

Un método rápido para la cuantificación de la autosemejanza de corto alcance en series RR

M. A. García González¹, M. Fernández Chimeno¹, J. Ramos Castro¹

¹ Grupo de Instrumentación Electrónica y Biomédica, Departamento de Ingeniería Electrónica, Universitat Politècnica de Catalunya, Barcelona, España, magarcia@eel.upc.edu

Resumen

Se propone un nuevo método para la cuantificación de la autosemejanza de corto alcance en series temporales. Los índices propuestos presentan alta correlación con los resultados del análisis de fluctuaciones tras eliminación de tendencias (DFA) y tienen una velocidad de cálculo mayor en más de dos órdenes de magnitud respecto al cálculo tradicional de exponentes de escala. En simulaciones con ruido fraccional gaussiano y en series RR, los exponentes de escala y los índices propuestos presentan una relación prácticamente lineal por lo que la metodología presentada puede emplearse como un método alternativo al DFA en series temporales con un gran número de muestras.

1. Motivación

El análisis de fluctuaciones tras eliminación de tendencias (DFA del inglés Detrended Fluctuation Analysis) trata de cuantificar la autosemejanza de series temporales y fue ya propuesto para el análisis de la variabilidad del ritmo cardíaco en 1995 [1]. Con este método, la dinámica de la serie RR (intervalos temporales entre complejos QRS sucesivos) puede ser cuantificada a partir del ajuste de exponentes para un determinado rango de escalas de interés sin necesidad que la señal sea estacionaria. El exponente de escala de corto alcance (short-range scaling exponent o α_1) ha recibido una atención especial por su capacidad de distinguir entre estados normales y patológicos [2].

No obstante, la estimación de α_1 presenta varios problemas. En primer lugar, el resultado puede estar influenciado por el tipo de eliminación de tendencia empleado. Generalmente se realiza una sustracción de tendencia lineal, pero otros métodos también son posibles [3]. Además, no existen recomendaciones para el rango de escalas en el que se realiza el ajuste de α_1 . Por otro lado, el coste computacional de la estimación puede ser muy elevado al analizar series RR de 24 horas. Es por ello que sería interesante encontrar un índice con alta correlación con α_1 pero cuyo cálculo fuera rápido.

En este trabajo se propone un índice rápido de alta correlación con α_1 en el rango de escalas de 4 a 16 (según definición en [1]) y se estudia también el efecto del método de eliminación de tendencia escogido.

2. Materiales y métodos

2.1. Estimación de α_1 mediante DFA

Sea $\mathbf{X}=\{x(1), x(2), \dots, x(N)\}$ una serie temporal con N muestras. El primer paso en el DFA es la eliminación del valor medio de la serie temporal y la obtención de la serie \mathbf{Y} como la suma acumulada según la expresión:

$$y(j) = \sum_{i=1}^j (x(i) - \bar{x}) \quad (1)$$

A continuación, la serie \mathbf{Y} se divide en subseries consecutivas de idéntica longitud n (escala). Para cada subserie se obtiene la tendencia ($y_n(k)$) y se calcula el siguiente estadístico:

$$F(n) = \sqrt{\frac{1}{n} \sum_{k=1}^n (y(k) - y_n(k))^2} \quad (2)$$

que dependerá de la subserie, de n y del tipo de ajuste de tendencia empleado. Finalmente, para la escala n se calcula el promedio de $F(n)$ empleado todas las subseries disponibles. Los exponentes de escala se obtienen a partir de la pendiente del ajuste lineal en un diagrama en el cual se representa el logaritmo de n respecto al logaritmo de la media de $F(n)$. En este trabajo, α_1 se obtiene empleando las escalas comprendidas entre 4 y 16. Nótese que para series temporales largas, el número de subseries (y , por consiguiente, el número de ajustes de tendencia y de estadísticos a promediar) puede ser muy elevado haciendo que el cálculo del índice sea lento.

Se han empleado dos métodos para la eliminación de tendencia: eliminación de tendencia lineal por ajuste de recta por mínimos cuadrados y eliminación de tendencia empleando el filtro de Hodrick-Prescott (HP) [4]. Este filtro depende de un parámetro empírico (λ). En este trabajo, cada exponente de corto alcance se ha estimado empleando el ajuste lineal ($\alpha_{1 \text{ lin}}$), aplicando el filtro HP con $\lambda = 10^4$ ($\alpha_{1 \text{ HP4}}$) y aplicando el filtro HP con $\lambda = 10^3$ ($\alpha_{1 \text{ HP3}}$). El filtro HP ha sido también empleado para el cálculo de los índices propuestos (véase siguiente sección). El objetivo de este triple cálculo es investigar si las diferencias en las estimaciones de α_1 pueden llegar a ser importantes.

2.2. Índice rápido para la cuantificación de la autosemejanza de corto alcance

Sea \mathbf{X}_d la serie temporal obtenida tras eliminación de tendencia. Se define la serie invertida como

$$x_{dr}(n) = x_d(N - n + 1) \quad \forall n \in [1, N] \quad (3)$$

A continuación se define la serie temporal de los signos de las diferencias invertidas como:

$$srd(n) = \text{sign}(x_d(n) - x_{dr}(n)) \quad (4)$$

El índice propuesto es la frecuencia relativa de cambios de signos en esta serie ($fscmd$). Se calcula a partir del número de cambios de signo presentes en la serie normalizados por $N-1$.

En este trabajo se ha empleado como método de eliminación de tendencia el filtro HP con $\lambda = 10^4$ ($fscmd_{HP4}$) y con $\lambda = 10^3$ ($fscmd_{HP3}$). En el apéndice hay un listado de MATLAB para el cálculo del índice una vez eliminada la tendencia.

2.3. Simulación, bases de datos empleadas y análisis de datos.

Para estudiar el efecto de la eliminación de tendencia y la relación entre α_1 y $fscmd$, se ha empleado un generador de ruido gaussiano fraccional [5]. Se han generado 100000 realizaciones con $N=300$, 10000 realizaciones con $N=3000$ y 1000 realizaciones con $N=30000$. En cada realización, el exponente de Hurst fue escogido aleatoriamente y se estimaron $\alpha_{1 \text{ lin}}$, $\alpha_{1 \text{ HP4}}$, $\alpha_{1 \text{ HP3}}$, $fscmd_{HP4}$ y $fscmd_{HP3}$. Estos cinco índices también se han calculado en series RR disponibles gratuitamente en www.physionet.org [6]. En concreto, se ha analizado la Normal Sinus Rhythm RR time series database (NSR) que contiene 54 registros de 24 horas, la Fantasia Database (FT) que consiste en 40 registros de unas dos horas de duración y la Congestive Heart Failure RR interval database (CHF) con 29 registros de 24 horas.

Debido a que las bases de datos presentan artefactos en la detección de los complejos QRS, previo a la estimación de los índices se realizó una corrección siguiendo los siguientes pasos: Si el cambio entre un intervalo RR y el siguiente es mayor que diez veces el margen intercuartil se decide que hay presente un artefacto. Éste se clasifica como falso positivo (FP), falso negativo (FN) o latido ectópico (EB). Para un FP, se añaden tantos intervalos RR consecutivos como haga falta hasta que el intervalo RR corregido esté próximo a la media de los anteriores 10 intervalos RR. Un FN se descompone en tanto intervalos RR de igual longitud como haga falta para que los intervalos añadidos tengan un valor próximo a la media de los anteriores 10 intervalos RR. Finalmente en un EB se sustituyen los dos intervalos RR afectados por dos intervalos de valor igual a la media de los dos intervalos considerados como EB.

Para las distintas combinaciones de pares de índices y de número de muestras o base de datos se obtuvieron las correspondientes regresiones lineales. El coeficiente de determinación (R^2) y la desviación estándar de los

residuos (σ_{res}) se escogieron como indicadores de la bondad del ajuste del modelo.

El tiempo de cálculo en el mismo ordenador fue empleado para mostrar la mayor velocidad en la estimación de $fscmd$ frente a α_1 .

Finalmente, para cada índice se ha realizado un test de suma de rangos de Mann-Whitney para hallar diferencias significativas entre las bases de datos NSR y CHF. La significación (p) del test fue empleada para comparar los diversos índices.

3. Resultados

La figura 1 es un ejemplo de comparación entre la metodología propuesta y los resultados del DFA en las series temporales simuladas. Como se observa, a mayor número de muestras menor dispersión. Por otro lado, el índice propuesto guarda una relación casi lineal con α_1 aunque con pendiente negativa.

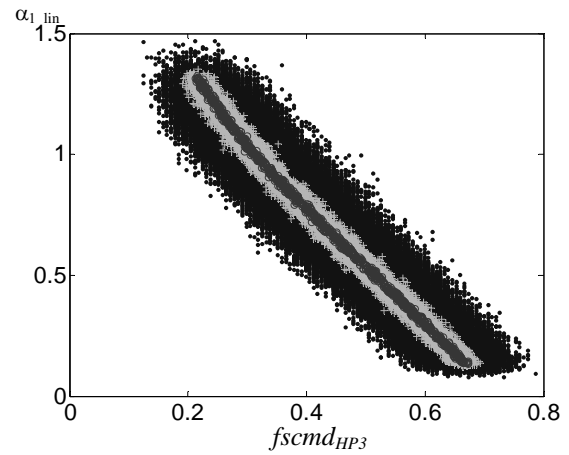


Figura 1. Índice $fscmd_{HP3}$ versus $\alpha_{1 \text{ lin}}$ en series temporales con autosimilaridad. Los puntos negros corresponden a $N=300$, cruces gris claro corresponde a $N=3000$ y círculos grises a $N=30000$

Al comparar los diversos índices α_1 entre sí para las series temporales simuladas se halló que para todas las combinaciones y número de muestras consideradas, R^2 es 1 y σ_{res} es inferior a 0,002. El modelo lineal encontrado indica que $\alpha_{1 \text{ lin}}$ es un 3% mayor que $\alpha_{1 \text{ HP3}}$ y un 0,3% mayor que $\alpha_{1 \text{ HP4}}$. La tabla 1 muestra los resultados de R^2 y σ_{res} en la comparación de $fscmd$ respecto α_1 en series temporales simuladas. En bases de datos se obtienen resultados similares respecto a la comparación entre índices obtenidos mediante DFA. Nuevamente, α_1 es sobrestimado empleando eliminación de tendencia lineal respecto a la eliminación de tendencia respecto HP. La tabla 2 muestra los resultados de comparación $fscmd$ respecto α_1 en bases de datos.

La figura 3 muestra el tiempo de computación de cada índice para diferentes números de muestras. Tal y como se observa, la estimación de $fscmd$ se realiza en menos del 1% del tiempo que se requiere para estimar α_1 .

X	Y		b	Y _o	R ²	σ _{res}
fscmd _{HP3}	α _{1 lin}	N=300	-2,203	1,707	0,916	0,096
		N=3000	-2,494	1,790	0,989	0,035
		N=30000	-2,528	1,806	0,996	0,021
fscmd _{HP4}	α _{1 lin}	N=300	-2,072	1,552	0,929	0,089
		N=3000	-2,200	1,602	0,991	0,033
		N=30000	-2,224	1,612	0,997	0,020
fscmd _{HP3}	α _{1 HP3}	N=300	-2,231	1,649	0,916	0,093
		N=3000	-2,416	1,729	0,989	0,035
		N=30000	-2,449	1,745	0,996	0,021
fscmd _{HP4}	α _{1 HP3}	N=300	-2,007	1,499	0,928	0,086
		N=3000	-2,131	1,548	0,991	0,032
		N=30000	-2,154	1,558	0,996	0,021
fscmd _{HP3}	α _{1 HP4}	N=300	-2,300	1,701	0,916	0,096
		N=3000	-2,486	1,784	0,989	0,035
		N=30000	-2,520	1,799	0,996	0,021
fscmd _{HP4}	α _{1 HP4}	N=300	-2,065	1,547	0,929	0,089
		N=3000	-2,193	1,597	0,991	0,033
		N=30000	-2,217	1,607	0,997	0,020
fscmd _{HP4}	fscmd _{HP3}	N=300	1,100	-0,069	0,967	0,028
		N=3000	1,132	-0,085	0,996	0,010
		N=30000	1,137	-0,087	1,000	0,004

Tabla 1. Resultados de regresiones lineales en series temporales autosimilares. El modelo lineal es $Y=b \cdot X+Y_o$

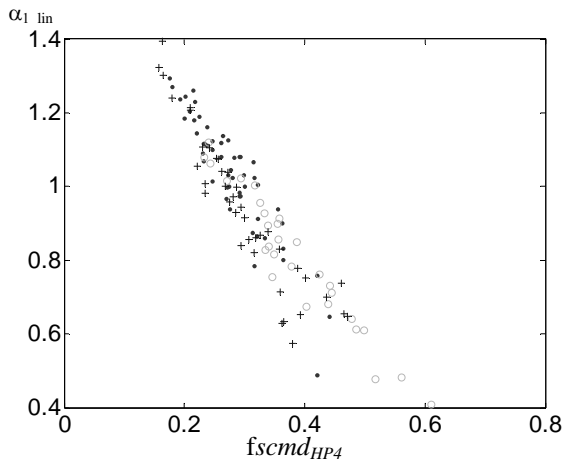


Figura 2. Índice fscmd_{HP4} versus α_{1 lin} en NSR (puntos), CHF (círculos) y FT (cruces)

X	Y		b	Y _o	R ²	σ _{res}
fscmd _{HP3}	α _{1 lin}	NSR	-2,077	1,709	0,736	0,082
		CHF	-2,027	1,702	0,846	0,073
		FT	-2,266	1,713	0,755	0,104
fscmd _{HP4}	α _{1 lin}	NSR	-2,435	1,709	0,841	0,063
		CHF	-1,897	1,533	0,932	0,048
		FT	-2,315	1,629	0,855	0,080
fscmd _{HP3}	α _{1 HP3}	NSR	-2,044	1,661	0,744	0,079
		CHF	-1,967	1,644	0,849	0,070
		FT	-2,223	1,664	0,762	0,100
fscmd _{HP4}	α _{1 HP3}	NSR	-2,392	1,660	0,846	0,061
		CHF	-1,840	1,478	0,932	0,047
		FT	-2,267	1,580	0,859	0,077
fscmd _{HP3}	α _{1 HP4}	NSR	-2,073	1,704	0,737	0,081
		CHF	-2,021	1,696	0,846	0,073
		FT	-2,261	1,708	0,756	0,103
fscmd _{HP4}	α _{1 HP4}	NSR	-2,431	1,704	0,841	0,063
		CHF	-1,891	1,528	0,932	0,048
		FT	-2,301	1,624	0,855	0,079
fscmd _{HP4}	fscmd _{HP3}	NSR	1,079	0,026	0,968	0,012
		CHF	0,865	0,111	0,942	0,020
		FT	0,936	0,063	0,950	0,018

Tabla 2. Resultados de regresiones lineales en bases de datos. El modelo lineal es $Y=b \cdot X+Y_o$

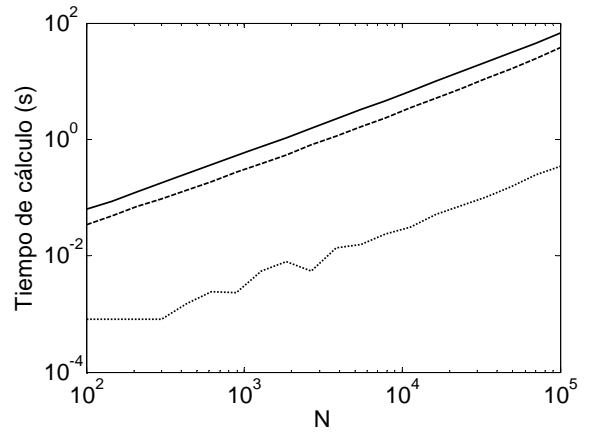


Figura 3. Tiempo de cálculo de los diferentes índices: Línea continua (α_{1 HP3} y α_{1 HP4}), línea discontinua (α_{1 lin}), línea punteada (fscmd_{HP4} y fscmd_{HP3})

Índice	α_1 lin	α_1 HP4	α_1 HP3	$fscmd_{HP4}$	$fscmd_{HP3}$
p	$5 \cdot 10^{-7}$	$5 \cdot 10^{-7}$	$4 \cdot 10^{-7}$	$4 \cdot 10^{-7}$	$3 \cdot 10^{-8}$

Tabla 3. Significación de comparación de medianas entre las bases de datos NSR y CHF para los diversos índices

Por último, la tabla 3 muestra los resultados de los tests Mann-Whithney al comparar los índices en las bases de datos NSR y CHF.

4. Discusión y conclusiones

La figura 1 muestra que en señales puramente autosimilares la relación entre $fscmd$ y α_1 es prácticamente lineal con lo cual la información aportada por ambos índices es prácticamente la misma, especialmente cuando el número de muestras es elevado. Por otro lado, la elección de un método de eliminación de tendencia determinado provoca que el valor absoluto en las pendientes cambie. La figura 2 muestra que para series RR esta relación sigue siendo aproximadamente lineal a pesar que R^2 sea menor. Este peor ajuste puede justificarse por el hecho que las series RR no son puramente autosimilares. La elección del método de eliminación de tendencia afecta a los modelos de la regresión lineal entre los diversos índices aunque en las tablas 1 y 2 se puede observar que R^2 en las diversas regresiones suele ser elevada con independencia del método empleado. Como era de esperar, σ_{res} disminuye a medida que N aumenta permitiendo estimar α_1 a partir de $fscmd$.

El $fscmd$ es un índice de cálculo rápido tal y como se desprende de la figura 3 y puede emplearse como método alternativo al DFA cuando se trata de cuantificar la autosemejanza de corto alcance en series temporales largas. Por lo tanto, se ha propuesto un nuevo método para la cuantificación de la autosemejanza de corto alcance en series temporales con alta correlación con los resultados del DFA y con una velocidad de cálculo mayor en más de dos órdenes de magnitud respecto al cálculo de α_1 .

Agradecimientos

Este trabajo ha sido financiado con el proyecto MEC PSI2008-06417-C03.

Apéndice

La siguiente función de MATLAB implementa el cálculo de $fscmd$ una vez se ha realizado la eliminación de tendencia pertinente.

```
function fscmd=fscmdcomput(rrd)
    rrr=rot90(rot90(rrd));
    d=sign(rrd-rrr);
    k=find(d==0);
    d(k)=-ones(size(k));
    fscmd=length(find(diff(d)))/(length(d)-1);
```

Referencias

- [1] Peng CK, Havlin S, Stanley HE, Goldberger AL. Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos*, vol:5, 1995, pp 82–87 (ISSN: 1054-1500)
- [2] Kleiger RE, Stein PK, Bigger JT. Heart rate variability: Measurement and clinical utility. *Annals of Noninvasive Electrophysiology*, vol: 10, 2005, pp 88–101 (ISSN: 1082-720X)
- [3] Hu K, Ivanov PCh, Chen Z, Carpena P, Stanley HE. Effect of trends on detrended fluctuation analysis. *Physical Review E*, vol: 64, 2001, pp 011114-1– 011114-19 (ISSN: 1539-3755)
- [4] Hodrick RJ Prescott EC. Postwar U.S. Business Cycles: An Empirical Investigation. *Journal of Money, Credit, and Banking*, vol: 29, 1997, pp 1–16 (ISSN: 0022-2879)
- [5] Paxson V. Fast, approximate synthesis of fractional gaussian noise for generating self-similar network traffic. *Computer Communication Review*, vol: 27, 1997, pp 5–18 (ISSN: 0146-4833)
- [6] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, vol: 101, 2000, pp: e215–e220 (ISSN: 0009-7322)