Ελληνικό Στατιστικό Ινστιτούτο
Πρακτικά 18ου Πανελληνίου Συνεδρίου Στατιστικής (2005) σελ.315-324

ΑΘΡΟΙΣΤΙΚΗ ΑΜΟΙΒΑΙΑ ΠΛΗΡΟΦΟΡΙΑ ΩΣ ΣΤΑΤΙΣΤΙΚΟ ΕΛΕΓΧΟΥ ΜΗ-ΓΡΑΜΜΙΚΟΤΗΤΑΣ ΧΡΟΝΟΣΕΙΡΩΝ

Παπάνα Αγγελική και Κουγιουμτζής Δημήτρης
Γενικό Τμήμα Πολυτεχνικής Σχολής Α.Π.Θ.
agpapana@gen.auth.gr, dkugiu@gen.auth.gr

ΠΕΡΙΛΗΨΗ

Στην ανάλυση χρονοσειρών θέλουμε να διερευνήσουμε την ύπαρξη συσχετίσεων μεταξύ των όρων της, δηλαδή εξετάζουμε αν η χρονοσειρά είναι ανεξάρτητη ή αν υπάρχουν γραμμικές ή μη-γραμμικές συσχετίσεις. Γι αυτόν το σκοπό εισάγουμε ένα νέο μέτρο μη γραμμικής συσχέτισης, την αθροιστική αμοιβαία πληροφορία $M(\tau_{\max})$, που ορίζεται ως το άθροισμα των τιμών της αμοιβαίας πληροφορίας $I(\tau)$ για ένα πλήθος υστερήσεων $\tau = 1,\ldots,\tau_{\max}$. Μπορεί να θεωρηθεί ως το αντίστοιχο μη-γραμμικό μέτρο του στατιστικού Portmanteau $Q(\tau_{\max})$, που ορίζεται ως το άθροισμα των αυτοσυσχετίσεων για ένα πλήθος υστερήσεων και χρησιμοποιείται στον έλεγχο ανεξαρτησίας χρονοσειρών. Το στατιστικό $M(\tau_{\max})$, μπορεί να χρησιμοποιηθεί για έλεγχο ανεξαρτησίας (ιδιαίτερα όταν η εναλλακτική υπόθεση αφορά μη-γραμμικές συσχετίσεις), αλλά κυρίως για έλεγχο μη-γραμμικότητας, δηλαδή για τη μηδενική υπόθεση H_0 ότι η χρονοσειρά προέρχεται από μια γραμμική στοχαστική διαδικασία. Προσομοιώσεις σε γνωστά μοντέλα ανέδειξαν την καταλληλότητα της $M(\tau_{\max})$ εξετάζοντας τη σημαντικότητα και την ισχύ του ελέγχου συγκριτικά με την $Q(\tau_{\max})$ για τον έλεγχο ανεξαρτησίας και με την $I(\tau)$ για τον έλεγχο μη-γραμμικότητας. Ειδικότερα για την περίπτωση μη- γραμμικών συσχετίσεων το στατιστικό $M(\tau_{\max})$ είναι καλύτερο για χρονοσειρές μικρού μήκους και για σχετικά μεγάλα ποσοστά θορύβου.

1. ΕΙΣΑΓΩΓΗ

Μια στάσιμη χρονοσειρά μπορεί να είναι μια σειρά ανεξάρτητων μεταβλητών με την ίδια κατανομή (iid), μπορεί να έχει μόνο γραμμικές συσχετίσεις, δηλαδή να παράγεται από μια γραμμική στοχαστική διαδικασία, ή να έχει και μη-γραμμικές συσχετίσεις, π.χ. στη μαθηματική έκφραση της στοχαστικής διαδικασίας να υπάρχουν όροι x_t^2 ή $x_t x_{t-1}$. Το πρώτο βήμα στην στατιστική ανάλυση μιας στάσιμης χρονοσειράς $\{x_t\}, t = 1,\ldots,n$, είναι ο έλεγχος ανεξαρτησίας, δηλαδή ο έλεγχος της μηδενικής υπόθεσης H_0 ότι η χρονοσειρά είναι iid. Ο πιο γνωστός έλεγχος

ανεξαρτησίας είναι ο έλεγχος Portmanteau που βασίζεται στη δειγματική συνάρτηση αυτοσυσχέτισης $r(\tau)$. Η $r(\tau)$ μετράει τις γραμμικές συσχετίσεις μεταξύ όρων της χρονοσειράς με χρονική υστέρηση τ :

$$r(\tau) = \frac{\frac{1}{n-\tau}\sum_{i=1+\tau}^{n}(x_i - \bar{x})(x_{i-\tau} - \bar{x})}{\frac{1}{n-1}\sum_{i=1+\tau}^{n}(x_i - \bar{x})^2}, \qquad (1)$$

όπου $\bar{x}$ η μέση τιμή της $\{x_t\}$. Αν η $\{x_t\}$ είναι iid, τότε $r(\tau) \sim \mathrm{N}(0,1/n)$ και το 95% των $r(\tau)$ θα πρέπει να ανήκουν στο διάστημα $\pm 1.96/\sqrt{n}$ [Box, Jenkins & Reinsel (1994), pg 188]. Χρησιμοποιώντας ως στατιστικό ελέγχου την $r(\tau)$, για κάποια υστέρηση τ, η H_0 απορρίπτεται σε στάθμη σημαντικότητας $\alpha = 0.05$ αν $|r(\tau)| < 1.96/\sqrt{n}$. Επειδή ο έλεγχος με το $r(\tau)$ γίνεται για κάθε τ ξεχωριστά, μπορεί να μην οδηγεί πάντα σε συνεπή αποτελέσματα, δηλάδη ανάλογα με την επιλογή του τ, να οδηγούμαστε άλλοτε σε απόρριψη και άλλοτε σε αποδοχή της H_0. Γι' αυτό στον έλεγχο Portmanteau χρησιμοποιείται το στατιστικό $Q(\tau_{\max})$ που ορίζεται από το άθροισμα των $r^2(\tau)$ για $\tau = 1,\ldots,\tau_{\max}$:

$$Q(\tau_{\max}) = n \cdot \sum_{\tau=1}^{\tau_{\max}} r^2(\tau), \qquad (2)$$

το οποίο κάτω από την H_0 ακολουθεί $\chi^2_{\tau_{\max}}$ κατανομή με $\tau_{\max}$ βαθμούς ελευθερίας και η απορριπτική περιοχή είναι $R = \{Q(\tau_{\max}) > X^2_{\tau_{\max},1-\alpha}\}$ [Box & Pierce (1970)].

Τα γραμμικά χαρακτηριστικά των χρονοσειρών, όπως η $r(\tau)$, χαρακτηρίζουν γραμμικές στοχαστικές διαδικασίες. Για τον έλεγχο μη-γραμμικότητας χρονοσειρών, δηλαδή όταν θέλουμε να ελέγξουμε την H_0 ότι η χρονοσειρά προέρχεται από μια γραμμική διαδικασία, πρέπει να χρησιμοποιήσουμε μη-γραμμικά στατιστικά μέτρα. Η Αμοιβαία Πληροφορία $I(\tau)$ μετράει την γραμμική και μη-γραμμική συσχέτιση μεταξύ των όρων της χρονοσειράς με χρονική υστέρηση τ και δίνεται:

$$I(\tau) = I(x_i, x_{i-\tau}) = \sum p_{X_i,X_{i-\tau}}(x_i, x_{i-\tau}) \log \frac{p_{X_i,X_{i-\tau}}(x_i, x_{i-\tau})}{p_{X_i}(x_i)\, p_{X_{i-\tau}}(x_{i-\tau})}, \qquad (3)$$

όπου το άθροισμα αναφέρεται σε διαμέριση του πεδίου τιμών των $(x_i, x_{i-\tau})$ και οι συναρτήσεις μάζας πιθανότητας (κοινή και περιθώριες) ορίζονται για κάθε περιοχή της διαμέρισης [Kantz & Schreiber (1997), Ch.9]. Η διαμέριση μπορεί να γίνει σε διαστήματα είτε ίσου μήκους είτε ίσης σχετικής συχνότητας [Moddemeijer (1999), Darbellay (1998)]. Για την εκτίμηση των πιθανοτήτων χρησιμοποιούμε τις σχετικές συχνότητες. Η $I(\tau)$ είναι μέτρο συσχέτισης με την ευρύτερη έννοια. Χρησιμοποιεί κατανομές και όχι μέσες τιμές όπως η $r(\tau)$ και υπάρχει κατάλληλος μετασχηματισμός της ώστε να ικανοποιεί τα αξιώματα Renyi [Granger & Lin

(1994)]. Η κατανομή που ακολουθεί γενικά η $I(\tau)$ δεν είναι γνωστή. Υπάρχουν κάποια αποτελέσματα για τις ασυμπτωτικές στατιστικές ιδιότητες της εκτίμησης της $I(\tau)$ [Pardo (1995), Hutter & Zaffalon (2004)].

Το νέο στατιστικό μέτρο που προτείνεται είναι η Αθροιστική Αμοιβαία Πληροφορία:

$$M(\tau_{\max}) = \sum_{\tau=1}^{\tau_{\max}} I(\tau)\,, \qquad (4)$$

η οποία ορίζεται ως συνάρτηση του $\tau_{\max}$, όπως και η $Q(\tau_{\max})$. Η μέγιστη τιμή που παίρνουμε για το $\tau_{\max}$ είναι η υστέρηση κατά την οποία το $I(\tau)$ αρχίζει να συγκλίνει προς κάποια μικρή τιμή και επομένως αυτό εξαρτάται κάθε φορά από την χρονοσειρά που εξετάζουμε. Επειδή γενικά δεν είναι γνωστή η κατανομή που ακολουθεί η $M(\tau_{\max})$ για τον έλεγχο χρησιμοποιείται η μέθοδος των υποκατάστατων (surrogate) δεδομένων, η οποία περιγράφεται παρακάτω.

2. ΣΤΑΤΙΣΤΙΚΟΣ ΕΛΕΓΧΟΣ ΜΕ ΥΠΟΚΑΤΑΣΤΑΤΑ ΔΕΔΟΜΕΝΑ

Η ιδέα είναι να δημιουργήσουμε χρονοσειρές (υποκατάστατα δεδομένα), που «μοιάζουν» με την υπό έλεγχο χρονοσειρά αλλά είναι συνεπείς με την H_0, δηλαδή ικανοποιούν τις συνθήκες που θέτει η H_0, αλλά όχι την ιδιότητα που θέλουμε να εξετάσουμε. Για τον έλεγχο ανεξαρτησίας τα υποκατάστατα δεδομένα δημιουργούνται με αναδιάταξη της αρχικής χρονοσειράς (random permutation) και έχουν την ίδια περιθώρια κατανομή με την αρχική χρονοσειρά, αλλά είναι iid χρονοσειρές. Για τον έλεγχο μη-γραμμικότητας τα υποκατάστατα πρέπει να έχουν τις ίδιες αυτοσυσχετίσεις και την ίδια περιθώρια κατανομή με την αρχική χρονοσειρά. Για την κατασκευή τους χρησιμοποείται ο αλγόριθμος Statistically Transformed Autoregressive Process (STAP) (για την περιγραφή του αλγορίθμου δες [Kugiumtzis (2002)]).

Το στατιστικό ελέγχου, ας το συμβολίσουμε q, πρέπει να μπορεί να διαχωρίσει τα υποκατάστατα δεδομένα από την αρχική χρονοσειρά όταν δεν είναι συνεπής με την H_0. Για τον έλεγχο ανεξαρτησίας χρησιμοποιούμε ως στατιστικό ελέγχου q το $Q(\tau_{\max})$ και το $M(\tau_{\max})$, ενώ για τον έλεγχο μη-γραμμικότητας χρησιμοποιούμε ως στατιστικό ελέγχου q το $I(\tau)$ και $M(\tau_{\max})$ (για υστερήσεις $\tau = 1,\cdots,10$). Για κάθε στατιστικό q υπολογίζουμε την τιμή του q_0 από την αρχική χρονοσειρά και από τα M υποκατάστατα, έστω $q_1,\cdots,q_M$ (χρησιμοποιούμε M=40). Η H_0 απορρίπτεται αν το q_0 δεν ανήκει στην κατανομή των $q_1,\cdots,q_M$ για κάποια στάθμη σημαντικότητας α. Θεωρώντας κανονική την κατανομή των $q_1,\cdots,q_M$, η απόρριψη δίνεται παραμετρικά από τη σημαντικότητα:

$$s = \frac{\left|q_0 - \bar{q}\right|}{\sigma_q}, \qquad (5)$$

όπου $\bar{q}$ είναι η μέση τιμή των $q_1, \cdots, q_M$ και σ_q η τυπική απόκλιση τους. Αν $s > 1.96$ τότε η H_0 απορρρίπτεται σε στάθμη σημαντικότητας $\alpha = 0.05$. Η σημαντικότητα και η ισχύς του ελέγχου εξετάστηκε με Monte Carlo προσομοιώσεις γνωστών συστημάτων, όπως δίνεται παρακάτω.

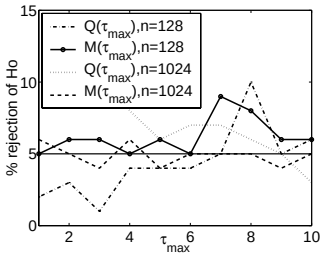
3. ΕΛΕΓΧΟΣ ΑΝΕΞΑΡΤΗΣΙΑΣ

Τα μοντέλα που χρησιμοποιήθηκαν για τις προσομοιώσεις είναι τα παρακάτω:

- λευκός θόρυβος που προέρχεται από λογαριθμική κανονική κατανομή,
- γραμμικό αυτοπαλινδρομούμενο $AR(1)$, με συντελεστή $\varphi = 0.25$,
- η απεικόνιση Henon [Kantz & Schreiber (1997), Ch.3].

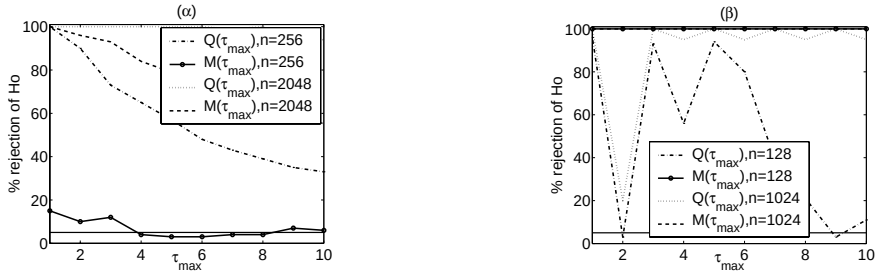
Οι χρονοσειρές λευκού θορύβου (idd) που κατασκευάζουμε χρησιμοποιούνται για την αξιολόγηση της σημαντικότητας του ελέγχου, που αναφέρεται και ως μέγεθος του ελέγχου. Το μέγεθος του ελέγχου είναι καλό αν η αναλογία απόρριψης της H_0 δεν υπερβαίνει τη στάθμη σημαντικότητας (εδώ $\alpha = 0.05$). Τα άλλα δύο μοντέλα, τα οποία δεν ικανοποιούν την H_0 (γραμμικό και μη-γραμμικό μοντέλο), χρησιμοποιήθηκαν για την αξιολόγηση της ισχύος του ελέγχου.

Για τις χρονοσειρές λευκού θορύβου βλέπουμε ότι τα ποσοστά απόρριψης της H_0 είναι χαμηλά και για τα δύο στατιστικά $Q(\tau_{\max})$ και $M(\tau_{\max})$ και άρα το πραγματικό μέγεθος του ελέγχου είναι καλό (δες Σχήμα 1).



Σχήμα 1 *Ποσοστά απόρριψης της H_0 με $Q(\tau_{\max})$, $M(\tau_{\max})$ για $\tau = 1, \cdots, 10$ από 100 προσομοιώσεις χρονοσειρών λευκού θορύβου (όπως δίνεται στο ένθετο).*

Οι προσομοιώσεις έδειξαν ότι για το γραμμικό μοντέλο $AR(1)$ ($\varphi = 0.25$) το $Q(\tau_{\max})$ έχει μεγαλύτερη ισχύ από τη $M(\tau_{\max})$, ενώ αντίθετα για την απεικόνιση Henon η $M(\tau_{\max})$ έχει πολύ μεγαλύτερη ισχύ από τη $Q(\tau_{\max})$ (δες Σχήμα 2).



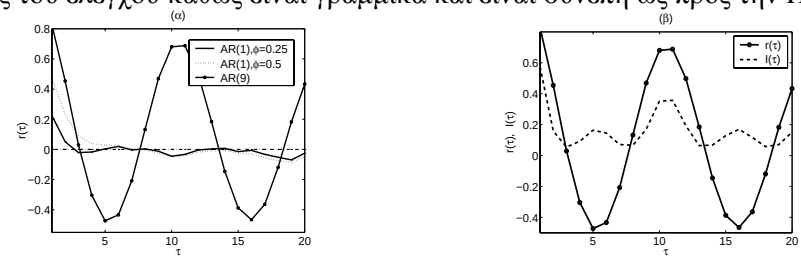
Σχήμα 2 *Όπως στο Σχημα 1 αλλά για μοντέλο* $AR(1)$, $\varphi = 0.25$ *στο (α) και την απεικόνιση Henon στο (β).*

Τα πρώτα αυτά αποτελέσματα δείχνουν τη χρησιμότητα της $M(\tau_{\max})$ σε προβλήματα, όπου τυχόν ασθενείς συσχετίσεις στη χρονοσειρά μπορεί να μην είναι γραμμικές.

4. ΕΛΕΓΧΟΣ ΜΗ-ΓΡΑΜΜΙΚΟΤΗΤΑΣ

Για τον έλεγχο μη γραμμικότητας πραγματοποιήθηκαν 100 προσομοιώσεις από γνωστά μοντέλα. Το μέγεθος και η ισχύς του ελέγχου εξετάστηκε για τα παρακάτω μήκη $n = 128,256,512,1024,2048$. Τα μοντέλα που χρησιμοποιήθηκαν είναι τα εξής:

- Δύο γραμμικά αυτοπαλλινδρομούμενα $AR(1)$ μοντέλα (linear autoregressive), με συντελεστές αυτοσυσχέτισης $\varphi = 0.25$, $\varphi = 0.5$ και ένα μοντέλο $AR(9)$. Στο Σχήμα 3α φαίνονται οι αυτοσυσχετίσεις $r(\tau)$ για μια πραγματοποίηση από καθένα από τα τρια αυτά μοντέλα. Τα μοντέλα αυτά χρησιμοποιήθηκαν για την αξιολόγηση του μεγέθους του ελέγχου καθώς είναι γραμμικά και είναι συνεπή ως προς την H_0.



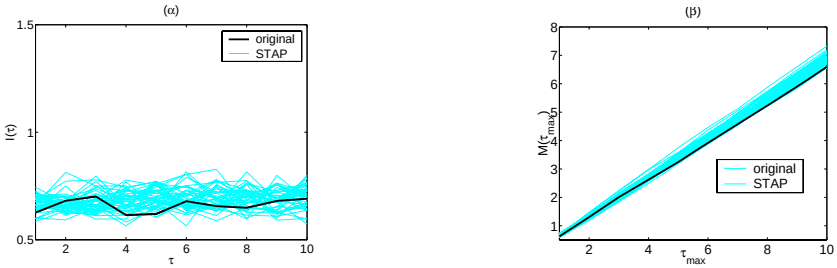
Σχήμα 3 *(α) Αυτοσυσχέτιση* $r(\tau)$ *των 3 γραμμικών μοντέλων όπως δίνεται στο ένθετο (β)* $r(\tau)$ *και* $I(\tau)$ *του μη-γραμμικού στοχαστικού μοντέλου* $NAR(3)$ *(*$\tau = 1,\cdots,20$*).*

Για τον έλεγχο της ισχύος του ελέγχου χρησιμοποιήθηκαν τα παρακάτω μη-γραμμικά μοντέλα:

• Ένα μη γραμμικό στοχαστικό μοντέλο $NAR(3)$ (Nonlinear Autoregressive) με εξίσωση: $X_t = 0.25X_{t-2} - 0.4e_{t-1}X_{t-1} + 0.2e_{t-2}X_{t-2} + 0.5e_{t-3}X_{t-3} + e_t$. Στο Σχήμα 3β φαίνονται τα $r(\tau)$ και $I(\tau)$ για μια πραγματοποίηση του μοντέλου.

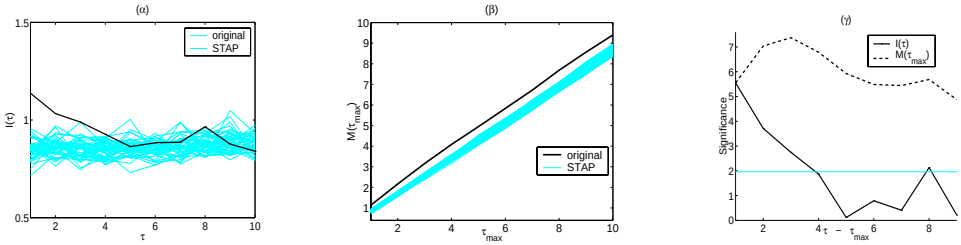
• η απεικόνιση Henon και Ikeda [Kantz & Schreiber (1997), Ch.6], που είναι διακριτά συστήματα 2 μεταβλητών με μικρή και μεγάλη πολυπλοκότητα αντίστοιχα και το σύστημα Lorenz [Kantz & Schreiber (1997), Ch.3] που προέρχεται από ροή (συνεχές σύστημα 3 διαφορικών εξισώσεων). Επειδή τα χαοτικά μοντέλα είναι ντετερμινιστικά, τα προσαρμόζουμε σε πραγματικές συνθήκες με την προσθήκη θορύβου σε ποσοστά: 4%, 8%, 16%, 32%, 64%.

Για τα τρια γραμμικά συστήματα, η αμοιβαία πληροφορία $I(\tau)$ και η αθροιστική αμοιβαία πληροφορία $M(\tau_{\max})$ της αρχικής χρονοσειράς βρίσκεται μέσα στην κατανομή των STAP υποκατάστατων χρονοσειρών για το $AR(1)$ για όλες τις υστερήσεις $\tau = 1,\cdots,10$ (δες Σχήμα 4α). Για μικρά μήκη χρονοσειρών τα ποσοστά απόρριψης της H_0 με τις $I(\tau)$ και $M(\tau_{\max})$ είναι κοντά στο 5% για $\tau = 1,\cdots,10$.

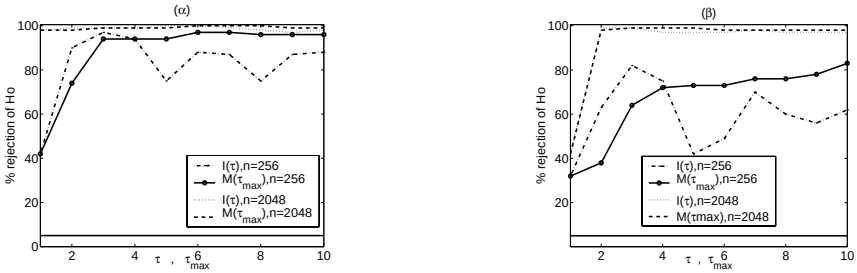


Σχήμα 4 *$I(\tau)$ στο (α) και $M(\tau_{\max})$ στο (β) για μια πραγματοποίηση του $AR(1)$ με $\varphi = 0.5$, $n = 128$ και για 40 STAP υποκατάστατα δεδομένα.*

Οι προσομοιώσεις στα μη-γραμμικά μοντέλα έδειξαν ότι η $M(\tau_{\max})$ δίνει γενικά μεγαλύτερη ισχύ ελέγχου από την $I(\tau)$, ειδικά όταν οι χρονοσειρές έχουν μικρό μήκος ή πολύ θόρυβο. Στο Σχήμα 5α παρατηρούμε ότι η διακριτική ικανότητα της $I(\tau)$ εξαρτάται από την υστέρηση τ, ενώ η $M(\tau_{\max})$ έχει διακριτική ικανότητα ανεξάρτητα της επιλογής του $\tau_{\max}$ (Σχήμα 5β). Η $M(\tau_{\max})$ δίνει σημαντικότητα $s > 5$ για όλες τις υστερήσεις, ένω η σημαντικότητα s για την $I(\tau)$ φθίνει με το τ και πέφτει κάτω από το όριο απόρριψης για $\alpha = 0.05$ για $\tau > 3$ (με εξαίρεση την οριακή απόρριψη για $\tau = 8$, δες Σχήμα 5γ).



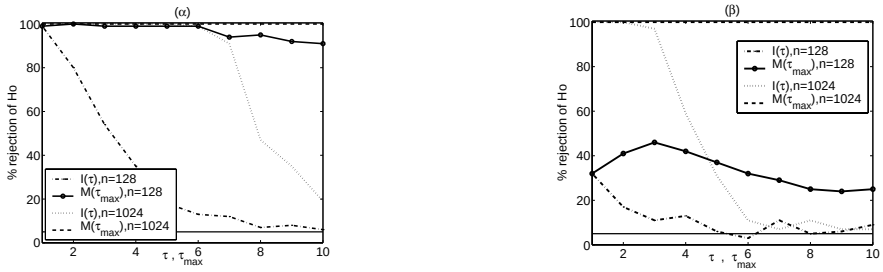
Σχήμα 5 *(α)* $I(\tau)$ *για μια πραγματοποίηση της απεικόνισης Henon με μήκος* $n = 128$ *και ποσοστό θορύβου 16% και για 40 STAP υποκατάστατες χρονοσειρές. (β)* $M(\tau_{\max})$ *για τις ίδιες χρονοσειρές. (γ) Σημαντικότητα* s *για τον έλεγχο μη-γραμμικότητας με υποκατάστατα δεδομένα με τις στατιστικές* $I(\tau)$ *και* $M(\tau_{\max})$.



Σχήμα 6 *Ποσοστά απόρριψης της* H_0 *για τον έλεγχο μη-γραμμικότητας με υποκατάστατα δεδομένα από τις 100 προσομοιώσεις του συστήματος Lorenz, για μήκη* $n = 256,2048$ *(α) χωρίς θόρυβο (β) με ποσοστό θορύβου 32%.*

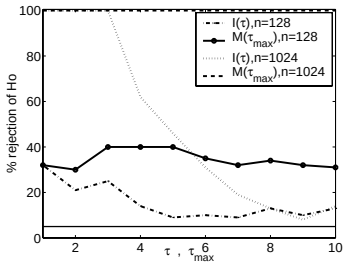
Γενικά διαπιστώνουμε ότι για την απεικόνιση Henon ο έλεγχος $M(\tau_{\max})$ έχει μεγαλύτερη ισχύ από τον έλεγχο $I(\tau)$ ειδικά όταν έχουμε μεγάλα ποσοστά θορύβου. Στα ίδια συμπεράσματα καταλήγουμε με τις προσομοιώσεις για το σύστημα Lorenz και Ikeda. Για το σύστημα Lorenz παρατηρούμε ότι η $M(\tau_{\max})$ έχει μεγαλύτερη ισχύ από την $I(\tau)$ για όλα τα μήκη που εξετάζουμε και για όλα τα ποσοστά θορύβου (δες Σχήμα 6). Τα ποσοστά απόρριψης της H_0 είναι σχεδόν 100% για όλες τις υστερήσεις για μεγάλο μήκος χρονοσειρών ακόμα και όταν έχουμε θόρυβο.

Οι διαφορές στην ισχύ του ελέγχου μεταξύ των $I(\tau)$, $M(\tau_{\max})$ είναι μεγαλύτερες για την απεικόνιση Ikeda. Ενώ για μεγάλα n, τα ποσοστά απόρριψης της H_0 είναι πολύ υψηλά και για τα δύο στατιστικά, για μικρά n τα ποσοστά απόρριψης είναι πολύ μικρότερα για την $I(\tau)$ (δες Σχήμα 7α). Για αρκετά μεγάλα ποσοστά θορύβου και μικρά μήκη χρονοσειρών βλέπουμε ότι η $M(\tau_{\max})$ απορρίπτει την H_0 σε ποσοστά 20%-40%, ενώ για την $I(\tau)$ το ποσοστό απόρριψης φθίνει με το τ και πέφτει στην στάθμη σημαντικότητας για $\tau > 4$. Για μεγαλύτερα n αυξάνονται τα ποσοστά απόρριψης και για τα δύο στατιστικά. Όπως φαίνεται στο Σχήμα 7β για $n = 1024$, με το $M(\tau_{\max})$ το ποσοστό απόρριψης είναι 100% ανεξάρτητα του $\tau_{\max}$.



Σχήμα 7 *Όπως στο Σχήμα 6 αλλά για την απεικόνιση Ikeda.*

Τα αποτελέσματα για το μη-γραμμικό μοντέλο $NAR(3)$ είναι συγκριτικά ίδια για τις $I(\tau)$ και $M(\tau_{\max})$ με αυτά για τα χαοτικά μοντέλα. Τα ποσοστά απόρριψης της H_0 είναι όμως πολύ μικρότερα, ειδικά για μικρό μήκος χρονοσειρών. Όπως φαίνεται στο Σχήμα 8, η $M(\tau_{\max})$ έχει πάλι καλύτερη διακριτική ικανότητα από την $I(\tau)$, δηλαδή δίνει μεγαλύτερη ισχύ ελέγχου από την $I(\tau)$.



Σχήμα 8 *Όπως το Σχήμα 6α αλλά για το σύστημα* $NAR(3)$.

5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Οι προσομοιώσεις με γνωστά μοντέλα έδειξαν σταθερή απόδοση του ελέγχου μη-γραμμικότητας με τη $M(\tau_{\max})$. Η διακριτική ικανότητα της $M(\tau_{\max})$ δε φαίνεται να επηρεάζεται από την επιλογή της υστέρησης σε αντίθεση με την $I(\tau)$. Παρ'όλα αυτά γίνεται προσπάθεια εύρεσης ενός αυτοποιημένου τρόπου επλογής του $\tau_{\max}$, ώστε να μην χρειάζεται σε κάθε έλεγχο να το ορίζουμε. Το πραγματικό μέγεθος του ελέγχου είναι καλό για την $I(\tau)$ και $M(\tau_{\max})$, ενώ για μεγάλο n το μέγεθος του ελέγχου είναι πολύ καλύτερο για την $M(\tau_{\max})$ (τα ποσοστά απόρριψης της Ho για $\alpha = 0.05$ μικρότερα του α). Όσον αφορά την ισχύ, η $M(\tau_{\max})$ έχει μεγαλύτερη ισχύ και για χρονοσειρές μικρού μήκους και για μεγάλα ποσοστά θορύβου. Τα αποτελέσματα αυτά αναδεικνύουν την καταλληλότητα του στατιστικού $M(\tau_{\max})$ για τον έλεγχο μη-γραμμικότητας.

ABSTRACT

This paper proposes a new non-linear measure the cumulative mutual information $M(\tau_{\max})$, defined as the sum of mutual information $I(\tau)$ for a number of delays τ. It is the non-linear analogue to Portmanteau test statistic $Q(\tau_{\max})$, which is the sum of autocorrelations $r(\tau)$ for a number of delays τ and is used for testing randomness of time series. $M(\tau_{\max})$ can be used both for the independence test (H_0: no correlations) and the non-linearity test (Ho: the data are generated by a linear process possibly undergoing a static transform). As the distribution of $M(\tau_{\max})$ is not known we use the surrogate data test (generation of data which are consistent with H_0) in order to investigate non-linearity in the data. The size and power of the test is examined with Monte Carlo realizations of well-known systems that are consistent to H_0 and others that are inconsistent to H_0, respectively. The models used for simulations are linear autoregressive models of certain orders, non-linear chaotic models (Henon map, Ikeda and Lorenz system), nonlinear autoregressive models (NAR) and white noise (only for the independence test). $M(\tau_{\max})$ is compared with $Q(\tau_{\max})$ for the independence test and to $I(\tau)$ for the non-linearity test. The performance of the test is examined for different lengths of time series and different noise levels (for the chaotic systems). The results from the simulations bring out the appropriateness of $M(\tau_{\max})$ as statistic for both the non-linearity and independence test.

ΑΝΑΦΟΡΕΣ

Box G. and D. Piere (1970). Distribution of residual autocorrelations in autoregressive Integrated moving average time series models. *J. Am. Statist. Assoc. 65*, 1509-1526

Box G., G. Jenkins & G. Reinsel (1994). *Time series Analysis: Forecasting and Control,* Prentice Hall

Darbellay G. (1998). An Adaptive Histogram Estimator for the Mutual Information, *UTIA Research Report*, no. 1889, Acad. Sc., Prague

Granger C. and J. Lin (1994). Using the Mutual Information Coefficient to Identify Lags in Nonlinear Models, *Journal of Time Series Analysis*, 15, 4, 371 -384.

Hutter M. and M. Zaffalon (2005). Distribution of mutual information from complete and incomplete data. *Computational Stat. & Data Analysis 48*, 633-657.

Kantz H. and T. Schreiber (1997). *Nonlinear Time Series Analysis*. Cambridge University Press, Cambridge.

Kugiumtzis D.(2002). Statistically transformed autoregressive process and surrogate data test for nonlinearity. *Physical Review* E66, 025201(R).

Moddemeijer R. (1999). A Statistic to estimate the Variance of the Histogram-Based Mutual Information Estimator on Dependent Pairs of Observations, *Signal Processing*, 75, 51-63.

Pardo J.A. (1995). Some Applications of the Useful Mutual Information. *Applied Mathematics and Computation* 72:33-50

Schreiber T. and A. Schmitz (2000). Surrogate time series, *Phys.* D142(3&4), 346-382.

Theiler J., S. Eubank, A. Longtin and B. Galdrikian (1992). