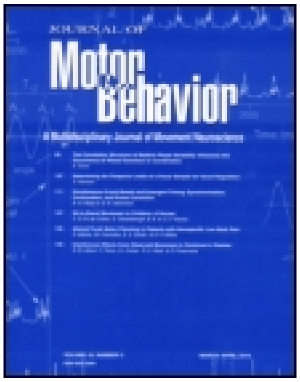

Journal of Motor Behavior
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/vjmb20

# Approximate Entropy Normalized Measures for Analyzing Social Neurobiological Systems

Sofia Fonseca [a], João Milho [b c], Pedro Passos [d e], Duarte Araújo [d e] & Keith Davids [f]

[a] Faculty of Physical Education and Sports, Lusófona University, Lisbon, Portugal

[b] IDMEC/IST-Institute of Mechanical Engineering/Instituto Superior Técnico, Technical University of Lisbon, Portugal

[c] ISEL-Instituto Superior de Engenharia de Lisboa, Polytechnic Institute of Lisbon, Portugal

[d] Faculty of Human Kinetics, Technical University of Lisbon, Portugal

[e] CIPER-Interdisciplinary Centre for the Study of Human Performance, Faculty of Human Kinetics, Portugal

[f] School of Human Movement Studies, Queensland University of Technology, Brisbane, Australia

Version of record first published: 18 Apr 2012.

To cite this article: Sofia Fonseca , João Milho , Pedro Passos , Duarte Araújo & Keith Davids (2012): Approximate Entropy Normalized Measures for Analyzing Social Neurobiological Systems, Journal of Motor Behavior, 44:3, 179-183

To link to this article: http://dx.doi.org/10.1080/00222895.2012.668233

Journal of Motor Behavior, Vol. 44, No. 3, 2012


***RESEARCH ARTICLE***

# Approximate Entropy Normalized Measures for Analyzing Social Neurobiological Systems

**Sofia Fonseca[1], João Milho[2,3], Pedro Passos[4,5], Duarte Araújo[4,5], Keith Davids[6]**
[1]Faculty of Physical Education and Sports, Lusófona University, Lisbon, Portugal. [2]IDMEC/IST–Institute of Mechanical Engineering/Instituto Superior Técnico, Technical University of Lisbon, Portugal. [3]ISEL–Instituto Superior de Engenharia de Lisboa, Polytechnic Institute of Lisbon, Portugal. [4]Faculty of Human Kinetics, Technical University of Lisbon, Portugal. [5]CIPER–Interdisciplinary Centre for the Study of Human Performance, Faculty of Human Kinetics, Portugal. [6]School of Human Movement Studies, Queensland University of Technology, Brisbane, Australia.



**ABSTRACT.** When considering time series data of variables describing agent interactions in social neurobiological systems, measures of regularity can provide a global understanding of such system behaviors. Approximate entropy (ApEn) was introduced as a nonlinear measure to assess the complexity of a system behavior by quantifying the regularity of the generated time series. However, ApEn is not reliable when assessing and comparing the regularity of data series with short or inconsistent lengths, which often occur in studies of social neurobiological systems, particularly in dyadic human movement systems. Here, the authors present two normalized, nonmodified measures of regularity derived from the original ApEn, which are less dependent on time series length. The validity of the suggested measures was tested in well-established series (random and sine) prior to their empirical application, describing the dyadic behavior of athletes in team games. The authors consider one of the ApEn normalized measures to generate the 95th percentile envelopes that can be used to test whether a particular social neurobiological system is highly complex (i.e., generates highly unpredictable time series). Results demonstrated that suggested measures may be considered as valid instruments for measuring and comparing complexity in systems that produce time series with inconsistent lengths.

