



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

**Nonlinear
Analysis**

Real World Applications

Nonlinear Analysis: Real World Applications 6 (2005) 858–873

www.elsevier.com/locate/na

A comparative simulation study on the IFS distribution function estimator

Stefano M. Iacus*, Davide La Torre

Department of Economics, University of Milan, Via Conservatorio 7, I-20122 Milan, Italy

Received 10 August 2004; accepted 15 November 2004

Abstract

In this paper, we do a comparative simulation study of the standard empirical distribution function estimator versus a new class of nonparametric estimators of a distribution function F , called the iterated function system (IFS) estimator. The target distribution function F is supposed to have compact support. The IFS estimator of a distribution function F is considered as the fixed point of a contractive operator T defined in terms of a vector of parameters p and a family of affine maps \mathcal{W} which can be both dependent on the sample (X_1, X_2, \dots, X_n) . Given \mathcal{W} , the problem consists in finding a vector p such that the fixed point of T is “sufficiently near” to F . It turns out that this is a quadratic constrained optimization problem that we propose to solve by penalization techniques. Analytical results prove that IFS estimators for F are asymptotically equivalent to the empirical distribution function (EDF) estimator. We will study the relative efficiency of the IFS estimators with respect to the empirical distribution function for small samples via the Monte Carlo approach.

For well-behaved distribution functions F and for a particular family of the so-called wavelet maps the IFS estimators can be dramatically better than the empirical distribution function in the presence of missing data, i.e. when it is only possible to observe data on subsets of the whole support of F .

© 2005 Elsevier Ltd. All rights reserved.

MSC: 41A; 65D15; 62G05

Keywords: Iterated function systems; Distribution function estimation; Nonparametric estimation; Missing data; Density estimation

* Corresponding author. Tel.: +39 02 50321461; fax: +39 02 50321505.

E-mail address: stefano.iacus@unimi.it (S.M. Iacus).

1. Introduction

Let X_1, X_2, \dots, X_n be an i.i.d. sample drawn from a random variable X with unknown distribution function F with compact support $[0, 1]$. The empirical distribution function (EDF)

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

is one commonly used estimator of the unknown distribution function F (here χ_A is the indicator function of the set A). The EDF has an impressive set of good statistical properties such as it is first-order efficient in the minimax sense (see [2,4,8,14,15]). 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 [9,10] and Efromovich [5] are two such examples. It is rather curious that a step-wise function can be such a good estimator and, in fact, Efromovich [5] shows that, for the class of analytic functions, for small sample sizes, the EDF is not the best estimator. In this paper, we study the properties of a new class of distribution function estimators based on iterated function systems (IFSs) introduced by the authors in a previous work [12]. IFSs have been introduced in [1,11]. The main idea on which this method is based consists of considering the estimation of F as the fixed point of a contraction T on a complete metric space. The operator T is defined in terms of a family of affine maps \mathcal{W} and a vector of parameters p . For a given family \mathcal{W} , T depends only on the choice p . The idea, known as *inverse approach* (see Section 2), is to determine p by solving a constrained quadratic optimization problem built in terms of sample moments. The nature of affine maps allow to derive easily the Fourier transform of F and, when available, an explicit formula for the density of F via anti-Fourier transform. In this way, given \mathcal{W} and p we have at the same time estimators for the distribution, characteristic and density functions.

The paper is organized as follows. In Section 2 some details concerning IFS techniques are recalled. In Section 3 numerical results and comparisons with classical estimators are shown for small samples via Monte Carlo analysis. Finally, we show an application of these estimators when the empirical distribution function (or the kernel density estimator) cannot be applied. We will consider situations of missing data when, for example, the data can only be observed on some windows of the support of F . This can be the case of directional data analysis when, for some reason, instruments are not able for technical or physical reasons to collect data in same range of angles say A and B , $A, B \subseteq [0, 2\pi]$. For x in A or B the EDF will be constant and, at the same time, the kernel density estimator will estimate a plurimodal distribution for these data. In this case we will show some examples in which the IFS estimator works better than classical estimators.

2. An IFS estimator

The theory of distribution function approximation via IFSs that we will use to derive estimators is due to Forte and Vrscay [6]. Results from this section, are from the cited

authors. Let $\mathcal{M}(X)$ be 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 (if not otherwise remarked, in this paper X will be $[0, 1]$ and d the Euclidean metric even if all the results can be extended to $X = [\alpha, \beta]$).

In the IFSs literature the following *Hutchinson metric* plays a crucial role:

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

where

$$\text{Lip}(X) = \{f : X \rightarrow \mathbb{R}, |f(x) - f(y)| \leq d(x, y), x, y \in X\},$$

thus $(\mathcal{M}(X), d_H)$ is a complete metric space (see [11]).

