APPROXIMATING DISTRIBUTION FUNCTIONS BY ITERATED FUNCTION SYSTEMS

STEFANO MARIA IACUS AND DAVIDE LA TORRE

An iterated function system (IFS) on the space of distribution functions is built with the aim of proposing a new class of distribution function estimators. One IFS estimator and its asymptotic properties are studied in detail. We also propose a density estimator derived from the IFS distribution function estimator by using Fourier analysis. Relative efficiencies of both estimators, for small and moderate sample sizes, are presented via Monte Carlo analysis.

1. Introduction

The iterated function systems (IFSs) were born in the mid eighties [2, 7] as applications of the theory of discrete dynamical systems and as useful tools for buildings' fractals and other similar sets. Some possible applications of IFSs can be found in image processing theory [4], in the theory of stochastic growth models [14], and in the theory of random dynamical systems [1, 3, 9]. Here we apply this methodology to estimation.

The fundamental result [2] on which the IFS method is based is the Banach contraction theorem. In practical applications, the crucial problem, usually called the *inverse problem* in the IFS literature, is formulated as follows. Given f in some metric space (S, d), find a contraction $T : S \to S$ from among some given set of contractions 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 exactly, it is possible to identify $\tilde{f}$ with the operator T which has it as fixed point.

The paper is organized as follows. Section 2 introduces a contractive operator T on the space of distribution functions and the definition of the inverse problem for T. Section 3 is devoted to estimation. We propose an IFS distribution function estimator and we study its properties. In particular, we will be able to establish a Glivenko-Cantelli theorem, a law of the iterated-logarithm-type theorem, and results concerning local asymptotic minimax efficiency. Then we derive a characteristic function estimator and, consequently, a density function estimator obtained via Fourier analysis. As the asymptotic results show good asymptotic properties of the IFS estimator, we will also investigate if there is any advantage in using the IFS estimator instead of the celebrated empirical distribution function (EDF)

estimator when the sample size is small ($n = 10$ and 30) or moderate ($n = 50$ and 100). Monte Carlo analysis seems to show some gain of the IFS over the EDF.

2. A contraction on the space of distribution functions

Given a distribution function F on $[0,1]$, that is, $F : [0,1] \rightarrow [0,1]$ such that $F(0) = 0$, $F(1) = 1$, F is nondecreasing and right continuous, we denote by $\mathcal{F}([0,1])$ the space of distribution functions and by $\mathcal{B}([0,1])$ the space of real bounded functions on $[0,1]$. We further define, for $F, G \in \mathcal{B}([0,1])$, $d_\infty(F,G) = \sup_{x \in [0,1]} |F(x) - G(x)|$. Then $(\mathcal{F}([0,1]), d_\infty)$ is a complete metric space.

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

    (i) $w_i : [a_i, b_i) \rightarrow [c_i, d_i) = w_i([a_i, b_i))$, $i = 1, \ldots, N-1$, $w_N : [a_N, b_N] \rightarrow [c_N, d_N]$, with $a_1 = c_1 = 0$ and $b_N = d_N = 1$; $c_{i+1} = d_i$, $i = 1, 2, \ldots, N-1$; $[a_i, b_i) \subset [0,1]$;
    (ii) w_i, $i = 1, \ldots, N$, 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 $[c_i, d_i) \cap [c_j, d_j) = \varnothing$;
    (v) $p_i \geq 0$, $i = 1, \ldots, N$, $\delta_i \geq 0$, $i = 1, \ldots, N-1$, $\sum_{i=1}^{N} p_i + \sum_{i=1}^{N-1} \delta_i = 1$.

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

$$
TF(x) = \begin{cases} p_1 F(w_1^{-1}(x)), & x \in [c_1, d_1), \\ p_i F(w_i^{-1}(x)) + \sum_{j=1}^{i-1} p_j + \sum_{j=1}^{i-1} \delta_j, & x \in [c_i, d_i), \ i = 2, \ldots, N-1, \\ p_N F(w_N^{-1}(x)) + \sum_{j=1}^{N-1} p_j + \sum_{j=1}^{N-1} \delta_j, & x \in [c_N, d_N], \end{cases} \quad (2.1)
$$

where $F \in \mathcal{F}([0,1])$. In many practical cases, w_i are affine maps. 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 of the maps w_i and the parameters δ_j as fixed whilst the parameters p_i have to be chosen. To put in evidence the dependence of the operator T on the vector $\mathbf{p} = (p_1, \ldots, p_N)$, we will write T_p instead of T. Notice that the ordering relations among the intervals $[c_i, d_i)$ are crucial in the definition of T_p.

In the first Remark below, 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 $T_p F(0) = 0$ and $T_p F(1) = 1$. Furthermore, if $x_1 > x_2$, without loss of generality, we can consider the following 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

$$
T_p F(x_1) = p_i F(w_i^{-1}(x_1)) + \sum_{j=1}^{i-1} p_j + \sum_{j=1}^{i-1} \delta_j \geq p_i F(w_i^{-1}(x_2)) + \sum_{j=1}^{i-1} p_j + \sum_{j=1}^{i-1} \delta_j = T_p F(x_2).
$$

$$(2.2)$$

In case (ii), we obtain

$$
\begin{aligned}
T_p F(x_1) - T_p F(x_2) &= p_i + \delta_i + p_{i+1} F(w_{i+1}^{-1}(x_1)) - p_i F(w_i^{-1}(x_2)) \\
&= p_i(1 - F(w_i^{-1}(x_2))) + p_{i+1} F(w_{i+1}^{-1}(x_1)) + \delta_i \geq 0
\end{aligned}
\tag{2.3}
$$

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

The following remark will be useful for the applications in Section 3.

Remark 2.2. Replace hypotheses (i), (ii), and (v) in the definition of T_p by the following: $(i' + ii')$ $w_i(x) = x$, $a_i = c_i$, $b_i = d_i$, $i = 1,\ldots,N$; $(v')p_i = p$, $\delta_i \geq -p$, $Np + \sum_{i=1}^{N-1} \delta_i = 1$. Then $T_p : \mathscr{F}([0,1]) \to \mathscr{F}([0,1])$.

THEOREM 2.3. *If $c = \max_{i=1,\ldots,N} p_i < 1$, then T_p is a contraction on $(\mathscr{F}([0,1]), d_\infty)$ with contractivity constant c.*

Proof. Let $F, G \in (\mathscr{F}([0,1]), d_\infty)$ and suppose that $x \in w_i([a_i, b_i))$. We have

$$
\left| T_p F(x) - T_p G(x) \right| = p_i \left| F(w_i^{-1}(x)) - G(w_i^{-1}(x)) \right| \leq c d_\infty(F, G).
\tag{2.4}
$$

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

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 $T_{p^*} F_2 = F_2$, $F_1, F_2 \in \mathscr{F}([0,1])$. Then*

$$
d_\infty(F_1, F_2) \leq \frac{1}{1-c} \sum_{j=1}^{N} | p_j - p_j^* |,
\tag{2.5}
$$

where c is the contractivity constant of T_p.

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

$$
\begin{aligned}
d_\infty(F_1, F_2) &= \max_{i=1,\ldots,N} \sup_{x \in [c_i, d_i)} \left\{ \left| p_i F_1(w_i^{-1}(x)) + \sum_{j=1}^{i-1} p_j - p_i^* F_2(w_i^{-1}(x)) - \sum_{j=1}^{i-1} p_j^* \right| \right\} \\
&\leq \sum_{i=1}^{N} | p_i - p_i^* | + c d_\infty(F_1, F_2),
\end{aligned}
\tag{2.6}
$$

since

$$\left| p_i F_1\left(w_i^{-1}(x)\right) + \sum_{j=1}^{i-1} p_j - p_i^* F_2\left(w_i^{-1}(x)\right) - \sum_{j=1}^{i-1} p_j^* \right|$$

$$\le \sum_{j=1}^{i-1} |p_j - p_j^*| + |p_i F_1\left(w_i^{-1}(x)\right) - p_i F_2\left(w_i^{-1}(x)\right)|$$

$$+ |p_i F_2\left(w_i^{-1}(x)\right) - p_i^* F_2\left(w_i^{-1}(x)\right)| \tag{2.7}$$

$$\le \sum_{j=1}^{i-1} |p_j - p_j^*| + p_i d_\infty (F_1, F_2) + |p_i - p_i^*|$$

$$\le c d_\infty (F_1, F_2) + \sum_{j=1}^{N} |p_j - p_j^*|. \qquad \square$$

Choose $F \in (\mathscr{F}([0,1]), d_\infty)$. The goal now consists of finding a contractive map $T : \mathscr{F}([0,1]) \to \mathscr{F}([0,1])$ which has a fixed point "near" to F. In fact if it is possible to solve the inverse problem exactly, 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 chosen in the following convex set:

$$C = \left\{ \mathbf{p} \in \mathbb{R}^N : p_i \ge 0, \ i = 1,\dots,N, \ \sum_{i=1}^{N} p_i = 1 - \sum_{i=1}^{N-1} \delta_i \right\}. \tag{2.8}$$

We have the following result that is trivial to prove.

PROPOSITION 2.5. *Choose $\epsilon > 0$ and $\mathbf{p} \in C$ such that all the $p_i > 0$ for some i. If $d_\infty(T_p F, F) \le \epsilon$, then*

$$d_\infty (F, \widetilde{F}) \le \frac{\epsilon}{1 - c}, \tag{2.9}$$

where $\widetilde{F}$ is the fixed point of T_p on $\mathscr{F}([0,1])$ and $c = \max_{i=1,\dots,N} p_i$ is the contractivity constant of T_p.

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

(P)

$$\min_{p \in C} d_\infty (T_p F, F). \tag{2.10}$$

It is clear that the ideal solution of (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 a fixed point. Indeed the use of Proposition 2.5 gives us only an approximation to F. This can be improved by increasing the number of parameters p_i (and maps w_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 2.6. *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

$$
\begin{aligned}
D(\lambda \mathbf{p}_1 + (1-\lambda)\mathbf{p}_2) &= \sup_{x \in [0,1]} \left| T_{\lambda p_1 + (1-\lambda)p_2} F(x) - F(x) \right| \\
&\leq \lambda \sup_{x \in [0,1]} \left| T_{p_1} F(x) - F(x) \right| + (1-\lambda) \sup_{x \in [0,1]} \left| T_{p_2} F(x) - F(x) \right| \quad (2.11) \\
&= \lambda D(p_1) + (1-\lambda) D(p_2).
\end{aligned}
$$

$\square$

Hence for solving problem (**P**), one can recall classical results about convex programming (see, e.g., [15]). A necessary and sufficient condition for $\mathbf{p}^* \in C$ to be a solution of (**P**) can be given by the Kuhn-Tucker conditions.

3. Distribution function estimation and applications

In this section, we focus on estimation problems. Instead of trying to solve exactly the problem (**P**), we will use the properties of distribution functions to obtain a *good* approximator of F that can be directly used in distribution function estimation. We will show that under suitable conditions on the maps w_i, the proposed IFS distribution function estimator is asymptotically efficient for large samples. Via Monte Carlo analysis, we will also show that, for small sample sizes, this IFS estimator is better than the celebrated empirical distribution function in several situations.

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 continuous distribution function $F \in \mathcal{F}([0,1])$, one can easily construct the (EDF) $\hat{F}_n$ given by

$$
\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \chi_{(-\infty, x]}(x_i), \quad x \in \mathbb{R}, \quad (3.1)
$$

where χ_A is the indicator function of the set A. Asymptotic optimality properties of $\hat{F}_n$ as an estimator of the unknown F when n goes to infinity are well known and studied; see [12, 13].

Remark 3.1. 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 $p_i = 1/n$, $i = 2, \ldots, n+1$, $p_1 = 0$ and $\delta_1 = (n-1)/n^2$, $\delta_i = -1/n^2$, then the following representation holds:

$$
T_p \hat{F}_n(x) = \begin{cases} 0, & i = 1, \\ \dfrac{1}{n} \hat{F}_n(x) + \dfrac{n-1}{n^2}, & i = 2, \\ \dfrac{1}{n} \hat{F}_n(x) + \dfrac{i-1}{n} + \dfrac{n-i+1}{n^2}, & i = 3, \ldots, n+1, \end{cases} \quad (3.2)
$$

for $x \in [x_{i-1}, x_i)$. It should be clear that any discrete distribution function can be represented exactly with an IFS by similar arguments.

From now on we assume $\delta_i = 0$, for all i. To produce an estimator, we should first provide a good approximator of F. So fix an $F \in \mathcal{F}([0,1])$ and choose $N+1$ ordered points $(x_1, \ldots, x_{N+1})$ such that $x_1 = 0$ and $x_{N+1} = 1$. Define the maps w_i and coefficients p_i as follows: for $i = 1, \ldots, N$,

$$p_i(F) = F(x_{i+1}) - F(x_i), \qquad w_i(x) : [0,1) \longrightarrow [x_i, x_{i+1}) = (x_{i+1} - x_i) \cdot x + x_i. \qquad (3.3)$$

The functional T_p can be denoted as T_N with this given set of maps and coefficients as it depends on the number of points and their values. For any $u \in \mathcal{F}([0,1])$, T_N can be written for $x \in \mathbb{R}$ as

$$T_N u(x) = \sum_{i=1}^{N} p_i u(w_i^{-1}(x)) = \sum_{i=1}^{N} (F(x_{i+1}) - F(x_i)) \cdot u\left(\frac{x - x_i}{x_{i+1} - x_i}\right). \qquad (3.4)$$

T_N is a contraction on $(\mathcal{F}([0,1]), d_{\sup})$ and $T_N u(x_i) = F(x_i)$, for all i.

This functional is indeed a function of F and cannot be used directly in statistical applications as F is unknown. To this end, take the points X_i to be the quantiles q_i of F, that is, choose $N+1$ points $u_1 = 0 < u_2 < \cdots < u_n < u_{N+1} = 1$ equally spaced on $[0,1]$ and set $q_i = F^{-1}(u_i)$. The function T_N becomes

$$T_N u(x) = \sum_{i=1}^{N} \frac{1}{N} u\left(\frac{x - q_i}{q_{i+1} - q_i}\right), \quad x \in \mathbb{R}, \qquad (3.5)$$

and T_N depends on F only through the quantiles q_i. Moreover in this way, it is assured that the profile of F is followed smoothly. In fact, if two quantiles q_i and q_{i+1} are relatively distant from each other, then F is slowly increasing in the interval (q_i, q_{i+1}) and vice versa. As the quantiles can be easily estimated from a sample, we now have a possible candidate for an IFS distribution function estimator. Thus, let $x_1, x_2, \ldots, x_n$ be a sample drawn from F and let $\hat{q}_i$, $i = 1, \ldots, N+1$, be the empirical quantiles of order $1/N$ such that $\hat{q}_1 = 0$ and $\hat{q}_{N+1} = 1$. Then we propose as IFS distribution function estimator the fixed point of the following IFS:

$$\hat{T}_N u(x) = \sum_{i=1}^{N} \frac{1}{N} u\left(\frac{x - \hat{q}_i}{\hat{q}_{i+1} - \hat{q}_i}\right), \quad x \in \mathbb{R}, \qquad (3.6)$$

with $u \in \mathcal{F}([0,1])$.