*Keywords*: analysis of regularity, entropy measures, social neurobiological systems, time series

Approximate entropy (ApEn) was first introduced in 1991 by Pincus as a nonlinear measure to quantify regularity in the behaviors of complex systems. The regularity of a signal relates to the complexity of the system generating it (Pincus, 1995), thus, the greater the value of ApEn, the lower the regularity of the time series, and the greater the complexity of the system under study. ApEn values vary between 0 and 2, with high values identifying data series with less regular and predictable patterns, and low values associated with data series containing many repetitive patterns (i.e., data which are more regular and more predictable). Since its introduction, ApEn has been established as a measure of regularity in a time series, with numerous applications in analysis of physiological time series such as heart rate variability, electrocardiogram measures, respiration, anesthesia, gene sequences, pulse waveform, and electroencephalography (Xu, Wang, & Wang, 2005).

A major interest when analyzing the complexity of physiological systems is to compare the regularity of a given time series between different groups, for example, compare the ApEn of pulse data records in healthy persons, inpatients with cardiovascular disease, and inpatients without any cardiovascular disorder (Wang et al., 2003). However, given that ApEn values are highly dependent on times series length, and are particularly unstable for short time series (e.g., Pincus & Golberger, 1994; Richman, 2007; Xu et al., 2005), the application of such a regularity measure is only recommended when considering signals of the same length, preferably with at least 50 data points (Stergiou, Buzzi, Kurz, & Heidel, 2004). To ensure such conditions, when considering physiological time series (e.g., heart rate variability, pulse), individuals are monitored during a fixed amount of time and data are collected at the same rate (Pincus, Padmanabhan, Lemon, Randolph, & Midgley, 1998; Pincus & Viscarello, 1992; Ryan, Goldberger, Pincus, Mietus, & Lipsitz, 1994; Wang et al., 2003).

When the previous conditions cannot be guaranteed, modified measures of the original ApEn can be applied (e.g., sample entropy [Richman & Moorman, 2000], Gaussian Kernel approximate entropy [Xu et al., 2005], modified sample entropy [Xie, He, & Lui, 2008], Fuzzy approximate entropy [Chen, Zhuang, Yu, & Wang, 2008]). These measures have been shown to be less dependent on record length and more stable for short series.

In the study of social neurobiological systems, such as flocking birds, schooling fish, herding animals, human societies, and sports teams (Couzin, 2007; Sumpter, 2006), unlike physiological systems, it may not be possible to ensure that all system output samples are of the same length. This is particularly difficult in studying social neurobiological systems because of the continuous interactions of system agents in tasks where a specific performance goal has to be achieved. Because the length of the captured time series is dependent on the time required by the agents to conclude a particular performance task (as exemplified by an attacking or defending performance subphase in a team game), the use of ApEn for assessing regularity is not advisable. Modified measures of regularity, such as those mentioned previously, could be applied here; however, we suggest in this article two normalized measures of the original ApEn. By applying these new measures it is possible to compute a straightforward normalization of any ApEn value in which the original ApEn

*Correspondence address: Sofia Fonseca, Faculty of Physical Education and Sports, Lusófona University, Campo Grande 376, 1749–024 Lisboa, Portugal. e-mail: sofiaccfonseca@gmail.com*

was used, which allows a reliable comparison of time series regularity in different complex systems.

## Method

Given a data series with N points, say $\{x_1, x_2, \ldots, x_N\}$, ApEn ($m$, $r$, N) can be used to measure the logarithmic likelihood that runs of patterns with $m$ points that are close, remain close within a tolerance factor $r$ in ensuing incremental comparisons (Pincus, 1991; i.e., to measure the predictability of the data series). To compute ApEn ($m$, $r$, N), the parameters $m$, the length of compared runs, and $r$, the tolerance factor, need to be fixed for all calculations to ensure reliable analysis (Pincus, & Goldberger, 1994). In our analysis, as suggested in studies of other neurobiological systems, we considered $m = 2$ and $r = 0.2$. All calculations were performed in Matlab (version 7.6.0, The MathWorks, Natick, MA, USA) using routines written for this purpose (Kaplan & Staffin, 2009).

The techniques for normalization considered here are based on the ratio between an observed ApEn value and a threshold reference ApEn value, for a specific data series length. This normalization allows the regularity of data series of different lengths to be compared.

Our first normalized measure, designated $\text{ApEn}_{\text{RatioRandom}}$, is given by

$$\text{ApEn}_{\text{RatioRandom}} = \frac{\text{ApEn}(2, 0.2, N)_X}{\sum_{i=1}^{100} \text{ApEn}(2, 0.2, N)_{U_i} \Big/ 100} \qquad (1)$$

Here, the regularity of the data series $X = \{x_1, x_2, \ldots, x_N\}$ is quantified by means of the ratio between its original ApEn value, ApEn (2, 0.2, N)$_X$, and the mean ApEn calculated in 100 random series $U_i$ with the same length N. Note that for each generated random series, $U_i$, the corresponding ApEn, ApEn(2, 0.2, $N$)$_{U_i}$, represents a maximum value of approximate entropy for that particular length. Hence, ApEn (2, 0.2, N)$_X$ is normalized with respect to a maximum value of ApEn of a series of length N.

Our second normalized measure, designated $\text{ApEn}_{\text{RatioShuffle}}$, is given by

$$\text{ApEn}_{\text{RatioShuffle}} = \frac{\text{ApEn}(2, 0.2, N)_X}{\sum_{i=1}^{100} \text{ApEn}(2, 0.2, N)_{S_i} \Big/ 100} \qquad (2)$$

Here, the regularity of the data series $X = \{x_1, x_2, \ldots, x_N\}$ is given by the ratio between its original ApEn value, ApEn (2, 0.2, N)$_X$, and the mean ApEn calculated in 100 shuffled replicas $S_i$ of the original data. Note that for each shuffled replica of X, $S_i$, the corresponding approximate entropy, ApEn(2, 0.2, $N$)$_{S_i}$, represents a maximum value of approximate entropy for that particular set of points. Hence, ApEn (2, 0.2, N)$_X$ is normalized with respect to a maximum value of ApEn of that particular set of points. In both methods described here, low values of the corresponding measures will indicate that the time series under study is generated by a social neurobiological system that is less predictable than random time series of the same length.

For testing the methods presented in this article, we considered data from a dyadic human movement system; more precisely, a rugby union attacker–defender system in which the attacker aims to score and the defender tries to prevent it. Results should be in accordance with findings in the literature that suggest that physical contact between an attacker and defender increases the complexity of this system (Passos et al., 2009), making the dyadic subsystem behaviors that emerge in try situations (success for the attacker) more predictable than in tackle situations (success for the defender) in which players do experience physical contact.

In this regard, the interactive behaviors that emerges in each trial of this social neurobiological system is accurately measured, across its duration, by a one-dimensional variable X defined in previous work by Passos et al. (2009) and designated as collective variable. This variable represents the vector connecting the agents in the dyad, and is formally given by the value of the angle between the defender–attacker vector and a horizontal line parallel to the try line with the origin in the defender. The values of X range from –90° to 90°, which occur when an attacker and defender are in the same vertical position, being 90° when the defender is closer to the try line and –90° when the attacker is closer to the try line. X is zero when attacker and defender are in the same horizontal position.

To assess the regularity of this collective variable, we considered 47 experimental dyadic trials in which participants were male rugby players aged 11–12 years, with an average of $4.0 \pm 0.5$ years of rugby practice. Treatment of participants was in accordance with the ethical standards of American Psychological Association. Trials were performed on a field of 5 m width × 10 m depth and two fixed digital video cameras at 25 Hz were used to capture players' movements. The angle given by the variable X was calculated from players' trajectory motion data extracted from the videos using the methodology described in detail in Passos et al. (2009). Figure 1 displays two examples of these data, one from a successful situation (try scored) and the other from an unsuccessful situation (try not scored).

