[11] Sample Entropy

By JOSHUA S. RICHMAN, DOUGLAS E. LAKE, and J. RANDALL MOORMAN

Introduction

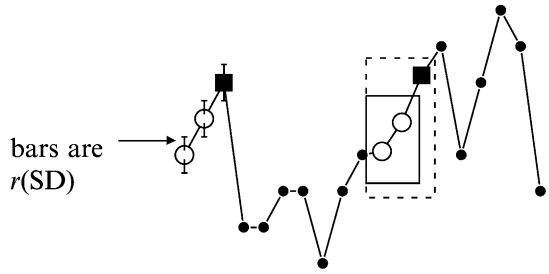
The findings of deterministic dynamics in seemingly random physical process have excited biological researchers who collect time series data. The best tools for this kind of analysis require extremely long and noise-free data sets that are not available from biological experiments. Nonetheless, one such tool has found widespread use. In 1991, Pincus adapted the notion of ''entropy'' for real-world use.[1] In this context, entropy means order or regularity or complexity, and has roots in the works of Shannon, Kolmogorov, Sinai, Eckmann and Ruelle, and Grassberger and coworkers. The idea is that time series with repeating elements arise from more ordered systems, and would be reasonably characterized by a low value of entropy. Were the data sets infinite and perfect, it would be possible to determine a precise value of entropy. Biological data sets are neither, but Pincus had the important insight that even an imperfect estimate of entropy could be used to rank sets of time series in their hierarchy of order. He introduced approximate entropy (ApEn), and many papers have appeared drawing conclusions about the relative order of physiological processes.

In principle, the calculation is simple enough, and is shown schematically in Fig. 1. ApEn quantifies the negative natural logarithm of the conditional probability (CP) that a short epoch of data, or template, is repeated during the time series. Having selected a template of length m points, one identifies other templates that are arbitrarily similar and determines which of these remain arbitrarily similar for the next, or $m + 1$st point. ''Arbitrarily similar'' means that points are within a tolerance r of each other, where r is usually selected as a factor of the standard deviation (SD). The negative logarithm of the conditional probability is calculated for each possible template and the results averaged. If the data are ordered, then templates that are similar for m points are often similar for $m + 1$ points, CP approaches 1, and the negative logarithm and entropy approach 0.

The concepts are solid and the potential utility is great. We found, however, that there are practical issues of great importance in implementing the algorithm. These findings motivated us to develop sample entropy

[1] S. M. Pincus, *Proc. Natl. Acad. Sci. USA* **88,** 2297 (1991).

0076-6879/04 $35.00



$$\text{ApEn} \approx \Sigma -\log (1+A_i)/(1+B_i)$$
$$\text{SampEn} = -\log ((\Sigma A_i)/(\Sigma B_i)) = -\log A/B$$

For regular, repeating data, A/B nears 1 and entropy nears 0.

FIG. 1. Schematic demonstration of entropy estimation using approximate entropy (ApEn) and sample entropy (SampEn). The time series begins with the ith template. In this example, m is 2. The tolerance for accepting matches is r times the standard deviation, and is shown by the error bars. Here, the template is matched by the 11 and 12th points (solid box), and the $m + 1$st points also match (dashed box). Thus quantities A and B both increment by 1.

(SampEn) as an alternative method for entropy estimation in real world data. In this chapter, we first overview the problems of ApEn and how SampEn addresses them. We next present a formal implementation of SampEn and discuss the practical issues of optimization of parameters and data filtering. This is followed by a discussion of the difficulties with short data sets and nonstationary data. We end with comments on interpretation of entropy estimates and a direct comparison of ApEn and SampEn. The algorithms discussed are available at www.Physionet.org. For full details, we refer the reader to our original papers.[2,3]

Motivation for SampEn Analysis

In our initial implementation of ApEn analysis of heart rate dynamics, we encountered practical questions.

1. What if some templates have no matches, and the CPs are not defined? Pincus follows the teaching of Eckmann and Ruelle and allows

[2] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, *Am. J. Physiol.* **283,** R789 (2002).
[3] J. S. Richman and J. R. Moorman, *Am. J. Physiol.* **278,** H2039 (2000).

templates to match themselves. Thus if there are no other matches, the CP is 1 and ApEn is 0, a report of perfect order. If there are only a few template matches, then the result is biased toward 0, and the bias resolves with lengthening data sets and more template matches. Pincus and co-workers have explicitly described the bias of ApEn and have contended that the important goal of reporting a correct hierarchy of order is preserved, a feature named ''relative consistency.'' Thus if time series A arises from a more ordered system than time series B, then ApEn of A will always be less than ApEn of B regardless of the number of template matches and the extent of the bias.

We have developed SampEn statistics to reduce the bias of ApEn statistics.[2,3] We have found that SampEn preserves relative consistency more often than ApEn.

2. How long should a template be, and how similar should ''arbitrarily similar'' be? That is, how does one pick m and r? The usual suggestion is that m should be 1 or 2, noting that there are more template matches and thus less bias for $m = 1$, but that $m = 2$ (or greater) reveals more of the dynamics of the data. The usual suggestion is that r should be 0.2 times the SD of the data set on empirical grounds.

We have proposed a systematic approach to selecting m and r based on evaluation of novel error metrics for SampEn that are discussed later.

3. Does a low value of entropy always mean increased order?

No differential diagnosis has been suggested for ApEn. We have found that time series with spikes have low values of ApEn and SampEn, a direct consequence of the practice of basing the tolerance r on the SD.[2] Spikes inflate the SD and allow many template matches of lengths m and $m + 1$ in the baseline. The high CP leads inevitably to a low value for the entropy, but it is not intuitively correct that a large number of matching templates in the baseline necessarily reflects order.

Sample Entropy Calculation

As a statistic, *SampEn(m,r,N)* depends on three parameters. The first, m, determines the length of vectors to be considered in the analysis. That is, given N data points $\{u(j): 1 \le j \le N\}$, form the $N-m+1$ vectors $x_m(i)$ for $\{i \mid 1 \le i \le N - m + 1\}$ where $x_m(i) = \{u(i + k): 0 \le k \le m - 1\}$ is the vector of m data points from $u(i)$ to $u(i + m - 1)$. The distance between two vectors, denoted $d[x_m(i), x_m(k)]$, is defined to be $\max\{|u(i + j) - u(k + j)|: 0 \le j \le m - 1\}$, the maximum difference between their corresponding scalar components.

The original formulation of SampEn closely followed the Grassberger–Procaccia correlation integral. However, a more intuitive and less notationally intensive approach simply considers *SampEn(m,r,N)* to be the negative

natural logarithm of the empirical probability that $d[x_{m+1}(i), x_{m+1}(k)] \leq r$ given that $d[x_m(i), x_m(k)] \leq r$. Where the values of the parameters are specified let B denote the number of pairs $x_m(i)$, $x_m(k)$ such that $d[x_m(i), x_m(k)] \leq r$, and let A then be the number of pairs of vectors $x_{m+1}(i)$, $x_{m+1}(k)$ such that $d[x_{m+1}(i), x_{m+1}(k)] \leq r$. Then $SampEn(m,r,N) = -\ln(A/B)$. For simplicity, we refer to a match of two vectors of length m as a *template match,* and the matching of two vectors of length $m + 1$ as a *forward match*.

The computationally intensive aspect of the algorithm is simply counting the numbers A and B, which at least requires checking the distance between the $[N(N-1)]/2$ pairs of points, then counting the number of vectors that match for m and $m + 1$ points. There are two additional computational considerations, both of which are fully discussed in the initial paper by Richman and Moorman.[3] First, we do not compare any vector with itself since this provides no new information. Second, although the vector $x_m(N - m + 1)$ exists, we do not use it for comparisons, since the vector $x_{m+1}(N - m + 1)$ is not defined. That is, if some $x_m(i)$ were to match $x_m(N - m + 1)$, they could not both be incremented, and thus could contribute only to B but not A. In practice, for a data set of reasonable size this will have little effect, but scenarios can be constructed where it could significantly alter results.

