Evaluation of mutual information estimators on nonlinear dynamic systems

Aggeliki Papana and Dimitris Kugiumtzis

General Department, Faculty of Technology, Aristotle University of Thessaloniki.

Abstract

Mutual information is a nonlinear measure used in time series analysis in order to measure the global correlations (linear and non-linear) between their terms with time lag τ. The aim of this study is to evaluate some of the most commonly used mutual information estimators, i.e. estimators based on histograms (with fixed or adaptive bin size), k-nearest neighbors and kernels. We assess the estimators by Monte-Carlo simulations on time series from nonlinear dynamical systems of varying complexity. As the true mutual information is generally unknown, we investigate the consistency of the estimators (convergence to a stable value with the increase of time series length), the rate of consistency and the degree of deviation among the estimators.

Definition of mutual information

The mutual information of a time series $\{X_t\}$, t=1,…,n, is:

$$\mathbf{I}(\tau) = I(x_i, x_{i-\tau}) = \Sigma p_{x_i,x_{i-\tau}}(x_i x_{i-\tau}) \log \frac{p_{x_i,x_{i-\tau}}(x_i x_{i-\tau})}{p_{x_i}(x_i)\, p_{x_{i-\tau}}(x_{i-\tau})}$$

where the sum is referred to the partition of the two-dimensional plane (X_i, $X_{i-\tau}$) and $p_{x_i}, p_{x_{i-\tau}}, p_{x_i x_{i-\tau}}$ are the marginal and joint probability distribution defined for each region of the partition.

Mutual information estimators

The distribution of mutual information is generally not known as joint and marginal probability density functions are unknown. For the estimation $I(\tau)$ of $\mathbf{I}(\tau)$, the theoretical probabilities are estimated in different ways. In this work the most common-used estimators are evaluated. These estimators depend on a parameter that has to be chosen appropriately. The choice of the optimal parameter is also examined.

1. Histogram-based estimators

• The first one and most naïve estimator partitions the range of values into a finite number b of discrete bins of equal length (***equidistant partitioning***). The density of each bin is estimated by the corresponding relative frequencies of occurrence of the samples within a bin.

• The second estimator partitions the range of values into ***equiprobable bins***, so that each bin has the same occupancy.

In each case the partitioning is made with the same way in each variable.

• The third histogram-based estimator uses an ***adaptive partitioning*** of the two-dimensional plane [Darbellay & Vajda (1999)]. Mutual information is estimated by calculating relative frequencies on appropriate partitions which achieve conditional independence on the rectangles. The advantage of this estimator is that is data-dependent and parameter free.

2. k-nearest neighbours

This method considers the probability distributions for the distance between the point at which the density is to be estimated and its k-th nearest neighbour [Kraskov et. al (2004)]. The free parameter is the number of neighbours k.

3. Kernel estimators

The kernel density estimators construct a smooth estimate of the unknown probability density by centering kernel functions at the data samples; kernels are used to obtain the weighted distances [Moon (1995)]. In this work we use the Gaussian kernel. The free parameter is the bandwidth h_1 for one-dimensional data (equivalent to bin width) and h_2 for two-dimensional data.

Evaluation of mutual information estimators

The evaluation of the estimators is assessed by Monte-Carlo simulations in the non-linear systems: ***Henon, Ikeda*** map and ***Mackey Glass*** delay differential system with delay ***Delta 17, 30, 100***. The factors considered are the time series length: ***n = 256, 512, 1024, 2048, 4096, 8192***, and the noise level: ***additive Gaussian noise of 20, 40*** and ***80%***.

$I(\tau)$ is computed using all methods on 1000 realizations from the above systems up to that lag τ that $I(\tau)$ converges to a non-negative constant value. For each method, the corresponding free parameter ranges as follows.

1. histogram-based estimators: ***b = 2, 4, 8, 16, 32, 64***.

2. k-nearest neighbour estimator: **k = 2, 4, 8, 16, 32, 64**.