The 47 data series analyzed, try scored ($n = 20$) and try not scored ($n = 27$), had a record length ranging from 69 to 230 data points ($112 \pm 36.3$). Both normalized measures of ApEn were calculated and comparative statistical analyses were performed using nonparametric tests (Mann-Whitney test) due to lack of normality in the data and the small sample size. The level of statistical significance was fixed at 5%.

## Results

The normalized measures of ApEn suggested in this article, $\text{ApEn}_{\text{RatioRandom}}$ and $\text{ApEn}_{\text{RatioShuffle}}$, were tested with regard to the series length effect. An application of these two well-known data series (sine and random) with different

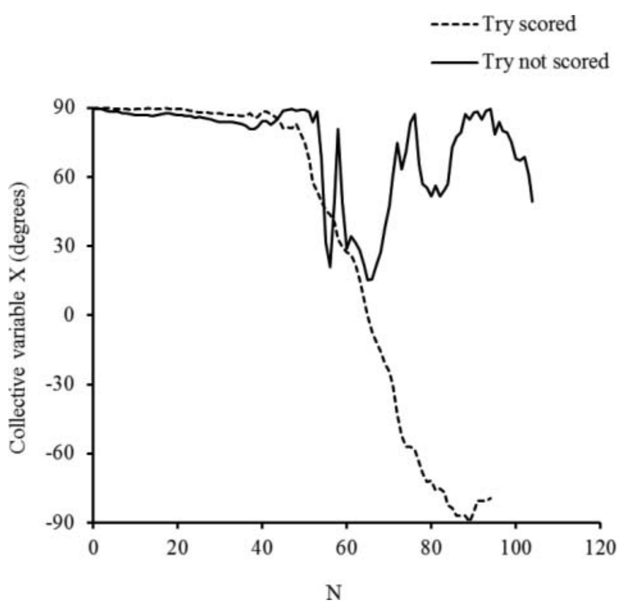


**FIGURE 1.** Example data for the collective variable X measured in a successful trial (try scored) and in a unsuccessful trial (try not scored).

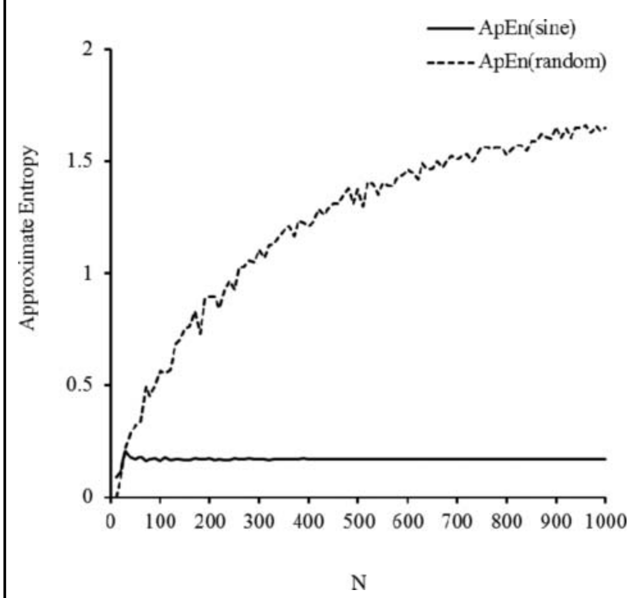


**FIGURE 3.** Original ApEn calculated for sine and random series data of different lengths (N).

lengths has shown the advantages of these (Figure 2) in comparison with the original ApEn measure (Figure 3).

Both normalized measures appeared to be less dependent on record length for both data series, reaching stability for small lengths. This observation reinforces the need of considering more reliable measures for analyzing complexity in systems that produce time series with inconsistent lengths, a typical occurrence when studying social neurobiological systems. Nevertheless, a minimum of 50 data points is also advised to allow reliable approximate entropy comparisons (Stergiou et al., 2004). In a specific application of these measures to a dyadic subsystem (1 vs. 1), interaction in the team sport of rugby union, in which physical contact is associated with less regular interaction behaviors, both ApEn normalized measures indicated, accordingly, greater unpredictability in situations with effective contact between the players (i.e., an attacker was tackled by an opposing defender (try not scored; see Figure 4).

