

ΑΞΙΟΛΟΓΗΣΗ ΕΚΤΙΜΗΤΩΝ ΑΜΟΙΒΑΙΑΣ ΠΛΗΡΟΦΟΡΙΑΣ ΑΠΟ ΙΣΤΟΓΡΑΜΜΑ ΣΕ ΧΡΟΝΟΣΕΙΡΕΣ

Παπάνα Αγγελική, Κουγιουμτζής Δημήτρης

Γενικό Τμήμα Πολυτεχνικής Σχολής Α.Π.Θ.

agrapana@gen.auth.gr, dkugiu@gen.auth.gr

ΠΕΡΙΛΗΨΗ

Η αμοιβαία πληροφορία είναι ένα μέτρο γραμμικής και μη-γραμμικής συσχέτισης μεταξύ δύο μεταβλητών που κάνει χρήση της περιθώριας και κοινής κατανομής τους και γενικεύεται σε χρονοσειρές όπως η συνάρτηση αυτοσυσχέτισης. Στην εργασία αυτή μελετούνται μέθοδοι εκτίμησης της αμοιβαίας πληροφορίας που χρησιμοποιούν το ιστόγραμμα για την εκτίμηση των κατανομών ορίζοντας ισομήκη ή ισοπίθανα διαστήματα. Γίνεται αξιολόγηση γνωστών κριτηρίων για τον ορισμό του πλήθους διαστημάτων με Monte Carlo προσομοιώσεις σε συστήματα λευκού θορύβου και γραμμικά στοχαστικά συστήματα. Η απόδοση των κριτηρίων μελετάται ως προς το μήκος της χρονοσειράς και την κατανομή του θορύβου στο σύστημα. Από τις προσομοιώσεις φαίνεται ότι ο ισοπίθανος εκτιμητής είναι καλύτερος από τον ισομήκη, ενώ τα βέλτιστα κριτήρια είναι αυτά που εξαρτώνται μόνο από το μήκος της χρονοσειράς στις γραμμικές διαδικασίες.

1. ΕΙΣΑΓΩΓΗ

Η αμοιβαία πληροφορία δυο τυχαίων συνεχών μεταβλητών X, Y ορίζεται ως:

$$I(X, Y) = \iint_{x, y} f_{X,Y}(x, y) \log \frac{f_{X,Y}(x, y)}{f_X(x)f_Y(y)} dx dy \quad (1)$$

όπου $f_X(x), f_Y(y), f_{X,Y}(x, y)$ είναι οι περιθώριες συναρτήσεις πυκνότητας η κοινή συνάρτηση πυκνότητας των X, Y [Shannon (1948), Scott (1992)]. Για μια χρονοσειρά $\{X_i\}, i=1, \dots, n$, η αμοιβαία πληροφορία δίνεται ως συνάρτηση της χρονικής υστέρησης τ , $I(\tau) = I(X_i, X_{i-\tau})$, όπου θέτουμε $X = X_i, Y = X_{i-\tau}$ και μετράει τις συνολικές συσχετίσεις μεταξύ των όρων της χρονοσειράς με χρονική υστέρηση τ [Kantz & Schreiber (1997), Ch. 9].

Για την εκτίμησή του $I(\tau)$ συχνά χρησιμοποιείται η προσέγγιση των ιστογραμμάτων και η γενική μορφή του εκτιμητή είναι:

$$I(\tau) = \sum_i \sum_{i-\tau} p_{X_i, X_{i-\tau}}(x_i, x_{i-\tau}) \log \frac{p_{X_i, X_{i-\tau}}(x_i, x_{i-\tau})}{p_{X_i}(x_i)p_{X_{i-\tau}}(x_{i-\tau})}, \quad (2)$$