3. kernel estimator:

A/A	h_1	h_2	Reference
1	$h_1=(4/3n)^{1/5}$	$h_2=(1/n)^{1/6}$	Silverman (1986)
2	$h_1=(4/3n)^{1/5}$	$h_2=(4/5n)^{1/6}$	
3	$h_1=a1.06n^{-1/5}, a=\begin{cases}1.8-r(1), n<200\\ 1.5, n\ge 200\end{cases}$	$h_2=an^{-1/6}$	Harrold et al. (2001)
4	$h_1=(8\sqrt{\pi}R/3\gamma_2^2 n)^{1/5}\min(s, IQ/1.349)$	$h_2=h_1$	Wand & Jones (1995)
5	$h_1=(8\sqrt{\pi}R/3\gamma_2^2 n)^{1/5}\min(s, IQ/1.349)$	$h_2=\sqrt{2}h_1$	
6	L-stage direct plug in	$h_2=h_1$	Wand & Jones (1995)
7	L-stage direct plug in	$h_2=\sqrt{2}h_1$	
8	Solve-the-equation plug in	$h_2=h_1$	Sheather & Jones (1991)
9	Solve-the-equation plug in	$h_2=\sqrt{2}h_1$	

r(1) : autocorrelation for lag 1.

The true mutual information $\mathbf{I}(\tau)$ is generally not known for non-linear chaotic systems. In order to evaluate the mutual information estimators, we examine their consistency and their dependence in the corresponding parameters for all systems and time series lengths. Therefore we compute $I(\tau)$ for a realization up to a length n=**10^6** or n=**10^7** for all systems. If the estimator is consistent then it will converge with n.

Simulations results for each estimator

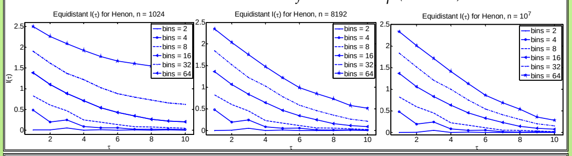
In the plots, $I(\tau)$ is the mean value of $I(\tau)$ from 1000 simulations.

Equidistant estimator

• $I(\tau)$ increases with b (for a fixed n).

• $I(\tau)$ depends on b, also for very large n.

We demonstrate these results for Henon map (no noise).

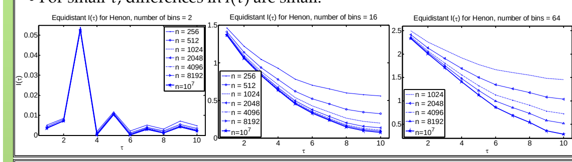


• $I(\tau)$ decreases with n (for fixed b).

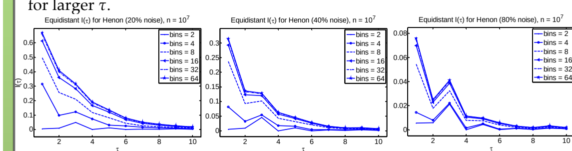
• For small b, $I(\tau)$ is rather stable for all n (but gives poor estimation).

• As τ increases, $I(\tau)$ varies with b and n.

• For small τ, differences in $I(\tau)$ are small.

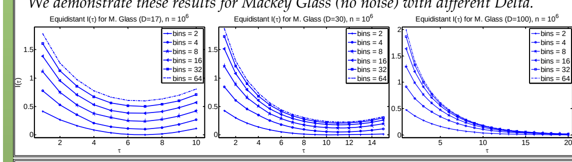


• As noise level increases, $I(\tau)$ values are smaller and converge to zero level for larger τ.

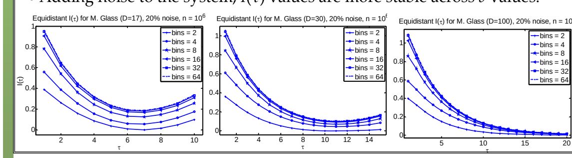


• As system complexity increases, $I(\tau)$ converges to zero level for larger τ.

We demonstrate these results for Mackey Glass (no noise) with different Delta.



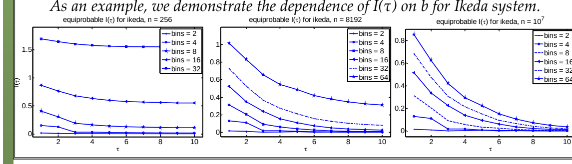
• Adding noise to the system, $I(\tau)$ values are more stable across b values.



Equiprobable estimator

• Equiprobable $I(\tau)$ estimator has exactly the same properties as the equidistant one.

As an example, we demonstrate the dependence of $I(\tau)$ on b for Ikeda system.