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

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

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

$$c = \max_{1 \leq i \leq N} |b_i| < 1.$$

Consider the following (usually called *Markov*) operator $M : \mathcal{M}(X) \rightarrow \mathcal{M}(X)$ 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 [11] it was shown that M is a contraction mapping on $(\mathcal{M}(X), d_H)$ i.e. for all $\mu, \nu \in \mathcal{M}(X)$, $d_H(M\mu, M\nu) \leq c d_H(\mu, \nu)$. 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. Associated to each measure $\mu \in \mathcal{M}(X)$, there exists a distribution function F . In terms of this the previous operator M can be rewritten as

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

2.1. Minimization approach

For affine IFSs there exists a simple and useful relation between the moments of probability 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 $f : X \rightarrow \mathbb{R}$,

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

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

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

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

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

Let μ and $\mu^{(j)} \in \mathcal{M}(X)$, $j = 1, 2, \dots$, with associated moments of any order g_k and

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

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

1. $g_k^{(j)} \rightarrow g_k$,
2. $\forall f \in \mathbf{C}(X), \int_X f d\mu^{(j)} \rightarrow \int_X f d\mu$, (weak* convergence),
3. $d_H(\mu^{(j)}, \mu) \rightarrow 0$.

(here $\mathbf{C}(X)$ is the space of continuous functions on X). This result gives a way for finding out an appropriate set of maps and probabilities by solving the so-called problem of moment matching. With the solution in hand, 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 [1]).

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

Collage theorem. *Let (Y, d_Y) be a complete metric space. Given $y \in Y$, 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 [6,7] that we will use to build one of the IFS estimators is that the inverse problem can be reduced 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

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

be the simplex of probabilities. Let $\mathbf{w} = (w_1, w_2, \dots, w_N)$, $N = 1, 2, \dots$, be subsets of $\mathcal{W} = \{w_1, w_2, \dots\}$ the infinite set of affine contractive maps on $X = [0, 1]$ and let \mathbf{g} be the set 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 $v_N = M\mu$, with associated moment vector of any order \mathbf{h}_N . The collage distance between the moment vector of μ and v_N

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

is a continuous function and attains an absolute minimum value Δ_{\min}^N on Π^N , where

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

Moreover, $\Delta_{\min}^N \rightarrow 0$ as $N \rightarrow \infty$. Thus, the collage distance can be made arbitrarily small by choosing a suitable number of maps and probabilities.

The above inverse problem can be posed as a quadratic programming one in the following notation:

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

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

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

$$h_k = \sum_{i=1}^N A_{ki} p_i, \quad k = 1, 2, \dots \quad \text{where} \quad A_{ki} = \sum_{j=0}^{\infty} \binom{k}{j} b_i^j a_i^{k-j} g_j. \tag{4}$$

Thus,

$$(QP) \quad S(\mathbf{p}) = \mathbf{p}' Q \mathbf{p} + \mathbf{B}' \mathbf{p} + C,$$

where

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

$$B_i = -2 \sum_{k=1}^{\infty} \frac{g_k}{k^2} A_{ki}, \quad i = 1, 2, \dots, N \quad \text{and} \quad C = \sum_{k=1}^{\infty} \frac{g_k^2}{k^2}. \tag{5}$$

The above series 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 .

In [6] the following two sets of wavelet-type maps are proposed for solving the inverse problem. Fixed and index $i^* \in \mathbb{N}$, define

$$\gamma_{ij} = \frac{x + (j - 1)}{2^i}, \quad i = 1, 2, \dots, i^*, \quad j = 1, 2, \dots, 2^i$$

and

$$\eta_{ij} = \frac{x + (j - 1)}{i}, \quad i = 2, \dots, i^*, \quad j = 2, \dots, i.$$

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:

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

and

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

respectively.

2.2. Fourier analysis results

We now recall some results concerning Fourier analysis of IFS operators [7]. These will be very useful for doing density estimation in the presence of missing data. We recall that, without loss of generality, the support of the measures is $X = [0, 1]$.

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

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

with the well-known properties $\phi(0) = 1$ and $|\phi(t)| \leq 1, \forall t \in \mathbb{R}$. It can be shown that the space of characteristic functions $\mathcal{F}\mathcal{T}(X)$ is complete with an opportune metric (see again [7]). Thus, given an N -maps affine IFS(\mathbf{w}, \mathbf{p}) it is possible to define a new linear operator $B : \mathcal{F}\mathcal{T}(X) \rightarrow \mathcal{F}\mathcal{T}(X)$ whose unique fixed point reads as

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

This $\bar{\phi}(t)$ is the FT of the fixed point of the N -maps affine IFS(\mathbf{w}, \mathbf{p}). Now (see e.g. [16]), suppose that the target distribution F admits a density f . It is possible to write the density f via Fourier expansion. In fact,

$$\phi(t) = \int_0^1 f(x) e^{-itx} dx = \int_0^1 e^{-itx} dF(x),$$

thus

$$f(x) = \frac{1}{2\pi} \sum_{k=-\infty}^{+\infty} B_k e^{ikx} \quad \text{where } B_k = \phi(k).$$

3. Estimation by IFSs

After showing some classical results concerning IFS methodology, we are now interested in applying this procedure to estimation. To do this, first of all we need to specify some computational details concerning:

- (a) the solution of the inverse problem by quadratic optimization and
- (b) the choice of the affine maps.