όπου το άθροισμα στη (2) αναφέρεται στη διαμέριση του πεδίου τιμών των $(X_i, X_{i-\tau})$ και οι συναρτήσεις μάζας πιθανότητας $p_{X_i, X_{i-\tau}}, p_{X_i}, p_{X_{i-\tau}}$ (κοινή και περιθώριες) ορίζονται για κάθε περιοχή της διαμέρισης και εκτιμώνται από τις αντίστοιχες σχετικές συχνότητες. Η διαμέριση μπορεί να είναι ίσου μήκους ή ίσης πιθανότητας, αλλά έχουν χρησιμοποιηθεί και τεχνικές προσαρμογής της διαμέρισης σε κελιά διαφορετικού μεγέθους, όπως για τον εκτιμητή των Darbellay & Vajda (1999) όπου τα κελιά σχηματίζονται με βάση κριτήριο δεσμευμένης ανεξαρτησίας. Άλλοι εκτιμητές βασίζονται στους πυρήνες (kernels) [Silverman (1986)], σε B-splines πολυώνυμα [Daub et al. (2002)] ή σε γειτονιές k-κοντινότερων σημείων [Kraskov et al. (2004)]. Όλοι οι παραπάνω εκτιμητές της αμοιβαίας πληροφορίας δεν ακολουθούν γενικά γνωστή κατανομή.

2. ΕΚΤΙΜΗΤΕΣ ΑΜΟΙΒΑΙΑΣ ΠΛΗΡΟΦΟΡΙΑΣ

Θα μελετήσουμε εκτιμητές της αμοιβαίας πληροφορίας $I(\tau)$ που χρησιμοποιούν διαμέριση του πεδίου τιμών της κάθε μεταβλητής. Η διαμέριση γίνεται σε b διαστήματα που είτε έχουν το ίδιο μήκος (ισομήκη διαμέριση) ή περιέχουν το ίδιο πλήθος στοιχείων n/b (ισοπίθανη διαμέριση με πιθανότητα b/n σε κάθε διάστημα). Η διαμέριση γίνεται με τον ίδιο τρόπο για κάθε μια από τις δυο μεταβλητές και η διδιάστατη διαμέριση αποτελείται από b^2 κελιά. Η ελεύθερη παράμετρος στην εκτίμηση $I(\tau)$ είναι το πλήθος b των διαστημάτων της διαμέρισης. Στην εργασία μελετούμε 10 γνωστά κριτήρια για την επιλογή του b , που δίνονται στον Πίνακα 1.

Πίνακας 1. Κριτήρια επιλογής πλήθους κελιών b .

Κριτήριο	Πλήθος διαστημάτων διαμέρισης	Αναφορές
1	$1 + \log_2 n$	Sturge (1926)
2	$1.87(n-1)^{0.4}$	Bendat & Piersol (1966)
3	$1 + \log_2 n + \log_2 \gamma$	Doane (1976)
4	$\sqrt{n}$	Mosteller & Tukey (1977)
5	$R\sqrt{n}/3.49s$	Scott (1979)
6	$R/2(IQ)n^{-1/3}$	Freedman & Diaconis (1981)
7	$\sqrt[3]{2n}$	Terrell & Scott (1985)
8	Ελάχιστο της συνάρτησης στοχαστικής πολυπλοκότητας $F(b)$	Rissanen (1992)
9	Επικρατούσα τιμή του λογαρίθμου της εκ των υστέρων συνάρτ. πιθανότ. του πλήθους κελιών του ιστογράμματος	Knuth (2006)
10	$\sqrt{n/5}$	Cohran (1954)

Στους ορισμούς των κριτηρίων n είναι το μήκος χρονοσειράς, R το εύρος, s η τυπική απόκλιση, IQ το ενδοτεταρτομοριακό εύρος, και γ η κανονικοποιημένη λοξότητα:

$$\gamma = \sqrt{(n+1)(n+3)} \sum_{i=1}^n (x_i - \bar{x})^3 / \sqrt{6(n-2) \sum_{i=1}^n (x_i - \bar{x})^2} \quad [\text{Doane (1976)}]. \quad \text{Η συνάρτηση}$$

στοχαστικής πολυπλοκότητας με $n_1, \dots, n_b$ το πλήθος των στοιχείων κάθε διαστήματος της διαμέρισης ορίζεται ως: $F(b) = n \log_2(R/b) + \log_2 \frac{n!}{n_1! n_2! \dots n_b!} + \log_2 \frac{(n+b-1)!}{n!(b-1)!}$. Τέλος,

η συνάρτηση της εκ των υστέρων πιθανότητας στο κριτήριο 9 υπολογίζεται χρησιμοποιώντας τη θεωρία πιθανοτήτων Bayes και εξάγεται ένας άμεσος αλγόριθμος που υπολογίζει την εκ των υστέρων πιθανότητα (posterior probability) του πλήθους των κελιών του ιστογράμματος για ένα σύνολο δεδομένων.

