New Methods for Measuring Independence with an Application to Chaotic Systems

Yang Chen

Department of Radio Engineering, Southeast University, Nanjing 210096, China
Phone: +86-25-3793267, Fax: +86-25-7712719
Email: cheny@seu.edu.cn

Abstract. Some new independence measures are proposed and shown to be good criteria beyond the classical mutual information (MI) for choosing delay in strange attractor reconstruction.

1. Introduction

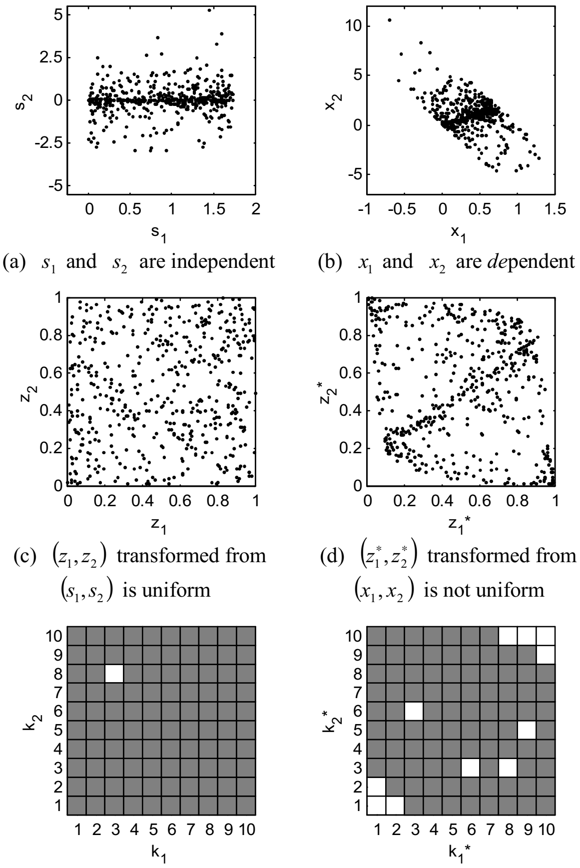
In many cases, the commonly seen one-dimensional scalar signals or time series are observations of a certain variable of a multivariate dynamical system. An interesting property of dynamical systems is their multivariate behavior can be approximately "regenerated" from observations of just one variable. The method is called state space reconstruction. For chaotic systems, it is also termed strange attractor reconstruction, which has become a fundamental approach in chaotic signal analysis during the ever-interesting research for chaos theory and applications in the past few decades. A basic way to reconstruction is by means of delay [3]. That is, we take the current observed value and the values after equally spaced time lags τ , 2τ , 3τ … to obtain "observations" of several reconstructed variables where τ is the (quantity of) delay. Then how to choose a proper τ ? Since reconstruction aims to release the information of the whole system "condensed" in one variable, we naturally do not want the new variables obtained by delay be simple repetitions of the original one, or their existence might be meaningless. So we want the reconstructed variables to be as independent as possible [4]. We know that n variables $r_1,\cdots,r_n$ are independent *if and only if* (iff) their joint *probability density function* (PDF) can be factored as the product of their individual (marginal) PDFs, i.e.,

$$p_{r_1\cdots r_n}(u_1,\cdots,u_n)=\prod_{i=1}^{n} p_{r_i}(u_i) \qquad \forall\,(u_1,\cdots,u_n) \tag{1}$$

However, in practice, we usually only have observations of variables. So how to conveniently estimate independence from observations is the main problem to be addressed in this paper. New independence measures named grid occupancy (GO), quasi-entropy (QE), and generalized mutual information (GMI) are proposed and shown to be good criteria for choosing delay.