In the previous section we saw that the inverse problem can be reduced to (QP). However, when practical cases are considered the series in the function S have to be truncated and so the matrix Q could not be definite positive. Standard numerical procedures for the minimization of constrained quadratic optimization problems involving positive definite quadratic forms cannot be used in this context. To solve this problem an approach is to build the following penalized function L_λ :

$$L_\lambda(\mathbf{p}) = \mathbf{p}^t Q \mathbf{p} + \mathbf{B}^t \mathbf{p} + C + \lambda \left(1 - \sum_{i=1}^N p_i \right)^2$$

and then to study the following problem:

$$(LOP) \quad \min L_\lambda(\mathbf{p}), \quad 0 \leq p_i \leq 1.$$

Trivially, an optimizer \mathbf{p}^* of (LOP) such that $\sum_{i=1}^N p_i^* = 1$ is also an optimizer for the problem

$$(OP) \quad \min S(\mathbf{p}), \quad \mathbf{p} \in \Pi^N$$

To solve (LOP) numerically, we use the L-BFGS-B method due to Byrd et al. [3] which allows to minimize a nonlinear function with box constraints, i.e. when each variable can be given a lower and/or upper bound. The initial value of this procedure must satisfy the constraints. This uses a limited-memory modification of the BFGS quasi-Newton method. The method “BFGS” is a quasi-Newton method (also known as a variable metric algorithm).

As for point (b): in the previous section we recalled two types of wavelet maps proposed by Forte and Vrscay. In [12] we proposed the following quantile-based maps:

$$\mathcal{Q}_1 = \{w_i(x) = (q_{i+1} - q_i)x + q_i, i = 1, 2, \dots, N\},$$

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 $[0, 1]$. With these maps, it has been shown that, there is no need to use a moment matching approach. In particular, given $p_i = 1/N$, the IFSs turn out to be a smoother of the EDF and so it has nice small sample and asymptotic statistical properties (see cited reference) even for noncompact support distribution functions F . Here we will also mix the quantile information with the moment matching idea. To distinguish the two cases (fixed $p_i = 1/N$ or p solution of (QP) in the following we will use the notation \mathcal{Q}_1 and \mathcal{Q}_2 .

Suppose we were to have an i.i.d. (independent and identically distributed) sample on n observations X_1, X_2, \dots, X_n with common unknown distribution function F with compact

support on $[0, 1]$ which has all the moments up to order M . An IFS estimator of F is the fixed point of the functional TF where the N maps are chosen in advance and the p_i are the solution of the (QP) quadratic programming problem where in A_{ik} , B_i and C we replace the true moments g_k with the sample moments m_k (see Eqs. (5) and (4)), $k = 0, 1, \dots, M$ for a fixed M and we consider the first M terms of the series involved. Sample moments are calculated as follows:

$$m_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k \geq 0$$

Given the solution of (QP), we have an estimator for F and an estimator for the characteristic function of F , say $\hat{\phi}$. Suppose that F possesses a density f then we also have a (Fourier) density estimator for f

$$\begin{aligned} \hat{f}(x) &= \frac{1}{2\pi} \sum_{k=-m}^{+m} \hat{B}_k e^{ikx} \\ &= \frac{1}{2\pi} + \frac{1}{\pi} \sum_{i=1}^m \left\{ \text{Re}(\hat{B}_k) \cos(kx) - \text{Im}(\hat{B}_k) \sin(kx) \right\}, \end{aligned}$$

where $\hat{B}_k = \hat{\phi}(k)$ and m , the number of Fourier terms, is chosen in the usual way, i.e.

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

(see again [16]).

3.1. Monte Carlo analysis

Tables 1 and 2 show some comparisons between the empirical cumulative distribution function \hat{F}_n and the IFS estimator, say \hat{T}_N , for some target distributions F , in terms of average mean square error (AMSE) and sup-norm (SUP) distance. For assessing the quality of the IFS estimator, we choose target distribution functions in the family of beta distribution functions. This allows to test the estimator against a wide family of shapes of distribution functions, i.e. symmetric and asymmetric, heavy tails or no tails, U-shaped and uniform. In particular we focus our attention on the following beta family members: beta(0.9,0.1), beta(0.1,0.9), beta(0.1,0.1), beta(2,2), beta(5,5), beta(3,5), beta(5,3) and beta(1,1). For each target distribution we run 100 simulations for different sample sizes, i.e. we draw 100 samples of size $n = 10, 20$ or 30 (small sample sizes, respectively, $n = 50, 100, 250$ for moderate sample sizes) from the corresponding beta distribution and we evaluate both the IFS estimator and the empirical distribution function on each sample. Then we measure, for each sample, the quality of the estimation using the AMSE and the SUP distance defined as follows:

$$\text{AMSE}_1 = \frac{1}{512} \sum_{i=1}^{512} (T_N(z_i) - F(z_i))^2$$

Table 1

Relative efficiency of IFS estimators with different set of maps $\mathcal{W}_1, \mathcal{W}_2, \mathcal{Q}_1$ and \mathcal{Q}_2 with respect to the empirical distribution function (i.e. IFS/EDF)