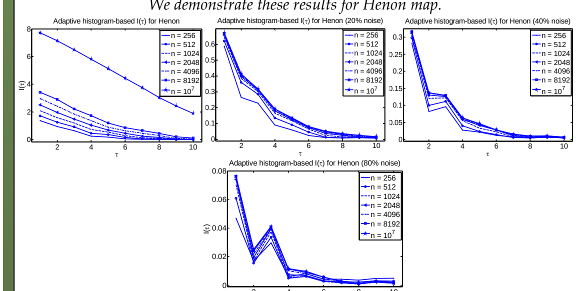
Adaptive histogram-based estimator

• $I(\tau)$ increases with n, opposite to fixed-bin methods (see over).

• The estimator is consistent only when adding noise.

• As noise level increases, $I(\tau)$ decreases and stabilizes for all n (fixed τ).

We demonstrate these results for Henon map.

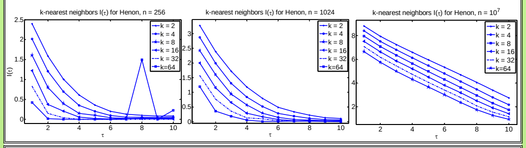


k–nearest neighbours mutual information

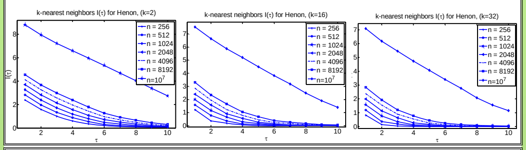
• $I(\tau)$ decreases with k (for fixed n) and depends on k, also for very large n.

• k= 64 is too large value for k (poor estimation).

We demonstrate these results for Henon map (no noise).

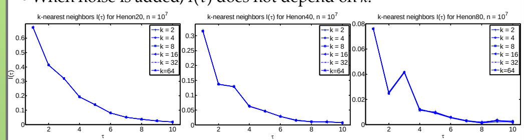


• $I(\tau)$ increases with n (for fixed k).



• $I(\tau)$ estimates decrease with noise level.

• When noise is added, $I(\tau)$ does not depend on k.

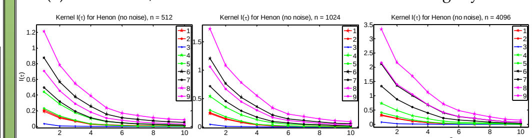


Kernel estimator of mutual information

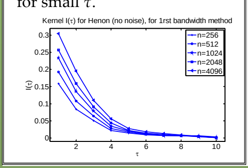
Kernel estimator has the highest computational cost and therefore the evaluation is not yet completed. From the simulations made so far we concluded to the following:

• $I(\tau)$ increases with n and is smaller for larger values of h_2.

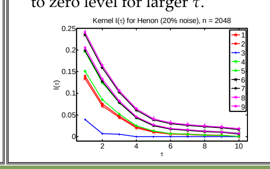
• $I(\tau)$ from the 1st, 2nd and 5th bandwidth methods differ slightly.



•$I(\tau)$ varies for different n (fixed bandwidth method) for small τ.



• As noise level increases, $I(\tau)$ values are smaller and converge to zero level for larger τ.

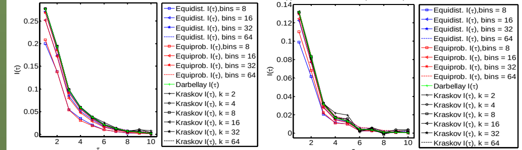


Conclusions

Mutual information estimators are not consistent for non-linear noise-free systems and the choice of parameters is crucial for all estimators. However we cannot find an optimal parameter choice as there is no consistency.

With added noise, the choice of the parameters is not that crucial as there is convergence of the estimated $I(\tau)$ values. k-nearest neighbor estimates of $I(\tau)$ varies less with the free parameter (k) compared to the other estimators.

$I(\tau)$ from all estimators for Ikeda system with 20% & 40% noise levels.



References

Darbellay G., Vajda I. (1999). *IEEE Transactions on Information Theory* 45 (4).

Kraskov A., Stogbauer H., Grassberger P. (2004). Phys. Rev. E 69.

Moon Y, Rajagopalan B, Lall U. (1995). Physical Review E 52, 2318–2321.

Silverman (1986) B. W. Chapman and Hall, London.

Wand M.P., Jones M.C. (1995). Chapman and Hall, London.

Sheather S. J., Jones M.C. (1991). J. of the Royal Statistical Society, Series B 53, 683–690.

Harrold, T.H. , Sharma, A., Sheather, S. (2001). Stochastic Environmental Research and Risk Assessment, 15(4), 310-324.

This research project (PENED) is co-financed by E.U.-European Social Fund (75%) and the Greek Ministry of Development-GSRT (25%).