Remark 3.2. The resulting IFS will only be an approximation of the target distribution function F as it depends on the values of the sample quantiles that in turn are functions of the observed values $x_1, x_2, \ldots, x_n$. At the same time, the quality of the approximation increases with n, the number of sample points, as the N sample quantiles converge, in probability, to the true quantiles. In the theorems below, we discuss the relationship between the number of quantiles N and the sample size n when both are varying. Also note that the fixed point of the IFS is a fractal object that does not share necessarily the same

smoothness properties of the target distribution function F. This does not really matter as we are mostly concerned with uniform convergence. Finally, our approach is different from the current literature on distribution function estimation as we propose a fixed point instead of an L^p development or projection techniques.

3.1. Asymptotic properties of the IFS distribution function estimator.

Let $N = N_n$ be a sequence depending on the sample size n. Denote the fixed point of $\widehat{T}_{N_n}$ by $\widehat{T}^*_{N_n}$. Then $\widehat{T}^*_{N_n}$ satisfies

$$\widehat{T}^*_{N_n}(x) = \sum_{i=1}^{N_n} \frac{1}{N_n} \widehat{T}^*_{N_n}\left(\frac{x - \widehat{q}_i}{\widehat{q}_{i+1} - \widehat{q}_i}\right), \quad x \in \mathbb{R}. \tag{3.7}$$

THEOREM 3.3 (Glivenko-Cantelli, see [16]). *Let $N_n \to \infty$ as $n \to \infty$. Then for any fixed F,*

$$\limsup_{n \to \infty} \sup_{x \in \mathbb{R}} \left| \widehat{T}^*_{N_n}(x) - F(x) \right| \overset{a.s.}{=} 0. \tag{3.8}$$

Proof. Write

$$\left| \widehat{T}^*_{N_n}(x) - F(x) \right| \le \left| \widehat{T}^*_{N_n}(x) - \widehat{F}_n(x) \right| + \left| \widehat{F}_n(x) - F(x) \right|. \tag{3.9}$$

The first term can be estimated by $1/N_n$ while the second one converges to 0 almost surely by the Glivenko-Cantelli theorem for the empirical distribution function. $\square$

We can also establish a law of iterated logarithm-type result. Recall that (see [18, 19, 20]) an estimator F_n of F is said to have the *Chung-Smirnov property* if

$$\limsup_{n \to \infty} \left(\frac{2n}{\log \log n}\right)^{1/2} \sup_{x \in [0,1]} \left| F_n(x) - F(x) \right| \le 1 \quad \text{with probability 1}. \tag{3.10}$$

THEOREM 3.4. *Let $N_n = O(n^\alpha)$, $\alpha \in (1/2, 1]$. Then $\widehat{T}^*_{N_n}$ has the Chung-Smirnov property.*

Proof. The result follows because, by assumptions,

$$\limsup_{n \to \infty} \left(\frac{2n}{\log \log n}\right)^{1/2} \frac{1}{N_n} = 0. \tag{3.11}$$

$\square$

We can also establish the local asymptotic minimax optimality of our estimator when F is in a *rich family* (in the sense of [6, 11] and [12, Section 6]) of distribution functions. For any estimator F_n of the unknown distribution function F, we define the integrated mean square error as follows:

$$R_n(F_n, F) = n\mathbf{E}_F \int_0^1 \left(F_n(x) - F(x)\right)^2 \lambda(\mathrm{d}x) = \mathbf{E}_F \| \sqrt{n}(F_n - F) \|_2^2, \tag{3.12}$$

where $\lambda(\cdot)$ is a fixed probability measure on $[0,1]$ and $\mathbf{E}_F$ is the expectation under the true law F. What follows is the minimax theorem in the version given in [6]. To state the

theorem, we first define the following quantity:

$$R_0(F) = \int_0^1 F(x)(1 - F(x))\lambda(dx), \quad F \in \mathcal{F}([0,1]). \tag{3.13}$$

THEOREM 3.5 (Gill and Levit [6]). *If $\mathcal{F}$ is a rich family, then for any estimator F_n of F,*

$$\lim_{V \downarrow F_0} \liminf_{n \to \infty} \sup_{F \in V} R_n(F_n, F) \geq R_0(F_0), \tag{3.14}$$

where $V \downarrow F_0$ denotes the limit in the net of shrinking neighborhoods (with respect to the variation distance) of F_0.

The above theorem states that, for any fixed F_0, it is impossible to do better than $R_0(F_0)$ when we try to estimate F_0. The empirical distribution function $\widehat{F}_n$ is such that $R_n(\widehat{F}_n, F) = R_0(F)$ for all n and so it is asymptotically efficient in the above-mentioned sense. The result follows from the continuity of R_n in the variation distance topology (see [6]). It is almost trivial to show that also the IFS estimator is asymptotically efficient in the sense of the minimax theorem. The only condition needed is the number of maps N_n as in the law of iterated logarithm result.