While SampEn is often used with just one fixed value of m, there is an implementation of the algorithm that efficiently calculates *SampEn(k,r,N)* for all k from 1 up to m. The basic idea is to build up runs of points matching within the tolerance r until there is not a match and keep track of template matches $A(k)$ and $B(k)$ of all lengths k up to some specified parameter m. If a particular run ends up being of length 4, for example, then that means that 1 is added to the count for template matches of length 4. In addition, there are 2 template matches of length 3, 3 of length 2, and 4 of length 1 that need to be added to the corresponding counts. A special distinction is needed when a run ends at the last point in the data where the $A(k)$ counters are incremented but the $B(k)$ counters are not.

Practical Issues with SampEn Calculation

Optimizing Parameters

Having decided how to manage the data the next task must be to optimize the parameters for *SampEn(m,r,N)* by some rational strategy. For most current applications the parameters must be fixed to allow valid comparisons. Circumstances that may indicate varying parameters will be discussed separately. The parameter N is usually taken as the size of the

data set. Since N determines the range of *SampEn(m,r,N)* by providing an upper limit for B, care must be used when comparing epochs of differing lengths, especially if *SampEn(m,r,N)* approaches its theoretical maximum for any epoch. Selecting m and r is more involved. Assuming that there is some underlying deterministic structure, as m grows larger, the conditional probability of a forward match should increase, since conditioning on a longer vector should increase predictive accuracy. On the other hand, due to the noise from measurement errors or system interaction the number of template matches will decrease as m increases simply because it is less likely that the noise level will permit long templates to match at all. In fact, beyond a certain length the templates will largely be due to chance rather than dynamic similarity. In addition, and likely before this point is reached, the matches will be so few that the statistics become unreliable. Conversely, when m is too small, template matches are plentiful (B is large), but not enough predictive information is contained in the short template leading to an underestimation of the probability of a forward match. The compromises involved in choosing r are similar. An r that is too small, smaller than the typical noise amplitude, will result in many vectors that are actually similar failing to match. On the other hand, when r is too large, *SampEn(m,r,n)* loses its discriminating power entirely since most templates will look similar to one another given sufficiently lax matching conditions.

The ideal choice then would be to make m as large and r as small as possible, while ensuring that B remains large enough to ensure precise statistics. Ideally, one could choose m based upon knowledge of the time scale of the underlying process, and r based upon knowledge of the scale of signal noise. In practice, we have chosen m by fitting an autoregressive model to the data and setting m to be the optimal order of the model. This method has the advantage of drawing upon a vast literature and established methods for optimizing such models. We then choose r to minimize the relative error of A/B and *SampEn(m,r,N)*.

Given the B template matches calculating the standard error of the conditional probability estimate $\mathrm{CP} = A/B$ is nontrivial. This is largely due to the fact that vector comparisons are not independent. Not only should there be some dependence due to the underlying process but some comparisons are formally dependent on others because they overlap and thus share data points. Nevertheless an approximation of this standard error is given by

$$\sigma_{CP}^2 = \frac{CP(1-CP)}{B} + \frac{1}{B^2}\left[K_A - K_B(CP)^2\right]$$

where K_A denotes the number of overlapping pairs or $m + 1$ point vectors, and K_B denotes the number of overlapping pairs of m point templates. That is, K_A is the number of pairs $[x_{m+1}(i), x_{m+1}(j)]$, $[x_{m+1}(k), x_{m+1}(l)]$ in which a vector from one pair shares at least one data point with a vector from the second pair and both pairs match within r. This would be the case if, for example, $x_{m+1}(i)$ and $x_{m+1}(k)$ had a point in common. K_B is defined analogously for vectors of length m. Counting K_A and K_B requires both care and a considerable amount of computer time. From this we can approximate the standard error of *SampEn(m,r,n)* by σ_{CP}/CP. Having chosen m we then advise choosing r to minimize the larger of σ_{CP}/CP and $\sigma_{CP}/-\log(CP)CP$, the relative errors of the empirical CP and *SampEn(m,r,n)*, respectively. The estimate of σ_{CP} is also useful in that it allows for the calculation of approximately 95% confidence intervals for *SampEn(m,r,n)* by 1.96 (σ_{CP}/CP).

At this point another consideration must be addressed regarding r. It is standard practice to set r to be some multiple of the standard deviation of the data. This effectively normalizes the data, adjusting for differences of scale. This is appropriate if the analysis is driven by the search for order in the dynamics. If, however, the goal is to efficiently distinguish various data sets the rescaling can make two data sets appear identical when they clearly are not. There can also be problems with interpretation since broad trends or transient phenomena can inflate the variance to the point where all local dynamics appear similar only because they have made the tolerance r too coarse to discern genuine similarity.

Having chosen a method for selecting parameters it is then necessary to choose either some or all of the data and calculate *SampEn(m,r,N)* for a range of the parameters. Since these calculations can be time consuming it is often advisable to select either a random sample from all the data or some characteristic data sets to find parameters that are broadly appropriate for all the data. This is not, of course, a definitive method for choosing parameters, but one that balances the requirements of discerning distinctive features of the data (sufficiently large m and small r) and allowing for accurate and precise estimates (adequate B).

Other suggested approaches involve choosing the parameters that maximize the deviation of *SampEn(m,r,N)* from *SampEn(0,r,N)*, the negative logarithm of the probability that a pair of points will match within r given no other information, i.e., the value of *SampEn(m,r,N)* if the data were in random order. This method focuses on locating the parameters that detect the most ''order'' in the data. Above all, the most important factor is choosing parameters appropriate for the data at hand. It cannot be overemphasized that prior to any meaningful analysis the parameters must be chosen to be suitable for the data. We have no reason to suspect that there

are universally optimal parameters suitable for any and all data sets. We therefore advocate that they be chosen either by optimizing their empirical properties on a subset of the data, or by choosing reasonable parameters based on prior knowledge and checking their performance by, for example, making sure that the resulting confidence intervals are not too wide.

Another strategy for distinguishing data sets is to view the identification of optimal parameters from another angle. We have so far considered identifying the finding parameters that maximize the usefulness of SampEn statistics as the characteristic measure of the time series. We could choose instead to optimize properties of *SampEn(m,r,N)* and use the resulting parameters to characterize the time series. This approach has the potential advantage of providing a categorical classification by m of predictive scale (or by a time delay τ, if using the time delay techniques discussed later) or a continuous measure of structural detail by r. Again, a reasonable approach is to maximize the difference between *SampEn(m,r,N)* and *SampEn(0,r,N)*. Another would be to maximize the difference between *SampEn(m,r,N)* for the time series and for a surrogate time series incorporating the linear correlation structure of the time series.[4]

Once the parameters have been chosen all the data may be analyzed and the values of *SampEn(m,r,N)* incorporated into regression models or tested via their confidence intervals as to whether they differ significantly from one another. Again, under these conditions SampEn statistics appear to provide a robust way to discriminate between time series. More care is needed to say what proportion of the measure is due to order and what is due to nonstationarity.

Filtering

Many time series techniques assume at least weak stationarity, so data are often subjected to filtering to render them approximately stationary prior to analysis. This alters the correlation structure of the data in a way that can change the SampEn analysis, and should be done judiciously. The accuracy and precision of $SampEn(m,r,N) = -\ln(A/B)$ is limited by the magnitude of B. Trends in the data can separate points whose local dynamics are similar but whose locations are far removed. When trends are understood, they may be removed, laying bare the unexplained dynamics. Nevertheless, there is always the risk that the most interesting features could be blunted or obliterated by filtering and smoothing. Leaving the data unfiltered removes some danger of manufacturing spurious matches at the expense of lowering B. Fortunately, only large data sets really have

[4] T. Schreiber and A. Schmitz, *Physica. D* **142,** 346 (2000).

enough points to display complicated trends, and these are the least likely, by virtue only of their size, to be hampered by a low value of B.

This discussion has so far assumed that the time scale of the studied dynamics are near to the scale of sampling. If, however, one is more interested in longer time scales, that is, when the data are effectively oversampled, the usual approach can be modified in several ways. These generally involve changing or altering the reconstruction vectors $x_m(i)$ to include time delays. This is accomplished by using vectors of the form $x_m^\tau(i) = \{u(i + \tau k)\colon 0 \le k \le m - 1\}$ for a time delay τ. Techniques for choosing τ are discussed extensively in the literature.[5] In this case there is little reason to use every possible vector since for an oversampled process $x_m^\tau(i)$ will match $x_m^\tau(i + 1)$ as often and uninterestingly as $x_m(i)$ will match $x_m(i + 1)$. Thus some consideration needs to be given as to which subsample of vectors should be used for analysis. Recently Costa *et al.*[6] developed multiscale entropy analysis (MSE), which compresses the data by constructing coarse time series of the form $y^\tau(j) = 1/\tau \sum_{i=(j-1)\tau+1}^{j\tau} u(i)$ for $1 \le j \le N/\tau$, that is, by dividing the series into a sequence of nonoverlapping windows of τ points and representing the window by its sample mean. The new point $y^\tau(j)$ corresponds to the mean of the jth window of τ points.