Parameters		AMSE				SUP			
<i>n</i>	Law	\mathcal{W}_1	\mathcal{W}_2	\mathcal{Q}_1	\mathcal{Q}_2	\mathcal{W}_1	\mathcal{W}_2	\mathcal{Q}_1	\mathcal{Q}_2
10	beta(0.9,0.1)	81.08	77.05	203.53	149.68	85.76	75.44	110.11	110.81
10	beta(0.1,0.9)	211.78	2024.68	203.39	258.88	175.32	441.32	114.51	161.55
10	beta(0.1,0.1)	118.27	416.17	182.88	104.07	114.87	192.94	119.57	106.56
10	beta(2,2)	56.47	80.53	67.68	112.46	53.31	69.24	70.36	98.21
10	beta(5,5)	52.77	57.90	110.35	152.29	53.99	54.83	81.61	125.67
10	beta(3,5)	55.95	71.07	99.92	142.52	51.93	60.58	81.72	116.79
10	beta(5,3)	52.50	57.34	91.75	131.37	51.74	52.47	77.97	109.84
10	beta(1,1)	73.35	119.04	79.01	102.04	65.63	90.40	70.89	90.85
20	beta(0.9,0.1)	94.69	85.25	201.85	169.78	90.30	79.92	105.02	123.28
20	beta(0.1,0.9)	388.83	4183.36	203.70	195.36	257.13	612.55	109.10	122.99
20	beta(0.1,0.1)	154.1	690.08	125.35	97.53	139.65	255.26	103.56	99.28
20	beta(2,2)	61.46	93.37	85.46	95.49	55.34	73.95	84.42	91.38
20	beta(5,5)	54.31	52.89	105.84	131.84	53.76	48.73	85.85	106.27
20	beta(3,5)	60.42	67.33	93.30	118.51	55.98	60.88	85.39	101.16
20	beta(5,3)	53.82	57.72	92.26	114.84	53.46	52.20	85.23	102.85
20	beta(1,1)	95.93	89.79	71.66	154.54	63.20	106.95	81.56	82.54
30	beta(0.9,0.1)	107.46	90.27	195.39	143.00	101.83	81.05	108.59	109.85
30	beta(0.1,0.9)	540.73	6462.03	190.82	213.45	107.80	137.26	314.53	759.57
30	beta(0.1,0.1)	112.66	97.04	233.50	1342.44	186.70	356.91	103.39	99.98
30	beta(2,2)	60.30	92.92	88.90	96.88	53.71	72.06	84.92	89.11
30	beta(5,5)	62.04	56.07	100.26	121.41	60.08	51.82	89.26	100.16
30	beta(3,5)	70.31	76.90	93.02	108.76	61.68	66.29	86.36	95.24
30	beta(5,3)	55.78	56.85	92.10	102.02	55.56	51.21	88.20	94.75
30	beta(1,1)	71.88	211.28	94.36	88.17	63.15	121.23	83.74	83.40

Based on 100 Monte Carlo simulations for each target beta(α, β) distribution. Small sample sizes $n = 10, 20$ and 30.

and the SUP distance

$$SUP_1 = \max_{i=1, \dots, 512} |T_N(z_i) - F(z_i)|,$$

where $z_i, i = 1, \dots, 512$ are equally spaced points on $[0, 1], F$ is one of the beta distributions and T_N is the IFS estimator. The same applies for the empirical distribution function, i.e. we calculate

$$AMSE_2 = \frac{1}{512} \sum_{i=1}^{512} (\hat{F}_n(z_i) - F(z_i))^2$$

and the SUP distance

$$SUP_2 = \max_{i=1, \dots, 512} |\hat{F}_n(z_i) - F(z_i)|.$$

Table 2

Relative efficiency of IFS estimators with different set of maps $\mathcal{W}_1, \mathcal{W}_2, \mathcal{Q}_1$ and \mathcal{Q}_2 with respect to the empirical distribution function (i.e. IFS/EDF)

Parameters		AMSE				SUP			
<i>n</i>	Law	\mathcal{W}_1	\mathcal{W}_2	\mathcal{Q}_1	\mathcal{Q}_2	\mathcal{W}_1	\mathcal{W}_2	\mathcal{Q}_1	\mathcal{Q}_2
50	beta(0.9,0.1)	132.67	115.10	163.33	129.24	109.18	88.77	103.37	101.90
50	beta(0.1,0.9)	1044.12	12573.16	181.99	180.42	421.49	991.33	104.37	123.39
50	beta(0.1,0.1)	306.49	1917.23	105.68	97.27	214.27	430.63	100.13	98.04
50	beta(2,2)	63.03	106.56	95.35	95.66	58.39	80.00	89.42	89.36
50	beta(5,5)	68.94	60.19	102.22	114.92	66.77	55.49	91.86	97.40
50	beta(3,5)	79.98	93.80	96.20	102.32	66.76	77.57	91.39	93.76
50	beta(5,3)	63.13	62.21	93.59	98.47	62.04	55.95	90.66	93.19
50	beta(1,1)	73.47	304.41	97.24	92.19	62.69	150.39	87.38	86.30
100	beta(0.9,0.1)	195.54	158.80	140.55	108.27	138.93	105.31	102.25	99.07
100	beta(0.1,0.9)	1557.30	20324.60	135.45	125.94	536.84	1267.81	103.87	106.05
100	beta(0.1,0.1)	554.11	3918.62	102.67	98.29	304.59	625.75	99.10	98.04
100	beta(2,2)	61.63	165.60	95.58	97.46	57.18	98.50	92.11	93.09
100	beta(5,5)	87.97	67.79	99.28	108.21	78.94	60.96	94.83	96.52
100	beta(3,5)	111.30	134.54	100.68	103.31	79.59	100.20	95.35	95.72
100	beta(5,3)	61.03	57.19	97.28	101.32	65.97	55.08	94.14	95.42
100	beta(1,1)	67.91	558.50	97.71	94.87	58.71	201.10	90.83	89.97
250	beta(0.9,0.1)	338.72	255.23	115.25	101.55	180.29	131.97	100.68	99.43
250	beta(0.1,0.9)	3979.61	50448.13	117.81	105.37	874.65	2045.15	100.82	99.73
250	beta(0.1,0.1)	1345.72	10051.20	100.60	98.97	480.12	977.30	99.16	98.73
250	beta(2,2)	79.01	275.93	98.59	98.30	67.14	132.87	95.50	95.24
250	beta(5,5)	163.68	99.35	99.07	100.54	111.38	78.48	96.40	96.83
250	beta(3,5)	212.17	228.58	99.45	99.69	113.70	142.21	96.57	96.32
250	beta(5,3)	91.32	73.31	99.05	99.20	88.87	67.13	96.84	97.24
250	beta(1,1)	69.03	1165.61	99.47	98.46	61.07	293.58	94.88	94.55

Based on 100 Monte Carlo simulations for each target beta(α, β) distribution. Moderate to large sample sizes $n = 50, 100, 250$.

Finally, as usual in statistics, the comparison between estimators is measured in terms of relative efficiency, which means that we actually consider the following ratios:

$$AMSE = 100 * \frac{AMSE_1}{AMSE_2}, \quad SUP = 100 * \frac{SUP_1}{SUP_2}.$$

Values lower than 100 in AMSE and SUP mean that the IFS estimator T_N is more efficient than the empirical distribution function \hat{F}_n in estimating the unknown F . Tables 1 and 2 report the average of AMSE and SUP over 100 replications for different sample sizes and for T_N using four different sets of maps $\mathcal{W}_1, \mathcal{W}_2, \mathcal{Q}_1$ and \mathcal{Q}_2 (for wavelet maps we make use of the first 30 sample moments). It is possible to notice that the IFS estimator based on maps \mathcal{W}_1 has good properties for symmetric bell-shaped distributions and distributions with not so heavy tails (see also Fig. 2). From our numerical results the asymptotic equivalence between IFSs and EDF is also evident when quantile maps are used. Note that we used

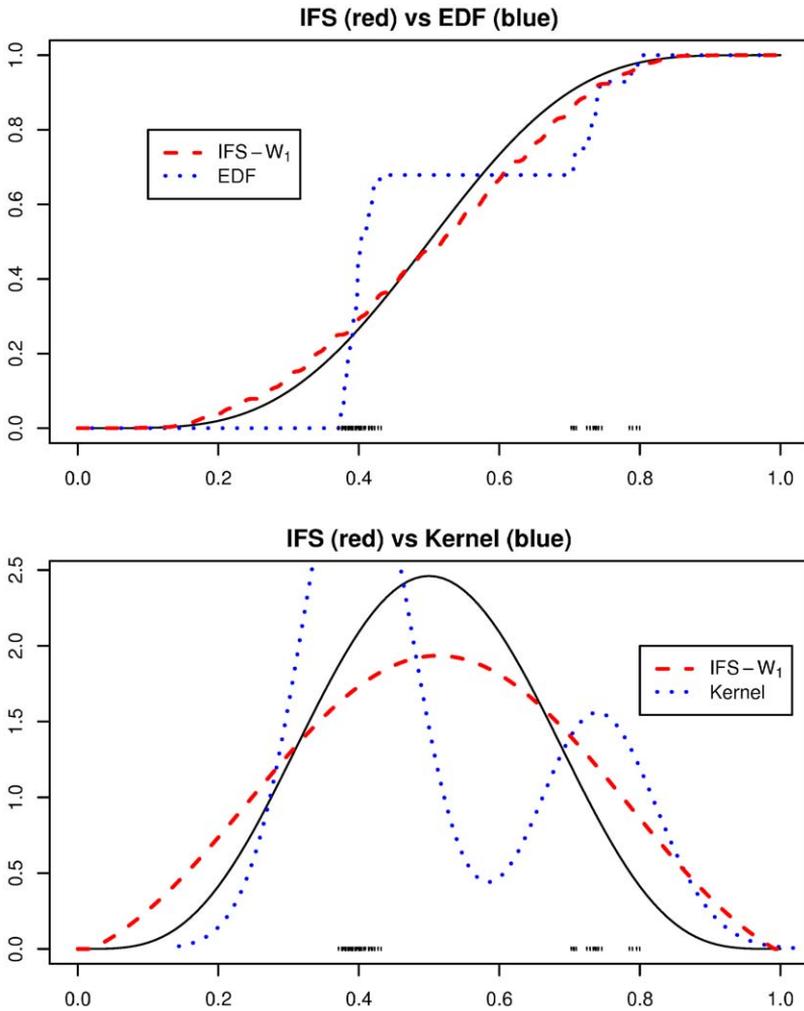


Fig. 1. Data from a Beta(2,2) distribution when only the observations in $(0.1, 0.15) \cup (0.37, 0.43) \cup (0.7, 0.8)$ are available to the observer all the others being truncated by the instrument. The observations are marked as vertical ticks. The IFS estimator with \mathcal{W}_1 maps seems to be able to reconstruct the underlying distribution and density function, whilst, for obvious reasons both the edf and the kernel estimators fail. Notice that the arbitrary choice of the window of observation can be changed without substantial loss or gain. In this example the relative efficiency (IFS/EDF) is 7% for the AMSE and 23% for the SUP-norm.