THEOREM 3.6. *Let $N_n = O(n^\alpha)$, $\alpha \in (1/2, 1]$. Then $\widehat{T}^*_{N_n}$ is asymptotically efficient under the hypotheses of Theorem 3.5.*

Proof. Note that $R_0(F_0)$ is a lower bound on the asymptotic risk of $\widehat{T}^*_{N_n}$ by Theorem 3.5. Moreover,

$$\begin{aligned}
R_n(\widehat{T}^*_{N_n}, F) &= \mathbf{E}_F \|\sqrt{n}(\widehat{T}^*_{N_n} - F)\|_2^2 \\
&\leq \mathbf{E}_F \|\sqrt{n}(\widehat{T}^*_{N_n} - \widehat{F}_n)\|_2^2 + \mathbf{E}_F \|\sqrt{n}(\widehat{F}_n - F)\|_2^2 \\
&\quad + 2\mathbf{E}_F(\|\sqrt{n}(\widehat{T}^*_{N_n} - \widehat{F}_n)\|_2 \cdot \|\sqrt{n}(\widehat{F}_n - F)\|_2) \\
&\leq \frac{n}{N_n^2} + R_0(F) + 2\frac{\sqrt{n}}{N_n}\sqrt{R_0(F)}
\end{aligned} \tag{3.15}$$

by the Cauchy-Schwartz inequality applied to the cross-product of the binomial expansion. As $N_n = O(n^\alpha)$, $\alpha \in (1/2, 1]$, we have the result. Note that $\alpha > 1$ is not admissible as at most $N_n = n$ quantiles are of statistical interest. $\qquad \square$

3.2. Application to density function estimation. We will derive now a density estimator by using the Fourier analysis. This is possible because an IFS with affine maps admits a simple representation of the Fourier transform. Let

$$\phi(t): \mathbb{R} \longrightarrow \mathbb{C}, \quad \phi(t) = \int_0^1 e^{-itx} f(x)dx, \quad t \in \mathbb{R}, \tag{3.16}$$

be the Fourier transform (FT) of a continuous distribution function F with density f. The FT is such that $\phi(0) = 1$ and $|\phi(t)| \leq 1$, for all $t \in \mathbb{R}$. We denote by $\mathcal{FT}([0,1])$ the set

of all FTs associated to the distribution functions in $\mathscr{F}([0,1])$. Given two elements ϕ and ψ in $\mathscr{F}\mathscr{T}(X)$, the following metric can be defined:

$$d_{FT}(\phi,\psi) = \left(\int_{\mathbb{R}} |\phi(t) - \psi(t)|^2 t^{-2} \mathrm{d}t \right)^{1/2}. \tag{3.17}$$

The above integral is always finite (see [5]). With this metric, $(\mathscr{F}\mathscr{T}([0,1]), d_{FT})$ is a complete metric space. The FT of the IFS operator T_p, with coefficients p_i and affine maps $w_i = s_i x + a_i$, using the same arguments in [5], is given by

$$\psi(t) = B\phi(t) = \sum_{k=1}^{N} p_k e^{-ita_k} \phi(s_k t), \quad t \in \mathbb{R}, \tag{3.18}$$

where ϕ is the FT of a distribution function F and ψ is the FT of $G = T_p F$. The operator $B : \mathscr{F}T([0,1]) \to \mathscr{F}T([0,1])$ is a contractive operator and can be shown to have a unique fixed point $\bar{\phi}$ that satisfies

$$\bar{\phi}(t) = \sum_{k=1}^{N} p_k e^{-ita_k} \bar{\phi}(s_k t), \quad t \in \mathbb{R}. \tag{3.19}$$

Moreover, $\bar{\phi}$ is the FT of the fixed point of T_N. The proof of the above results is trivial (by the linearity of the operators and the maps) and can be derived from [5]. Given the FT of F, the density f can always be derived by means of

$$f(x) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} B_k e^{ikx}, \quad B_k = \int_0^1 f(x) e^{-ikx} \mathrm{d}x = \phi(k). \tag{3.20}$$

Denote by $\hat{\phi}^*$ the fixed point of the operator B where the maps and coefficients are the same as those of $\hat{T}_{N_n}$. Then a density function estimator is given by

$$\hat{f}_{FT}(x) = \frac{1}{2\pi} \sum_{k=-m}^{+m} \hat{\phi}^*(k) e^{ikx}, \tag{3.21}$$

where m can be chosen according to the following rule of thumb (see [17]):

$$\text{if } |\hat{\phi}^*(m+1)|^2 \text{ and } |\hat{\phi}^*(m+2)|^2 < \frac{2}{n+1}, \text{ then use the first } m \text{ coefficients,} \tag{3.22}$$

where n is the sample size. Note that by the properties of the Fourier expansion, it is also possible to estimate the first derivative of f by differentiating $\hat{f}_{FT}$ or have another distribution function estimator, this time a smooth one, by integrating $\hat{f}_{FT}$.

3.3. Monte Carlo analysis for small samples. We have seen in the previous sections that the IFS estimator is as efficient as the empirical distribution function in the large sample

Table 3.1. Relative efficiency of IFS-based estimator with respect to the empirical distribution function and the kernel density estimator for small and moderate sample sizes with ten thousand replications for each distribution and sample size.

n	Law	AMSE $\hat{T}_N$ w.r.t. $\hat{F}_n$ (std. err.)	SUP-NORM $\hat{T}_N$ w.r.t. $\hat{F}_n$ (std. err.)	AMSE $\hat{f}_{FT}$ w.r.t. kernel (std. err.)	MAE $\hat{f}_{FT}$ w.r.t. kernel (std. err.)
10	beta(0.9,0.1)	110.05 (66.3)	82.50 (15.3)	—	—
10	beta(0.1,0.9)	108.58 (65.5)	82.43 (15.1)	—	—
10	beta(0.1,0.1)	81.28 (14.4)	83.18 (11.2)	—	—
10	beta(2,2)	97.98 (50.9)	73.00 (16.7)	132.96 (150.4)	107.55 (57.4)
10	beta(5,5)	135.67 (72.7)	76.11 (16.5)	288.39 (381.9)	176.53 (94.4)
10	beta(5,3)	122.24 (64.7)	76.57 (17.4)	224.50 (283.9)	146.66 (76.6)
10	beta(3,5)	120.97 (63.8)	76.53 (17.3)	217.67 (257.7)	145.18 (74.0)
10	beta(1,1)	81.99 (35.6)	74.54 (17.8)	84.01 (56.9)	91.40 (36.6)
30	beta(0.9,0.1)	102.90 (40.4)	91.36 (12.4)	—	—
30	beta(0.1,0.9)	100.88 (37.3)	91.15 (12.0)	—	—
30	beta(0.1,0.1)	88.69 (13.9)	91.89 (9.9)	—	—
30	beta(2,2)	104.12 (38.0)	82.66 (13.2)	98.92 (73.4)	96.36 (37.0)
30	beta(5,5)	126.88 (51.0)	85.47 (13.5)	152.12 (109.3)	131.06 (43.1)
30	beta(5,3)	118.49 (47.5)	85.40 (13.7)	133.69 (89.6)	115.89 (33.4)
30	beta(3,5)	118.31 (47.9)	85.29 (13.8)	133.02 (87.5)	115.82 (33.6)
30	beta(1,1)	93.50 (29.1)	82.85 (13.9)	120.62 (65.6)	105.93 (28.4)
50	beta(0.9,0.1)	101.58 (31.9)	94.13 (10.9)	—	—
50	beta(0.1,0.9)	99.60 (29.7)	93.90 (10.7)	—	—
50	beta(0.1,0.1)	92.31 (13.9)	94.19 (9.1)	—	—
50	beta(2,2)	102.63 (30.8)	86.60 (11.2)	83.50 (45.4)	89.39 (26.1)

Table 3.1. Continued.

n	Law	AMSE $\hat{T}_N$ w.r.t. $\hat{F}_n$ (std. err.)	SUP-NORM $\hat{T}_N$ w.r.t. $\hat{F}_n$ (std. err.)	AMSE $\hat{f}_{FT}$ w.r.t. kernel (std. err.)	MAE $\hat{f}_{FT}$ w.r.t. kernel (std. err.)
50	beta(5,5)	119.08 (39.7)	88.91 (11.4)	100.99 (44.5)	107.73 (25.2)
50	beta(5,3)	113.30 (36.4)	88.92 (11.6)	107.62 (48.2)	106.32 (22.3)
50	beta(3,5)	112.80 (36.8)	88.79 (11.7)	106.55 (48.2)	105.99 (22.2)
50	beta(1,1)	95.88 (24.7)	86.02 (11.7)	128.60 (59.5)	115.00 (24.1)
100	beta(.9,.1)	101.07 (23.3)	96.50 (8.5)	—	—
100	beta(.1,.9)	99.38 (20.6)	96.60 (8.6)	—	—
100	beta(.1,.1)	96.51 (12.8)	96.29 (7.6)	—	—
100	beta(2,2)	101.49 (22.4)	90.86 (8.7)	82.93 (38.0)	87.93 (21.2)
100	beta(5,5)	110.66 (28.2)	92.73 (8.8)	73.08 (30.6)	91.48 (22.0)
100	beta(5,3)	107.86 (26.4)	92.86 (9.1)	93.66 (32.0)	100.50 (17.3)
100	beta(3,5)	107.90 (26.6)	92.88 (9.2)	93.69 (32.7)	100.36 (17.6)
100	beta(1,1)	97.32 (18.5)	89.86 (9.0)	126.04 (50.7)	116.07 (23.5)

