How well do practical information measures estimate the Shannon entropy?

Ulrich Speidel
Department of Computer Science
The University of Auckland
Auckland, New Zealand
Email: ulrich@cs.auckland.ac.nz

Mark Titchener
Department of Computer Science
The University of Auckland
Auckland, New Zealand
Email: mark@tcode.tcs.auckland.ac.nz

Jia Yang
Peking University
Beijing, China
Email: yangj@pku.edu.cn

***Abstract*— Estimating the entropy of finite strings has applications in areas such as event detection, similarity measurement or in the performance assessment of compression algorithms. This report compares a variety of computable information measures for finite strings that may be used in entropy estimation. These include Shannon's n-block entropy, the three variants of the Lempel-Ziv production complexity, and the lesser known T-entropy. We apply these measures to strings derived from the logistic map, for which Pesin's identity allows us to deduce corresponding Shannon entropies (Kolmogorov-Sinai entropies) without resorting to probabilistic methods.**

I. Introduction

The term "entropy" is used in both physics and information theory to describe the amount of uncertainty or information inherent in an object or system. Clausius introduced the notion of entropy into thermodynamics in order to explain the irreversibility of certain physical processes in thermodynamics. Boltzmann quantified this as $S = k \log W$. Shannon recognized that a similar approach could be applied to information theory. In his famous 1948 paper [1], he introduced a probabilistic entropy measure $H_{S,n}$:

$$H_{S,n} = -\sum_{a_1,a_2,\ldots,a_n} P(a_1,a_2,\ldots,a_n)\log_2 P(a_1,a_2,\ldots,a_n) \quad (1)$$

where $P(a_1,a_2,\ldots,a_n)$ is the probability of occurrence of the pattern $a_1,a_2,\ldots,a_n$ in the output of an information source. This entropy measure is known as the *n-block entropy*. The Shannon entropy rate of a process is then given by

$$h_S = \lim_{n\to\infty} \frac{H_{S,n}}{n}. \quad (2)$$

Computation of the n-block entropy is straightforward – provided the $P(a_1,a_2,\ldots,a_n)$ are known. In many practical applications, however, one is interested in the entropy inherent in a finite object, which can usually be represented in the form of a finite string x of length $|x| = N$. Such applications include, e.g., similarity measurement [2] and the detection of denial-of-service attacks [3], [4].

However, finite strings imply an absence of genuine probabilities, thus leaving the entropy of a finite object undefined from a probabilistic perspective. If one regards $|x|$ as representative output from some source process, one may estimate $P(a_1,a_2,\ldots,a_n)$ from the pattern frequencies observed in x. However, even for well-behaved sources, only estimates for $n < \log N$ are sensible, which implies a severe trade-off between N and estimation accuracy. This is the chief motivation behind non-probabilistic approaches to entropy estimation.

Non-probabilistic approaches have been proposed by a number of of authors and include the works by Kolmogorov [5], [6], Solomonoff [7], and Chaitin [8], as well as the various parsing algorithms of the Lempel-Ziv family. Among the latter, Lempel and Ziv's original parsing algorithm for the computation of a *production complexity*, called LZ76 [9] in our paper, was explicitly designed to address the question of finite sequence complexity. It measures the production complexity of a string as the number of steps required by the parsing algorithm.

The two other algorithms discussed here, LZ77 [10] and LZ78 [11], were mainly intended for data compression and both restrict the pattern search space of LZ76 to achieve linear processing time. However, these algorithms also perform a number of successive parsing steps. This number may be used as an estimate for the LZ production complexity but can never be smaller than the latter. Lempel and Ziv showed that their production complexity is asymptotically equivalent to Shannon's entropy as N goes to infinity. However, for the reasons already mentioned, it is not possible to show this for finite strings. Evaluating entropy measures for finite and, in particular, short strings thus requires a different approach.

Comparing entropy estimates for strings with a known entropy may supply corroborative evidence for the suitability of both probabilistic and non-probabilistic entropy measures. One source for such strings that is often proposed in literature is the partitioning of the logistic map with biotic potential r. Its non-negative Lyapunov exponents for a given r are equal to the Kolmogorov-Sinai (Pesin) entropy of the corresponding string [12].

This paper compares Shannon's n-block entropy, entropies from three implementations of the Lempel Ziv complexity measure (LZ-76, LZ-77 with varying window sizes, and number of steps in LZ-78), and the T-entropy [13] against the non-negative Lyapunov exponents for the logistic map.

II. The logistic map as a reference information source

The logistic map is defined by the recurrence relation $x_{t+1} = rx_t(1-x_t)$. The coefficient r (referred to as the "biotic potential") is given a value in the range $0 \le r \le 4$. For $0 < x_0 < 1$, $x_t \in (0,1)$ for all t. With increasing t, the

values of the series either become periodic or chaotic, i.e., unpredictable depending on the choice of r. One may derive strings of symbols from the values of the logistic map by partitioning the map's state space into subspaces known as *Markov cells*. These Markov cells are then labeled with symbols. The real-valued x_t are then encoded by their labels to yield a string. Different choices of partition thus yield different symbolic representations of the series. The Shannon entropy of the resulting string depends on this choice of partition. The supremum of the corresponding Shannon entropies over all possible partitions (finite or infinite) is known as the *Kolmogorov-Sinai entropy* (KS-entropy) and is a characteristic of the dynamical system. For the logistic map, the binary partition (bipartition) is well known to achieve the KS-entropy [14]. The bipartition maps x_t to the binary alphabet, i.e., 0 for $x_t < 0.5$ and to 1 otherwise.

Pesin's identity [12] proves that for certain classes of dynamical system (including the logistic map), the KS-entropy equals the sum of the positive Lyapunov exponents for the dynamical system. The Lyapunov exponent for the logistic map may be computed from the series $[x_t]$ to numerical accuracy. The Shannon entropy for the strings produced from the logistic map may thus be computed directly by way of Pesin's identity, without reference to source probabilities.

The logistic map has another useful property at the *Feigenbaum accumulation point* $r = r_\infty \approx 3.569945670$, which corresponds to the onset of chaos. It is known [15] that by adding white noise ξ with amplitude ϵ, i.e.,

$$x_{t+1} = r_\infty x_t(1 - x_t) + \epsilon\xi_t \text{ with } \xi_t \in [-1, 1], \quad (3)$$

results in a KS-entropy proportional to ϵ.

III. Lempel-Ziv parsing

Lempel and Ziv's original 1976 algorithm [9] defines a production complexity as the minimum number of parsing steps of a self-learning automaton. LZ-77 [10], primarily known as a compression algorithm, may similarly be used to measure complexity in terms of the vocabulary size. It achieves a speed improvement by restricting parsing to patterns within a window of a restricted size. LZ-77 defaults to LZ-76 for window sizes that match or exceed the length of the string being measured. The vocabulary size is also used as the measure of complexity in LZ-78 [11], the fastest of the three algorithms.

IV. T-entropy

T-entropy is an information measure derived from a recursive parsing process known as *T-decomposition* [16], [17], [18], [19]. T-decomposition is not unlike Lempel-Ziv parsing in that it produces a production-style complexity [20], [21], [22], [23] known as the T-complexity. This is subsequently linearised by the inverse logarithmic integral [24] to yield *T-information* [20], [21], [22], [23]. The T-entropy for the string is the average T-information per symbol, i.e., the total T-information divided by the length of the string. It has already been observed [13] that T-entropy exhibits a correspondence with the KS-entropy.

V. Experiments

In the first part of our experiments, we computed

- Shannon's n-block entropy, computed from Eqn. (1),
- LZ-76 complexity,
- LZ-77 complexity with a selection of window sizes,
- LZ-78 complexity,
- T-entropy,
- and the KS-entropy

for 4000 values of r.

Each of the first five sets of entropies/complexities was plotted against the respective KS-entropy values. As the KS-entropy ranges across several orders of magnitude, logarithmic axes were chosen for all plots. A perfectly matched entropy measure (i.e., one for which the computed entropy is exactly equals the KS-entropy) would thus be rendered as a set of points on the dashed line shown in the plots. Two types of deviation may be observed in the plots: scatter around the dashed line and systematic deviations. Scatter is caused by random errors in the observation and/or deviations of the sample string's entropy from the associated Lyapunov exponent, which are a consequence of the truncated nature of the string. Systematic deviations, on the other hand, result from *systematic* under- and/or overestimation of the parameters being plotted. They may be observed as ensembles that are not scattered around the dashed line.

Figure 1 shows the Shannon n-block entropies for $n = 1$, 4, and 15 versus the corresponding KS-entropy values. As expected [25], the Shannon n-block entropy approaches the KS-entropy from above as n increases. However, as n approaches the logarithm of the sample string length, Shannon n-block entropy starts to seriously underestimate higher entropies, while still overestimating lower entropies. The plots are indicative of the difficulties inherent in using Shannon's n-block entropy as a practical entropy measure.

Figure 2 shows LZ-77 complexities for selected window sizes. The performance of the LZ-77 algorithm is better than that of the Shannon n-block entropy. In order to obtain entropy estimates from Lempel-Ziv complexities, a further normalisation step is required. This omitted here.

The accuracy of the LZ-77 estimates improves substantially with increasing window size. If the chosen window size is large enough to cover the sample string, LZ-77 is equivalent to LZ-76, shown as the bottom scatter diagram in the plot. The time efficiency of LZ-77 is $O(N \times M)$ for strings of length N and windows of size M, i.e., $O(N^2)$ in the LZ-76 case. As in data compression, the window size in LZ-77 thus represents a compromise between speed and accuracy.

LZ-78 is an $O(N \log N)$ algorithm permitting faster complexity measurement suitable for longer strings. Figure 3 shows that LZ-78 also severely overestimates lower entropies, even if the sample string size is increased to 1,000,000 bits. Note that the spread of LZ-78 complexity values for a given KS-entropy values seems generally much reduced compared to LZ-77. This can most likely be attributed to the difference in string length.

Figure 4 similarly depicts the T-entropy values for 1,000,000 bit strings. T-entropy may be computed in

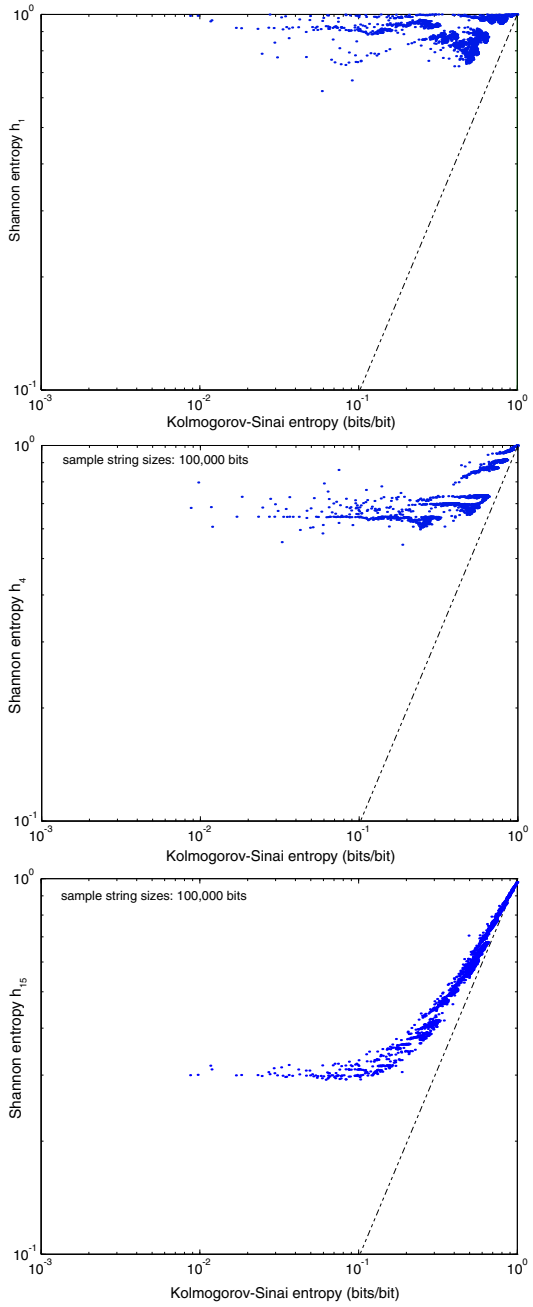

Fig. 1. Shannon n-block entropies for $n = 1$, 4, and 15, for sample strings produced from the logistic map.

$O(N \log N)$ [26]. T-entropy behaves similarly to LZ-76 in Fig. 2. As for LZ-76, the graph suggests that there may be a degree of overestimation for smaller entropy values. It is an open question whether this is a feature of LZ-76 or T-entropy, or perhaps at least in part attributable to the KS-entropy measurements.

The second part of our experiments utilizes the fact that adding noise to the logistic map at the Feigenbaum point gives us access to an extended range of entropy values. Figure 5 shows that LZ-76 gives a linear response across the range, consist with the results by Crutchfield and Packard [15].

The result for LZ-78 in Fig. 6 confirms the earlier observation of significant overestimation at low entropies. In fact, the measure seems to be complete insensitive

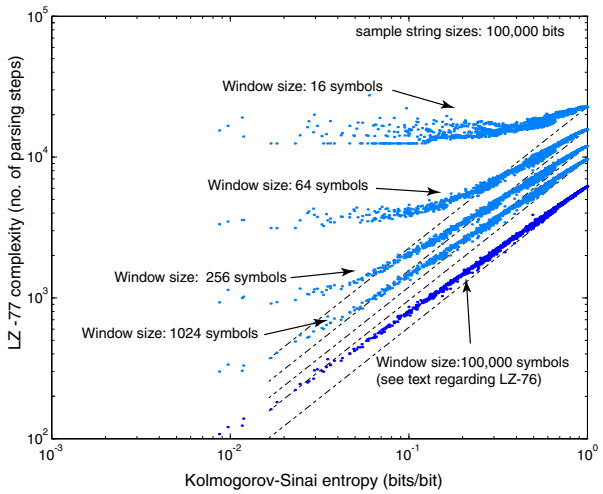

Fig. 2. LZ-77 complexities for selected window sizes as indicated. Note that a window size of 100,000 covers the entire string. LZ-77 is equivalent to LZ-76 in this case.

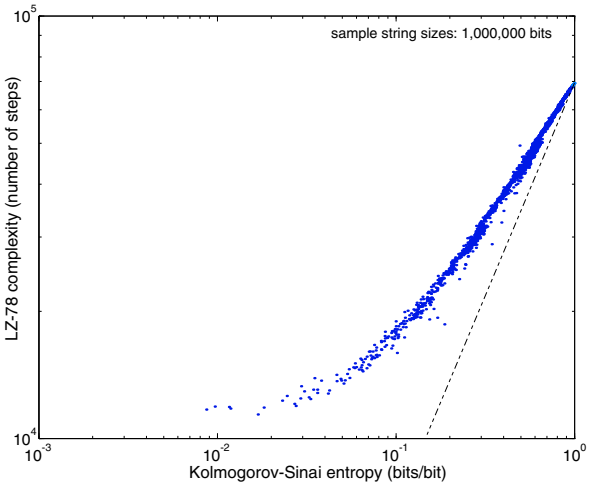

Fig. 3. LZ-78 complexities versus corresponding KS-entropy values.

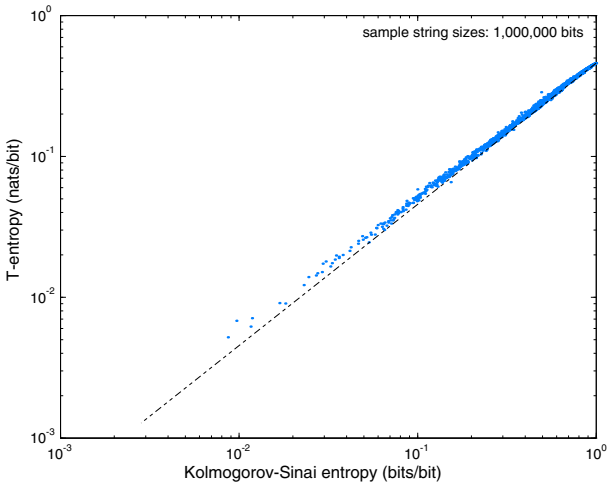

Fig. 4. T-entropies versus corresponding KS-entropy values.

below the top decade of entropies.

T-entropy in Fig. 7 once again reflects very much the characteristics of LZ-76, albeit at a fraction of the computational effort. This may be seen from Fig. 8, which

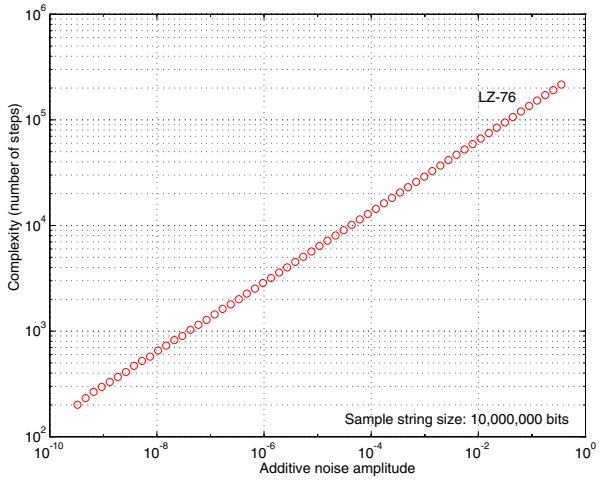

Fig. 5. LZ-76 complexity as a function of additive noise amplitude.