2. Grid Occupancy

Hereafter, we denote the *cumulative distribution function* (CDF) of variable r by $q_r(\cdot)$,



(a) s_1 and s_2 are independent (b) x_1 and x_2 are *de*pendent

(c) (z_1,z_2) transformed from (s_1,s_2) is uniform (d) (z_1^*,z_2^*) transformed from (x_1,x_2) is not uniform

(e) GO of s_1 and s_2 is –0.99 (f) GO of x_1 and x_2 is –0.89

Fig. 1. Illustration of Grid Occupancy (GO).

$q_r(u)=\text{Prob}(r\le u)=\int_{-\infty}^{u} p_r(v)\,dv$.

Consider two continuous variables r_1 and r_2. We transform them by their individual CDFs respectively to get two continuous variables

$$z_1 \equiv q_{r_1}(r_1), \qquad z_2 \equiv q_{r_2}(r_2) \tag{2}$$

Clearly $(z_1,z_2)\in[0,1]\times[0,1]$. And from

$$p_{z_1z_2}(u,v)=\frac{p_{r_1r_2}(w,h)}{p_{r_1}(w)p_{r_2}(h)} \tag{3}$$

where $u=q_{r_1}(w)$ and $v=q_{r_2}(h)$, we know that r_1 and r_2 are independent iff $p_{z_1z_2}(u,v)=1$, $\forall(u,v)\in[0,1]\times[0,1]$, i.e., z_1 and z_2 are jointly uniformly distributed in $[0,1]\times[0,1]$.

Fig. 1 illustrates this fact. Fig. 1 (a) and (b) are 500 observed points of two independent variables, s_1 and s_2, and two *de*pendent variables, x_1 and x_2, respectively. The points in Fig. 1 (c) and (d) are 500 observations of (z_1, z_2) and (z_1^*, z_2^*), respectively, where $z_1 \equiv q_{s_1}(s_1)$, $z_2 \equiv q_{s_2}(s_2)$, $z_1^* \equiv q_{x_1}(x_1)$, and $z_2^* \equiv q_{x_2}(x_2)$. Clearly the points in Fig. 1 (c) distribute uniformly whereas those in Fig. 1 (d) do not. Now we partition the region $[0,1]\times[0,1]$ in Fig. 1 (c) or (d) into an $l \times l$, say, 10×10, same-sized squares. A square is said to be occupied if there is at least one point in it. Intuition tells that being uniform is the most efficient way to occupy maximum number of squares and this is confirmed in Fig. 1 (e) and (f) where the situations of Fig. 1 (c) and (d) are shown respectively. (The occupied squares are shaded.) We define an independence measure of r_1 and r_2, denoted by $\alpha(r_1, r_2)$, as minus the ratio of occupied squares, and call it the *grid occupancy* (GO). As Fig. 1 (e) and (f) clearly show, $\alpha(s_1, s_2) = -0.99 < \alpha(x_1, x_2) = -0.89$. So s_1 and s_2 are more independent than x_1 and x_2. In practice, analytical forms of CDFs are usually unknown. But this does not matter since the CDF can be well simulated by mapping the data to samples from a uniform distribution while preserving the ordering.

3. Quasi-Entropy

Let l be an integer greater than 1. Define an *l*-level uniform quantization operator $D_l(\cdot)$ on the interval $[0,1]$ as

$$D_l(u) = \begin{cases} \lceil ul \rceil & 0 < u \le 1 \\ 1 & u = 0 \end{cases} \tag{4}$$

where $\lceil v \rceil$ denotes the least integer not less than v. Clearly $D_l(u) \in \{1, 2, \cdots, l\}$ for $u \in [0,1]$.