case. We will now give empirical evidence that it can be even better in some situations for the small sample case. The main difference between the EDF and the IFS estimator is that the empirical distribution function is a stepwise function whilst the IFS is somewhat "smooth" in the sense that the IFS jumps have several orders of magnitude smaller then the ones of the empirical distribution function. Remember that we assume that the underlying distribution function F is a continuous one. We will also compare the performance of the density function estimator with respect to the kernel density estimator with optimal bandwidth. Table 3.1 reports the results of a Monte Carlo analysis for both distribution and density function estimation. At each Monte Carlo step, we have drawn samples of $n = 10, 30, 50, 100$ replications for several types of distribution. For each distribution and sample size n, we have done 10 000 Monte Carlo simulations. We have chosen the beta family of distribution functions because they allow very good and well-tested random number generators, different kinds of asymmetry (beta(3,5) and beta(5,3)), bell-shaped distributions with (beta(5,5)) or without (beta(2,2)) tails, and
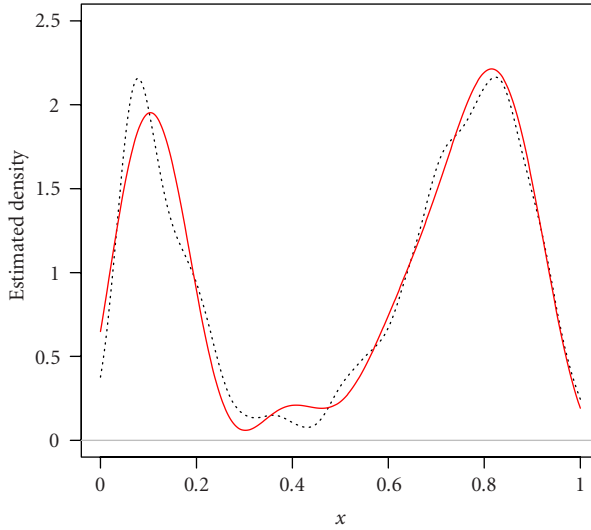
Figure 3.1. Old Faithful geyser data rescaled on [0,1]. Dotted line is the kernel density estimator (bw= 0.03, kernel=Gaussian), solid line is the IFS-Fourier expansion estimator (iterated 2 times, 26 Fourier coefficients).

also U-shaped distributions (beta(0.1,0.1)). For distribution function estimation, we have considered the d_{sup} (SUP-NORM) distance and the average mean square error (AMSE) both for $\hat{T}_N$ and $\hat{F}_n$, then we have reported in the table only the ratio of the indexes. Thus, each entry in the table reports the percentage of error of $\hat{T}_N$ with respect to $\hat{F}_n$, the error of $\hat{F}_n$ being 100. For density function estimation, we have compared $\hat{f}_{FT}$ and the kernel estimator with optimal bandwidth and Gaussian kernel. We have reported again the AMSE and the mean absolute error (MAE). The table reports the value of the index for $\hat{f}_{FT}$ with respect to the same value for the kernel density estimator, 100 being the value of the index of the last estimator. All the distances are evaluated in 512 points equally spaced on [0,1]. For the distributions like beta(0.1,0.1) that have unbounded density in the endpoints of the interval, both kernel and our Fourier estimators become really unstable, so we decided to omit the value of the index in the tables. The software used is R [8], freely available on http://cran.R-project.org, using a beta package IFS available as an additional contributed package.

For the estimator $\hat{T}_N$, we have chosen to take $N_n = n/2$ that fits the conditions of Theorems 3.4 and 3.6 on the law of iterated logarithm and asymptotic efficiency.

In the small sample size case, $n = 10, 30$, it can be noticed that $\hat{T}_N$ is sometimes better (from 10% to 20%) than the empirical distribution function in the sup-norm distance. This behavior is not shown by the density function estimator $\hat{f}_{FT}$. For moderate sample sizes, $n = 50, 100$, the distance between $\hat{T}_N$ and $\hat{F}_n$ decreases and consequently the gain in using the IFS estimator is not so evident (on the average 5% to 10% in sup-norm distance). The density estimator performs a little better in this case, but the associated standard errors are too high to lead to sharp conclusions.

4. 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: are there maps other than the ones used in T_N that can improve the performance of the corresponding IFS estimator? 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 [4], this problem is incidentally touched on, but not in a statistical context.

We have tested our density estimator on real data and we have chosen the Old Faithful geyser data. This classical textbook data set is used to show the power of the kernel estimator in discriminating subpopulations by adjusting the bandwidth. We have used the automatic procedure to select the number of Fourier coefficients as explained previously. Figure 3.1 shows that this ability of discriminating subpopulation curves is maintained by the $\hat{f}_{FT}$ estimator.

5. Acknowledgments

References

[1]   L. Arnold and H. Crauel, *Iterated function systems and multiplicative ergodic theory*, Diffusion Processes and Related Problems in Analysis, Vol. II (Charlotte, NC, 1990), Stochastics Flows (Pinsky M. A. and Vihstutz V., eds.), Progress in Probability, vol. 27, Birkhäuser Boston, Massachusetts, 1992, pp. 283–305.

[2]   M. F. Barnsley and S. Demko, *Iterated function systems and the global construction of fractals*, Proc. Roy. Soc. London Ser. A **399** (1985), no. 1817, 243–275.

[3]   J. H. Elton and M. Piccioni, *Iterated function systems arising from recursive estimation problems*, Probab. Theory Related Fields **91** (1992), no. 1, 103–114.

[4]   B. Forte and E. R. Vrscay, *Solving the inverse problem for function/image approximation using iterated function systems. I. Theoretical basis*, Fractals **2** (1994), no. 3, 325–334.