3. ΑΞΙΟΛΟΓΗΣΗ ΕΚΤΙΜΗΤΩΝ ΚΑΙ ΚΡΙΤΗΡΙΩΝ

Η αξιολόγηση των εκτιμητών και των κριτηρίων γίνεται με Monte-Carlo προσομοιώσεις σε κάποια γνωστά συστήματα χρονοσειρών. Θέλουμε να συγκρίνουμε τα 10 κριτήρια επιλογής του $I(\tau)$ σε εκτιμητές με ισομήκη και ισοπίθανη διαμέριση. Οι παράγοντες που λαμβάνουμε υπόψη είναι το μήκος της χρονοσειράς και η κατανομή του θορύβου εισόδου στα 14 συνολικά διαφορετικά συστήματα (δες Πίνακα 2), τα οποία επιλέχθηκαν όπως στην εργασία των Harrold et al. (2001).

Πίνακας 2. Συστήματα προσομοιώσεων, όπου ϕ είναι ο συντελεστής του αυτοπαλινδρομούμενου μοντέλου τάξης 1, $AR(1)$, θ είναι ο συντελεστής του μοντέλου κινούμενου μέσου τάξης 1, $MA(1)$, και τα ϕ και θ χρησιμοποιούνται επίσης στο κοινό μοντέλο $ARMA(1,1)$.

Συστήματα	Τύποι και παράμετροι συστημάτων	
Λευκός θόρυβος	Κανονικός	Μέση τιμή $\mu=0$, Διασπορά $s^2=1$
	Γάμμα	$\mu=0$, $s^2=1$, $\gamma=0.5$
Γραμμικά συστήματα με κανονικό θόρυβο	$AR(1)$	$\phi = 0.5, 0.9, -0.5, -0.9$
	$ARMA(1,1)$	$\phi = 0.9$, $\theta = 0.6$ και $\phi = 0.7$, $\theta = 0.3$
Γραμμικά συστήματα με Γάμμα θόρυβο	$AR(1)$	$\phi = 0.5, 0.9, -0.5, -0.9$
	$ARMA(1,1)$	$\phi = 0.7$, $\theta = 0.3$ και $\phi = 0.3$, $\theta = 0.1$

Στα συστήματα αυτά υπολογίζεται μόνο η $I(1)$, καθώς όλα είναι έως και τάξης 1. Η $I(1)$ υπολογίζεται για 1000 πραγματοποιήσεις από τα παραπάνω συστήματα, με ισοπίθανη και ισομήκη διαμέριση και για b από 2 έως κάποια μέγιστη τιμή που μεταβάλλεται με το μήκος της χρονοσειράς n , όπως δίνεται στον Πίνακα 3.

Πίνακας 3. Μέγιστο πλήθος διαστημάτων διαμέρισης για 9 διαφορετικά μήκη n .

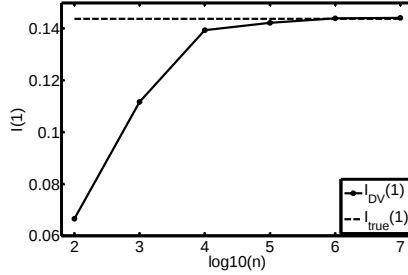
Μήκος n	32	64	128, 256, 512	1024, 2048, 4096, 8192
Μέγιστο b	8	16	32	64

