\left(\frac{x - x_{(i)}}{x_{(i+1)} - x_{(i)}}\right), \tag{2}$$

for real x. The following (Glivenko-Cantelli) theorem states that $\tilde{F}^*_{(n)}$ has the same properties of an *admissible perturbation of the e.d.f.* in the sense of [15–17].

Theorem 4.2. *Let $\tilde{F}^*_{(n)}$ be as in* (2)*. If F is continuous, then*

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left|\tilde{F}^*_{(n)}(x) - F(x)\right| \overset{a.s.}{=} 0\,.$$

Proof. We can write

$$\left|\tilde{F}^*_{(n)}(x) - F(x)\right| = \left|\tilde{F}^*_{(n)}(x) - \hat{F}_n(x)\right| + \left|\hat{F}_n(x) - F(x)\right|$$

and the first term can be estimated by $1/n$ while the second one converges to 0 almost surely by the Glivenko-Cantelli theorem. □

Notice that $\tilde{F}^*_{(n)}$ cannot be directly reduced to an admissible perturbed e.d.f. $F^*_n = n^{-1} \sum_{i=1}^n u_n(x - x_{(i)})$ as defined in [15]. In fact, in our case we have a sequence of families $\{u_n^i, i = 1, 2, \dots, n\}_{n>1}$ instead of a simple

number of points drawn from F	$d_\infty\left(\tilde{F}^{(4)}_{(k)}u, F\right)$ (a)	$d_\infty\left(\hat{F}_n, F\right)$ (b)	$\frac{(a)}{(b)} \cdot 100\%$	distribution F
10	0.20232	0.24103	83.94%	Beta(2,2)
50	0.09376	0.10241	91.56%	Beta(2,2)
100	0.06989	0.07131	98.01%	Beta(2,2)
500	0.02884	0.02917	98.87%	Beta(2,2)
1000	0.02475	0.02506	98.78%	Beta(2,2)
10	0.18747	0.19472	96.27%	Beta(3,3)
50	0.09945	0.09777	101.72%	Beta(3,3)
100	0.07103	0.07521	94.44%	Beta(3,3)
500	0.03077	0.03061	100.52%	Beta(3,3)
1000	0.01993	0.02018	98.74%	Beta(3,3)
10	0.20842	0.22220	93.80%	Beta(5,3)
50	0.10615	0.10517	100.93%	Beta(5,3)
100	0.06881	0.07096	96.96%	Beta(5,3)
500	0.02959	0.02971	99.60%	Beta(5,3)
1000	0.02176	0.02194	99.17%	Beta(5,3)
10	0.23054	0.23301	98.94%	Beta(3,5)
50	0.08993	0.089347	100.66%	Beta(3,5)
100	0.06541	0.06515	100.40%	Beta(3,5)
500	0.02978	0.03015	98.80%	Beta(3,5)
1000	0.01978	0.02003	98.77%	Beta(3,5)
10	0.20522	0.24492	83.79%	Beta(1,1)
50	0.10456	0.11990	87.20%	Beta(1,1)
100	0.07621	0.08124	93.81%	Beta(1,1)
500	0.02938	0.02974	98.77%	Beta(1,1)
1000	0.02382	0.02428	98.09%	Beta(1,1)

Table 1. Simulation results. Values are the arithmetic means over 30 trials. The functional is iterated 4 times starting with the uniform distribution on [0,1] as initial point. Functions are evaluated at 20 equally spaced points on [0,1]. The proposed estimator can be said to be almost equivalent as $\hat{F}_n$, the best estimator of F.

sequence $\{u_n\}_{n>1}$ even if, for each fixed $i \leq n$, we have

$$u_n^i = \tilde{F}^*_{(n)}\left(\frac{x}{x_{(i+1)} - x_{(i)}}\right) \xrightarrow{n} u_0 = \begin{cases} 0, & x < 0 \\ 1, & x > 0 \end{cases}$$

by the properties of the order statistics.

Remark 4.3. If we let $k = k(n)$, with $k(n) \to \infty$ as $n \to \infty$ in the construction of $\tilde{F}_{(k)}$, then a similar Glivenko-Cantelli result can be obtained. In fact, if $\tilde{F}^*_{(k)}(x)$ is the fixed point of $\tilde{F}_{(k)}$, we have

$$\left|\tilde{F}^*_{(k)}(x) - F(x)\right| = \left|\tilde{F}^*_{(k)}(x) - \hat{F}_n(x)\right| + \left|\hat{F}_n(x) - F(x)\right|$$

and the first term is bounded by $1/k(n)$.

