33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Improved Entropy Rate Estimation in Physiological Data

D. E. Lake

Abstract— **Calculating entropy rate in physiologic signals has proven very useful in many settings. Common entropy estimates for this purpose are sample entropy (*SampEn*) and its less robust elder cousin, approximate entropy (*ApEn*). Both approaches count matches within a tolerance *r* for templates of length *m* consecutive observations. When physiologic data records are long and well-behaved, both approaches work very well for a wide range of *m* and *r*. However, more attention to the details of the estimation algorithm is needed for short records and signals with anomalies. In addition, interpretation of the magnitude of these estimates is highly dependent on how *r* is chosen and precludes comparison across studies with even slightly different methodologies. In this paper, we summarize recent novel approaches to improve the accuracy of entropy estimation. An important (but not necessarily new) alternative to current approaches is to develop estimates that convert probabilities to densities by normalizing by the matching region volume. This approach leads to a novel concept introduced here of reporting entropy rate in *equivalent Gaussian white noise* units. Another approach is to allow *r* to vary so that a pre-specified number of matches are found, called the *minimum numerator count*, to ensure confident probability estimation. The approaches are illustrated using a simple example of detecting abnormal cardiac rhythms in heart rate records.**

I. Introduction

PHYSIOLOGIC time series often are characterized by practitioners using a myriad of surrogate terms for entropy such as regularity, order, predictability, nonlinearity, non-Gaussianity, and complexity. For time series data, entropy rate is a more precise term than entropy, but both terms will be used here. In this context, the meaning of entropy has been around for over 60 years and follows the work of Shannon [1], Kolmogorov [2, 3], Sinai [4], Grassberger and Procaccia [5], Eckmann and Ruelle [6], and others, who conceived of entropy rate as a measure of the degree to which template patterns repeat themselves. Repeated patterns imply order, and lead to reduced values of entropy. There are often variations in how all these concepts are interpreted among physicists, engineers, and mathematicians, but there still remain many applications where these characteristics are extremely important in determining, for example, disease versus non-disease states.

One particularly good recent example of the application of entropy rate is detecting the abnormal cardiac rhythm atrial fibrillation (AF) [7]. AF predominantly occurs in older adults and can lead to stroke and other clinical deterioration. The hallmark of AF is its irregularity and clinicians sometimes even use the nonsensical descriptor "irregularly irregular" to underscore this fundamental difference from normal sinus rhythm (NSR). Figure 1 shows examples of AF and NSR for two heart records of RR (inter-beat) intervals. While both series have identical mean and standard deviation (original signals from the same patient were transformed to achieve this), there is a clear difference in the dynamics of the two signals that can be well-characterized by entropy rate.

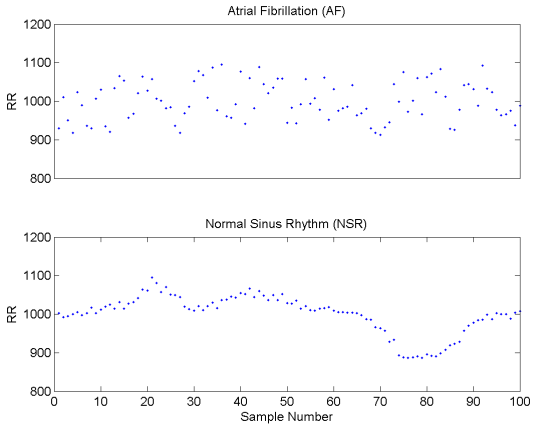

Fig. 1. Examples of RR interval heart rate records for AF and NSR. Both signals have n=100 samples, mean of 1000 and standard deviation of 50 milliseconds. The AF signal is clearly less regular than the NSR signal and will have higher entropy rate.

For long physiologic data records, entropy estimates can be extended to more detailed multi-scale entropy (MSE) analysis [8, 9] . This method has been successfully used to distinguish cardiac rhythms, including AF, in records of 30,000 samples with scales up to 20 giving an effective sample size as low as 1500. The methods presented here allow improved estimation at much larger scales of MSE analysis up to 500 to 3000, where the number of values used to estimate sample entropy can be on the order of 10-60.

An important step in improving the estimation and interpretation of entropy rates is to develop a solid theoretical understanding of the mathematics behind the problem. A framework of Renyi entropy (or *q*-entropy) rate estimation has been presented recently that is a step in this direction [10]. Among other results, this work provides a setting that includes both *ApEn* (*q*=1=> entropy rate) and *SampEn* (*q*=2=> quadratic entropy rate) and notions of differential and conditional entropy rate which are equivalent only when *q*=1.. To be more precise, sample

Manuscript received April 15, 2011. This work was supported in part by American Heart Association Grant-in-Aid 5R01HD048562-05.

D. E. Lake is with Department of Internal Medicine (Cardiovascular Division) and Department of Statistics, University of Virginia. Box 800158, Charlottesville, VA 22908 USA (phone: 434-243-9367; fax: 434-982-1998; e-mail: dlake@virginia.edu).

entropy is a measure related to quadratic differential entropy rate of order q=2. Approximate entropy is related to both differential and conditional entropy rate of order q=1. This framework also suggests a (yet to be named) third measure equivalent to the conditional quadratic entropy rate.

While the principles behind the approaches to entropy estimation presented here can be applied to all Renyi entropy rate measures, *SampEn* will be exclusively used to describe the algorithms. The methods will be applied to the AF and NSR example signals to illustrate their relative advantages.

II. Novel Approaches to Entropy Estimation

A. Densities versus probabilities

Sample entropy was introduced by Richman and Moorman in 2000 [11, 12, 13] as a less-biased and more robust alternative to approximate entropy which was introduced by Pincus in 1991 [14, 15]. Estimates of sample entropy rely on counts of templates of length m matching within a tolerance r that also match at the next point, and have found utility in many applications including the prediction of infection and death in premature infants [16, 17, 18, 19, 20, 21]. Sample entropy is the negative natural logarithm of the conditional probability that any two sequences of length m that match within tolerance r will also match at the subsequent point number $m+1$. Counting the number of times that templates find matches is the central activity of entropy estimation. The number of matches of length $m+1$ is denoted by A and the number of matches of length m is denoted by B. The next intermediate result is a proportion or conditional probability $p=A/B$. More matches means more confident estimation of this probability, and, up to a point, better entropy estimation. From a statistical point of view, inaccurate estimates of a p occur when either A or the difference B-A are small.

In long data records, when matches abound, entropy measures are accurately estimated this way even for larger m. If m is too large or r too small, then the number of template matches will be too small for confident estimation of the conditional probability. If, on the other hand, m is too small and r too large, then all templates will match each other, and there is no ability to discriminate physiologic signals. Strategies have been suggested, but there still remains large inter-study variability in parameter selection.

Of these, the larger problem addressed here in implementing entropy estimation is picking the value of the tolerance r. The original recipe has been to select r as 20% of the standard deviation σ of each time series segment, based largely on original conclusions in 1991 for implementing approximate entropy (*ApEn*) calculations [14]. Twenty years later, this convention is too often blindly applied to any and all studies with physiological data. Systematic approaches to picking r have been presented based on analysis of relative errors or discrimination capability in large data sets [7, 19].

An alternative approach to this issue arises from treating all signals as continuous-valued stochastic processes and applying concepts of probability density estimation [7, 10].

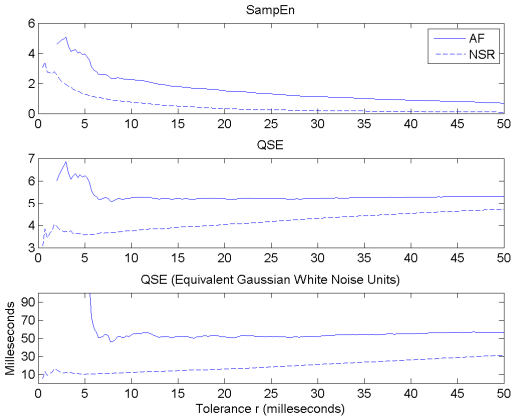

Fig. 2. *SampEn* and *QSE* as a function of tolerance for 2 signals in Figure 1. *Sampen* starts to diverge while *QSE* converges for small r. Interpreting y-axis units for first two graphs is not straight-forward. In contrast, the equivalent white noise standard deviations are approximately 50 milliseconds for AF and 10 milliseconds for NSR.