62 maps for \mathcal{W}_1 , 28 maps for \mathcal{W}_2 and $n/2$ quantiles for the quantile maps \mathcal{Q}_1 and \mathcal{Q}_2 (Figs. 1–4).

3.2. What if data are missing?

Suppose now that for some reason, the sample of n observations from F are in fact a subset of a larger sample, of unknown size. In practice we do not observe the data on the

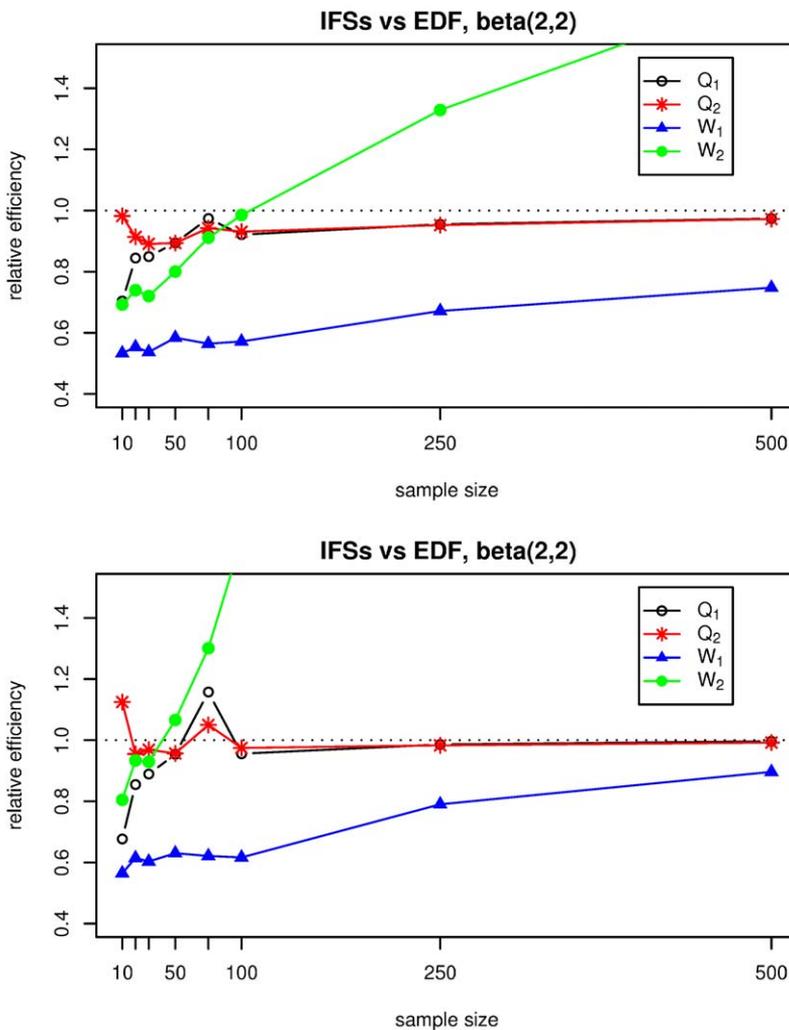


Fig. 2. Relative efficiency of IFS estimator for different set of maps \mathcal{W}_1 , \mathcal{W}_2 , \mathcal{Q}_1 and \mathcal{Q}_2 with respect to the empirical distribution function. Based on 100 Monte Carlo simulations. SUP-norm up, AMSE bottom. Values lower than 1.0 mean that the IFS estimator performs better than the EDF.

whole support of F but only on some windows. This sample reduction is due to some sort of censoring. So we are in the presence of missing data when we do not know how many data are missing and where exactly they were missed, i.e. we are not in a classical censoring setup. A motivation for this scheme of (non)-observation is the following: suppose one wants to estimate the distribution of the angle of the wind registered by some instruments in degrees $(0,360)$. For some reason, data from angles $(15,37)$ and $(62,79)$ are missing for technical failures or physical obstacles. In this case the empirical distribution function will be flat on