4.1. **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 $h(t) = \lim_{\Delta t \to 0} P(t \le T < t + \Delta t | T \ge T)/\Delta t = f(t)/S(t)$ and in particular the cumulative hazard function $H(t) = \int_0^t h(s)\,\mathrm{d}s = -\log S(t)$. 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 [8] or for a modern account [5]). 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 no 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 $k + 1$ quantiles, being $\hat{q}_1 = 0$ and $\hat{q}_{k+1} = 1$. One possible definition of the empirical quantile of order m/k is obtained by the formula

$$\hat{q}_{m+1} = t_i + \frac{t_{i+1} - t_i}{\hat{\mathcal{H}}(t_{i+1}) - \hat{\mathcal{H}}(t_i)} \cdot \left(\frac{m}{k} - \hat{\mathcal{H}}(t_i) \right), \quad \text{if} \quad \hat{\mathcal{H}}(t_i) \le \frac{m}{k} < \hat{\mathcal{H}}(t_{i+1}), \tag{3}$$

for $i = 0, 1, \ldots, d - 1$ and $m = 1, 2, \ldots, k - 1$. Now set $p_i = 1/k$, $i = 1, 2, \ldots, k$ and $\hat{q}_i$, $i = 1, 2, \ldots, k + 1$ as in (3). An IFS estimator of H is $\hat{H}(1) \cdot \tilde{H}(t)$ where $\tilde{H}(t)$ is the following IFS:

$$\tilde{H}(t) = \tilde{H}u(t) = \frac{1}{k} \sum_{i=1}^{k} u \left(\frac{t - \hat{q}_i}{\hat{q}_{i+1} - \hat{q}_i} \right),$$

and u is any member of the space of distribution function on $[0, 1]$. In (3) 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

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

5. Final remarks about the method

There is at least one open issue in this topic as this is a first attempt to introduce IFS in distribution function estimation: how to choose the maps ? We have suggested a quantile approach but some other good partition of the space, like a dyadic sequence, can be used at the cost of the need to solve some optimization problems. In [6] this problem is touched incidentally but not in a statistical perspective.

The authors are thanksfull to Anestis Antoniadis for his *conseils* and for having brought to their attention the papers of Winter [15, 16] and Yukich [17].

References

[1] L. Arnold and H. Crauel, Iterated function systems and multiplicative ergodic theory, *Diffusion processes and related problems in analysis* (Pinsky M.A., Vihstutz V. eds.), vol. II, Stochastics flows, Progress in probability 27, Birkäuser, Boston, 283-305, 1992.

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

[3] C.A. Cabrelli, B. Forte, U.M. Molter and E.R. Vrscay, Iterated fuzzy set systems: a new approach to the inverse problem for fractals and other sets, *J. Math. Anal. Appl.*, **171**(1992), 79-100.

[4] J.H. Elton and M. Piccioni, Iterated function systems arising from recursive estimation problems, *Probab. Theory Related Fields*, **91**(1992), 103-114.

[5] T.R. Fleming and D.P. Harrington, Counting processes and survival analysis, Wiley, New York, 1991.

[6] B. Forte and E.R. Vrscay, Solving the inverse problem for function/image approximation using iterated function systems, I. Theoretical basis, *Fractal*, **2**(1994), 3, 325-334.

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

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

[9] A.A. Kwiecinska and W. Slomczynski, Random dynamical systems arising from iterated function systems with place-dependent probabilities, *Stat. and Prob. Letters*, **50**(2000), 401-407.

[10] D. La Torre and M. Rocca, Iterated function systems and optimization problems, *Rend. Sem. Mat. Messina SER II*, **6**(1999), 21, 165-173.

[11] P.W. Millar, Asymptotic minimax theorems for the sample distribution function, *Z. Wahrscheinlichkeitstheorie Verw. Geb.*, **48**(1979), 233-252.

[12] P.W. Millar, The minimax principle in asymptotic statistical theory, *Lecture Notes in Math.*, **976**(1983), 76-265.

[13] L. Montrucchio and F. Privileggi, Fractal steady states in stochastic optimal control models, *Ann. Op. Res.*, **88**(1999), 183-197.

[14] R.T. Rockafellar and R.J-B. Wets, *Variational analysis*, Springer, New York, 1998.

[15] B.B. Winter, Strong uniform consistency of integrals of density estimators, *Can. J. Statist.*, **1**(1973), 247–253.

[16] B.B. Winter, Convergence rate of perturbed empirical distribution functions, *J. Appl. Prob.*, **16**(1979), 163–173.

[17] J.E. Yukish, A note on limit theorems for perturbed empirical processes. *Stoch. Proc. Appl.*, **33**(1989), 163–173.