

# Estudio del Efecto del Umbral en la Entropía Muestral: Aplicación a la Variabilidad de Frecuencia Cardíaca

R. Goya Esteban<sup>1,2</sup>, J.P. Marques de Sá<sup>1,3</sup>, J.L. Rojo Álvarez<sup>2</sup>, O. Barquero Pérez<sup>1,2</sup>

<sup>1</sup>Departamento de Teoría de la Señal y Comunicaciones, Universidad Rey Juan Carlos, Fuenlabrada, Madrid, España, {rebegoya,obarquero}@gmail.com, {joseluis.rojo}@urjc.es

<sup>2</sup>Instituto de Engenharia Biomédica, Porto, Portugal, {meb07028,meb07027}@fe.up.pt

<sup>3</sup>Faculdade de Engenharia, Universidade do Porto, Porto, Portugal, {jmsa}@fe.up.pt

## Resumen

Las medidas basadas en el cálculo de la entropía de las señales, tales como la entropía muestral (*Sample Entropy*, *SampEn*), han sido ampliamente utilizadas para cuantificar la Variabilidad de la Frecuencia Cardíaca (VFC) en estratificación de riesgo cardíaco, bajo la hipótesis de que valores bajos de entropía son indicativos de enfermedades o de perturbaciones en los complejos mecanismos fisiológicos. Sin embargo, en diversos estudios se han obtenido valores mayores de entropía en pacientes con determinadas patologías que en sujetos sanos. En este trabajo, utilizamos la *SampEn* para medir la VFC de sujetos sanos y de pacientes con Insuficiencia Cardíaca (IC), con el propósito de discriminar entre ambos grupos y de caracterizar la pérdida de variabilidad debida a la edad. Concluimos que el ajuste de un valor umbral  $r$  fijo en el algoritmo, en lugar su ajuste convencional como porcentaje de la desviación estándar de la señal, mejora las capacidades de discriminación entre sujetos sanos y pacientes con IC, y además permite cuantificar con mayor nitidez la pérdida de VFC debida a la edad en sujetos sanos.

## 1. Introducción

La Variabilidad de la Frecuencia Cardíaca (VFC) es un potente marcador del estado del control cardiovascular por parte del sistema nervioso autónomo. Dado el alto contenido informativo de esta señal, se han llevado a cabo numerosas tentativas de caracterizarla mediante índices medidos en la misma utilizando técnicas de procesamiento digital. Durante los últimos años, se han propuesto en este escenario un gran número de métodos y técnicas no lineales, tanto basados en la Teoría del Caos, como en el análisis fractal de las series temporales y en principios de Teoría de la Información. En este último grupo se incluyen algoritmos representativos tales como la entropía aproximada (*Approximate Entropy*, *ApEn*) y la entropía muestral (*Sample Entropy*, *SampEn*). La primera fue propuesta por Pincus [1] para medir el nivel de irregularidad de las series temporales, y ha sido extensamente utilizada para caracterizar la VFC, tanto en adultos como en fetos [2, 3, 4]. La segunda es una modificación del algoritmo de cálculo original de la *ApEn*, que pretende superar sus limitaciones [5].

Sin embargo, los resultados numéricos obtenidos en la literatura no siempre son coincidentes con la idea subyacente de que valores bajos de índices de entropía corresponderían a estados patológicos o a trastornos de la regulación homeostática [6, 7]. Hasta la fecha, la mayor parte de los trabajos en los que se han aplicado estas medidas de entropía fijan los parámetros libres de los algoritmos de cálculo a valores previamente indicados en la literatura, en concreto, el parámetro umbral  $r$  se fija a una cantidad relativa a la varianza de cada señal bajo estudio. En este trabajo, hemos estudiado la influencia de la elección de estos parámetros en la evaluación de la pérdida de VFC debida a la edad, así como la caracterización de la VFC de los pacientes con Insuficiencia Cardíaca (IC), teniendo como objetivo una discriminación fiable entre sujetos sanos y sujetos patológicos utilizando la *SampEn*.

La estructura del trabajo es como sigue. En primer lugar, se introducen los algoritmos de la *ApEn* y de la *SampEn*. A continuación, se presentan las bases de datos utilizadas en el estudio y los experimentos realizados utilizando la *SampEn*. Tras presentar los resultados de discriminación y efecto de edad, se incluyen las conclusiones.

## 2. Algoritmos ApEn y SampEn

Si bien teóricamente la caracterización de la VFC de las series de intervalos RR podría llevarse a cabo utilizando medidas clásicas de entropía, tales como la entropía Kolmogorov-Sinai y variantes, estas medidas presentan algunos inconvenientes en su aplicación práctica. De una parte, exhiben un comportamiento divergente, y de otra no siempre son capaces de distinguir algunos procesos con complejidad diferenciada [1]. Con el propósito de superar estas limitaciones, Pincus formuló una familia de estadísticos, conocida como *ApEn*, para el análisis eficiente de la información contenida en series temporales cortas y ruidosas.

Para el cálculo de la *ApEn* es necesario la especificación previa de dos parámetros: la dimensión embebida  $m$ , que es la longitud de los vectores a ser comparados, y el valor umbral del ruido  $r$ .

El procedimiento para el cálculo de la  $ApEn$ , dadas  $N$  muestras de una serie temporal, es como sigue:

Se obtienen los vectores  $\mathbf{x}(1), \dots, \mathbf{x}(N - m + 1)$ , definidos por  $\mathbf{x}(i) = [u(i), \dots, u(i + m - 1)]$ , para  $i = 1, \dots, N - m + 1$ . La distancia entre dos vectores  $\mathbf{x}(i)$  y  $\mathbf{x}(j)$ ,  $d[\mathbf{x}(i), \mathbf{x}(j)]$ , se define como la máxima diferencia, en módulo, entre sus respectivas componentes escalares, esto es:

$$d[\mathbf{x}(i), \mathbf{x}(j)] = \max_{k=1, \dots, m} (|u(i + k - 1) - u(j + k - 1)|)$$

Basada en esta distancia, se define la siguiente medida de correlación:

$$C_i^m(r) = N^m(i) / (N - m + 1)$$

donde  $N^m(i)$  cuenta, para un vector dado  $\mathbf{x}(i)$ , el número de veces que  $d[\mathbf{x}(i), \mathbf{x}(j)] \leq r$  para  $j = 1, \dots, N - m + 1$ . A continuación, se calcula la media del logaritmo natural de  $C_i^m(r)$  para todo  $i$ :

$$\Phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln C_i^m(r)$$

Finalmente, se define la  $ApEn(m, r)$  y su estadístico asociado  $ApEn(m, r, N)$  como

$$ApEn(m, r) = \begin{cases} \lim_{N \rightarrow \infty} [\Phi^m(r) - \Phi^{m+1}(r)] & \text{para } m > 0 \\ -\Phi^1(r) & \text{para } m = 0 \end{cases}$$

$$ApEn(m, r, N) = \begin{cases} \Phi^m(r) - \Phi^{m+1}(r) & \text{para } m > 0 \\ -\Phi^1(r) & \text{para } m = 0 \end{cases}$$

La  $ApEn$  cuantifica la verosimilitud de que series de patrones que son parecidos para  $m$  observaciones, no continúen siendo parecidas en la siguiente comparación creciente. Por lo tanto, series con patrones repetitivos obtendrán valores de  $ApEn$  más bajos, mientras que series menos predecibles obtendrán valores de  $ApEn$  más altos.

Varias propiedades de la  $ApEn$  hacen que este estadístico sea apropiado para el análisis de datos fisiológicos [3]:

1. No es necesario un número grande de muestras en la señal para obtener estimaciones razonables ( $30^m$  muestras son suficientes).
2. Es robusta frente a datos atípicos.
3. Se ve poco afectada por ruido de magnitud inferior a  $r$ .
4. Es aplicable tanto a procesos estocásticos, como deterministas o mixtos, ya que se obtienen valores finitos en todos los casos.

Sin embargo, tal y como el propio Pincus relató [2], debido a la comparación de cada vector consigo mismo y para evitar la ocurrencia del  $\ln(0)$  en el algoritmo de cálculo, la  $ApEn$  es un estadístico sesgado. Este sesgo hace que la  $ApEn$  sea dependiente de la longitud de los datos, efecto que tiende a 0 cuando  $N \rightarrow \infty$ . Además, en algunos casos, la  $ApEn$  carece de consistencia relativa, si bien preserva esta propiedad en el rango estadísticamente válido de los

pares  $(m, r)$ . Por esta razón, su utilización ha proporcionado buenos resultados en numerosos estudios.

Tratando de mejorar el algoritmo de la  $ApEn$  para superar sus limitaciones, Richman y Moorman [5] desarrollaron el estadístico  $SampEn$ , el cual no realiza autocomparaciones. Asimismo, la  $SampEn$  presenta dos diferencias adicionales respecto a la  $ApEn$ : en primer lugar, únicamente son considerados los  $N - m$  vectores de longitud  $m$  en ambas etapas del proceso de cálculo; en segundo lugar, la medida de probabilidad se obtiene como el logaritmo natural de la probabilidad condicionada, y no como la razón de las sumas logarítmicas.

El algoritmo de la  $SampEn$  se calcula siguiendo los siguientes pasos:

$B_i^m(r)$  se define como  $(N - m - 1)^{-1}$  veces el número de vectores  $\mathbf{x}_m(j)$  parecidos a  $\mathbf{x}_m(i)$  (en relación a  $r$ ) donde  $j = 1 \dots N - m$  con  $j \neq i$ . La media de  $B_i^m(r)$  para todo  $i$  se calcula como

$$B^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} B_i^m(r)$$

De forma similar,  $A_i^m(r)$  se define como  $(N - m - 1)^{-1}$  veces el número de vectores  $\mathbf{x}_{m+1}(j)$  parecidos a  $\mathbf{x}_{m+1}(i)$  (en relación a  $r$ ) donde  $j = 1 \dots N - m$  con  $j \neq i$ . La media de  $A_i^m(r)$  para todo  $i$  se calcula como

$$A^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} A_i^m(r)$$

Finalmente,  $SampEn(m, r)$  y su estadístico asociado  $SampEn(m, r, N)$  se definen como sigue:

$$SampEn(m, r) = \lim_{N \rightarrow \infty} \{-\ln [A^m(r) / B^m(r)]\}$$

$$SampEn(m, r, N) = -\ln [A^m(r) / B^m(r)]$$

La  $SampEn$  ofrece mayor independencia respecto a la longitud de los datos, así como consistencia relativa en más situaciones que la  $ApEn$ . Al mismo tiempo, mantiene las características originales de la  $ApEn$  que la hacen apropiada para el estudio de señales fisiológicas.

### 3. Bases de Datos y Análisis con SampEn

El estadístico  $SampEn$  se utilizó para evaluar la VFC a partir de señales de intervalos RR obtenidas en registros Holter de 24 horas, procedentes de sujetos sanos y de pacientes con IC. Ambos conjuntos de registros fueron obtenidos de la base de datos Physionet [8]. Los datos procedían de registros pertenecientes a 72 sujetos sanos, de edades comprendidas entre los 20 y los 76 años, y de 44 pacientes con IC, de edades comprendidas entre los 22 y los 79 años. Todos los registros fueron preprocesados para eliminar artefactos, latidos mal identificados y latidos ectópicos. Además, los intervalos RR menores de 200 ms y mayores de 2000 ms fueron excluidos, así como aquellos que diferían más de un 20 % del intervalo RR previo o subsecuente [9].

	SampEn(3,15)	SampEn(1,0.1DE)
Sanos	1.00±0.23	1.09±0.27
IC	0.51±0.26*	1.37±0.36*

**Tabla 1.** Media  $\pm$  DE de la *SampEn* para ambos métodos. El símbolo \* indica variación significativa ( $p < 0.001$ ) entre sujetos sanos y sujetos con IC.

Dado que la elección de los parámetros libres en los métodos de entropía es crucial para su funcionamiento, en primer lugar se analizó la influencia de la selección de estos parámetros en el comportamiento de la *SampEn*, con el objetivo de maximizar la separación entre sujetos sanos y pacientes con IC. Para este propósito, se evaluaron un conjunto de posibles valores para  $m = 1, 2, 3$ . En relación al parámetro  $r$ , la *SampEn* se calculó de acuerdo con dos métodos:

1. El parámetro  $r$  como un cierto porcentaje de la desviación estándar (DE) de cada serie de datos, que es el método utilizado convencionalmente en la literatura [2, 5, 6, 7] (con  $r = r' * DE$ , donde  $r' = [0.1, 0.15, 0.2]$ ).
2. El parámetro  $r$  fijado como un porcentaje de la desviación estándar media de todos los registros (de los sujetos sanos y de los pacientes), lo cual implica emplear un valor de  $r$  fijo para todos los registros (utilizando  $r = 10, 15, 20$ ).

Este último método se justifica en [4], donde se obtuvieron mejores resultados en la caracterización de la VFC Fetal con la elección de un parámetro  $r$  fijo, independiente de la DE de cada registro.

También se estudiaron las capacidades discriminativas del estadístico para distinguir entre sujetos jóvenes y sujetos mayores, dividiendo para ello las bases de datos de sujetos sanos en dos grupos (grupo joven, de 20 a 50 años, y grupo mayor, de 51 a 80 años).

Finalmente, se hizo una división de todos los registros de los sujetos sanos en seis grupos diferentes, de acuerdo con la edad de los pacientes, y se calculó la *SampEn* para cada uno de los grupos con el propósito de analizar la evolución del estadístico en personas sanas. La relación del estadístico con la edad fue analizada utilizando la regresión lineal de la *SampEn* respecto a la edad, y obteniendo la pendiente (variación del estadístico por año) y su coeficiente de determinación, en este caso tanto para los sujetos sanos como para los sujetos con IC.

#### 4. Resultados

La Tabla 1 muestra la combinación de los parámetros  $m$  y  $r$  que consiguieron mayor discriminación para cada método. La utilización de un valor umbral  $r$  fijo para todos los registros, en lugar del uso más extendido de  $r$  como porcentaje de la DE de cada señal, proporcionó una mejor discriminación entre los sujetos sanos y los sujetos con IC. Además, se obtuvieron valores mayores de la *SampEn* para sujetos sanos que para sujetos con IC, lo cual concuerda con la idea de la pérdida de VFC en condiciones patológicas, mientras que utilizando  $r$  como un porcentaje

	DE
Sanos	136.84±32.67
Sanos-Jóvenes	136.20±35.51
Sanos-Mayores	137.20±31.36
IC	66.98±38.87

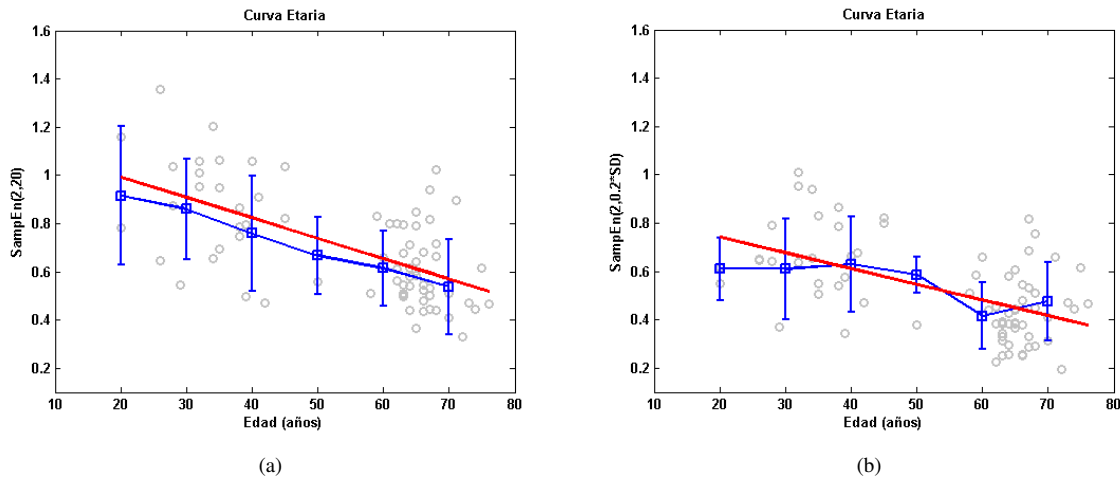
**Tabla 2.** Desviación estándar de los diferentes grupos. Media  $\pm$  DE.

de la DE de cada serie de datos se obtuvieron mayores valores de *SampEn* para los pacientes con IC que para los sujetos sanos. La razón de estos resultados podría ser que los pacientes con IC poseen una desviación estándar menor que los pacientes sanos (ver Tabla 2), y por lo tanto, al dividir cada registro por su desviación estándar, las diferencias relativas entre las muestras podrían no quedar equitativamente preservadas. Fijando  $r$  de forma relativa a la DE de cada registro en el algoritmo de cálculo, los registros pertenecientes a sujetos sanos son escalados por valores mayores que los registros de los sujetos con IC, lo cual no afecta únicamente a la amplitud de los datos, sino también a la diferencia relativa entre las muestras. Posteriormente, tanto los registros de sujetos sanos, como los registros de pacientes con IC son comparados respecto al mismo valor fijo  $r'$ . Por otro lado, debido a la diferencia considerable de valores entre las DE de ambos grupos, se obtuvo una discriminación similar utilizando la DE ( $p < 0.001$ ) y el índice *SampEn* con un valor umbral  $r$  fijo.

Para los grupos de jóvenes y mayores, las combinaciones de parámetros  $m$  y  $r$  que consiguieron una mayor discriminación, para cada método, se muestran en la Tabla 3. En este caso, ambos métodos proporcionaron valores mayores de la *SampEn* para los sujetos jóvenes que para los sujetos mayores, debido a que los valores de DE para ambos grupos son similares (ver Tabla 2), y por lo tanto, en el método 1 de obtención del parámetro  $r$ , la DE ya no interfiere en el cálculo del grado de irregularidad de las señales. Además, en este caso, no fue posible discriminar entre ambos grupos utilizando únicamente la DE de cada grupo ya que es similar para ambos, mientras que sí fue posible con el índice *SampEn*. La Tabla 4 y la Figura 1 muestran los resultados del índice *SampEn* para cada grupo de edad. Se encontró que con el parámetro  $r$  obtenido por el método 2 era posible cuantificar la pérdida de variabilidad debida a la edad en sujetos sanos, lo cual no fue posible con el parámetro  $r$  obtenido por el método convencional, debido a que no mostró ninguna una tendencia fija (creciente o decreciente), esta cuantificación tampoco fue posible utilizando únicamente la DE de cada grupo de edad. La regresión lineal (ver Tabla 5) no mostró correlación entre la edad y la variación de los valores de la *SampEn* para los sujetos con IC.

#### 5. Conclusiones

Utilizando un valor umbral fijo para  $r$  en el algoritmo de la *SampEn* se ha obtenido una representación más clara de la pérdida de la VFC para pacientes con IC, así como del decrecimiento gradual de la VFC y de la *SampEn* con



**Figura 1.** Evolución de la *SampEn* con la edad. Los círculos representan el valor de la *SampEn* para cada sujeto, los cuadrados la *SampEn* media para grupo de edad, las barras la desviación estándar de cada grupo de edad y la línea recta es la recta de regresión. (a) Para *r* fijo. (b) Para *r* como porcentaje de la desviación estándar de cada registro.

	SampEn(2,20)	SampEn(2,0.2DE)
Jóvenes	0.87±0.23	0.66±0.18
Mayores	0.61±0.13*	0.43±0.14*

**Tabla 3.** Media ± DE de la *SampEn* para ambos métodos. El símbolo \* indica variación significativa ( $p < 0.001$ ) entre sujetos jóvenes y sujetos mayores.

Edad(años)	SampEn(2,20)	SampEn(2,0.2DE)
20-30	0.91±0.29	0.61±0.13
31-40	0.86±0.21	0.68±0.21
41-50	0.76±0.24	0.63±0.20
51-60	0.67±0.16	0.58±0.07
61-70	0.61±0.15	0.42±0.14
71-80	0.54±0.20	0.47±0.16

**Tabla 4.** Evolución de la *SampEn* con la edad. Media ± DE para ambos métodos.

	Var. índice/año	$r^2$
<b>Sanos</b>		
SampEn(2,20)	-0.0085	0.3653*
SampEn(2,0.2DE)	-0.0065	0.2935*
<b>IC</b>		
SampEn(2,20)	-0.0011	0.0036
SampEn(2,0.2DE)	-0.0125	0.1857

**Tabla 5.** Resultados de la regresión lineal de la *SampEn* respecto a la edad. El símbolo \* indica variación significativa ( $p < 0.001$ ).

la edad. Asimismo, se han obtenido indicios de que podría construirse una curva de pérdida de la VFC debida a la edad para sujetos sanos. Para este último propósito serán necesarios futuros estudios con un número mayor de registros clínicos.

### Agradecimientos

Este trabajo ha sido parcialmente financiado por el Proyecto de Investigación TEC-2007-68096-C02-TCM del Ministerio de Educación y Ciencia.

### Referencias

- [1] Pincus SM. Approximate entropy as a measure of system complexity Proc. Natl. Acad. Sci. 1991;88:2297-2301.
- [2] Pincus SM, Goldberger AL. Physiological time-series analysis: what does regularity quantify? Am J Physiol Heart Circ Physiol 1994;35:1643-1656.
- [3] Pincus SM. Assessing serial irregularity and its implications for health. Annals New York Academy of Sciences 2001; 954:245-267.
- [4] Marques-de S JP. Characterization of fetal heart rate using approximate entropy. IEEE Computers in Cardiology 2005;32:671-673.
- [5] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy. Am J Physiol Heart Circ 2000;278:2039-2049.
- [6] Signorini, M. and Ferrario, M. and Marchetti, M. and Mar-seglia, A. Nonlinear analysis of heart rate variability signal for the characterization of cardiac heart failure patients. Conf Proc IEEE Eng Med Biol Soc 2006;1(1):3431-3434.
- [7] Costa, Madalena and Goldberger, Ary L. and Peng, C.-K. Multiscale entropy analysis of biological signals. Phys. Rev. E 2005; 71:021906.
- [8] Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, 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 2000 101(23):e215-e220.
- [9] Malik M, Cripps T, Farrell T, Camm A. Prognostic value of heart rate variability after myocardial infarction. a comparison of different data-processing methods. Medical and Biological Engineering and Computing 1989;27:603-621.