İstatistikçiler Dergisi:
İstatistik&Aktüerya 7 (2014) 14-19

İstatistikçiler Dergisi:
İstatistik&Aktüerya

Dağılım fonksiyonlarının yinelenmiş fonksiyon sistemleri ile tahmini

Figen Çilingir
Çankaya Üniversitesi
Fen-Edebiyat Fakültesi
Eskişehir Yolu Ankara, Türkiye
cilingirfigen@gmail.com

Özet

Bir dağılım fonksiyonunun parametrik olmayan tahmin edicisi yinelenmiş fonksiyon sistemleri kullanılarak elde edilebilmektedir. Bu yönteme göre, bir dağılım fonksiyonunun tahmin edicisi, $(X_1, X_2,\ldots,X_n)$ örneklemine bağlı olan bir p parametre vektörü ve w afin dönüşümleri ailesine göre tanımlanan T daralma operatörünün bir sabit noktası olarak düşünülmektedir. Döviz kuru verisi üzerinde yapılan uygulamadan elde edilen sonuçlar bir örnek olarak gösterilmiştir.

Anahtar Sözcükler: Fraktallar, yinelenmiş fonksiyon sistemleri, dağılım fonksiyonu, parametrik olmayan tahmin, yoğunluk tahmini

Abstract

Estimating the distribution functions by iterated function system

A class of nonparametric estimators of a distribution function base on the theory of iterated function systems (IFS). The estimator 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 which can be both depend on the sample $(X_1, X_2,\ldots,X_n)$. *An application to exchange data is showed as an example.*

Key Words: *Fractals, iterated function systems, distribution function estimation, nonparametric estimation, density estimation.*

1. Giriş

Yinelenmiş fonksiyon sistemleri (YFS) fraktalların oluşması için çok kullanışlı bir araçtır. Latincede "fraktus" dan gelen fraktal terimi 1975 yılında Benoit Mandelbrot tarafından ortaya atılmıştır[1]. Fraktal, düzensizlik içinde düzen arz eden şekillerdir; tıpkı kırılmış bir taş parçası gibi düzenli olmayan geometrik biçimlerdir. Cantor kümesi ilk ve en basit fraktal örneğidir. İki boyutlu fraktallara diğer bir örnek ise Sierpinski üçgenleridir [2].

Seksenli yılların ortasında, kesikli dinamik sistemler kuramının uygulamaları, birbirine benzer kümelerin ve fraktalların oluşması sonucunda yinelenmiş fonksiyon sistemleri kavramı ortaya atılmıştır [3].

Tanım 1. $0<\beta<1$ olmak üzere, düzlemde n tane $\boldsymbol{P_1}, \boldsymbol{P_2},\ldots, \boldsymbol{P_n}$ noktası verildiğinde $\forall\, i=1,2,\ldots,n$ için $A_i(\boldsymbol{P}) = \beta\,(\boldsymbol{P}-\boldsymbol{P_i}) + \boldsymbol{P_i}$ olacak biçimde $\{A_1, A_2,\ldots, A_n\}$ kümesi elde edilir. Tanım kümesi iki boyutlu düzlem olan $\{A_1, A_2,\ldots, A_n\}$ kümesine "yinelenmiş fonksiyon sistemi" denir [4].

Örnek: Düzlemde bir tane $P_0 = (x_0, y_0)$ noktası olsun. $\beta \in (0,1)$ olmak üzere P_0 noktası için $\boldsymbol{A}(\mathbf{P}) = \beta(\mathbf{P} - \boldsymbol{P_0}) + \boldsymbol{P_0}$, $\mathbf{A}(\boldsymbol{P_0}) = \boldsymbol{P_0}$ olduğundan $\boldsymbol{P_0}$, $\boldsymbol{A}$ fonksiyonun bir sabit noktasıdır. $0 < \beta < 1$ olduğundan $\boldsymbol{A}$, düzlemde $\boldsymbol{P_0}$ 'a yakın herhangi bir noktaya hareket eder. Gerçekten, $P = (x, y)$ iken $\boldsymbol{A}(\boldsymbol{P}) - \boldsymbol{A}(\boldsymbol{P_0}) = \beta(\boldsymbol{P} - \boldsymbol{P_0})$ olur ki burada $\boldsymbol{P}$ ile $\boldsymbol{P_0}$ arasındaki uzaklık β ile daraltılmıştır. Bu sebeple, β'ya daraltma çarpanı denir. $\boldsymbol{A}$ fonksiyonun birinci yinelenmesi ile düzlemdeki bir $\boldsymbol{P}$ noktasından $\boldsymbol{P_0}$ noktasına yaklaşılır. Birinci yineleme sonucu $\boldsymbol{P}$ noktasından $\boldsymbol{A}(\boldsymbol{P})$ noktasına gelinir. İkinci yineleme için yeni başlangıç noktası $\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}(\boldsymbol{P}))$ alınmak suretiyle işleme devam edilir; ikinci yineleme sonucunda yeni nokta $\boldsymbol{A}^2(\boldsymbol{P})$ ye gelinir ve $\boldsymbol{P}_0$ noktasına bir adım daha yaklaşılır. Bilinen bir örnek olarak, Cantor kümesinde $\boldsymbol{P_0}$ noktası, $P_0 = (1, 0)$ olup daraltma faktörü $\beta = 1/3$ tür.

Yinelenmiş fonksiyon sistemleri bir fonksiyona yaklaşım için de kullanılabilir. (E,d) tam metrik uzayında bir nokta olarak görülen bir f fonksiyonuna yaklaşmak istendiğinde, amaç $d(f, f^*)$ uzaklığı en küçük olacak şekilde, $T: E \rightarrow E$, $Tf^* = f^*$ biçiminde, tek bir f^* sabit noktalı T daraltma operatörü kurmaktır. Burada metrik, metrik uzay ve tam metrik tanımları şöyledir:

Tanım 2: E boş olmayan bir küme ve $d: ExE \rightarrow$ □ , *i)* $\forall x,y \in E$ *için* $d(x,y) \geq 0$ *ii)* $d(x,y) = 0 \Leftrightarrow x = y$ *iii)* $\forall x,y \in E$ *için* $d(x,y) = d(y,x)$ *iv)* $\forall x,y,z \in E$ *için* $d(x,y) \leq d(x,z) + d(z,y)$ şartlarını sağlıyorsa, d fonksiyonu E üzerinde bir metrik olur ve (E,d) ikilisi de bir metrik uzay olarak tanımlanır. Bir (E,d) metrik uzayı içinde her Cauchy dizisi bir limite yakınsıyorsa bu metrik uzaya tam metrik uzay denir.

Burada yakınsak her dizi bir Cauchy dizisi olduğundan ve yinelenmiş fonksiyonların değerleri de bir dizi gösterdiğinden her bir yinelenmiş fonksiyonun kendisi tam metrik uzayda bir nokta olarak tanımlanabilir. Daraltma operatörü olarak tanımlanan T fonksiyonunun bir f^* gibi sabit noktası vardır ve $T f^* = f^*$ dir. Burada verilen T operatörü YFS tanımındaki β ile bire bir benzerlik göstermektedir.

YFS' nin, görüntü işleme kuramında, olasılıksal büyüme modellerinde ve rastgele dinamik sistemler kuramında da bazı önemli uygulamaları bulunmaktadır. Özellikle YFS 'nin görüntü sıkıştırma yönteminde kullanılması ile konu güncellik kazanmıştır [5].

2. Dağılım fonksiyonunun tahmini

Son yıllarda, yinelenmiş fonksiyon sistemlerinin istatistik kuramında, dağılım fonksiyonları ve dolayısıyla olasılık yoğunluk ve karakteristik fonksiyonların tahmininde de kullanıldığı görülmektedir [6]. Bu kullanımda, E, bir $[0,1]$ kompakt kümesi üzerinde tanımlı dağılım fonksiyonlarının uzayını ve d ise sup-norm uzaklığını göstermektedir. Bu durumda $G \in E$ olan herhangi bir dağılım fonksiyonu G'nin YFS T operatörü altındaki görüntüsü,

$$TG(x) = \sum_{i=1}^{N} p_i \, G(w_i^{-1}(x)) \qquad (1)$$

biçiminde tanımlansın. Burada, w_i ler, $w_i(x) = a_i + s_i x$ biçimine tek boyutlu *afin* dönüşümler ve w_i^{-1} ler, w_i'lerin ters fonksiyonu olup sürekli ve artandır. YFS 'nin katsayıları olan p_i'ler olasılık dağılımıdır ($\Sigma p_i = 1$ ve $p_i \geq 0$). w_i ve p_i'ler hedef olarak verilen F dağılım foksiyonuna bağlıdır. *Afin* dönüşümlerin ve p_i lerin sayısını gösteren N tamsayısı, istenilen yakınsamanın kalitesine bağlı olarak T'nin bir F^* sabit noktası için $d(F^*, F) < \varepsilon$ olacak biçimde, $N = N_\varepsilon$ olarak seçilebilir[8].

YFS yönteminde temel sorun, w_i dönüşümlerini ve p_i katsayılarını belirlemektir. Bir F hedefi için w_i afin dönüşümlerinin kümesi ve p_i katsayıları bir minimizasyon işleminin sonucunda bulunabilir. Ancak çoğu zaman, hedef F bilinmediğinden, w_i dönüşümleri ve p_i katsayıları $(x_1, x_2, x_3, \ldots, x_n)$ örneklem verisinden bulunabilir. Bu yaklaşımda, önce bir N için dalgacık dönüşümleri seçilir ve ardından

$$\mathbf{p}^t Q \mathbf{p} + \mathbf{b}^t \mathbf{p} + \mathbf{c}$$

karesel biçimini kurmak için $M>N$ olmak üzere $\mathbf{m}=\langle m_1, m_2, m_3,\ldots,m_M\rangle$ örneklem momentleri kullanılır. Burada $\boldsymbol{Q}=Q(\mathbf{m})$, $\mathbf{b}=b(\mathbf{m})$ ve $\mathbf{c}=c(\mathbf{m})$ dir. $\boldsymbol{Q}$ için minimizasyon probleminin çözümü p_i katsayılarından oluşan $\mathbf{p}=\langle p_1, p_2,\ldots,p_N\rangle$ vektörünü verir. Bu karesel biçim, YFS momentleri ile **m** örneklem momentleri vektörü arasındaki uzaklığı ölçmektedir.

w_i afin dönüşümlerine dalgacık dönüşümleri de denir. Dalgacık dönüşümleri

$$W=\left\{w_1=h_{11}, w_2=h_{12}, w_3=h_{21},\ldots,w_6=h_{24}, w_7=h_{31},\ldots,w_N=h_{i^*2^{i^*}}\right\} \tag{2}$$

biçiminde kurulur. Burada $h_{ij}=(x+j-1)/(2^i)$, $i=1,2,\ldots,i^*$ $j=1,2,\ldots,2^i$ olup i^*

$$N=\sum_{i=1}^{i^*}2^i \tag{3}$$

olacak biçiminde tanımlanır.

Tam metrik uzaya *afin* dönüşümler kullanılarak kurulan YFS'ler basit ve kolay işlenebilir. Gerçekten eğer T, $\tilde{F}$ sabit noktasına sahip bir YFS ise, T nin Fourier dönüşümü de bir sabit noktaya sahip olur. Bu sabit nokta, $\tilde{F}$ sabit noktasının bir Fourier dönüşümüdür. Böylece YFS'nin Fourier dönüşümü de bir YFS dönüşümüdür ve aşağıdaki gibi yazılabilir:

$$B\phi(t)=\sum_{k=1}^{N}p_k e^{-ita_k}\phi(s_k t)\,. \tag{4}$$

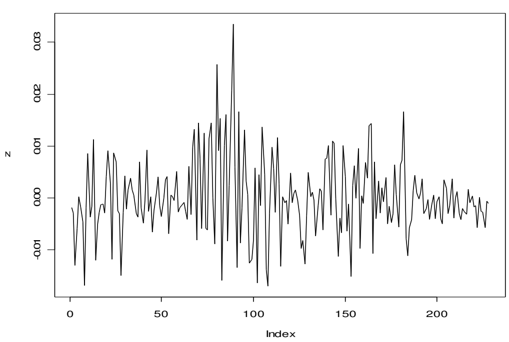
Burada B karakteristik fonksiyonlar uzayı üzerindeki bir operatördür. Bir kez w_i ve p_i 'ler elde edildiğinde dağılım fonksiyonuna ve Fourier dönüşümünün tahmin edicisine aşağıdaki gibi sahip olunur:

$$\hat{f}(x)=\frac{1}{2\pi}\sum_{k=-m}^{+m}\tilde{\phi}(k)e^{ikx}$$

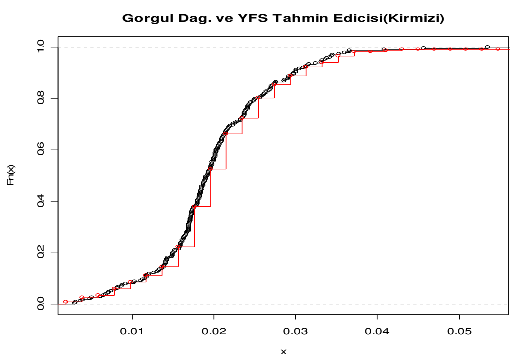
Burada $\tilde{\phi}$, B'nin sabit noktasıdır [6].

3. Sayısal Örnek

T.C.Merkez Bankası ABD Doları satış kurunun 2.1.2004 ile 29.11.2004 tarihleri arasındaki 228 günlük kapanış fiyatı TL olarak x_t ile gösterilsin. Doların TL karşındaki getirisini ifade eden yüzde değişimler dizisi $z_t=(\log x_t - \log x_{t-1})$ Tablo 1'de verilmiş ve Şekil 1'de gösterilmiştir.



Şekil 1. Dolar kuru yüzde değişimleri



Şekil 2 Dağılım fonksiyonunun görgül ve YFS (merdiven) tahmin edicileri

Ekonomik kriz sonrası dalgalı kur rejiminin uygulandığı bu dönemdeki değişimlerin olasılık dağılımının tahmini olarak önce görgül dağılım elde edilmiş buna YFS tahmin edicisinin nasıl uyum sağladığı Şekil 2'de gösterilmiştir. Büyük örneklem için görgül dağılımın, bilinmeyen dağılım fonksiyonuna yakınsaması beklenen bir durumdur[8]. Şekil 2'de YFS tahmin edicisinden elde edilen grafiğin de görgül dağılım ile çakıştığı gözlenmektedir. YFS tahminlerinin hesaplanmasından açık sistem R dili için yazılmış olan IFS programı kullanılmıştır [7]. IFS programı aşağıdaki altı adımlık işlem dizisini uygulamaktadır:

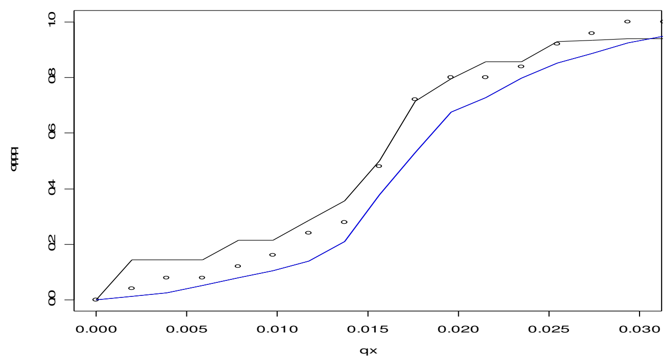
A1) Örneklem momentlerinin hesaplanması,
A2) ***w*** dönüşüm ailesinin seçilmesi,
A3) Karesel biçimin kurulması ve ve ***p*** için çözülmesi,
A4) *T* yinelemesine başlamak için bir dağılım fonksiyonun alınması (tekdüze dağılım),
A5) En fazla 5 yinelemeden sonra yinelemenin sona erdirilmesi,
A6) *T*'nin sabit noktasının $F(x)$'in tahmini olarak alınması.

Bilinmiyen bir F dağılım fonksiyonuna sahip X rastlantı değikenden alınan n tane bağımsız ve aynı dağılıma sahip X_1 X_2,..., X_n rastlantı değişkenleri verilsin. Görgül dağılım fonksiyonu bilişnmeyen F dağılım fonksiyonunun uyaygın olarak kullanılan bir tahmin edicisidir ve

$$\hat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} 1\ (X_i \leq x) \qquad (5)$$

ile ifade edilir. Burada **1**(.) ile gösterilen gösterge fonksiyonudur. Sonuçların görgül dağılımla karşılaştırılması için x'in tanım aralığı (0.01-0.005) olarak verilmiştir. Görgül birikimli bağılım fonksiyonu her gözlemde *1/n* kadar adım artan bir adım fonksiyonudur.

Küçük örneklem durumu için, bilinen bazı kuramsal dağılım fonksiyonlarının YFS tahminlerinin görgül dağılıma göre daha iyi sonuçlar verdiği Iacus ve La Torre tarafından göstermiştir [9]. Kısa bir seride de YFS tahminlerinin kullanılabilir olduğunu bir örnekle göstermek için, 228 uzunluğundaki günlük ABD Dolar kuru değişimleri serisinden sistematik olarak 9 günlük dönem atlatılarak 25 noktalı yeni bir seri elde edilmiştir (Tablo2). Bu küçük örneklem için elde edilen YFS tahminleri ile hem küçük örneklemden hem de büyük örneklemden (228 noktadan) elde edilen görgül dağılımlar Şekil 3'de verilmiştir. Büyük örneklemin görgül dağılımına yaklaşma bakımından, küçük örneklem YFS tahmin edicisinin, diğer küçük örneklemin görgül dağılımından daha uzak olmadığı gözlenmektedir. Uzaklık ölçüsü olarak kullanılan ortalama karesel fark değerleri 0.00022 ile 0.00041 olarak hesaplanmıştır. Bu sonuç YFS tahminlerinin kullanılabilir olduğu konusunda bir fikir vermektedir. Küçük örneklemlerde YFS tahminlerinin görgül dağılım fonksiyonlarından daha iyi olduğu Iacus ve La Torre, tarafından yapılan kapsamlı Monte Carlo çalışmaları ile gösterilmiştir [10][11]. Bu örnekte de aynı sonuç doğrulanmaktadır.



Şekil 3 . Küçük örneklem YFS tahmini (noktalı), küçük örneklem görgül dağılım (en üstte) ve büyük örneklem görgül dağılım(en altta)

Tablo 1. Dolar kuru değişimleri (Sütunlar halinde)

1-25	25-50	51-75	76-100	101-125	126-150	151-175	176-200	201-228
-0.00183	-0.00245	0.00005	0.01144	0.00576	-0.00976	-0.00635	-0.00308	0.00027
-0.00287	-0.00308	0.00352	0.01448	-0.01632	-0.00822	-0.00124	0.00641	-0.00400
-0.01297	-0.01495	0.00412	-0.00002	0.00449	-0.01282	-0.01513	0.00065	-0.00501
-0.00584	-0.00285	-0.00691	-0.00893	-0.00151	-0.00347	0.00240	-0.00568	0.00343
0.00020	0.00424	0.00060	0.02572	0.01362	0.00491	0.00623	0.00638	0.00190
-0.00172	-0.00216	0.00019	0.00916	0.00455	0.00026	-0.00015	0.00723	-0.00305
-0.00470	0.00153	-0.00059	0.01528	-0.01387	0.00104	0.00950	0.01664	-0.00156
-0.01687	0.00376	0.00506	-0.01598	-0.01702	-0.00039	-0.00971	-0.00791	0.00374
-0.00196	0.00153	-0.00267	0.01012	-0.00304	-0.00733	0.00049	-0.01120	-0.00385
0.00855	0.00053	-0.00179	0.01602	0.00978	-0.00149	-0.00105	-0.00570	0.00000
-0.00363	-0.00308	-0.00150	-0.00838	0.00567	0.00179	0.00679	-0.00426	0.00118
-0.00165	-0.00366	-0.00081	0.00240	-0.00279	0.00132	0.00379	0.00145	-0.00300
0.01130	0.00696	-0.00264	0.02080	0.01166	-0.00615	0.01387	0.00439	-0.00426
-0.01207	-0.00197	-0.00416	0.03352	0.00268	0.00749	0.01427	0.00098	-0.00210
-0.00505	-0.00492	0.00610	0.00596	-0.01328	0.00758	-0.01074	-0.00023	-0.00256
-0.00276	-0.00043	-0.00322	-0.01342	0.00019	0.01007	0.00697	0.00100	-0.00311
-0.00126	0.00918	0.00977	0.01661	-0.00099	-0.00337	-0.00401	0.00371	0.00164
-0.00120	-0.00256	0.01327	-0.00871	-0.00049	0.01096	0.00329	-0.00304	-0.00102
-0.00288	0.00018	-0.00816	-0.00194	-0.00510	0.01052	-0.00292	-0.00202	0.00031
0.00417	-0.00666	0.01446	0.01311	0.00480	-0.00019	0.00199	-0.00029	-0.00177
0.00913	-0.00246	0.00752	0.00322	-0.00100	-0.01130	-0.00074	-0.00407	-0.00154
0.00286	0.00109	-0.00600	0.00065	0.00086	-0.00387	0.00386	-0.00131	-0.00572
-0.01180	0.00397	0.01244	-0.01254	0.00156	-0.00676	-0.00503	0.00055	0.00010
0.00865	-0.00155	-0.00585	-0.01180	-0.00039	0.01006	-0.00166	-0.00398	-0.00262
0.00698	-0.00354	-0.00613	-0.00817	-0.00337	0.00412	-0.00475	-0.00073	-0.00273
								-0.00579
								-0.00061
								-0.00112

Tablo 2 Dolar kuru değişimleri kısa seri (Satırlar halinde)

-0.001955	-0.001199	-0.003079	-0.003075	-0.006656	-0.006907	-0.002644	-0.006002	0.009162
0.005959	-0.011802	-0.017024	-0.000993	-0.009761	-0.001488	0.010522	-0.015135	0.003788
-0.000735	0.006377	0.000983	0.000549	-0.001556	-0.003111	-0.002733		

4. Sonuç ve öneriler

Bu makalede, son yıllarda ortaya atılan, YFS ile dağılım fonksiyonun tahmini konusu kısaca tanıtılmış ve gerçek veri kümesi kullanılarak sayısal bir örnek verilmiştir. Yöntemin bilgisayar yazılım programlarının R açık sistem dili altında geliştirilmiş olması konunun sayısal örneklerle incelenmesini kolaylaştırmaktadır. Gerçek bir veri kümesi ile yapılan uygulamada, hem büyük hem de küçük örneklem için, YSF tahmininin, görgül dağılımdan farklı olmadığı grafik inceleme ile görülmüştür. YSF ile yapılan dağılım fonksiyonu tahminlerinin başarımını, çarpık ve basık kalın kuyruklu dağılımlar dahil olmak üzere, büyük ve küçük örneklemler ayırımında, çekirdek tipi parametrik olmayan yöntemlere karşı test etmek için yeni Monte Carlo çalışmalarının yapılması önerilmektedir.

Kaynaklar

[1] Maldrbrot, B. (1982) *The Fractal Geometry of Nature*, W.H. Freeman and Co., San Francisco.

[2] Barnsley, M. F. (1993), *Fraktals Everywhere*, Academic Press.

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

[4] Devaney, R, L.(1993), *A First Course in Chaotic Dynamical Systems*, Persaus Books Pub.

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

[6] S.M. Iacus and D. La Torre., "*Approximating distribution functions by iteration functions systems and applications*", Proceedings the S.I.M.A.I. Confarence, Chia Laguna, Italy, 2002.

[7] S.M. Iacus and D. La Torre (2002), *Fractals and Statistics: an R package 'Distributed Statistical Computing'* (DSC 2003).

[8] S.M. Iacus and D. La Torre (2002), *On fractals distribution function estimation and applications,* Departemental Working Papers 2002-07, Dep. of Economics, University of Milan, Italy.

[9] S.M. Iacus and D. La Torre (2002), *Nonparametric estimation of distribution and density function in presence of missing data: an IFS approach*, Departemental Working Papers 2002-25, Dep. of Economics, University of Milan, Italy.

[10] S.M. Iacus and D. La Torre (2005), Approximating distribution functions by iterated function systems, *J.Appl.Math. Dec.Sci.*, 1, 334-345.

[11] S.M. Iacus and D. La Torre (2005), A comparative simulation study on the IFS distribution function estimator, *Nonlinear Analysis: Real World Applications*, 6, 774-785.