SampEn analysis was then carried out on these new series to give information regarding the dynamics on larger time scales. Clearly this is easily generalized to a family of such techniques. A more general version less sensitive to the window locations would be to replace the series $\{u(j)\colon 1 \le j \le N\}$ by a moving averaged series $\{u^\tau(j)\colon \tau \le j \le N\}$ where each point is the mean of the previous τ points of the original time series. SampEn analysis could then be carried out on the delayed vectors $\tilde{x}_m^\tau(i) = u(i + k\tau)\colon 0 \le k \le m - 1\}$. Doing this for various values of τ would broaden the MSE approach. It may also be desirable to utilize weighted averaging rather than the simple window average depending on the desired smoothing properties.

If the data are fairly precise and merely oversampled, it may be worthwhile to use usual m-vectors and $m + 1$ vectors of the form

$$x_{m+1,\tau}(i) = \{u(i + k\tau), \quad u[i + (m - 2) + \tau]\colon 0 \le k \le m - 2\}$$

Thus the first m points of the vector are the usual m-dimensional reconstruction vector while the last point has been sampled at the delay τ. This allows the oversampled (and dynamically uninteresting) portion of the data to fix the system's location in phase space more precisely, while the

[5] H. D. I. Abarbanel, "Analysis of Observed Chaotic Data." Springer-Verlag, New York, 1996.

[6] M. Costa, A. L. Goldberger, and C. K. Peng, *Phys. Rev. Lett.* **89,** 068102 (2002).

comparisons at the next τth point allows the system enough time to evolve so that matches will be meaningful. Experimentation will be required to select the optimum choice of τ.

SampEn analysis does not require that any assumptions be made regarding the stationarity of the data. However, in the absence of approximate stationarity care must be taken with the interpretation of SampEn statistics. Indeed, as we shall see throughout this discussion, care must be taken in all circumstances when interpreting SampEn statistics as indicative of order in the data.

Difficult Data

Short Data Sets

The analysis of very short data sets may call for extra care. *SampEn(m,r,N)* is bounded by 0 and $\ln[(N-m)(N-m-1)]/2$. Thus short data sets will have a decreased range. They will also have smaller values of B and thus less precise statistics. This will be exacerbated by the fact that in a small data set a higher proportion of comparisons involves overlapping templates, a factor that will tend to inflate the variance of *SampEn(m,r,N)*. We have also shown previously that for short sets of i.i.d. random numbers, overlapping templates lead to an average underestimation of A/B. Scenarios can be imagined wherein for oversampled data, the underlying process does not have much time to evolve along any particular template. In these cases overlapping templates will match each other with high probability. Unfortunately, the high estimated CP will be due to the local similarity due to slow evolution of the system rather than self-similarity in phase space.

One obvious remedy would be to disregard comparisons involving overlapping $m+1$ vectors. There are generally $[(N-m)(N-m-1)]/2$ distinct pairs of templates. Discarding those pairs of vectors that would overlap when incremented to length $m+1$, there are $[(N-2m)(N-2m-1)]/2$ distinct pairs. The fraction of matches discarded is then

$$\frac{mN\left(2-3\frac{m}{N}-\frac{1}{N}\right)}{N^2\left(1-\frac{m}{N}\right)\left(1-\frac{(m+1)}{N}\right)} \simeq \frac{2m}{N}$$

for $m << N$. Thus for large N in exchange for a modest reduction in the number of comparisons, we can ensure that the vectors are at least formally, if not dynamically, independent, which should lead to a reduction in SampEn's variance as well as increased confidence that the CP is correctly estimated.