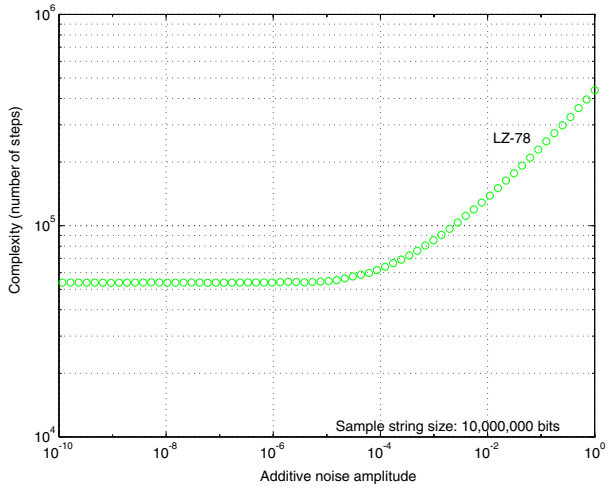

Fig. 6. LZ-78 complexity as a function of additive noise amplitude.

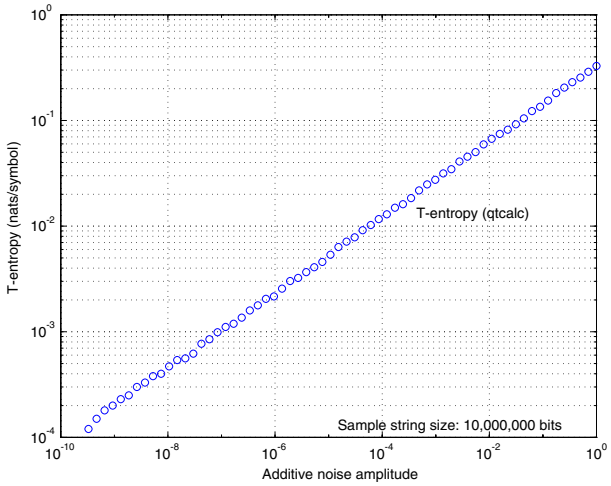

Fig. 7. T-entropy as a function of additive noise amplitude.

shows a time comparison of the LZ-76, LZ-78, and T-entropy measures as a function of entropy (additive noise amplitude at the Feigenbaum point).

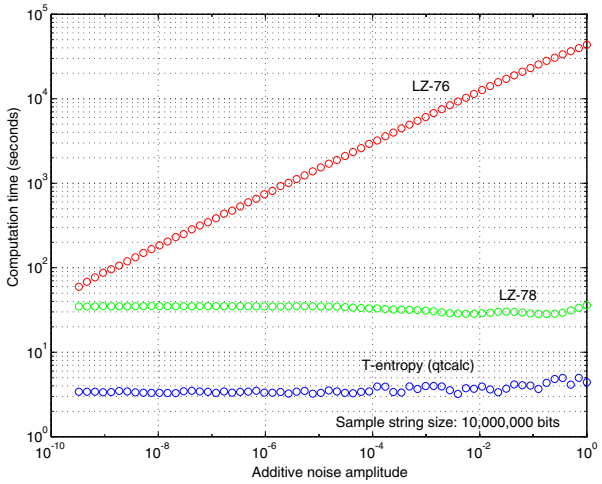

Fig. 8. A comparison of computation times a function of additive noise amplitude.

VI. Conclusions

Both LZ-76 and T-entropy seem to deliver consistent performance across the range of values tested and exhibit close correspondence with KS-entropy. T-entropy may be implemented as an $O(N \log N)$ algorithm. Its time performance seems to be largely independent of entropy. LZ-76 on the other hand is $O(N^2)$ and its running time seems to be proportional to entropy. The popular accelerations, LZ-77 and LZ-78 can achieve up to $O(N)$, but incur a noticeable penalty in terms of accuracy at low entropies.

There are a number of open problems associated with our experiments. Among others, the sources of scatter and systematic deviation need to be investigated for all complexity and entropy measures presented here.

References

[1] C. E. Shannon: *A mathematical theory of communication*, The Bell System Tech. J.27, 1948

[2] J. Yang: *Fast String Parsing and its Application in Information and Similarity Measurement*, PhD thesis, The University of Auckland, 2005.

[3] A. B. Kulkarni and S. F. Bush and S. C. Evans: *Detecting Distributed Denial-of-Service Attacks Using Kolmogorov Complexity Metrics*, Technical Information Series,GE Development & Research Center, February 2002.