Η πραγματική (θεωρητική) τιμή της αμοιβαίας πληροφορίας $I(\tau)$ είναι γνωστή μόνο για κάποιες διαδικασίες. Ειδικότερα για κανονικές διαδικασίες (π.χ. γραμμικά συστήματα με κανονικό θόρυβο) με συνάρτηση αυτοσυσχέτισης $\rho(\tau)$ [Darbellay & Vajda (1999)]:

$$I(\tau) \approx -0.5 \log(1 - \rho^2(\tau)). \quad (3)$$

Για μη κανονικές διαδικασίες γενικά η πραγματική τιμή της δεν είναι γνωστή. Για τα γραμμικά συστήματα προσομοιώσεων με Γάμμα θόρυβο θεωρήσαμε τη συνέπεια του εκτιμητή $I_{DI}(\tau)$ των Darbellay & Vajda (1999) και εκτιμήσαμε την πραγματική $\mathbf{I}(\tau)$ από την τιμή $I_{DI}(\tau)$ για $n=10^7$ (για AR(1) με $\varphi=0.5$ φαίνεται η συνέπεια αυτού του εκτιμητή στην Εικόνα 1). Ο τρόπος αυτός προσέγγισης της $\mathbf{I}(\tau)$ έχει χρησιμοποιηθεί και σε άλλες εργασίες [Harrold et al. (2001), Celluci et al. (2005)]. Συμβολίζω με I_{true} το $\mathbf{I}(\tau)$ όταν αυτό είναι γνωστό ή αντίστοιχα το $I_{DI}(\tau)$, όταν το $\mathbf{I}(\tau)$ δεν είναι γνωστό.

Εικόνα 1. Εκτιμητής $I_{DI}(1)$ για αυξανόμενο n μιας πραγματοποίησης AR(1) με $\varphi=0.5$, όπου $\mathbf{I}(1)=0.144$ (δίνεται από την οριζόντια γραμμή).



Στη συνέχεια ορίζουμε ένα συγκεντρωτικό δείκτη για κάθε τύπο διαμέρισης και κριτήριο επιλογής του b ως προς όλες τις περιπτώσεις (πλήθος συνολικών περιπτώσεων: (πλήθος συστημάτων)*(πλήθος n)=4*9=126). Για κάθε περίπτωση l , υπολογίζουμε τη μέση τιμή του εκτιμητή από τις Monte Carlo πραγματοποιήσεις $\bar{I}_c$, όπου $c=1,..10$ αντιστοιχεί στο κριτήριο επιλογής του b . Δηλώνουμε την ισομήκη (ισοπίθανη) διαμέριση με $\bar{I}_c^d$ ($\bar{I}_c^p$). Θεωρώντας το μέσο από όλα τα κριτήρια για κάθε περίπτωση l : $\bar{\bar{I}}(l) = \frac{1}{10} \sum_{c=1}^{10} \bar{I}_c(l)$, ορίζουμε δύο συγκεντρωτικούς δείκτες (scores) για κάθε κριτήριο c από ένα πλήθος L περιπτώσεων ($L \leq 126$):

$$E_1 = \frac{1}{L} \sum_{l=1}^L \frac{(\bar{I}_c(l) - I_{true}(l))^2}{(\bar{\bar{I}}(l) - I_{true}(l))^2}, \quad E_2 = \frac{\sum_{l=1}^L (\bar{I}_c(l) - I_{true}(l))^2}{\sum_{l=1}^L (\bar{\bar{I}}(l) - I_{true}(l))^2}. \quad (4)$$

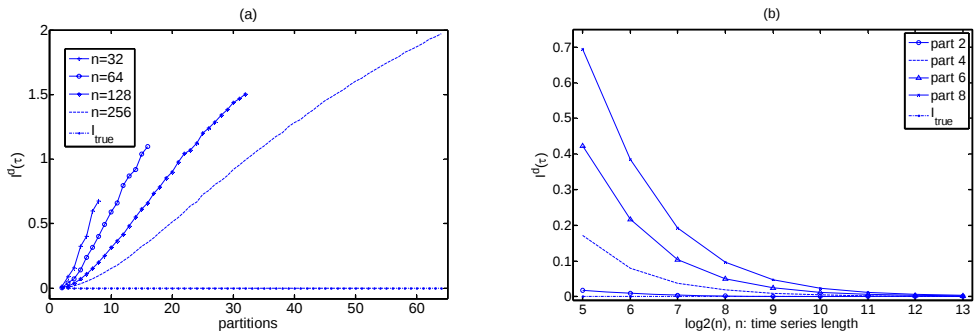
Ο E_2 έχει παρονομαστή το άθροισμα των τετραγώνων των I_{true} , αντί των μέσων αποκλίσεων από το I_{true} που όρισε ο Hamilton (1964) και που είχε χρησιμοποιηθεί ξανά για σύγκριση εκτιμητών αμοιβαίας πληροφορίας [Celluci et al. (2005)]. Επειδή σε ανεξάρτητες χρονοσειρές ($\mathbf{I}(\tau)=0$) μηδενίζεται ο παρονομαστής, θεωρήσαμε τις μέσες αποκλίσεις στους ορισμούς των E_1 και E_2 .