Nonstationary Data

We noted previously that there is no requirement that the data be stationary. The main concern is that the nonstationarity of the data could inflate the tolerance parameter *r,* and thereby induce spurious matches. Given large amounts of data, choosing *r* to be as small as possible should provide a reasonable safeguard. Nevertheless, particularly when confronted with nonstationary data one may decide to average the values of *SampEn(m,r,n)* for windows of length $n < N$, or adopt some other resampling approach. *SampEn(m,r,n)* provides a global measure of the probabilistic self-similarity of the data. Since the hallmark of nonstationarity is a varying probabilistic structure, there may be cases in which it is advisable to perform a piecewise analysis, dividing the time series into sections with homogeneous structure and calculating SampEn for each segment. One could then report the average of the individual SampEn statistics, where each is weighted according to their proportion of the entire series.

Interpretation of SampEn

SampEn was originally intended as a measure of the order in a time series. We have noted, however, that low SampEn statistics, indicative of high CP estimates, cannot be assumed to imply a high degree of order. There are in general two distinct mechanisms for generating high CP estimates. The first is that genuine order has been detected. The second derives from the fact that *r* is usually taken as a proportion of the standard deviation of the series, thus rendering the analysis scale free. When nonstationary features, especially transient ''spikes,'' inflate the variance and thus coarsen the criterion for matching it can happen that virtually all recorded matches are similar only because their dynamic scale is dwarfed by the spikes. This has two ramifications. First, when the aim is to quantify order with SampEn statistics, additional scrutiny is required to ascertain how much of the statistic's value is, in fact, due to order. Second, when the goal is simply to numerically distinguish between data sets, SampEn is very adept at detecting such spikes and generally discerning which epochs are atypical.

When order detection is desired we suggest comparing *SampEn(m,r,N)* to *SampEn(0,r,N)* to see whether conditioning on the templates significantly increased the probability of a forward match. If *SampEn(0,r,N)* lies outside the confidence interval for *SampEn(m,r,N)*, then we are reasonably confident that a significant amount of order was detected. If the goal remains chiefly to discriminate among series, but it is desirable to lean toward detecting order, *r* can be taken to be a proportion of the mean

(or variance) of the series of first differences, that is, of the average change from one point to the next. This effectively rescales the series to have similar dynamic scales, and will generally give less weight to spikes. One could of course simply perform SampEn analysis on the series of first differences as well, but this acts as a high-pass filter and will obscure long-range dynamics. Another approach would be to use r as an absolute number, thereby allowing SampEn to distinguish between series that differ only in their scaling. Unfortunately when the scales of the time series vary widely, no r may be suitable to provide comparisons across all sets with a meaningful interpretation for the statistic per se. However, the SampEn statistics will likely still efficiently detect which data sets are atypical.

ApEn and SampEn

We now summarize the differences between sample entropy and approximate entropy and discuss possible bridges between the two approaches. Let B_i denote the number of template matches with $x_m(i)$ and A_i denote the number of template matches with $x_{m+1}(i)$. The number $p_i = A_i/B_i$ is an estimate of the conditional probability that the point x_{j+m} is within r of x_{i+m-1} given that $x_m(j)$ matches $x_m(i)$. ApEn is calculated by

$$ApEn(m, r, N) = -\frac{1}{N-m}\sum_{i=1}^{N-m} \log\left(\frac{A_i}{B_i}\right)$$

and is the negative average natural logarithm of this conditional probability. Self-matches are included in the ApEn algorithm to avoid the $p_i = 0/0$ indeterminate form, but this convention leads to a conditional probability estimate of 1. This necessarily overestimates the true value and leads to a noticeable bias especially for smaller N and larger m. In contrast to the above, SampEn is calculated by

$$SampEn(m, r, N) = -\log\left(\sum_{i=1}^{N-m} A_i \Big/ \sum_{i=1}^{N-m} B_i\right) = -\log(A/B)$$

which is just negative the logarithm of an estimate of the conditional probability CP of a match of length $m + 1$ given a match of length m.

There are advantages in using SampEn over ApEn. ApEn is more sensitive to bias from short time series and to problems arising from outliers. We discovered that ApEn could give misleading or contradictory results, particularly for very short or noisy time series. Sample entropy was developed to improve some of these properties while maintaining the spirit

of measuring the probability that two vectors that are close for m points remain close. In all cases studied so far SampEn appears to be a more robust and less biased statistic than ApEn. SampEn's formulation is more amenable to the construction of confidence intervals that give approximate guidelines for significance tests.