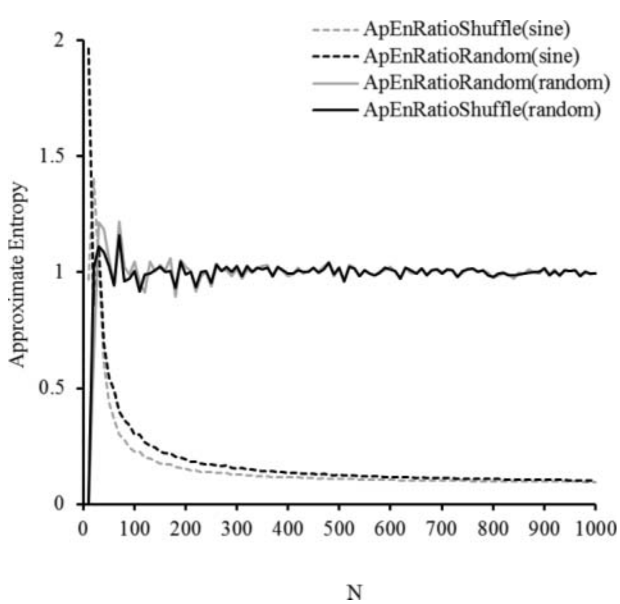


**FIGURE 2.** Normalized entropy measures calculated for sine and random series data of different lengths (N).

Using the nonparametric Mann-Whitney test, significant differences were found between the two task outcomes for $\mathrm{ApEn}_{\mathrm{RatioRandom}}$ ($p = .0196$) and $\mathrm{ApEn}_{\mathrm{RatioShuffle}}$ ($p = .0185$), confirming that behavioral outcomes in try situations are more regular than tackle situations.

Given the similarity of both measures, we considered the $\mathrm{ApEn}_{\mathrm{RatioRandom}}$ to determine the 95th percentile envelope of this normalized measure, calculated from 100 simulations of random data series of length from 50 to 1000 (Figure 5).

The logarithm curves fitted to the upper (U) and lower (L) bounds of the 95th percentile of the $\mathrm{Apen}_{\mathrm{RatioRandom}}$ for random time series with length greater than 50 are given by

$$\mathrm{ApEn}_{\mathrm{RatioRandom}}\Big|_{U}^{95th} = -0.09 \ln(N) + 1.6089 \quad (3)$$

$$\mathrm{ApEn}_{\mathrm{RatioRandom}}\Big|_{L}^{95th} = 0.0845 \ln(N) + 0.4233 \quad (4)$$

with a corresponding $R^2$ for the logarithm fitting of .752 and .742, respectively.

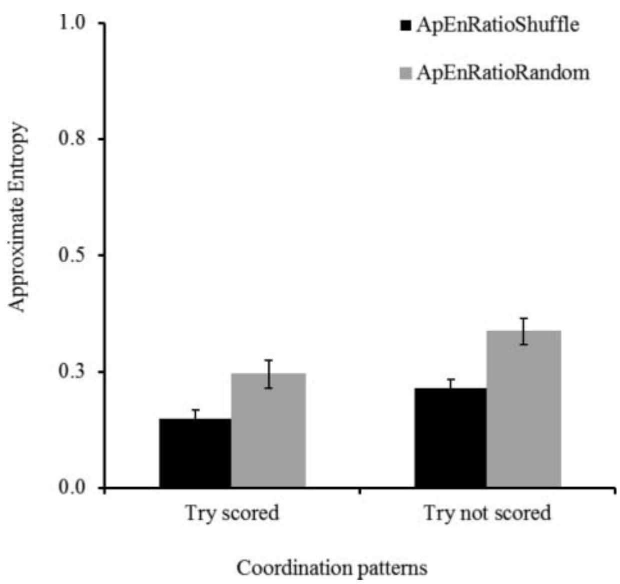


**FIGURE 4.** Mean approximate entropy for each of the two task outcomes using $Apen_{RatioRandom}$ and $ApEn_{RatioShuffle}$.

Given these, deviations from complete behavioral randomness (i.e., high unpredictability) observed in a specific social neurobiological system could be tested by computing the median $ApEn_{RatioRandom}$ for a sample of time series of that system to verify whether the obtained value is within the envelopes estimated for N equal to the median of dimension of the time series considered. For the social neurobiological system considered in this study, the median of the time series dimension was 98 and 105 for try and no-try situations and therefore the envelopes are [0.81, 1.2] and [0.82, 1.19], respectively. The median $ApEn_{RatioRandom}$ values in try and no-try situations were 0.23 and 0.33, with both below the respective lower reference value. This finding suggests that, regardless of the outcome, the dyadic system behavior under study is more predictable than would be expected in the case of complete randomness. Nevertheless, results suggested that the level of system output regularity was significantly different between the try and no-try performance situations, being more predictable for try situations.

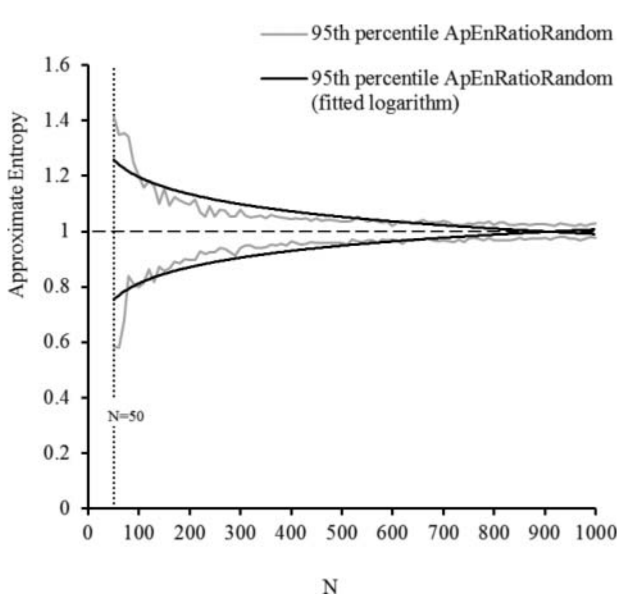


**FIGURE 5.** The 95th percentile envelopes of $ApEn_{RatioRandom}$ for random series of different lengths (N) and the fitted logarithm curves for the upper and lower bounds.

## Discussion

In this article we presented two normalized measures based on the original ApEn for quantifying and comparing regularity in the interactions of agents in social neurobiological systems, particularly in those that produce time series with inconsistent lengths. The limitations associated with the application of the original ApEn to time series of varying lengths, have been previously addressed by other authors (Chen et al., 2008; Richman & Moorman, 2000; Xie et al., 2008; Xu et al., 2003) introducing modified measures of the original ApEn. Alternatively, the measures here presented consider the same limitations but are based on the use of the original ApEn.

We considered two well-known data series (sine and random) with different lengths, for testing the advantages of these normalized measures in comparison with the original ApEn measure. For the normalized measures we calculate the 95th percentile envelopes, which can be interpreted as reference values for testing deviations from complete randomness (i.e., low predictability) in social neurobiological time series of any length greater than 50. An application of these measures to empirical data from a dyadic system behavior in rugby union suggested that the emergent behavior of this particular social neurobiological system is more regular than expected in the case of complete randomness, given that the agents in this system have a specific performance goal. Additionally, the analysis of regularity indicates that the complexity of this system was significantly lower when physical contact between the two players occurred, as suggested by Passos et al. (2009). Overall, the application of the normalized ApEn measures to theoretical (sine and random) and empirical data suggest that they can be regarded as reliable measures for quantifying and comparing regularity of time series with different lengths. These findings could be used to reinterpret previous work on behaviors of social neurobiological systems (e.g., Araújo, Davids, Bennett, Button, & Chapman, 2004) with criteria to compare the regularity of time series of different lengths, something that was not possible previously beyond simple visual inspection. Moreover, an exciting possibility for future researchers is to study complex daily social interaction behaviors to identify different patterns, without concerns over the possible loss of explanatory power.