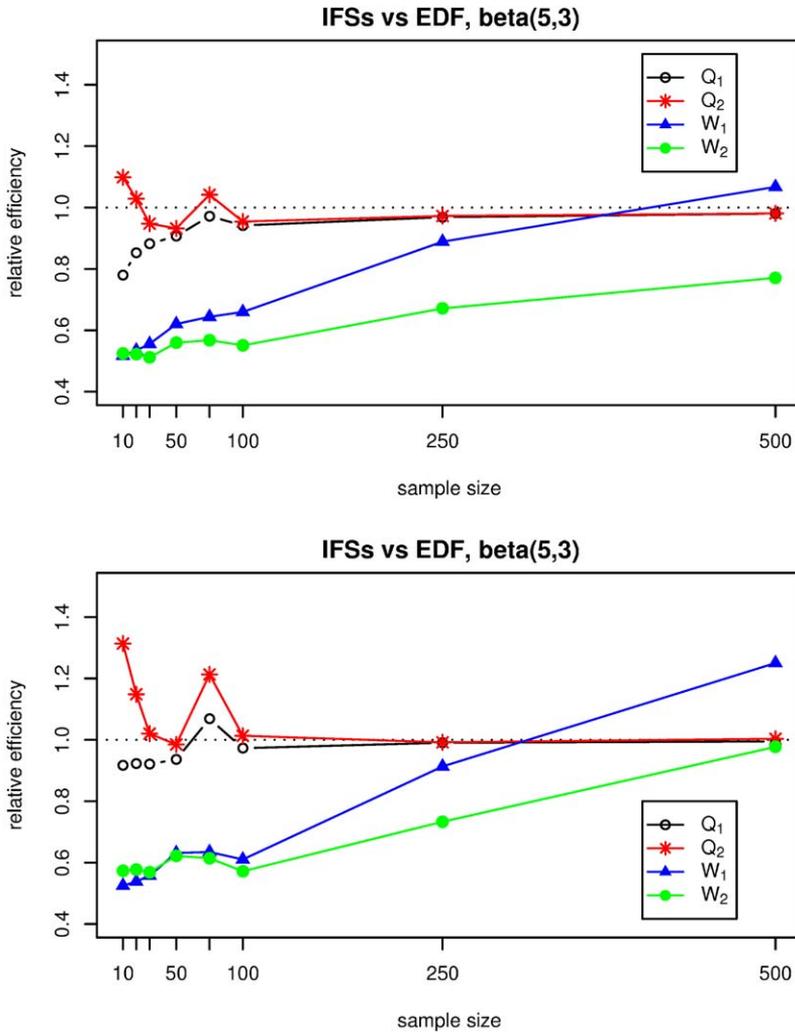


Fig. 3. Relative efficiency of IFS estimator for different set of maps \mathcal{W}_1 , \mathcal{W}_2 , \mathcal{Q}_1 and \mathcal{Q}_2 with respect to the empirical distribution function. Based on 100 Monte Carlo simulations. SUP-norm up, AMSE bottom. Values lower than 1.0 mean that the IFS estimator performs better than the EDF.

these windows and a kernel density estimator will probably show a multimodal shape. This is due to the fact that quantile estimation is inappropriate in this context. At the same time, moment estimation tends to be more robust, in particular if the distribution is symmetric. We only report a graphical example of what can happen. Fig. 1 is about a sample from a Beta(2,2) distribution when only the observations in $(0.1, 0.15) \cup (0.37, 0.43) \cup (0.7, 0.8)$ are available to the observer all the others being truncated by the instrument (we have chosen this interval by hazard). The IFS estimator with \mathcal{W}_1 maps seems to be able to reconstruct

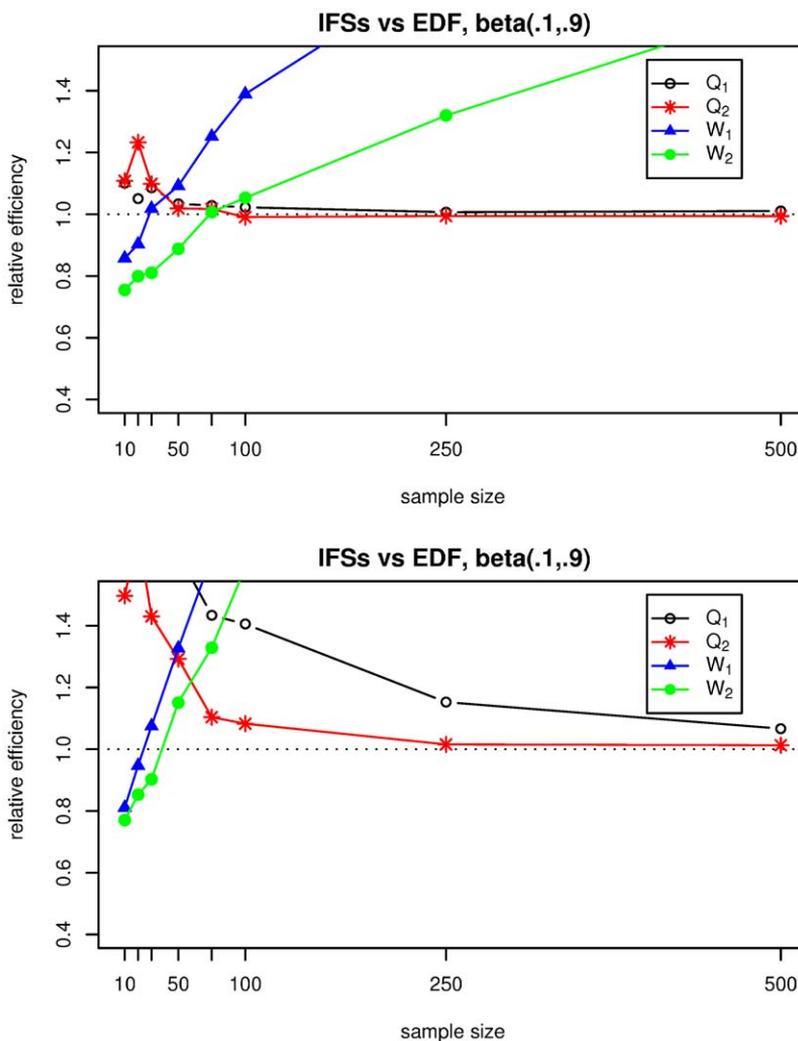


Fig. 4. Relative efficiency of IFS estimator for different set of maps \mathcal{W}_1 , \mathcal{W}_2 , \mathcal{Q}_1 and \mathcal{Q}_2 with respect to the empirical distribution function. Based on 100 Monte Carlo simulations. SUP-norm up, AMSE bottom. Values lower than 1.0 mean that the IFS estimator performs better than the EDF.

the underlying distribution and density function, while, for obvious reasons, both the EDF and the kernel estimators fail. In this example the relative efficiency (IFS/EDF) is 7% for the AMSE and 23% for the SUP-norm which is dramatically better than expected!

3.3. Algorithm flow for estimation

1. Choose the family of maps \mathcal{W} .

2a. If $\mathcal{W} = \mathcal{W}_1$ or \mathcal{W}_2 ,

- (i) calculate sample moments,
- (ii) build the quadratic form and solve it for p .

2b. $\mathcal{W} = \mathcal{Q}_1$ or \mathcal{Q}_2

- (i) estimate the sample quantiles.

3. If you want to estimate F at point x : take any distribution function, for example the uniform over $[0, 1]$ and start to iterate T .
4. Stop after a few iterations (normally 5 is enough).
5. The “fixed point” of T evaluated in x is the estimate of $F(x)$.

In case the support of F is not known one case uses the range of the sample but the resulting IFS estimator will then try to approximate a distribution function which has exactly that support. If any hints on the shape of the distribution F is available, use it to choose the maps. All the examples, tables and graphics have been done by some software developed by the authors. In particular, a package called `ifs` is freely available for the R environment system [13] in the CRAN (Comprehensive R Archive Network) <http://cran.R-project.org> under the *contributed* section.

4. Conclusions

It seems that this kind of approach can be used to make nonparametric inference when data are missing or sample size are small. Note that with this method it is only possible to work with distributions with compact support. Moreover, a knowledge of the support itself it is needed. Nevertheless, it seems a promising approach and the use of different sets of maps can be the object of further investigations.

References

- [1] M.F. Barnsley, S. Demko, Iterated function systems and the global construction of fractals, *Proc. R. Soc. London, Ser. A* 399 (1985) 243–275.
- [2] R. Beran, Estimating a distribution function, *Ann. Stat.* 5 (1977) 400–404.
- [3] R.H. Byrd, P. Lu, J. Nocedal, C. Zhu, A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Comput.* 16 (1995) 1190–1208.
- [4] A. Dvoretzky, J. Kiefer, J. Wolfowitz, Asymptotic minimax character of the sample distribution function and of the classical multinomial estimators, *Ann. Math. Stat.* 27 (1956) 642–669.
- [5] S. Efromovich, Second order efficient estimating a smooth distribution function and its applications, *Methods Comput. Appl. Probab.* 3 (2001) 179–198.
- [6] B. Forte, E.R. Vrscay, Solving the inverse problem for function/image approximation using iterated function systems, I. Theoretical basis, *Fractal* 2 (3) (1995) 325–334.
- [7] B. Forte, E.R. Vrscay, Inverse problem methods for generalized fractal transforms, in: Y. Fisher (Ed.), *Fractal Image Encoding and Analysis*, NATO ASI Series F, vol. 159, Springer, Heidelberg, 1998.
- [8] R.D. Gill, B.Y. Levit, Applications of the van Trees inequality: a Bayesian Cramér–Rao bound, *Bernoulli* 1 (1995) 59–79.

- [9] G.K. Golubev, B.Y. Levit, On the second order minimax estimation of distribution functions, *Math. Methods Stat.* 5 (1996) 1–31.
- [10] G.K. Golubev, B.Y. Levit, Asymptotic efficient estimation for analytic distributions, *Math. Methods Stat.* 5 (1996) 357–368.
- [11] J. Hutchinson, Fractals and self-similarity, *Indiana Univ. J. Math.* 30 (5) (1981) 713–747.
- [12] S.M. Iacus, D. La Torre, Approximating distribution functions by iterated function systems, *J. Appl. Math. Decision Sci.*, accepted for publication.
- [13] R. Ihaka, R. Gentleman, R: a language for data analysis and graphics, *J. Comput. Graphical Stat.* 5 (1996) 299–314.
- [14] B.Y. Levit, Infinite-dimensional information inequalities, *Theory Probab. Appl.* 23 (1978) 371–377.
- [15] P.W. Millar, Asymptotic minimax theorems for sample distribution functions, *Z. Warschein. Verb. Geb.* 48 (1979) 233–252.
- [16] M.E. Tarter, M.D. Lock, *Model Free Curve Estimation*, Chapman & Hall, New York, 1993.