[5]   ——, *Inverse problem methods for generalized fractal transforms*, Fractal Image Encoding and Analysis (Y. Fisher, ed.), NATO ASI Series F, vol. 159, Springer Verlag, Heidelberg, 1998.

[6]   R. D. Gill and B. Y. Levit, *Applications of the Van Trees inequality: a Bayesian Cramér-Rao bound*, Bernoulli **1** (1995), no. 1-2, 59–79.

[7]   J. E. Hutchinson, *Fractals and self-similarity*, Indiana Univ. Math. J. **30** (1981), no. 5, 713–747.

[8]   R. Ihaka and R. Gentleman, *R: A Language for Data Analysis and Graphics.*, J. Comput. Graph. Statist. **5** (1996), 299–314.

[9]   A. A. Kwiecinska and W. Slomczynski, *Random dynamical systems arising from iterated function systems with place-dependent probabilities*, Statist. Probab. Lett. **50** (2000), no. 4, 401–407.

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

[11] B. Y. Levit, *Infinite-dimensional informational inequalities*, Theory Probab. Appl. **23** (1978), 371–377.

[12] P. W. Millar, *Asymptotic minimax theorems for the sample distribution function*, Z. Wahrsch. Verw. Gebiete **48** (1979), no. 3, 233–252.

[13]   ⸻ , *The minimax principle in asymptotic statistical theory*, Eleventh Saint Flour Probability Summer School—1981 (Saint Flour, 1981), Lecture Notes in Math., vol. 976, Springer, Berlin, 1983, pp. 75–265.

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

[15]   R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Fundamental Principles of Mathematical Sciences, vol. 317, Springer, Berlin, 1998.

[16]   G. R. Shorack and J. A. Wellner, *Empirical Processes with Applications to Statistics*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, New York, 1986.

[17]   M. E. Tarter and M. D. Lock, *Model-Free Curve Estimation*, Monographs on Statistics and Applied Probability, vol. 56, Chapman & Hall, New York, 1993.

[18]   B. B. Winter, *Strong uniform consistency of integrals of density estimators*, Canad. J. Statist. **1** (1973), no. 2, 247–253.

[19]   ⸻ , *Convergence rate of perturbed empirical distribution functions*, J. Appl. Probab. **16** (1979), no. 1, 163–173.

[20]   J. E. Yukich, *A note on limit theorems for perturbed empirical processes*, Stochastic Process. Appl. **33** (1989), no. 1, 163–173.

Stefano Maria Iacus: Department of Economics, Business and Statistics, Faculty of Political Science, University of Milan, 20122 Milan, Italy

E-mail address: stefano.iacus@unimi.it

Davide La Torre: Department of Economics, Business and Statistics, Faculty of Political Science, University of Milan, 20122 Milan, Italy

E-mail address: davide.latorre@unimi.it

Special Issue on
Intelligent Computational Methods for Financial Engineering

Call for Papers

As a multidisciplinary field, financial engineering is becoming increasingly important in today's economic and financial world, especially in areas such as portfolio management, asset valuation and prediction, fraud detection, and credit risk management. For example, in a credit risk context, the recently approved Basel II guidelines advise financial institutions to build comprehensible credit risk models in order to optimize their capital allocation policy. Computational methods are being intensively studied and applied to improve the quality of the financial decisions that need to be made. Until now, computational methods and models are central to the analysis of economic and financial decisions.

However, more and more researchers have found that the financial environment is not ruled by mathematical distributions or statistical models. In such situations, some attempts have also been made to develop financial engineering models using intelligent computing approaches. For example, an artificial neural network (ANN) is a nonparametric estimation technique which does not make any distributional assumptions regarding the underlying asset. Instead, ANN approach develops a model using sets of unknown parameters and lets the optimization routine seek the best fitting parameters to obtain the desired results. The main aim of this special issue is not to merely illustrate the superior performance of a new intelligent computational method, but also to demonstrate how it can be used effectively in a financial engineering environment to improve and facilitate financial decision making. In this sense, the submissions should especially address how the results of estimated computational models (e.g., ANN, support vector machines, evolutionary algorithm, and fuzzy models) can be used to develop intelligent, easy-to-use, and/or comprehensible computational systems (e.g., decision support systems, agent-based system, and web-based systems)

This special issue will include (but not be limited to) the following topics:

- **Computational methods**: artificial intelligence, neural networks, evolutionary algorithms, fuzzy inference, hybrid learning, ensemble learning, cooperative learning, multiagent learning

- **Application fields**: asset valuation and prediction, asset allocation and portfolio selection, bankruptcy prediction, fraud detection, credit risk management

- **Implementation aspects**: decision support systems, expert systems, information systems, intelligent agents, web service, monitoring, deployment, implementation

Authors should follow the Journal of Applied Mathematics and Decision Sciences manuscript format described at the journal site http://www.hindawi.com/journals/jamds/. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at http://mts.hindawi.com/, according to the following timetable:

Manuscript Due	December 1, 2008
First Round of Reviews	March 1, 2009
Publication Date	June 1, 2009

Guest Editors

Lean Yu, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong; yulean@amss.ac.cn

Shouyang Wang, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; sywang@amss.ac.cn

K. K. Lai, Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong; mskklai@cityu.edu.hk