The direct result is to convert the measured conditional probability to a density by normalizing the match count to the volume of the matching region, or $(2r)^m$. As shown in the appendix, this operation reduces to adding a quantity $\log(2r)$ to the entropy estimate where log is the natural logarithm base e. The result is *quadratic sample entropy*, or *QSE*, and is related to *SampEn* and the conditional probability p by

$$QSE = -\log(p/2r) = -\log(p)+\log(2r) = SampEn+\log(2r)$$

The measure *QSE* is not new, per se, but returns to notions of measuring limiting values as the volume of the matching region tends to 0 and is a quantity independent of r.

A comparison of *QSE* with *SampEn* with m=1 is made in Figure 2. The match counts for templates of length 2 and 1 were calculated for all tolerance values up to 50 milliseconds. With *SampEn*, lower tolerance generally leads to lower probability of matching and appears to diverge as r approaches zero. Conversely, *QSE* approaches a finite, albeit noisy, limit as r gets small. For *SampEn*, the entropy of AF is clearly higher than for NSR for any *fixed value of r*. However, the QSE of AF is clearly higher than for NSR for the entire range of *all values of r*.

Historically, the motivation behind fixing r in the approximate entropy was to discriminate deterministic or low-dimensional systems where *QSE* would always diverge while *SampEn* would converge to a finite limit. This is almost never a consideration for real physiological signals like heart rate and makes *QSE* a good practical choice.

This approach allows r to vary, as needed, based on the dynamics of individual physiologic signals. Also, with this formulation, estimates made with different values of r measure the same inherent quantity and can be compared directly across varying data sets and studies. Another advantage to this approach is that that the tolerance r can be

optimally varied for each individual data record. This is analogous to varying the bin-widths of histograms to optimally depict the distribution of a particular data set. For example, Figure 2 suggests that AF might be best estimated with r=10 milliseconds where the curve starts to get noisy while the estimate for NSR could use a smaller tolerance about r=5 milliseconds. The curves for *QSE* also suggest that if computation time is not an issue, some sort of line-fitting could be done to estimate the intercept at r=0. Care should be taken in this process, however, because assumptions of standard linear regression are clearly violated here. Most notably, the observations are not independent and do not have the same error variance, (i.e., heteroscedasticity).

One issue that arises is how to interpret the numerical quantity of an entropy estimate. After all, the units of *SampEn* are logarithm of probability and the units of QSE in Figure 2 are an even more obscure, the logarithm of probability per millisecond. One advantage of the theoretical framework behind Renyi entropy rates is that there are theoretical known values for certain stochastic processes. For example, for Gaussian white noise with standard deviation σ, the value of the quadratic entropy rate is $\frac{1}{2}\log(4\pi)+\log(\sigma)$ or $1.266+\log(\sigma)$. This allows the conversion of QSE to the standard deviation of Gaussian white noise with the equivalent entropy. This conversion has units equal to the physiologic measurement and is shown in the last panel of Figure 2. The interpretation of this plot is that AF and NSR have entropies equivalent to white noise with standard deviations of approximately 50 and 10 milliseconds respectively. Since both series already have equivalent overall standard deviations of 50, one could conclude that AF is very similar to Gaussian white noise without order while NSR has considerable order (reduces standard deviation by 80%).

The differences between AF and NSR in records of length 100 are readily discernible with most any current method. However, as shown in Figure 3, where entropy estimates are made using just the first n points of the signals, this does not necessarily remain the case as record lengths get shorter. To improve the accuracy of entropy estimation in this setting a new approach is described in the next section. This method provides stable QSE estimates and AF/NSR discrimination for records as short as n=8.

B. Minimum numerator count

An important distinction of *SampEn* is that self-matches are not counted [11, 12, 13]. This significantly reduces bias, but contributes to the problem of falling counts of template matches to the point that A (and even B) could be 0 leading to infinite or indeterminate estimates. This becomes an increasing concern for short records. Some methods have been proposed to remedy this situation [20], but this will often remain an issue when the exact same r is applied to a large population of possibly heterogeneous physiologic signals.