## REFERENCES


Araújo, D., Davids, K., Bennett, S., Button, C., & Chapman, G. (2004). Emergence of sport skills under constraints. In A. M. Williams & N. J. Hodges (Eds.), *Skill acquisition in sport: Research, theory and practice* (pp. 409–433). London, England: Routledge, Taylor & Francis.

Chen, W., Zhuang, J., Yu, W., & Wang, Z. (2008). Measuring complexity using FuzzyEn, ApEn and SampEn. *Medical Engineering & Physics*, *31*(1), 61–68.

Couzin, I. D. (2007). Collective minds. *Nature*, *445*, 715.

Kaplan, D., & Staffin, P. (2009). Heart rate variability [Software]. Retrieved from http://www.macalester.edu/∼kaplan/hrv/doc/download.html

Passos, P., Araújo, D., Davids, K., Gouveia, L., Serpa, S., Milho, J., & Fonseca, S. (2009). Interpersonal pattern dynamics and adaptive behavior in multi-agent neurobiological systems: A conceptual model and data. *Journal of Motor Behavior*, *41*, 445–459.

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences*, *88*, 2297–2301.

Pincus, S. M. (1995). Approximate entropy (ApEn) as a complexity measure. *Chaos*, *5*, 110–117.

Pincus, S. M., & Goldberger, A. (1994). Physiological time-series analysis: What does regularity quantify? *American Journal of Physiology–Heart and Circulatory Physiology*, *266*, H1643–H1656.

Pincus, S. M., Padmanabhan, V., Lemon. W., Randolph, J., & Midgley, A. R. (1998). Follicle stimulating hormone is secreted more irregularly than luteinizing hormone in both humans and sheep. *Journal of Clinical Investigation*, *101*, 1318–1324.

Pincus, S. M., & Viscarello, R. (1992). Approximate entropy: A regularity measure for fetal heart rate analysis. *Obstetrics & Gynecology*, *79*, 249–255.

Richman, J. S. (2007). Sample entropy statistics and testing for order in complex physiological signals. *Communications in Statistics–Theory and Methods*, *36*, 1005–1019.

Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology–Heart and Circulatory Physiology*, *278*, H2039–H2049.

Ryan, S. M., Goldberger, A. L., Pincus, S. M., Mietus, J., & Lipsitz, L. A. (1994). Gender- and age-related differences in heart rate dynamics: Are women more complex than men? *Journal of the American College of Cardiology*, *24*, 1700–1707.

Stergiou, N., Buzzi, U., Kurz, M., & Heidel, J. (2004). Nonlinear tools in human movement. In N. Stergiou (Ed.), *Innovative analyses of human movement* (pp. 63–87). Champaign, IL: Human Kinetics.

Sumpter, D. (2006). The principles of collective animal behaviour. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *361*, 5–22.

Wang, K., Xu, L., Li, Z., Zhang, D., Li, N., & Wang, S. (2003). Approximate entropy based pulse variability analysis. *Proceedings of the 16th IEEE Symposium on Computer-Based Medical Systems*, 236–241.

Xie, H. B., He, W. X., & Lui, H. (2008). Measuring time series regularity using nonlinear similarity-based sample entropy. *Physics Letters A*, *372*, 7140–7146.

Xu, L. S., Wang, K. Q., & Wang, L. (2005). Gaussian kernel approximate entropy algorithm for analyzing irregularity of time series. *Proceedings of the 4th International Conference of Machine Learning and Cybernetics*, *9*, 5605–5608.


*Received December 27, 2010*
*Revised February 5, 2012*
*Accepted February 16, 2012*