Despite the advantages of using SampEn, there may be theoretical and personal preference issues that necessitate using ApEn. ApEn statistics are much more closely related to Shannon's definition of information-theoretic entropy that spurred the field of information theory more than 50 years ago and they possess an additive property that is not rigorously shared by sample entropy. Rukhin[7] gives a clear statement of the relationship between Shannon's entropy and ApEn. In addition, it is formally very closely related to the entropy of statistical mechanics. There exists a vast literature exploring the properties of such functions from both perspectives. In particular, Rukhin has recently proven that for discrete series, ApEn statistics are asymptotically distributed as a χ^2 random variable. This suggests a future practice of applying ApEn to discretized signals. The trade-off here is that while information is lost to discretization, the parameter r can be discarded since sequences of discrete values can be considered to match either exactly or not at all. We note that while there are, as yet, no theoretical proofs of SampEn's asymptotic properties, Monte Carlo studies have given no reason to suppose that SampEn's asymptotic properties are not well behaved.

Approximate entropy also is optimally suited to measure the Gaussianity of a distribution and a process. Berg's theorem established that the maximum entropy for a random process with finite variance is attained by a Gaussian process. Thus ApEn values departing from the theoretical maximum indicate a lack of Gaussianity. SampEn can also be effectively used as a measure of Gaussianity though its maximum occurs for non-Gaussian random processes. When ApEn is used, modifications of how ApEn handles zero and small number of matches can help minimize its bias and begin to approach the statistical stability of sample entropy. This is an important open area for research. One of the major sources of ApEn bias results when a template does not match any others, whereupon ApEn estimates a conditional probability of 1. Pincus has suggested correcting this by incorporating a factor ε into calculation, thus setting the CP estimate in the absence of matches.

In any implementation of ApEn, we advocate not counting self-matches. When $A_i = B_i = 0$ (or near 0), the estimate A_i/B_i can be replaced with some value ε that is sufficiently smaller than 1 to indicate the likely

[7] A. L. Rukhin, *J. Appl. Prob.* **37,** 88 (2000).

uncertainty of matching if no matching templates exist.[6,8,9] The value cannot be too small or $-\log(\varepsilon)$ is prohibitively large and can lead to an upward bias. Clearly, selecting the right value of ε is dependent on the process being studied. One approach is to use estimates *ApEn(k,r,N)* with $k < m$ based on shorter templates (and thus less biased) to approximate $-\log(A_i/B_i)$ when matches are lacking. This is a generalization of the method used by Porta and co-workers[6,9] in the calculation of *corrected conditional entropy* (CCE) where the Shannon entropy of the process is used to approximate $-\log(A_i/B_i)$ when $B_i = 0$. We suggest that when this strategy is employed that the ε factor should correct toward the SampEn CP estimate. This novel and promising approach would take advantage of the statistical value of sample entropy by setting $\varepsilon = CP = A/B$ from the SampEn calculation and lead to a hybrid of the two algorithms.

There remain several unresolved issues in the use of SampEn. Much work needs to be done to evaluate and rigorously establish the statistical properties of SampEn statistics. Most desirable would be proof of the statistics' asymptotic distributions. It will also be important to develop tests for the proportional contributions of order and nonstationarity. Studies also need to be undertaken to ascertain how the various optimization methods lead to the selection of parameters, and how sensitive those selections are to outliers in the data. Given the parameter dependence of the statistics and its relation to comparing datasets, we are also working to develop related parameter-free statistics.

Thus we conclude the following:

1. A low value of ApEn is due to bias, order, or spikes. The relative contributions cannot be quantified.
2. A low value of SampEn is due to order or spikes.

Our practice is to use SampEn with optimized choices of m and r. Work remains on confident parsing of the results to order and spikes.

[8] S. M. Pincus and A. L. Goldberger, *Am. J. Physiol.* **266,** H1643 (1994).
[9] A. Porta, G. Baselli, F. Lombardi, N. Montano, A. Malliani, and S. Cerutti, *Biol. Cybern.* **81,** 119 (1999).