Now let us define discrete variables

$$k_1 \equiv D_l(z_1), \qquad k_2 \equiv D_l(z_2) \tag{5}$$

where z_1 and z_2 are defined as in (2). Let $p_{k_1k_2}(i,j) = \text{Prob}(k_1 = i \text{ and } k_2 = j)$ be the joint probability of k_1 and k_2 where $(i,j) \in \{1,\cdots,l\}\times\{1,\cdots,l\}$. If r_1 and r_2 are independent, then (z_1, z_2) is uniform in $[0,1]\times[0,1]$ and thus (k_1, k_2) is uniform in $\{1,\cdots,l\}\times\{1,\cdots,l\}$. This leads us to propose the second independence measure

$$\beta(r_1, r_2) = \sum_{i=1}^{l}\sum_{j=1}^{l} f\left(p_{k_1k_2}(i,j)\right) \tag{6}$$

where the function $f(\cdot)$ is strictly convex on $[0,1]$. Due to the Jensen's inequality, we have

$$\beta(r_1, r_2) \ge l^2 f\left(1/l^2\right) \tag{7}$$

with equality iff (k_1, k_2) is uniform in $\{1,\cdots,l\}\times\{1,\cdots,l\}$. Indeed, we can prove [5]

Theorem 1: If r_1 and r_2 are independent, then $\beta(r_1, r_2)$ reaches its minimum; if r_1 and r_2 are not independent, then there exists l_0 such that $\beta(r_1, r_2)$ cannot reach its minimum for any $l > l_0$.

We know that the entropy in information theory $H(p) = -\sum_i p(i)\log p(i) = \sum_i f(p(i))$ where $f(u) = -u\log u$ is strictly concave. So we may call $\beta(r_1, r_2)$ *quasi-entropy* (QE) due to its similar form to entropy.

Since there are infinitely many strictly convex functions on $[0,1]$, QE is actually a class of infinitely many independence measures. For example, $f(\cdot)$ can be $f(u) = u^a$ ($a > 1$), $f(u) = -u^a$ ($0 < a < 1$), $f(u) = \left(u - 1/l^2\right)^2$, $f(u) = -\sqrt{1-u^2}$, $f(u) = 1/(1+u)$, $f(u) = 1/\left(1+\sqrt{u}\right)$, $f(u) = -\log(1+u)$, $f(u) = -u\log(1+1/u)$, $f(u) = u\log u$, $f(u) = \exp(u)$, $f(u) = \exp(-u)$, $f(u) = a^u$ ($a > 0$ and $a \ne 1$), $f(u) = -\sin u$, and $f(u) = \tan u$, etc.

Interestingly enough, we can prove that the expectation of GO is also a kind of QE. It is the QE when the convex function $f(u) = -[1-(1-u)^N]/l^2$ where N is the number of total observed points satisfying $N > 1$. Inserting $f(u) = -[1-(1-u)^N]/l^2$ into (7), we have the expectation of GO

$$E[\alpha(r_1, r_2)] \ge \left(1 - 1/l^2\right)^N - 1 \tag{8}$$

For instance, in Fig. 1, $N = 500$, $l = 10$, by (8) we have $\min E[\alpha] = -0.993$. So we can judge rather confidently that s_1 and s_2 are independent whereas x_1 and x_2 are not.

Therefore, GO provides a way to estimate a subclass of QE. Besides, it is not difficult either to estimate other QE's. Using the approach at the end of last section and (5), we can get estimates of observations of (k_1, k_2) from observations of (r_1, r_2). The ratio that (i,j) occurs among them is an estimate of $p_{k_1k_2}(i,j)$. Insert it into (6) and an estimate of QE is obtained.

4. Generalized Mutual Information

Naturally, we can also use a strictly convex function to directly measure the uniformity of the joint density of z_1 and z_2. Similar to (6), define

$$\gamma(r_1, r_2) = \int_0^1\int_0^1 f\left(p_{z_1z_2}(u,v)\right) du\, dv \tag{9}$$

where $f(\cdot)$ is strictly convex on $[0,\infty)$. Easily shown that

$$\gamma(r_1, r_2) \ge f(1) \tag{10}$$

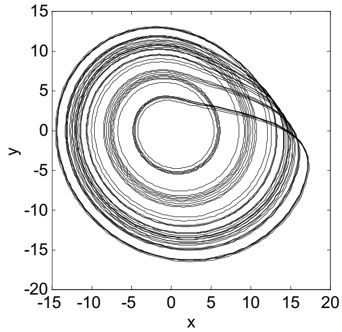


Fig. 2. Rossler attractor.

with equality iff r_1 and r_2 are independent. Of course, γ is also a class of infinitely many independence measures. Then how to estimate γ? Easily proved that

$$\lim_{l\to\infty}\sum_{i=1}^{l}\sum_{j=1}^{l} f\left(p_{k_1k_2}(i,j)l^2\right)/l^2 = \int_0^1\int_0^1 f\left(p_{z_1z_2}(u,v)\right)dudv \quad (11)$$

Especially, we let $l=2^m$, $m=0,1,2,\cdots$, and define

$$\gamma_m = \sum_{i=1}^{2^m}\sum_{j=1}^{2^m} f\left(p_{k_1k_2}^m(i,j)4^m\right)/4^m \quad (12)$$

where the superscript m of $p_{k_1k_2}^m$ represents that k_1 (k_2)'s quantization level $l=2^m$. Obviously

$$\gamma_m \to \gamma(r_1,r_2) \qquad \text{as} \quad m\to\infty \quad (13)$$

Partition $[0,1]\times[0,1]$ into a $2^m\times 2^m$ uniform grid. Then the square in the i th column from left and the j th row from bottom (cf. Fig. 1)

$$\xi_m(i,j)=\left[\frac{i-1}{2^m},\ \frac{i}{2^m}\right]\times\left[\frac{j-1}{2^m},\ \frac{j}{2^m}\right] \quad (14)$$

Now we have the following recursive algorithm for computing $\gamma(r_1,r_2)$, which uses different quantization levels according to the bumpiness of $p_{z_1z_2}$ in different local regions:

$$\gamma(r_1,r_2)=F(\xi_0(1,1)) \quad (15)$$

where if $p_{z_1z_2}$ is uniform in $\xi_m(i,j)$, then

$$F(\xi_m(i,j))=f\left(4^m p_{k_1k_2}^m(i,j)\right) \quad (16)$$

and if $p_{z_1z_2}$ is not uniform in $\xi_m(i,j)$, then

$$F(\xi_m(i,j))=\frac{1}{4}\sum_{h=0}^{1}\sum_{w=0}^{1} F(\xi_{m+1}(2i-h,2j-w)) \quad (17)$$

Easily verified that (9) can be written in another form

$$\gamma(r_1,r_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{r_1}(u)p_{r_2}(v) f\left(\frac{p_{r_1r_2}(u,v)}{p_{r_1}(u)p_{r_2}(v)}\right)dudv \quad (18)$$

From (18), clearly the *mutual information* (MI) of r_1 and r_2

$$I(r_1,r_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p_{r_1r_2}(u,v)\log\frac{p_{r_1r_2}(u,v)}{p_{r_1}(u)p_{r_2}(v)}dudv \quad (19)$$

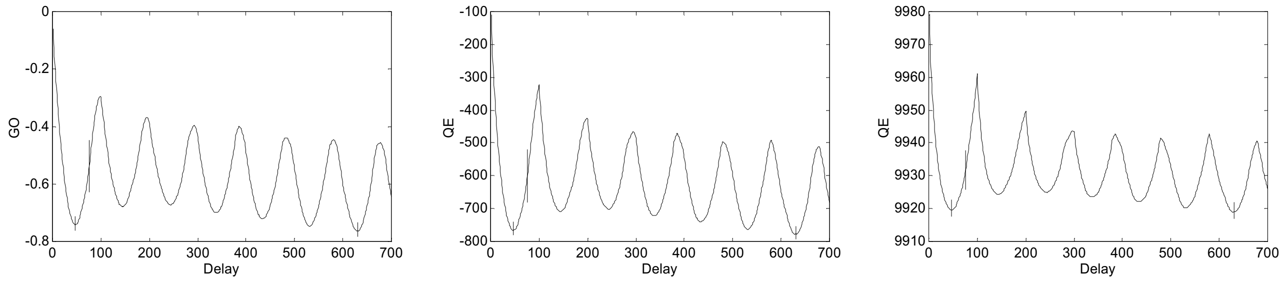
is the $\gamma(r_1,r_2)$ when $f(u)=u\log u$, so the independence measure γ may be called *generalized mutual information* (GMI).

