

1

2 **ENTROPY IN THE ANALYSIS OF GAIT COMPLEXITY: A STATE OF THE ART**3 **Abstract**

4 Human gait is a non-linear complex process requiring appropriate mathematical measuring
5 tools, one of which is entropy. Entropy is a measure that quantifies regularity in time series:
6 the more predictable a series is the lower the entropy value.
7 The mathematical methods used to estimate entropy have evolved over time. At present,
8 three algorithms are used to study physiological signals: the approximate entropy (ApEn), the
9 sample entropy (SampEn), and the multi-scale entropy (MSE).
10 Most studies on human gait complexity have been conducted on elderly subjects or subjects
11 with specific disorders affecting gait patterns and they used ApEn; but, because of a set of
12 conceptual errors, the ApEn is not the most appropriate algorithm for the analysis of biological
13 signals. Very few studies use SampEn or MSE to analyze human gait variability, but they agree
14 that these algorithms might contribute new perspectives in the analysis of human gait and that
15 MSE seems to be the most sensitive algorithm to changes in gait in healthy subjects.

16

17 **KEYWORDS:** entropy, human gait, complexity, multiscale.

18

19 **Introduction**

20 Physiological signs present complex and irregular fluctuations that cannot be analyzed with
21 conventional statistical analysis methods, as these methods yield very limited information on
22 behavior patterns.

23 There is not any clear definition of “complexity”. Generally, a complex signal accounts for one
24 (normally more than one) of the following characteristics (1): a) *Non-linearity*; complex
25 systems are composed of multiple subunits that should not be separately analyzed since they
26 may interact with each other; b) *non-stationarity*: the statistical properties of the system input
27 change with time; c) *time irreversibility* or *asymmetry*: the system operation is not balanced; d)
28 *multi-scale variability*: the system exhibits space-time patterns in a range of scales.

29 Human gait is considered a complex, non-linear process (2-4) by which the locomotor system
30 incorporates input from the cerebellum, the motor cortex, and the basal ganglia, as well as
31 feedback from visual, vestibular, and proprioceptive sensors. Traditionally, under healthy
32 conditions, the locomotor system is thought to produce a remarkably stable walking pattern;
33 the kinetics, kinematics, and muscular activity of gait appear to remain relatively constant from
34 one step to the next, even during unconstrained walking (4,9). For this reason, most
35 conventional biomechanical studies are based on the thorough analysis of a walking cycle. The
36 data obtained are then extrapolated into the whole walking process. However, a range of
37 studies, conducted using the non-linear dynamic approach, have revealed that gait patterns
38 present fluctuations even under apparently stable conditions (10-13). Thus, human gait
39 dynamics have a complex behavior that many studies have attempted to elucidate (14) using
40 practical applications mainly focused on aging and pathologies affecting human walking.

41 There are many mathematical methods that have proven useful in examining the complexity of
42 biological signs such as stride intervals (15), Detrended Fluctuation Analysis (DFA) (16-18),
43 power law scaling by Fourier’s method (19), the Lyapunov exponent (20) or entropy (22).

44

45 **Concept of Entropy. Algorithms for Estimating Entropy**

46 The concept of “entropy” was first developed in classic thermodynamics as a measure of the
47 molecular disorder within a closed system.

[Escribir texto]

48 Claude Shannon (1916 – 2001) applied the concept of entropy to the theory of information.
49 Shannon was an electrical engineer and a mathematician, and he is considered the father of
50 the so-called “Theory of Information”, which states that entropy –known as entropy of
51 information or Shannon’s entropy– is “a measure of the information that is lost before being
52 perceived by the receptor”. It is useful in measuring uncertainty or variability.

53 In the field of non-linear dynamic systems, entropy quantifies the regularity of a system: the
54 more predictable a series is the lower the entropy value. Thus, a range of studies have
55 reported that aging reduces entropy and impaired systems exhibit lower entropy than healthy
56 systems (24).

57 The more regular a series, the more predictable and the less complex it will be, which is
58 indicative of a less adaptive system. Therefore, in any time series accounting for a system
59 output variable, entropy is a measure of its uncertainty.

60 The mathematical methods used to estimate entropy have evolved over time. At present,
61 there are three algorithms: the approximate entropy (ApEn), the sample entropy (SampEn),
62 and the multi-scale entropy (MSE).

63 In 1991, Pincus introduced ApEn (22) as a tool to measure regularity in a time series under the
64 assumption that a series is regular when it contains repetitive patterns (25,26). The ApEn
65 shows the probability that “similar” observation patterns are not accompanied by “similar”
66 observations. A time series containing many repetitive patterns yield lower values of ApEn,
67 while a complex series (i.e. without repetitive patterns and, consequently, poorly regular and
68 less predictable) will yield greater values of ApEn.

69 The SampEn was incorporated by Richmann and Moorman (27) to correct a set of errors to
70 which the ApEn leads. As the ApEn counts each sequence as matching itself, in a time series it
71 yields a degree of similarity greater than the real one. Thus, the bias of the ApEn makes the
72 results it yields inconsistent. The SampEn is defined as the negative natural average logarithm
73 of the conditional probability that two sequences that are similar for m points remain similar
74 when the number of points is increased to $m+1$.

75 The calculation is as follows:

76 Given a sequence of N measures, $U_N = \{u_1, u_2, \dots, u_N\}$, where $x_m(i)$ y $x_m(j)$ are two vectors of
77 U_N , both with a length of m . In the sequence $x_m(i)$ the vector starts at the u_i element of the
78 series, and in the sequence $x_m(j)$, the vector starts at the u_j element. The $d[x_m(i), x_m(j)]$ distance

[Escribir texto]

79 between two vectors of $x_m(i)$ and $x_m(j)$ is defined as the maximum difference between their
 80 respective components. Consequently, the two vectors will be similar if $d[x_m(i), x_m(j)] < r$, where
 81 r is the parameter defining the criterion of similarity.

82 Let X_m be the set of all vectors with a length of m within U_N (i.e. $x_m(1), x_m(2), \dots, x_m(N-m+1)$).
 83 Given a vector $x_m(i)$, count the number of vectors $x_m(j)$, where $1 \leq j \leq N-m$, so $d[x_m(i), x_m(j)] < r$.
 84 Let B_i be the number of vectors. Thus, define the function $1 \leq i \leq N-m$ as:

$$85 \quad B_m(i) = \frac{B_i}{N-m}$$

86 The probability that two vectors of m points match will be:

$$87 \quad B_m = \frac{1}{N-m} \sum_{i=1}^{N-m} B_m(i)$$

88 The same procedure applies to $m+1$ to obtain $A_m(i)$ and A_m

89 Let *SampEn* be defined as:

$$90 \quad \text{SampEn}(m, r) = \lim_{N \rightarrow \infty} \left\{ -\ln \left[\frac{A_m}{B_m} \right] \right\}$$

91

92 As N is a finite number:

$$93 \quad \text{SampEn}(m, r) = -\ln \left[\frac{A_m}{B_m} \right]$$

94

95 The selection of m and r is key to estimating the *SampEn*. The best values are $m=1$ or $m=2$
 96 (preferably $m=2$) and r between 10 and 25% of the standard deviation of the time series. This
 97 is due to the fact that r must be at least greater than the noise contaminating the signal;
 98 furthermore, it cannot have a very high value, as we would miss a great part of the signal
 99 information content.

100 As it is said above, *ApEn* and *SampEn* measure the degree of regularity of a time series.
 101 However, when the calculations above are applied to physiological data, impaired systems
 102 sometimes yield greater values of entropy than healthy systems (28). This is illogical, since an
 103 impaired system is less adaptive, less complex and more regular than a healthy system and it
 104 should yield a lower value of entropy. An explanation to this might be that both, *ApEn* and

[Escribir texto]

105 SampEn, are calculated on a single scale (29) where the structure and organization in higher
106 scales of the series is not considered.

107 To solve this, Madalena Costa introduced the concept of MSE (30). Given a discrete time
108 series, we generate new series which points are average values of k sequential elements of the
109 original series (where k is the order of the scale) with no overlapping. Thus, for a time scale =
110 1, we have the original series; for a scale = 2 the new series will be composed of the average
111 value of the elements taken in pairs on each scale. Finally, we calculate the SampEn for each of
112 the new series generated accounting for values obtained against the scale factor. Therefore,
113 entropy is highly dependent on the time scale.

114 The maximum value of the scale will depend on the length of the series. Previous studies (31)
115 take a maximum scale factor of 20 with time series of a length of 20 000 data points. The MSE
116 method uses the same statistical formula as SampEn. Therefore, the results lose consistency as
117 the number of data points decreases.

118

119 **Applications of Entropy to the Analysis of Human Gait Complexity**

120 Most studies on human gait complexity have been conducted on elderly subjects, patients
121 with disorders affecting walking or after anterior cruciate ligament reconstruction.

122 Kurz and Stergiou³² used the ApEn to elucidate whether the neurophysiological changes
123 associated with aging have an impact on the nervous system operation in order to give the gait
124 more stability and functionality. The yielded results supported the hypothesis that aging is
125 associated with a loss in the ability of the neuromuscular system to move the lower limbs
126 during walking. They hypothesized that such changes might yield inappropriate input from
127 visual, vestibular, and somatosensory sensors (proprioceptive, cutaneous and articular).
128 Therefore, an aged neuromuscular system provides inappropriate input that hinders correct
129 walking. This might be the cause of the frequent falls in the elderly.

130 Later, Khandoker et al (33) used the ApEn to study the risk for falls in the elderly by analyzing
131 the variability of gait. On such purpose, they compared input from a group of healthy elderly
132 (control group) against the input from a group of elderly subjects with a history of falls. The
133 study demonstrated that the ApEn of elderly subjects with a history of falls was significantly
134 greater than that of the control group (0.18 and 0.13, respectively), which is indicative of an
135 increase in irregularities and randomness in their gait patterns and an indication of loss of gait
136 control mechanisms. Therefore, they concluded that gait variability –as analyzed by ApEn–

[Escribir texto]

could be useful for the early diagnosis of at-risk gait and for the detection and prevention of falls in elderly subjects.

In the field of pathology, Moraiti et al (34) studied gait variability after anterior cruciate ligament reconstruction (ACL). The results obtained showed that subjects undergoing ACL surgery presented ApEn values significantly greater than those of the control group (0.24 and 0.30, respectively).

The same authors (21) analyzed gait variability in subjects undergoing surgery for ACL tear. They found that the ApEn values were significantly greater in the injured leg as compared to the healthy leg. This comparative study was performed on a treadmill at different speeds and, although the difference between the operated leg and the healthy leg remained, the authors observed that ApEn values increased as speed increased.

As it was said above, because of a set of conceptual errors, the ApEn is not the most appropriate algorithm for the analysis of biological signals. In fact, some studies (35) yield consistent changes in ApEn values from a physiological perspective, but give numerical values that are characteristic of a period time series.

We only found three studies where either the SampEn or the MSE were used in the analysis of gait variability. Firstly, Tochigi et al (36) applied the SampEn to analyze the cycle-to-cycle variability in leg acceleration signals during walking in elderly subjects and in adults with symptomatic knee osteoarthritis. The authors found that the study group showed a lower variability than the control group. Secondly, Costa et al (37) used the MSE to analyze human gait variability by a pressure sensor installed in the footwear in different experimental situations (normal spontaneous walking at higher and lower speed, and walking controlled with a metronome). On such purpose, they asked a sample of ten healthy young subjects aged between 18-29 years to walk for one hour on the floor. The authors observed that normal, spontaneous free walking has the highest complexity as compared to the other forms of walking.

Sánchez et al studied different forms of gait and running using the three entropy algorithms described above (38). In particular, they analyzed time series of stride interval in ten healthy young subjects in three experimental situations: walking at a comfortable speed across the floor for 25 minutes in a form chosen by the subject (Situation A); walking on a treadmill at the same speed as in Situation A for 25 minutes (Situation B), and running on a treadmill at a 8 Km/h speed for 20 minutes (Situation C). The authors found (table 1) that ApEn yielded values

[Escribir texto]

169 significantly lower than those obtained by the SampEn or MSE and no significant changes were
170 observed in ApEn when the experimental situations were compared (A vs B and B vs C).

171

172 In short, entropy –especially SampEn and MSE algorithms– might contribute new perspectives
173 in the analysis of human gait and might be useful in the evaluation of pathological situations of
174 the locomotor system. The MSE seems to be the most sensitive algorithm to changes in gait in
175 healthy subjects.

176

177

	Situation A		Situation B		Situation C		Significance level (p)	
	MEAN	SD	MEAN	SD	MEAN	SD	A vs B	B vs C
ApEn	0.529	0.109	0.403	0.362	0.377	0.451	0.146	0.449
SampEn	1.666	0.094	1.559	0.182	1.306	0.245	0.070	0.003
MSE 2	1.607	0.050	1.517	0.064	1.482	0.082	0.002	0.035
MSE 3	1.497	0.082	1.390	0.074	1.316	0.081	0.006	0.001
MSE 4	1.408	0.098	1.305	0.059	1.247	0.068	0.002	0.007
MSE 5	1.363	0.110	1.253	0.074	1.158	0.063	0.006	0.001

178

179 **Table 1:** ApEn, SampEn, MSE 2, MSE3 MSE4 and MSE5 values expressed as the mean and
180 standard deviation of the time series for each of the experimental situations: walking at a
181 comfortable speed across the floor for 25 minutes in a form chosen by the subject
182 (Situation A); walking on a treadmill at the same speed as in Situation A for 25 minutes
183 (Situation B), and running on a treadmill at a 8 Km/h speed for 20 minutes (Situation C).
184 Significance level $p < 0.05$. From (38) with permission.

185

186

187

188 REFERENCES

- 189 1. Costa M, Peng C-K, Goldberger AL. Multiscale analysis of heart rate dynamics: entropy
190 and time irreversibility measures. *Cardiovasc Eng* 2008; 8 (2): 88-93.
- 191 2. Hausdorf JM, Peng CK, Ladin Z, Wei JY, Goldberger AL. Is walking a random walk?
192 Evidence for long range correlations in the stride interval of human gait. *J Appl Physiol*
193 1995; 78:349-58.
- 194 3. Hausdorf JM, Purdon PL, Peng CK, Ladin Z, Wei JY, Goldberger AL. Fractal Dynamics of
195 human gait: Stability of long range correlations in stride interval fluctuations. *J Appl*
196 *Physiol* 1996; 80:1448-57.
- 197 4. Decker LM, Cignetti F, Stergiou N. Complexity and human gait. *Rev Andal Med Deporte*
198 2010; 3(1): 2-12.
- 199 5. Inman VT, Ralston HJ, Todd F. *Human Walking*. Baltimore, MD: Williams & Wilkins,
200 1981.
- 201 6. Kadaba, MP, Ramakrishnan HK, Wootten ME, Gainey J, Gorton G, Cochran GVB.
202 Repeatability of kinematic, kinetic and electromyographic data in normal adult gait. *J*
203 *Orthop Res* 1989; 7: 849-860.
- 204 7. Pailhous J, Bonnard M. Steady-state fluctuations of human walking. *Behav Brain Res*
205 1992; 47: 181-190.
- 206 8. Palta AE. Some characteristics of EMG patterns during locomotion: implications for
207 locomotor control processes. *J Motor Behav* 1985; 17:443-461.
- 208 9. Winter DA. Kinematic and kinetic patterns in human gait: variability and compensating
209 effects. *Hum Mou Sci* 1984; 3: 51-76.
- 210 10. Gabel A, Nayak USL. The effect of age on variability in gait. *J Gerontol* 1984; 39: 662 -
211 666.
- 212 11. Guimares RM, Isaacs B. Characteristics of the gait in old people who fall. *Int Rehab*
213 *Med* 1980; 2: 177-180.
- 214 12. Yamasaki M, Sasaki T, Torii M. Sex difference in the pattern pattern of lower limb
215 movement during treadmill walking. *Eur J Appl Physiol Occup Physiol* 1991; 62: 99-103.
- 216 13. Yamasaki M, Sasaki T, Tsuzki S, Torii M. Stereotyped pattern of lower limb movement
217 during level and grade walking on treadmill. *Ann Physiol Anthropol* 1984; 3: 291-296.
- 218 14. Van Emmerik REA, Rosenstein MT, McDermott WJ, Hamill J. Nonlinear Dynamical
219 Approaches to Human Movement. *J App Biomech* 2004; 20:396-420.

- 220 15. Stergiou N, Buzzi UH, Kurz MJ, Heidel J. Nonlinear Tools in Human Movement. In:
221 Stergiou N, editor. Innovative Analyzes of Human Movement. Champaign, IL: Human
222 Kinetics; 2004. p. 63-90.
- 223 16. Peng CK, Mietus J, Hausdorff JM, Havlin S, Stanley HE, Goldberger AL. Long-range
224 anticorrelations and non-Gaussian behavior of the heartbeat. *Phys Rev Lett*
225 1993;70:1343–6.
- 226 17. Hausdorff JM, Purdon PL, Peng CK, Ladin Z, Wei JY, Goldberger AL. Fractal dynamics of
227 human gait: stability of long-range correlations in stride interval fluctuations. *J Appl*
228 *Physiol* 1996; 80:1448-1457.
- 229 18. Hausdorff JM. Gait dynamics, fractals and falls: Finding meaning in the stride-to-stride
230 fluctuations of human walking. *Hum Mov Sci.* 2007; 26:555-89.
- 231 19. Hausdorff JM, Peng CK, Ladin Z, Wei JY, Goldberger AL. Is walking a random walk?
232 Evidence for long-range correlations in stride interval of human gait. *J Appl Physiol*
233 1995; 78:349-358.
- 234 20. Buzzi UH, Stergiou N, Kurz MJ, Hageman PA, Heidel J. Nonlinear dynamics indicates
235 aging affects variability during gait. *Clin Biomech* 2003; 18(5): 435-43.
- 236 21. Moraiti C, Stergiou N, Ristanis S, Georgoulis AD. ACL deficiency affects stride-to-stride
237 variability as measured using nonlinear methodology. *Knee Surg Sports Traumatol*
238 *Arthros* 2007; 15(12): 1406-1413.
- 239 22. Pincus SM. Aproximate entropy as a measure of system complexity. *Proc Natl Acad Sci*
240 1991; 88: 2297-2301.
- 241 23. Stergiou N, editor. Innovate Analyzes of Human Movement. Analytical Tools for Human
242 Movement Research. Champaign (Illinois): Human Kinetics Publ; 1999.
- 243 24. Goldberger AL, Peng CK, Lipsitz, LA. What is physiologic complexity and how does it
244 change with aging and disease? *Neurobiol Aging* 2002; 23(1): 23-6.
- 245 25. Pincus SM, Goldberger AL. Physiological time-series analysis: what does regularity
246 quantify?. *Am J Physiol* 1994; 266 (Heart Circ Physiol 35): H1643-H1656.
- 247 26. Pincus SM. Assessing serial irregularity and its implications for health. *Ann NY Acad Sci*
248 2001; 954:245-67.
- 249 27. Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy
250 and sample entropy. *Am J Physiol Heart Circ Physiol* 2000; 278 (6): 2039-2049.
- 251 28. Mäkikallo TH, Seppänen T, Niemelä M, Airaksinen KEJ, Tulppo M, Huikuri HV.
252 Abnormalities in beat to beat complexity of heart rate dynamics in patients with a
253 previous myocardial infarction. *J Am Coll Cardiol* 1996; 28:1005-1011.
- 254 29. Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of biological signals. *Phys*
255 *Rev E Stat Nonlin Soft Matter Phys* 2005;71:021906.

- 256 30. Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic
257 time series. *Phys Rev Lett* 2002; 89: 068102.
- 258 31. Costa M, Goldberger AL, Peng CK. Multiscale Entropy Analysis: a new measure of
259 complexity loss in heart failure. *J Electrocardiol* 2003; 36 Suppl:39-40.
- 260 32. Kurz M, Stergiou N. The aging neuromuscular system expresses less certainty for
261 selecting joint kinematics during gait in humans. *Neurosci Lett* 2003; 348(3):155-158.
- 262 33. Khandoker AH, Palaniswami R, Begg RK. A comparative study on approximate entropy
263 measure and poincaré plot indexes of minimum foot clearance variability in the elderly
264 during walking. *J Neuroeng Rehabil* 2008; 5: 4.
- 265 34. Moraiti C, Stergiou N, Ristanis S, Vasiliadis HS, Patras K, Lee A, Georgoulis AD. The
266 effect of anterior cruciate ligament reconstruction on stride-to-stride variability.
267 *Arthroscopy* 2009; 25 (7): 742-749.
- 268 35. Georgoulis AD, Moraiti C, Ristanis S, Stergiou N. A novel approach to measure
269 variability in the anterior cruciate ligament deficient knee during walking: the use of
270 the approximate entropy in orthopaedics. *J Clin Monit Comput* 2006; 20(1):11-18.
- 271 36. Tochigi Y, Segal NA, Vaseenon T, Brown TD. Entropy analysis of tri-axial leg acceleration
272 signal waveforms for measurement of decrease of physiological variability in human
273 gait. *J Orthop Res* 2012; 30:879-904.
- 274 37. Costa M, Peng C-V, Goldberger AL, Hausdorff JM. Multiscale entropy analysis of
275 human gait dynamics. *Physica A* 2003; 313: 60-69.
- 276 38. Sánchez MD, De la Cruz B, Algaba J, Lagos MA, Naranjo J. Influencia del tapiz rodante
277 sobre la variabilidad de la marcha en sujetos jóvenes y sanos. *Rev Andal Med Deporte*
278 2012; 5 (2): 53-56.
- 279