In addition, the accuracy of a probability estimate A/B is dependent on the magnitude of the numerator A and the denominator B. More precisely, the important number to be large is $\min(A, B-A)$. For example an estimate of 0.1 with 100/1000 is more accurate than one with 1/10. Because QSE allows the flexibility to vary r, inaccurate probability estimates can often be avoided. As introduced in [10], one approach to accomplish this, called the *minimum numerator count* method. This approach generally looks to make r as small as possible, but varies r as needed until a pre-specified number of matches A are observed. Other additional restrictions, such as minimum denominator count on B, can also be used to control accuracy.

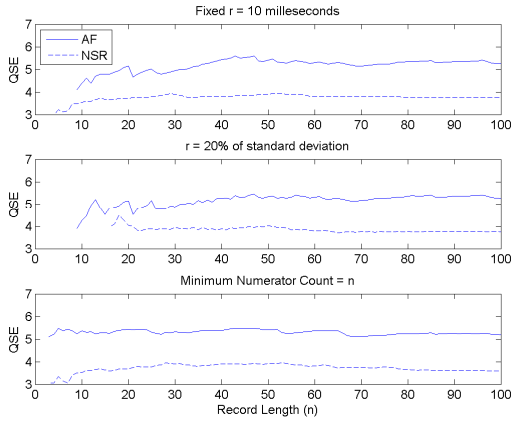

Fig. 3. Methods of QSE as a function of first n point for 2 signals in Figure 1. The minimum numerator count method provides stable estimates of QSE for lengths as short as n~8. Standard methods degrade starting for n~50 and degenerate starting at n~15.

In Figure 3, a minimum numerator count equal to the signal length was employed. This is a reasonable convention, but is by no means an optimal approach. For example, as shown in [7], a minimum numerator count of 5 was found to give optimally accurate estimates detecting AF in short records of length n=12. In general the minimum numerator count should increase and the tolerance decrease as n becomes larger.

C. Other approaches

Another new recent approach involves noting that counting matches is a special case of kernel density estimation [22]. In particular, density estimation in *QSE* uses a uniform kernel. Drawing upon the vast amount of theoretical work in this field, formulas for optimal asymptotic selection of r (related to the bandwidth in kernel density jargon) as a function of n can be found. The MATLAB function KSDENSITY implements some of these results in the univariate case and multivariate versions are feasible. One aspect of this MATLAB implementation that worth noting is a robust estimate of standard deviation as a basis for selecting matching tolerance. As noted in [19], *SampEn* is extremely sensitive to outliers or spikes in

physiologic data. It turns out this is a very good thing for detecting sepsis in neonatal infants, but in most cases a measure robust in the presence of outliers is desired.

III. Conclusion

Accurate algorithms and careful interpretation of entropy rate estimates are needed to optimally characterize and discriminate physiological signals. A minimum numerator algorithm works well for short records. Reporting entropy estimates in equivalent white noise units may provide better interpretation and understanding for clinical practice.

Appendix

The derivation of QSE and its relationship to differential quadratic entropy rate is discussed in more detail here. A data record consists of a series of n consecutive equally-spaced observations assumed to come from a continuous-valued stationary stochastic process. The *template* $\boldsymbol{x}_m(i)$ is the vector containing the m consecutive intervals $x_i, x_{i+1}, \ldots, x_{i+m-1}$. Note that $\boldsymbol{x}_m(i)$ has a well-defined marginal density, say f_m. For a matching tolerance $r>0$, an instance where all the components of $\boldsymbol{x}_m(i)$ are within a distance r of another template $\boldsymbol{x}_m(j)$ is called a *template match.* Let B_i denote the number of template matches of length m with $\boldsymbol{x}_m(i)$ and A_i denote the number of template matches of length $m+1$ with $\boldsymbol{x}_{m+1}(i)$. Also let $A=\sum A_i$ and $B=\sum_i B_i$ denote respectively the total number of matches of length $m+1$ and m. Then the ratio $p=A/B$ is the conditional probability that subsequent points of a set of closely matching m intervals also remain close and match. The sample entropy is the negative logarithm of this probability

$$SampEn = -\log(p) = -\log(A/B) = \log(B) - \log(A)$$

Note that B_n can be removed from B to reflect the lack of a subsequent point and allow the possibility of $p=1$.