In the above recursive algorithm for computing GMI, we need to judge if $p_{z_1z_2}$ is uniform in $\xi_m(i,j)$ or not. This can be done using the χ^2 test proposed in [4]. However, the 20% confidence level in [4] is chosen arbitrarily and lacks evidence. Experiments show that the uniformity test with 20% confidence level is too stringent and causes the estimate of GMI to be too large. Here we propose a simple method for choosing confidence level. We mix two independent variables s_1 and s_2 with identical distributions by a 2×2 rotation matrix with rotation angle θ to get two variables r_1 and r_2 whose MI is then a function of θ, i.e. $I(\theta)=I(r_1,r_2)$. According to the χ^2 distribution table, we choose five confidence levels 20%, 10%, 5%, 2%, and 1% to estimate $I(\theta)$. If s_1 and s_2 are with some standard distributions, then standard values of $I(\theta)$ are easy to obtain. We use the two cases where s_1 and s_2 are both uniform variables and both Laplacian variables respectively and spot that 5% confidence level produces minimum estimation error in both cases. Therefore, 5% is a better choice.

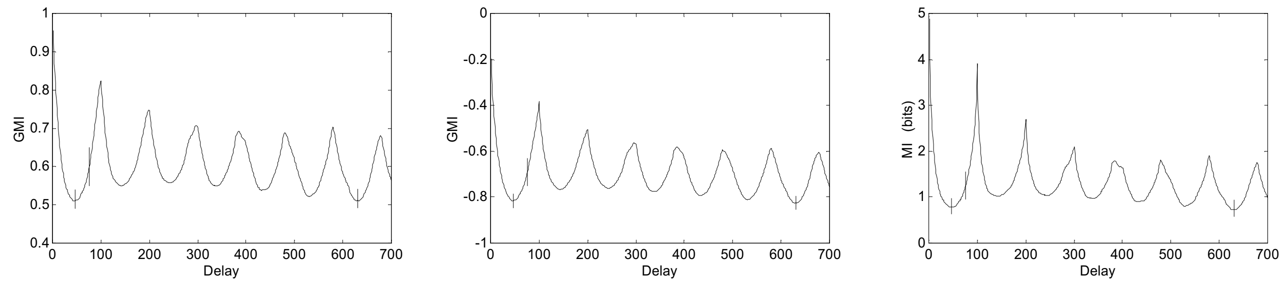
5. Experiments and Conclusion

This example is adapted from [4]. We apply GO, QE, and GMI to the Rossler system [2]. Fig. 2 is the original $(x(t),y(t))$ plot. The observed sequence $e(t)$ is obtained by adding noise to $x(t)$. Fig. 3 is the estimates of GO, QE, and GMI of the reconstructed variables versus delay where several different convex functions $f(\cdot)$'s are used for QE and GMI. All these plots look similar and the positions of minima coincide. Fig. 4 (a), (b), and (c) are the delay portraits using the first (leftmost) minimum, a non-minimum, and the seventh minimum of the curves in Fig. 3, respectively. Compared with Fig. 2, clearly Fig. 4 (a) is best. Experiments with the Lorenz system [1] yield similar results.

Fraser and Swinney used the first minimum of MI for choosing delay according to Shaw's suggestion. They proposed a recursive algorithm for computing MI and got good results with it. Their paper [4] has been highly cited till now by researchers in various disciplines since its publication, seeming that using MI to analyze nonlinear series has become a standard approach. However, our results show that MI is not the only choice. At least, the proposed measures are all good criteria for choosing delay. The algorithm for computing MI in [4] is a special case of our algorithm for computing GMI when the convex function $f(u)=u\log u$. We believe that more freedom as

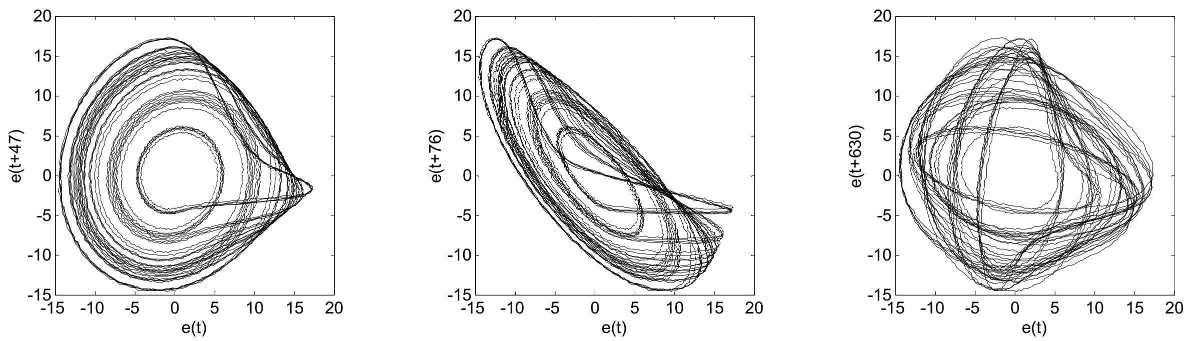


(a) GO (b) QE with $f(u) = -\sqrt[4]{u}$ (c) QE with $f(u) = 1/\left(1+\sqrt{u}\right)$



(d) GMI with $f(u) = \exp(-u)$ (e) GMI with $f(u) = -\sqrt{u}$ (f) GMI with $f(u) = u \log_2 u$ (i.e. MI)

Fig. 3. Estimates of GO, QE, and GMI of the reconstructed variables for Rossler attractor versus time delay τ. All the plots are calculated using 65536 sample points. $l = 100$ quantization levels are used for GO and QE. The delay is measured in sampling number. The three short vertical stems in each plot located at $\tau = 47$, $\tau = 76$, and $\tau = 630$ indicate the positions of the first minimum, a non-minimum, and the seventh minimum, respectively. Only the last plot is the classical MI.



(a) $\tau = 47$ (b) $\tau = 76$ (c) $\tau = 630$

Fig. 4. Delay reconstruction portraits of Rossler attractor using time delays corresponding to (a) the first minimum, (b) a non-minimum, and (c) the seventh minimum of the curves in Fig. 3, respectively.

well as possibility for improvement are obtained since there are infinitely many other choices than MI. Besides, GO and QE honor other applications such as blind source separation (independent component analysis) due to their high computational efficiency [5]. Concluding, the new independence measures enrich the tools for nonlinear signal analysis and enjoy good prospects.

References

[1] E. N. Lorenz, "Deterministic non-periodic flow," J. Atmos. Sci., Vol. 20, pp. 130-135, 1963.

[2] O. E. Rossler, "An equation for continuous chaos," Physics Letters A, Vol. 57, No. 5, pp. 397-398, 1976.

[3] F. Takens, "Detecting strange attractors in turbulence," in Warwick 1980 Lecture Notes in Math., Vol. 898, pp. 366-381, 1981.

[4] A. M. Fraser and H. L. Swinney, "Independent coordinates for strange attractors from mutual information," Physical Review A, Vol. 33, No. 2, pp. 1134-1140, 1986.

[5] Yang Chen, "On theory and methods for blind information extraction," Ph.D. dissertation, Southeast University, Mar. 2001 (in Chinese).