[4] R. Eimann, U. Speidel, N. Brownlee: *A T-Entropy Analysis of the Slammer Worm Outbreak*, Proceedings of the 8th Asia-Pacific Network Operations and Management Symposium (APNOMS), Okinawa, Japan, September 27-30, 2005, pp. 434–445

[5] A. N. Kolmogorov: *A new metric invariant of transitive dynamical systems and automorphisms in Lebesgue space*, Dokl. Acad. Nauk. SSSR 119 (1958)

[6] A. N. Kolmogorov: *Three approaches to the quantitative definition of information*, Probl. Inform. Transmis., 1, 1965, pp. 4–7

[7] R. J. Solomonoff: *A formal theory of inductive interference*, Inform. Contr., 7, 1964, pp. 1–22 (Part I), 224-254 (Part II).

[8] G. J. Chaitin: *On the lengths of programs for computing finite binary sequences*, J. Ass. Comput. Mach., 13, pp. 547–569, 1966

[9] A. Lempel and J. Ziv: *On the complexity of finite sequences*, IEEE Trans. Inform. Theory 22 (1976) 75-81.

[10] J. Ziv and A. Lempel: sl A Universal Algorithm for Sequential Data Compression, IEEE Trans. Inform. Theory, Vol 23, No. 3, May 1977, pp. 337-343.

[11] J. Ziv and A. Lempel: sl Compression of Individual Sequences via Variable-Rate Coding, IEEE Trans. Inform. Theory, Vol 24, No. 5, September 1978, pp. 530-536.

[12] J. B. Pesin: *Characteristic Lyapunov exponents and smooth ergodic theory*, Russ. Math. Surveys 32 (1977) 355.

[13] W. Ebeling, R. Steuer, and M. R. Titchener: *Partition-Based Entropies of Deterministic and Stochastic Maps*, *Stochastics and Dynamics*, 1(1), p. 45., March 2001.
[14] J. P. Eckmann and D. Ruelle: *Ergodic theory of chaos and strange attractors*, Rev. Mod. Phys. 57, 1985
[15] J. P. Crutchfield and N. H. Packard: *Symbolic dynamics of noisy chaos*, Physica D7, 1983
[16] R. Nicolescu: *Uniqueness Theorems for T-Codes*. Technical Report. Tamaki Report Series no.9, The University of Auckland, 1995.
[17] R. Nicolescu and M. R. Titchener: *Uniqueness theorems for T-codes*, Romanian J. Inform. Sci. Tech. 1 (1998).
[18] U. Guenther, P. Hertling, R. Nicolescu, and M. R. Titchener: *Representing Variable-Length Codes in Fixed-Length T-Depletion Format in Encoders and Decoders*, Journal of Universal Computer Science, 3(11), November 1997, pp. 1207–1225. `http://www.iicm.edu/jucs_3_11`.
[19] U. Guenther: *Robust Source Coding with Generalized T-Codes*. PhD Thesis, The University of Auckland, 1998. `http://www.tcs.auckland.ac.nz/~ulrich/phd.pdf`.
[20] M. R. Titchener, *Deterministic computation of string complexity, information and entropy*, *International Symposium on Information Theory*, August 16-21, 1998, MIT, Boston.
[21] M. R. Titchener: *A Deterministic Theory of Complexity, Information and Entropy*, *IEEE Information Theory Workshop*, February 1998, San Diego.
[22] M. R. Titchener, *A novel deterministic approach to evaluating the entropy of language texts*, *Third International Conference on Information Theoretic Approaches to Logic, Language and Computation*, June 16-19, 1998, Hsi-tou, Taiwan.
[23] U. Guenther: *T-Complexity and T-Information Theory – an Executive Summary*. CDMTCS Report 149, Centre for Discrete Mathematics and Theoretical Computer Science, The University of Auckland, February 2001. `http://www.tcs.auckland.ac.nz/CDMTCS/researchreports/149ulrich.pdf`.
[24] M. Abramowitz and I. A. Stegun (eds.): *Handbook of Mathematical Functions*, Dover, 1970
[25] P. Grassberger: *Finite sample corrections to entropy and dimension estimates*, Phys. Lett. A128 (1988).
[26] Jia Yang, Ulrich Speidel: *A T-Decomposition Algorithm with $O(n \log n)$ Time and Space Complexity*. Proceedings of the IEEE International Symposium on Information Theory, 4-9 September 2005, Adelaide, pp. 23–27.