4. ΑΠΟΤΕΛΕΣΜΑΤΑ MONTE CARLO ΠΡΟΣΟΜΟΙΩΣΕΩΝ

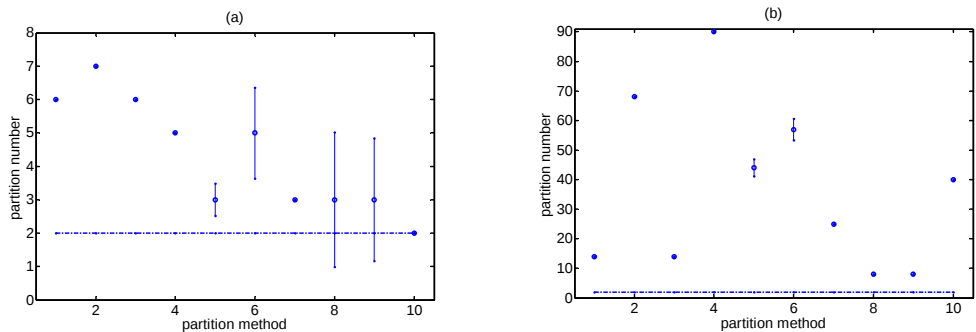
Αρχικά παρατηρούμε ότι η εκτίμηση της αμοιβαίας πληροφορίας για κανονικό λευκό θόρυβο αυξάνεται με το b (Εικόνα 2(a)), άρα το βέλτιστο πλήθος κελιών είναι 2, όπου προσεγγίζει περισσότερο το $\mathbf{I}(\tau)=0$. Επίσης παρατηρούμε ότι όσο μεγαλώνει το

μήκος της χρονοσειράς, τόσο η εκτίμηση συγκλίνει στην πραγματική τιμή (Εικόνα 2(b)), ακόμα και για μεγάλο πλήθος κελιών. Αναφορικά με τα κριτήρια επιλογής, το βέλτιστο b στην περίπτωση του λευκού κανονικού θορύβου είναι αυτό που δίνει το μικρότερο αριθμό κελιών. Παρατηρούμε ότι το n επηρεάζει σημαντικά την επιλογή του κριτηρίου (Εικόνα 3).

Εικόνα 2. Μέση αμοιβαία πληροφορία $I^d(1)$ από 1000 προσομοιώσεις λευκού κανονικού θορύβου, **(a)** ως προς b , για n όπως δίνεται στο ένθετο, **(b)** ως προς n , για b όπως δίνεται στο ένθετο για ισομήκη διαμέριση (αντίστοιχα αποτελέσματα προκύπτουν για ισοπίθανη διαμέριση). Η οριζόντια γραμμή δηλώνει το $I_{\text{true}}(1)=0$.



Εικόνα 3. Μέσο πλήθος διαστημάτων ισομήκης διαμέρισης και τυπική απόκλιση (με κατακόρυφες ράβδους) από τις προσομοιώσεις για κάθε κριτήριο και για **(a)** $n = 32$, **(b)** $n = 8192$. Η οριζόντια γραμμή είναι για το βέλτιστο $b=2$.



Η κατάταξη των κριτηρίων επιλογής του b για τους δυο εκτιμητές με βάση την απόκλιση τους από το $I_{\text{true}}(1)=0$ δίνεται στον Πίνακα 4 για κάθε μήκος χρονοσειρών λευκού κανονικού θορύβου (τα αποτελέσματα των εκτιμητών είναι όμοια).

Πίνακας 4. Κατάταξη των βέλτιστων κριτηρίων επιλογής b των δυο εκτιμητών (ισομήκη και ισοπίθανο) για κάθε n , για χρονοσειρές λευκού θορύβου.

Μήκος	Κατάταξη κριτηρίων									
32	10	7	5	9	8	6	4	1	3	2
64	10	9	5	7	8	6	1	3	4	2
128	10	9	7	8	5	1	3	6	4	2

256	7	10	9	8	1	3	5	6	4	2
512, 1024	9	8	1	3	7	10	5	6	2	4
2048, 4096, 8192	8	9	1	3	7	10	5	6	2	4

Στα συγκεντρωτικά αποτελέσματα για το σύστημα λευκού κανονικού θορύβου για

όλα τα μήκη χρονοσειρών (Πίνακας 5), παρατηρούμε ότι με το E_1 η κατάταξη είναι ίδια για ισομήκη και ισοπίθανη διαμέριση με εξίσου καλή εκτίμηση, ενώ με το E_2 έχουμε πάλι παρόμοια κατάταξη αλλά με τον εκτιμητή $\bar{I}_c^p$ καλύτερο.

Πίνακας 5. Κατάταξη των 5 βέλτιστων κριτηρίων επιλογής του b για κανονικό λευκό θόρυβο με τους δείκτες E_1 , E_2 (ως προς όλα τα n) με τις αντίστοιχες τιμές των E_1 , E_2 για ισομήκεις και ισοπίθανους εκτιμητές.

A/A	Αποτελέσματα από E_1				Αποτελέσματα από E_2			
	Κριτήριο	$\bar{I}_c^d$	Κριτήριο	$\bar{I}_c^p$	Κριτήριο	$\bar{I}_c^d$	Κριτήριο	$\bar{I}_c^p$
1	9	0.01	9	0.01	10	0.11	10	0.07
2	8	0.01	8	0.02	7	0.29	7	0.17
3	7	0.03	7	0.03	9	0.25	9	0.24
4	10	0.04	10	0.04	8	0.40	5	0.36
5	1	0.07	1	0.06	5	0.43	8	0.42

Στον Πίνακα 6 δίνονται τα συγκεντρωτικά αποτελέσματα για όλα τα συστήματα (λευκού θορύβου και γραμμικά) και μήκη χρονοσειρών. Παρατηρούμε ότι με το E_1 η κατάταξη είναι διαφορετική για ισομήκη και ισοπίθανη διαμέριση και με αρκετά μικρότερες τις τιμές του E_1 για τον ισομήκη εκτιμητή, ενώ το E_2 δίνει παρόμοια κατάταξη κριτηρίων αλλά με μικρότερες τιμές του E_2 για τον ισοπίθανο εκτιμητή.

Πίνακας 6. Ομοίως με πίνακα 5 αλλά για όλα τα γραμμικά συστήματα.

A/A	Αποτελέσματα από E_1				Αποτελέσματα από E_2			
	Κριτήριο	$\bar{I}_c^d$	Κριτήριο	$\bar{I}_c^p$	Κριτήριο	$\bar{I}_c^d$	Κριτήριο	$\bar{I}_c^p$
1	10	10.37	5	5.26	7	0.76	7	0.45
2	5	18.87	6	66.40	8	0.82	8	0.50
3	7	34.94	1	68.53	9	0.93	9	0.53
4	9	52.58	3	68.53	5	0.97	5	0.77
5	8	59.73	7	74.21	3	1.29	10	0.80

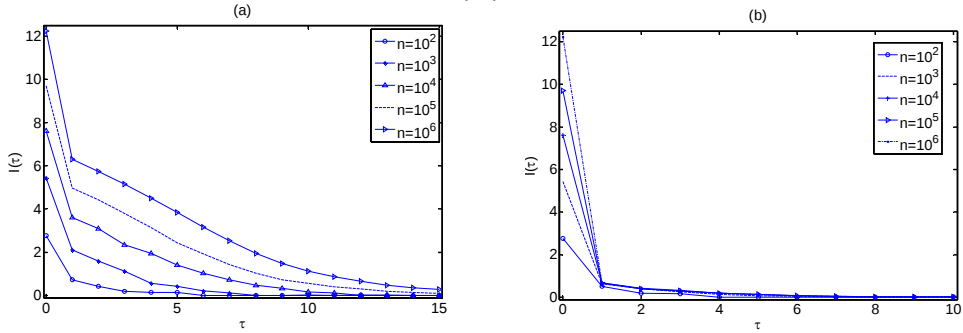
5. ΣΥΝΕΠΕΙΑ ΕΚΤΙΜΗΤΩΝ $I(\tau)$ ΣΕ ΜΗ-ΓΡΑΜΜΙΚΑ ΣΥΣΤΗΜΑΤΑ

Στα μη-γραμμικά συστήματα που δεν είναι γνωστή η πραγματική αμοιβαία πληροφορία παρατηρήσαμε ότι ο εκτιμητής $I_{DI}(\tau)$ δεν είναι συνεπής, καθώς φαίνεται να αυξάνει με το n , και συγκλίνει σε κάποια τιμή μόνο αν το σύστημα έχει θόρυβο (δες Εικόνα 4 για τη χαοτική απεικόνιση Henon $x_{t+1}=y_t+1-1.4x_t^2, y_{t+1}=0.3x_t$ [Kantz & Schreiber (1997), Ch.3]). Για τους δυο εκτιμητές $I^d(\tau)$, $I^p(\tau)$ βλέπουμε ότι όσο μεγαλώνει το n συγκλίνουν για σταθερό μόνο b . Δεν είναι όμως πάλι συνεπείς, καθώς δεν υπάρχει σύγκλιση της εκτιμώμενης $I(\tau)$ με το n για διαφορετικά b (Εικόνα 5).

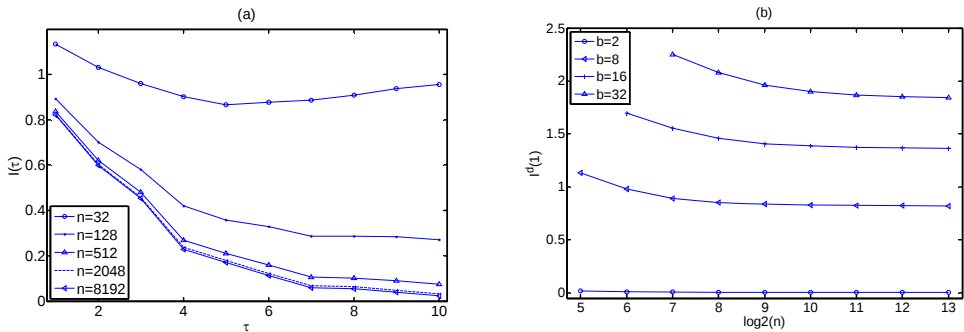
5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Από τις προσομοιώσεις καταλήξαμε στα παρακάτω συμπεράσματα. Πρώτον, ενώ η εκτίμηση της αμοιβαίας πληροφορίας αυξάνει με το b , τα «βέλτιστα» κριτήρια είναι αυτά που δίνουν μικρό b . Ο δείκτης E_2 φαίνεται ότι είναι πιο κατάλληλος από τον E_1 καθώς οι τιμές του είναι συγκρίσιμες στους δυο εκτιμητές με ισοπίθανη και ισομήκη διαμέριση και υπάρχει μια σταθερότητα ως προς την κατάταξη των κριτηρίων. Με βάση το E_2 καταλήγουμε ότι ο εκτιμητής με ισοπίθανη διαμέριση είναι καλύτερος από τον εκτιμητή με ισομήκη διαμέριση. Βέλτιστο κριτήριο από όλα τα γραμμικά συστήματα φαίνεται να είναι το κριτήριο $\sqrt[3]{2n}$, το οποίο εξαρτάται μόνο από το μήκος της χρονοσειράς. Επίσης οι εκτιμητές φαίνονται να είναι συνεπείς μόνο για τα γραμμικά συστήματα. Για τα μη-γραμμικά συστήματα δεν ήταν δυνατή η σύγκριση των εκτιμητών αφού δε μπορούμε να ορίσουμε την τιμή αναφοράς I_{true} .

Εικόνα 4. Εκτιμητής $I_{DV}(1)$ των Darbellay-Vajda ως προς τ , για μια πραγματοποίηση του συστήματος Henon, για διαφορετικά n όπως δίνονται στο ένθετο **(a)** χωρίς θόρυβο, **(b)** με 20% προσθετικό λευκό κανονικό θόρυβο.