When the average match counts A/n and B/n are suitably normalized by their matching volume and r is sufficiently small, the negative logarithms are respectively estimates of the quadratic entropy of the marginal densities f_{m+1} and f_m. The difference of these two estimates defines *QSE*

$$QSE = -\log((A/n)/(2r)^{m+1}) - (-\log((B/n)/(2r)^m))$$

This reduces to

$$QSE = SampEn + \log(2r)$$

and thus estimates the differential quadratic entropy rate.

References

[1] Shannon CE. A mathematical theory of communication. Bell System Technical Journal 27: 379-423, 1948.

[2] Kolmogorov AN. Entropy per unit time as a metric invariant of automorphism. Doklady of Russian Academy of Sciences 124: 754-755, 1959.

[3] Kolmogorov AN. New Metric Invariant of Transitive Dynamical Systems and Endomorphisms of Lebesgue Spaces. Doklady of Russian Academy of Sciences 119: 861-864, 1958.

[4] Sinai YG. On the Notion of Entropy of a Dynamical System. Doklady of Russian Academy of Sciences 124: 768-771, 1959

[5] Grassberger P, and Procaccia I. Estimation of the Kolmogorov entropy from a chaotic signal. Physical Review A 28: 2591-2593, 1983.

[6] Eckmann JP, and Ruelle D. Ergodic theory of chaos and strange attractors. Reviews of Modern Physics 57: 617-654, 1985.

[7] Lake DE, Moorman JR. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. Am J Physiol Heart Circ Physiol. 300(1):H319-25 2011.

[8] Costa M, Goldberger AL, and Peng CK. Multiscale entropy analysis of biological signals. Phys Rev E StatNonlinSoftMatter Phys 71: 021906, 2005.

[9] Ferrario M, Signorini MG, Magenes G. Estimation of long-term correlations from fetal heart rate variability signal for the identification of pathological fetuses. Conf Proc IEEE Eng Med Biol Soc. 2007: 295-8, 2007.

[10] Lake DE. Renyi entropy measures of heart rate Gaussianity. IEEE TransBiomed Eng 53: 21-27, 2006.

[11] Richman JS, and Moorman JR. Physiological time series analysis using approximate entropy and sample entropy. American Journal of Physiology 278: H2039-H2049, 2000.

[12] Richman JS. Sample entropy statistics and testing for order in complex physiological signals. Communications in statistics - theory and methods 36: 2006.

[13] Richman JS, Lake DE, and Moorman JR. Sample entropy. Methods Enzymol 384: 172-184, 2004.

[14] Pincus SM. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Science 88: 2297-2301, 1991.

[15] Pincus SM, and Goldberger AL. Physiological time-series analysis: what does regularity quantify? American Journal of Physiology 266: H1643-H1656, 1994.

[16] Griffin MP, Lake DE, Bissonette EA, Harrell FE, Jr., O'Shea TM, and Moorman JR. Heart rate characteristics: novel physiomarkers to predict neonatal infection and death. Pediatrics 116: 1070-1074, 2005.

[17] Griffin MP, O'Shea TM, Bissonette EA, Harrell FE, Jr., Lake DE, and Moorman JR. Abnormal heart rate characteristics are associated with neonatal mortality. Pediatr Res 55: 782-788, 2004.

[18] Griffin MP, O'Shea TM, Bissonette EA, Harrell FE, Jr., Lake DE, and Moorman JR. Abnormal heart rate characteristics preceding neonatal sepsis and sepsis-like illness. Pediatr Res 53: 920-926, 2003.

[19] Lake DE, Richman JS, Griffin MP, and Moorman JR. Sample entropy analysis of neonatal heart rate variability. American Journal of Physiology 283: R789-R797, 2002.

[20] Porta A, Baselli G, Liberati D, Montano N, Cogliati C, Gnecchi-Ruscone T, Malliani A, and Cerutti S. Measuring regularity by means of a corrected conditional entropy in sympathetic outflow. Biol Cybern 78: 71-78, 1998.

[21] Ramdani S, Seigle B, Lagarde J, Bouchara F, and Bernard PL. On the use of sample entropy to analyze human postural sway data. Med Eng Phys 31: 1023-1031, 2009.

[22] Lake DE. Nonparametric entropy estimation using kernel densities. Methods Enzymol 467: 531-546, 2009.