



Entropy and compression: two measures of complexity

Teresa Henriques MSc,^{1,2,3} Hernâni Gonçalves PhD,³ Luís Antunes PhD,^{2,4} Mara Matias MSc,² João Bernardes MD PhD^{5,6} and Cristina Costa-Santos PhD^{3,7}

¹Student, Health Information and Decision Sciences Department, Faculty of Medicine, University of Porto, Porto, Portugal

²Researcher, Instituto de Telecomunicações, Porto, Portugal

³Researcher, Centre for Research in Health Technologies and Information Systems – CINTESIS, Porto, Portugal

⁴Professor, Computer Science Department, Faculty of Science, University of Porto, Porto, Portugal

⁵Professor, Obstetrics and Gynecology Department, Faculty of Medicine, University of Porto, Porto, Portugal

⁶Researcher, Instituto de Engenharia Biomédica, Porto, Portugal

⁷Professor, Health Information and Decision Sciences Department, Faculty of Medicine, University of Porto, Porto, Portugal

Keyword

medical informatics

Correspondence

Ms Teresa Henriques

CINTESIS

Piso 2, edifício nascente

Centro de Investigação Médica, Faculdade de Medicina da

Universidade do Porto

Rua Dr. Plácido da Costa, s/n

4200-450

Porto

Portugal

Email: teresasarhen@gmail.com

Accepted for publication: 22 May 2013

doi:10.1111/jep.12068

Abstract

Rationale, aims and objectives Traditional complexity measures are used to capture the amount of structured information present in a certain phenomenon. Several approaches developed to facilitate the characterization of complexity have been described in the related literature. Fetal heart rate (FHR) monitoring has been used and improved during the last decades. The importance of these studies lies on an attempt to predict the fetus outcome, but complexity measures are not yet established in clinical practice. In this study, we have focused on two conceptually different measures: Shannon entropy, a probabilistic approach, and Kolmogorov complexity, an algorithmic approach. The main aim of the current investigation was to show that approximation to Kolmogorov complexity through different compressors, although applied to a lesser extent, may be as useful as Shannon entropy calculated by approximation through different entropies, which has been successfully applied to different scientific areas.

Methods To illustrate the applicability of both approaches, two entropy measures, approximate and sample entropy, and two compressors, paq8l and bzip2, were considered. These indices were applied to FHR tracings pertaining to a dataset composed of 48 delivered fetuses with umbilical artery blood (UAB) pH in the normal range ($\text{pH} \geq 7.20$), 10 delivered mildly academic fetuses and 10 moderate-to-severe academic fetuses. The complexity indices were computed on the initial and final segments of the last hour of labour, considering 5- and 10-minute segments.

Results In our sample set, both entropies and compressors were successfully utilized to distinguish fetuses at risk of hypoxia from healthy ones. Fetuses with lower UAB pH presented significantly lower entropy and compression indices, more markedly in the final segments.

Conclusions The combination of these conceptually different measures appeared to present an improved approach in the characterization of different pathophysiological states, reinforcing the theory that entropies and compressors measure different complexity features. In view of these findings, we recommend a combination of the two approaches.

Introduction

Researchers and clinicians recognize that many unsolved medical problems, as the prediction of fetal outcome, are due to the application of conventional mathematics methods to describe biological complex systems [1]. More recently, a different approach based on non-linear dynamics, chaos and complexity has been considered,

which recognizes irregularity, subjectivity and uncertainty as intrinsic and fundamental [2].

Complexity is a property of every system that quantifies the amount of structured information. Shannon demonstrated how the information within a signal could be quantified with absolute precision [3] as the amount of unexpected data contained in the message (designated 'entropy'). Subsequently, the Kolmogorov

complexity was proposed to quantify information on individual objects as the size of its smallest representation [4]. The Shannon information theory measures the average information from a random source, unlike Kolmogorov complexity that presents a form of absolute information [5].

Increasing use of entropy in medicine has accompanied theoretical improvements over the years. In 1991, Pincus suggested the use of Approximate Entropy (ApEn) to classify complex systems [6]. Following its application to quantify the creation of information in a time series, ApEn has since been used in the analysis of a wide range of signals, such as electroencephalography (EEG) and electrocardiography (ECG). The sample entropy (SampEn) concept appeared later in 2000 with the objective of reducing the ApEn bias [7], and was subsequently employed in the analysis of biomedical signals. In 2004, Ohmeda [8] developed an anaesthesia EEG analyser with an entropy module that calculates the characteristics of the biosignal with analysis of time-frequency balanced spectral entropy. The index, based on EEG and electromyography (EMG) activity (Entropy), is an indicator of the hypnotic effects of propofol, thiopental, isoflurane, sevoflurane and desflurane [9]. In 2002, Costa *et al.* [10] proposed the multiscale entropy (MSE) technique applicable in the analysis of physiologic time series.

Compression is a measure of system complexity, but has been used to a lower extent in the analysis of biomedical signals. Many types of compressors exist. Traditionally, the different available compressors are divided in two classes, specifically, lossless and lossy compressors. With lossless compression, every bit of data originally in the file remains after the file is uncompressed, whereas in lossy compression, the file is reduced by permanently eliminating certain information, particularly redundant information. The Lempel–Ziv is a compression algorithm introduced in 1977 [11], used as a measure of complexity in EEG and ECG. Jean-Loup Gailly and Mark Adler subsequently developed gzip [12], which is a combination of Lempel–Ziv and Huffman coding [13]. Julian Seward developed bzip2 using the Burrows–Wheeler block sorting text compression algorithm [14] and Huffman coding. PAQ is a series of lossless data compressors that uses a context-mixing algorithm. PAQ8L was released on 8 March 2007 by Matt Mahoney [15].

Cardiotocography is a technical means of recording the heart rate of the fetus (FHR) and uterine contractions, and is widely used for fetal monitoring during labour. The value of cardiotocography lies in its ability to predict newborn outcomes. Timely prediction of newborn outcome in the intrapartum period, that is, immediately before delivery, may lead to a decrease in perinatal mortality and morbidity. The FHR is measured in beats per minute and an external sensor attached on the mother's abdomen acquires it. FHR alterations are used to evaluate the fetal health condition and allow the obstetrician to intervene to prevent potential compromise and irreversible damage. However, studies have shown significant differences in inter- and intra-observer in FHR analysis and interpretation [16–18]. On the other hand, new signal processing and pattern recognition techniques have paved the way towards automated approaches.

One of the most widely accepted measures of newborn outcome is umbilical artery blood (UAB) pH, as it represents an active measure of fetal oxygenation. A low UAB pH indicates the presence of acidemia occurring during labour and delivery, presenting higher risk of perinatal death or neurological injuries from hypoxia [19].

In the particular case of fetal heart rate (FHR) tracings, several entropy measures have been widely used to detect different pathologies. ApEn and SampEn statistics are the most used entropy metrics, followed more recently by MSE. Despite the successful application of Lempel–Ziv and gzip compressors to FHR tracings to detect pathologies, compressors have been used only to a limited extent in the analysis of biological signals to date [20,21].

In the current study, we aimed to show that compression can be effectively applied as an alternative measure of complexity to the widely used entropy in biological signals. An example of FHR tracings was applied to demonstrate the utility of these non-linear indices [22].

Methods

Complexity is a property of every system that quantifies the amount of structured information. The quantification of information within a signal may be achieved using entropy or compression [7–9]. In this section, we briefly describe the entropy and compression approaches and their related indices as well as the dataset composed by FHR tracings used to evaluate the two approaches.

Entropy

In 1948, Shannon introduced the first approach for measuring 'information' [3]. This new measure, known as Entropy (Shannon entropy), attempts to determine how random a message is expected to be within a given distribution, and takes into account the minimum number of bits to transmit a message from a random source of known characteristics through an error-free channel (Grunwald, P. D. & Vitanyi, P. M. B unpublished data).

Let X be a random variable, taking values in Y with distribution $P(X = x) = p_x$. The Shannon entropy of the random variable X is given by

$$H(X) = \sum_{x \in Y} p_x \log \frac{1}{p_x}$$

The logarithm is base 2, so that entropy is measured in bits and $\lim_{x \rightarrow 0} x \log x = 0$; and thus, traditionally it is conventioned that $0 \log 0 = 0$.

With the goal of 'quantification of the amount of regularity in (heart rate) time-series data', Pincus presented the ApEn statistic in 1991 [6]. It can be estimated through the family of statistics ApEn (N, m, r), given N points, and is approximately equal to the negative average natural logarithm of the conditional probability that two sequences similar for m points remain similar within tolerance, r , at the next point. Accordingly, a low ApEn value is associated with high degree of regularity.

SampEn was introduced in 2000 with the same objective as ApEn by Richman and Moorman [7] to evaluate biological time series, particularly heart rate. The authors highlighted two drawbacks in ApEn properties, stating that 'First, ApEn is heavily dependent on the record length and uniformly lower than expected for short records. Second, it lacks relative consistency. That is, if ApEn of one dataset is higher than that of another, it should, but does not remain higher for all conditions tested' [7]. In order to overcome these limitations, the group proposed a new family of statistics, SampEn (m, r), which, among other differences, eliminates self-matches.

Regarding the necessary parameters to the evaluation of ApEn and SampEn, Pincus and Goldberger [23] concluded that $m = 2$, r values between 0.1 and 0.25 of the standard deviation and N value of 10^m , or preferably 30^m , will yield statistically reliable and reproducible results. In the particular case of FHR analysis, segments of 5 or 10 minutes (i.e. 600 or 1200 points for 2 Hz sampling) are adequate and can make a difference in sensitivity and specificity [22], since in smaller segments some patterns may not be detected, whereas in larger segments some patterns may get lost.

Compression

In 1958 [24], Kolmogorov connected the dynamic systems theory with information theory. Based on his work, Sinai [25] introduced the Kolmogorov–Sinai entropy that generalizes the probabilistic definition of entropy, allowing the use of entropy in time-series data.

The ‘algorithm information theory’ was proposed years later independently by Solomonoff in 1960/1964 [26,27], Kolmogorov in 1965 [4] and Chaitin in 1969 [28], which considers the amount of information in an object as the length of the smallest description of the object. The Kolmogorov complexity attempts to answer how ‘random’ an individual object is in terms of the number of bits necessary to describe it.

In [29] Kolmogorov claimed ‘The ordinary definition of entropy uses probability concepts, and thus does not pertain to individual values, but random values, i.e. probability distributions within a given group of values’.

The Kolmogorov complexity of a finite object is defined as the length of its shortest effective description (Grunwald, P. D. & Vitanyi, P. M. B unpublished data). This measure is not computable, but can be easily approximated using compressors. Good results were obtained using this approach in different areas, including languages tree, genomics, optical character recognition, literature [30], music [31], computer virus and Internet traffic analysis [32] and FHR anomaly detection [21]. In fact, Keogh [33] showed that when clustering heterogeneous data and anomaly detection in time sequences, the compression approach outperforms every known data-mining method.

In 1952, Huffman [13] developed an algorithm to use a short bit stream for characters that appear more often. In 1977, Ziv and Lempel introduced the Lempel–Ziv algorithm based on ‘the concept of encoding future segments of the source output via maximum-length copying from a buffer containing the recent past output’ [11]. Recently, the Lempel–Ziv algorithm has been widely used in the field of medicine. However, numerous compressors can be used with the same purpose.

The main point of data compression is the encoding of information using fewer bits than the original data. Gailly and Adler created the first version of gzip [12], representing a combination of Lempel–Ziv and Huffman coding. Seward developed bzip2 [14] in 1996, which was more effective than Lempel–Ziv and gzip, but considerably slower. A similar compressor was successfully used in 2006 to cluster FHR [21]. PAQ8 [15] represents a series of lossless compressors with the world’s highest compression ratio. PAQ8L, based on Dynamic Markov compression [34], was released in 2007. We believe that these compressors can be successfully used in the medical field as well.

Example of fetal heart rate

Most clinical decisions in the intrapartum period, in countries with the best maternal and perinatal indicators, are strongly based on FHR monitoring [16,18,19]. However, conventional visual analysis of FHR recordings suffers from unacceptable intra- and inter-observer variation [17–19]. To overcome this shortcoming, computer systems for bedside analysis of FHR recordings have been developed [35]. These systems have provided important progresses in FHR monitoring, but there is still room for their improvement, namely using methods for complexity analysis of FHR variability [22,36], which remains one of the most challenging tasks in the analysis of FHR recordings [17,19,37].

In this study, we analysed 68 FHR intrapartum traces consecutively selected from a pre-existing database of term singleton gestations, with at least 60 minutes of tracing. Of the 68 cases, 48 delivered fetuses with UAB pH in the normal (N) range ($\text{pH} \geq 7.20$), 10 delivered with UAB pH between 7.10 and 7.20, mildly acidemic (MA) fetuses and 10 moderate-to-severe acidemic (MSA) fetuses with $\text{UAB pH} \leq 7.10$. All traces were resampled at a frequency of 2 Hz after pre-processing, based on an algorithm described in previous studies. A more detailed description of the data and pre-processing algorithm is presented in [22].

Results

The non-linear methods, entropy (ApEn and SampEn) and compression (paq8l and bzip2), were calculated in the first and last 5 and 10 minutes of the tracings. The median and interquartile (IQ) range, as well as the statistical significance of the non-parametric Kruskal–Wallis test, are presented in Tables 1 and 2.

The mean running times for entropy (ApEn), paq8l compressor and bzip2 compressor in 10-minute segments were 4.2, 0.9 and 0.006 seconds, while those in 5-minute segments were 1.2, 0.5 and 0.003 seconds, respectively.

Considering the outcome MSA and the last 5-minute segments, entropy measures tend to have less discriminatory power (lower areas under the ROC curve ranging from 0.584 to 0.705) than compressors (areas under the ROC curve between 0.773 and 0.797). Considering the last 10-minute segments, entropy measures tend to have greater discriminatory power (higher areas under the ROC curve ranging from 0.797 to 0.855) than compressors (areas under the ROC curve between 0.676 and 0.699).

A comparison between the initial and final 5 minutes of FHR tracings was performed for each complexity measure and each group of fetuses (MSA, MA and N). The same procedure was performed for comparison between the initial and final 10 minutes of FHR tracings, and significance results of the non-parametric Mann–Whitney test are presented in Table 3.

As expected, SampEn was highly correlated with ApEn, with the correlation coefficient ranging between 0.928 and 0.955 (for similar values of r). Moreover, the two compressors were highly correlated with a correlation coefficient of 0.888. However, the correlation values between the considered compression and entropy metrics were lower, ranging between 0.316 and 0.414.

Figure 1 depicts a non-linear relationship between entropy (ApEn) and compression measures (paq8l). The correlation

Table 1 Median, first quartile (Q₁) and third quartile (Q₃) of complexity measures of fetal heart rate tracings from moderate-to-severe acidemic (MSA), mildly acidemic (MA) and normal (N) fetuses in the final 5-minute segments

	MSA		MA		N		<i>P</i>
	Median	(Q ₁ , Q ₃)	Median	(Q ₁ , Q ₃)	Median	(Q ₁ , Q ₃)	
Entropy							
ApEn(2,0.1)	0.585	(0.525, 0.733)	0.738	(0.686, 0.774)	0.682	(0.555, 0.739)	0.207
ApEn(2,0.15)	0.496	(0.291, 0.738)	0.642	(0.561, 0.792)	0.607	(0.490, 0.702)	0.304
ApEn(2,0.2)	0.351	(0.251, 0.553)	0.582	(0.469, 0.795)	0.516	(0.420, 0.627)	0.044
SampEn(2,0.1)	0.476	(0.325, 0.658)	0.598	(0.540, 0.985)	0.541	(0.402, 0.615)	0.149
SampEn(2,0.15)	0.309	(0.172, 0.636)	0.459	(0.403, 0.632)	0.434	(0.320, 0.549)	0.338
SampEn(2,0.2)	0.231	(0.172, 0.307)	0.369	(0.308, 0.637)	0.341	(0.256, 0.404)	0.036
Compression							
paq8l	234.0	(211.0, 279.0)	355.0	(306.0, 393.0)	335.0	(293.5, 372.5)	0.009
bzip2	283.5	(270.0, 382.0)	444.0	(404.0, 501.0)	426.5	(362.5, 488.0)	0.017

Further details on the complexity measures may be found in the Entropy and Compression sections. Boldfaced numerals correspond to *P* values < 0.05.

Table 2 Median, first quartile (Q₁) and third quartile (Q₃) of complexity measures of fetal heart rate tracings from moderate-to-severe acidemic (MSA), mildly acidemic (MA) and normal (N) fetuses in the final 10-minute segments

	MSA		MA		N		<i>P</i>
	Median	(Q ₁ , Q ₃)	Median	(Q ₁ , Q ₃)	Median	(Q ₁ , Q ₃)	
Entropy							
ApEn(2,0.1)	0.533	(0.376, 0.611)	0.837	(0.727, 0.930)	0.751	(0.604, 0.877)	0.003
ApEn(2,0.15)	0.392	(0.298, 0.541)	0.687	(0.667, 0.907)	0.613	(0.506, 0.739)	0.002
ApEn(2,0.2)	0.328	(0.243, 0.403)	0.593	(0.553, 0.662)	0.516	(0.408, 0.613)	0.001
SampEn(2,0.1)	0.342	(0.234, 0.397)	0.682	(0.534, 0.815)	0.572	(0.375, 0.678)	0.009
SampEn(2,0.15)	0.215	(0.152, 0.316)	0.498	(0.425, 0.770)	0.422	(0.286, 0.522)	0.002
SampEn(2,0.2)	0.194	(0.114, 0.238)	0.398	(0.349, 0.489)	0.317	(0.226, 0.406)	0.002
Compression							
paq8l	511.0	(452.0, 537.0)	648.5	(570.0, 692.0)	592.0	(515.5, 647.5)	0.019
bzip2	658.0	(616.0, 719.0)	814.0	(726.0, 850.0)	716.0	(667.5, 814.0)	0.032

Further details on the complexity measures may be found in the Entropy and Compression sections. Boldfaced numerals correspond to *P* values < 0.05.

	Initial versus final 5 minutes			Initial versus final 10 minutes		
	MSA	MA	N	MSA	MA	N
Entropy						
ApEn(2,0.1)	0.074	0.028	<0.001	0.050	0.017	<0.001
ApEn(2,0.15)	0.386	0.203	<0.001	0.028	0.386	<0.001
ApEn(2,0.2)	0.285	0.386	<0.001	0.017	0.028	<0.001
SampEn(2,0.1)	0.059	0.114	<0.001	0.005	0.009	<0.001
SampEn(2,0.15)	0.285	0.169	<0.001	0.005	0.203	<0.001
SampEn(2,0.2)	0.203	0.203	<0.001	0.005	0.022	<0.001
Compression						
paq8l	0.114	0.878	0.272	0.878	0.541	0.024
bzip2	0.575	0.333	0.001	0.241	0.139	<0.001

Values provided in the table correspond to *P* values of the Mann–Whitney test. Further details on the complexity measures may be found in the Entropy and Compression sections. Boldfaced numerals correspond to *P* values < 0.05.

Table 3 Comparison of the complexity measures computed in the initial and final minutes of fetal heart rate tracings with respect to moderate-to-severe acidemic (MSA), mildly acidemic (MA) and normal (N) fetuses

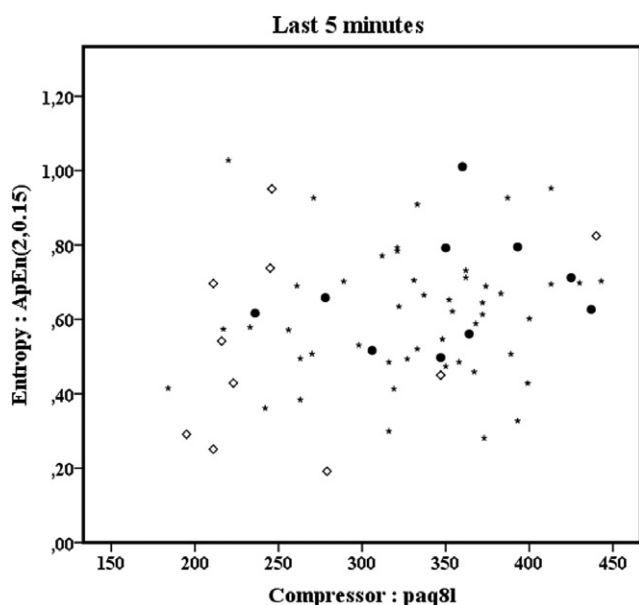


Figure 1 Scatterplot of indices ApEn(2,0.15) and paq8l for the final 5-minute segments, comparing normal fetuses (*), mildly acidemic fetuses (●) and moderate-severe acidemic fetuses (◇).

coefficient between ApEn (2,0.15) and paq8l for all cases was 0.414, and the same measure for the MSA, MA and N groups were 0.648, 0.219 and 0.214, respectively. Similar results were obtained with other possible combinations in the final 5- and 10-minute segments. Despite both being complexity measures, the poor correlation between them shows that entropy and compression are using algorithms that seek and emphasize different characteristics and patterns of each time series. Therefore, the combination of both measures may improve the classification of a FHR tracing.

Conclusions

We have shown that entropy and compression measures allow for the characterization of different pathophysiological conditions of the fetus – distinguishing fetuses at risk of hypoxia from their healthy counterparts and between different stages of labour – through the analysis of the FHR signal.

The use of compression as a measure of complexity has rarely been applied in the analysis of biological signals. However, we have shown that compressors can be effectively used as an alternative to the widely used entropy measures to quantify complexity in biological signals.

By using entropy and compression approaches, one can quantify different features of a system complexity, as shown by the low/moderate correlations between the entropy and compression measures. Accordingly, further research is required in order to study the different complexity features captured by either entropy or compression, as well as in exploring different combinations of the two strategies.

The small computational time associated with both measures, particularly the compressors, allows for considering their inclusion

in existing FHR monitoring systems. In this way, the information on the fetus complexity obtained from the FHR signal may provide important auxiliary information to clinicians, in the diagnostic of the fetal state and help in consequent clinical decisions.

Acknowledgements

This work was supported by the National Science Foundation, Fundação para a Ciência e Tecnologia (FCT), through FEDER funds through Programa Operacional Factores de Competitividade – COMPETE through the project CSI2 with the reference PTDC/EIA-CCO/099951/2008, through the project with the reference PEST-C/SAU/UI0753/2011 and through the PhD grant with the reference SFRH/BD/70858/2010. Hernâni Gonçalves is financed by a post-doctoral grant (SFRH/BPD/69671/2010) from FCT, Portugal.

References

- Alexander, C. (2010) Complexity and Medicine: The Elephant in the Waiting Room. Sheffield: Nottingham University Press.
- Rambihar, V. S. (2000) Science, evidence, and the use of the word scientific. *Lancet*, 355 (9216), 1730.
- Shannon, C. E. (1948) A mathematical theory of communication. *The Bell System Technical Journal*, 27, 379–423, 623–656.
- Kolmogorov, A. (1965) Three approaches to the quantitative definition of information. *Problems in Information Transmission*, 1, 1–7.
- Li, M. & Vitanyi, P. (2008) An Introduction to Kolmogorov Complexity and Its Applications, 3rd edn. New York: Springer.
- Pincus, S. M. (1991) Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 88 (6), 2297–2301.
- 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 (6), H2039–H2049.
- Viertio-Oja, H., Maja, V., Sarkela, M., Talja, P., Tenkanen, N., Tolvanen-Laakso, H., Paloheimo, M., Vakkuri, A., Yli-Hankala, A. & Merilainen, P. (2004) Description of the entropy algorithm as applied in the datex-ohmeda S/5 entropy module. *Acta Anaesthesiologica Scandinavica*, 48 (2), 154–161.
- Aho, A. J., Yli-Hankala, A., Lytikainen, L. P. & Jantti, V. (2009) Facial muscle activity, Response Entropy, and State Entropy indices during noxious stimuli in propofol-nitrous oxide or propofol-nitrous oxide-remifentanyl anaesthesia without neuromuscular block. *British Journal of Anaesthesia*, 102 (2), 227–233.
- Costa, M., Goldberger, A. L. & Peng, C. K. (2002) Multiscale entropy analysis of complex physiologic time series. *Physical Review Letters*, 89 (6), 068102.
- Ziv, J. & Lempel, A. (1977) A universal algorithm for sequential data compression. *Information Theory, IEEE Transactions on*, 23 (3), 337–343.
- Gailly, J.-L. & Adler, M. (2003) The gzip home page. Available at: <http://www.gzip.org> (last accessed June 2013).
- Huffman, D. A. (1952) A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40 (9), 1098–1101.
- Seward, J. (1996) bzip2 and libbzip2. Available at: <http://www.bzip.org/> (last accessed June 2013).
- Mahoney, M. (2007) Data compression programs. Available at: <http://mattmahoney.net/dc/index.html> (last accessed June 2013).

16. MacDonald, D., Grant, A., Sheridan-Pereira, M., Boylan, P. & Chalmers, I. (1985) The Dublin randomized controlled trial of intrapartum fetal heart rate monitoring. *American Journal of Obstetrics and Gynecology*, 152 (5), 524–539.
17. Bernardes, J., Costa-Pereira, A., Ayres-de-Campos, D., van Geijn, H. P. & Pereira-Leite, L. (1997) Evaluation of interobserver agreement of cardiotocograms. *International Journal of Gynaecology and Obstetrics*, 57 (1), 33–37.
18. Ayres-de-Campos, D., Bernardes, J., Costa-Pereira, A. & Pereira-Leite, L. (1999) Inconsistencies in classification by experts of cardiotocograms and subsequent clinical decision. *British Journal of Obstetrics and Gynaecology*, 106 (12), 1307–1310.
19. Bernardes, J. & Ayres-de-Campos, D. (2010) The persistent challenge of foetal heart rate monitoring. *Current Opinion in Obstetrics and Gynecology*, 22 (2), 104–109.
20. Ferrario, M., Signorini, M. G. & Magenes, G. (2009) Complexity analysis of the fetal heart rate variability: early identification of severe intrauterine growth-restricted fetuses. *Medical and Biological Engineering and Computing*, 47 (9), 911–919.
21. Santos, C. C., Bernardes, J., Vitanyi, P. M. B. & Antunes, L. (2006) Clustering fetal heart rate tracings by compression. Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems. 1153035: Washington, DC: IEEE Computer Society. pp. 685–690.
22. Goncalves, H., Rocha, A. P., Ayres-de-Campos, D. & Bernardes, J. (2006) Linear and nonlinear fetal heart rate analysis of normal and academic fetuses in the minutes preceding delivery. *Medical and Biological Engineering and Computing*, 44 (10), 847–855.
23. Pincus, S. M. & Goldberger, A. L. (1994) Physiological time-series analysis: what does regularity quantify? *The American Journal of Physiology*, 266 (4 Pt 2), H1643–H1656.
24. Kolmogorov, A. (1958) New metric invariant of transitive dynamical systems and endomorphisms of lebesgue spaces. *Doklady of Russian Academy of Sciences*, 119 (5), 861–864.
25. Sinai, Y. (1959) On the concept of entropy for a dynamic system.
26. Solomonoff, R. J. (1964) A formal theory of inductive inference. Part I. *Information and Control*, 7 (1), 1–22.
27. Solomonoff, R. J. (1964) A formal theory of inductive inference. Part II. *Information and Control*, 7 (2), 224–254.
28. Chaitin, G. J. (1969) On the length of programs for computing finite binary sequences: statistical considerations. *Journal of ACM*, 16 (1), 145–159.
29. Kolmogorov, A. (1968) Logical basis for information theory and probability theory. *Information Theory, IEEE Transactions on*, 14 (5), 662–664.
30. Cilibrasi, R. & Vitanyi, P. M. B. (2005) Clustering by compression. *Information Theory, IEEE Transactions on*, 51 (4), 1523–1545.
31. Cilibrasi, R., Vitanyi, P. & de Wolf, R. (eds) (2004) Algorithmic clustering of music. Web delivering of music, 2004 WEDELMUSIC 2004. Proceedings of the Fourth International Conference, Barcelona, Spain, 2004 September 13–14.
32. Wehner, S. (2007) Analyzing worms and network traffic using compression. *Journal of Computer Security*, 15 (3), 303–320.
33. Keogh, E., Lonardi, S. & Ratanamahatana, C. A. (2004) Towards parameter-free data mining. Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, 206–215.
34. Cormack, G. V. & Horspool, R. N. S. (1987) Data compression using dynamic markov modelling. *The Computer Journal*, 30 (6), 541–550.
35. Nunes, I., Ayres-de-Campos, D., Figueiredo, C. & Bernardes, J. (2013) An overview of central fetal monitoring systems in labour. *Journal of Perinatal Medicine*, 41 (1), 93–99.
36. Ferrario, M., Signorini, M. G. & Magenes, G. (2009) Complexity analysis of the fetal heart rate variability: early identification of severe intrauterine growth-restricted fetuses. *Medical & Biological Engineering & Computing*, 47 (9), 911–919.
37. Ayres-de-Campos, D., Bernardes, J., Kwee, A., Westerhuis, M. & Visser, G. H. (2007) Computer quantification of short-term variability as an adjunct to fetal electrocardiographic monitoring. *BIOG*, 114 (11), 1445–1446.