Εικόνα 5. Μέση αμοιβαία πληροφορία $\bar{I}^d(\tau)$ για το σύστημα Henon (χωρίς θόρυβο) **(a)** ως προς τ για $b = 8$, για n (δες ένθετο), **(b)** ως προς n , για b (δες ένθετο).



ABSTRACT

Mutual information is a measure of linear and non-linear correlation between two variables that uses their joint and marginal distributions. It can be generalized to time series like autocorrelation function. In this work we study methods of estimation of mutual information that use the histogram for the determination of the probabilities of distributions by taking

either equidistant or equiprobable bins of the partition. We also assess some well-known criteria for the number of the bins of the partition by Monte-Carlo simulations on systems of white noise and stochastic linear systems. The performance of the criteria is studied with respect to the time series length and the distribution of noise in the systems. The simulation results suggest that equiprobable partitioning gives better estimates and the best criteria depend only on time series length for all linear systems.

Το έργο αυτό (ΠΕΝΕΔ) συγχρηματοδοτείται κατά 90% κοινά από την Ε.Ε.: Ε. Κοινωνικό Ταμείο (75%), Υπ. Ανάπτυξης, Γ.Γ.Ε.Τ. (25%) και Rikshospitalet, Νορβηγίας (10%), στο πλαίσιο του Μέτρου 8.3 του Ε.Π. Ανταγωνιστικότητα – Γ΄ Κ.Π.Σ..

ΑΝΑΦΟΡΕΣ

- Bendat S.J., Piersol A.G. (1966) Measurements and analysis of random data. John Wiley, N.Y.
- Celluci C.J., Albano A.M., Rapp P.E. (2005) Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms. *Ph. Rev. E* **71**, 066208.
- Cohran W.G. (1954) Some Methods for strengthening the common X² test, *Biometrics* **10**, 417.
- Darbellay G., Vajda I. (1999). Estimation of information by an adaptive partitioning of the observation space. *IEEE Transactions on Information Theory* **45**, 4.
- Daub C.O., Steuer R., Kurths J., Weise J., Selbig J. (2002) The mutual information: detecting and evaluating dependencies between variables. *Bioinformatics* **18**(2).
- Doane D.P. (1976). Aesthetic frequency classifications. *The Am. Statist.* **30**, 181-183.
- Freedman D., Diaconis P. (1981). On the histogram as a density estimator: L₂ theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **57**, 453-476.
- Hamilton J.D. (1964). Time Series Analysis. University Press, Princeton, NJ.
- Harrold T.I., Sharma A., Sheather S. (2001). Selection of a kernel bandwidth for measuring dependence in hydrologic time series using the mutual information criterion. *Stochastic Environmental Research and Risk Assessment* **15**, 310-314.
- Kantz H., Schreiber T. (1997). Nonlinear Time Series Analysis. Cambridge Univ. Press, Cambridge.
- Kraskov A., Stogbauer H., Grassberger P. (2004) Estimating mutual information. *Ph. Rev. E* **69**, 066138.
- Knuth K.H. (2006). Optimal Data-Based Binning for Histograms, Draft paper.
- Mosteller F., Tukey J.W. (1977). Data Analysis and Regression: A Second Course in Statistics. Addison-Wesley, Reading, MA.
- Rissanen J.(1992) Stochastic Complexity in Statistical Inquiry. World Scientific, Singapore.
- Scott D.W. (1979). On optimal and data-based histograms. *Biometrika* **66**, 605-610.
- Scott D.W. (1992). Multivariate density estimation. John Wiley, N.Y.
- Shannon C.E. (1948). A mathematical theory of communication. *The Bell System Technical Journal* **27**, 379-423. *ibid*, 623-656.
- Silverman B. (1986). Density Estimation for Statistics and Data Analysis, Chapman and Hall, London.
- Sturge H.A. (1926). The choice of a class interval. *J. of the Am. Stat. Ass.* **21**, 65-66.
- Terrell G.R., Scott D.W. (1985). Oversmooth nonparametric density estimates. *J. of the Am. Stat. Ass.* **80**, 209-214.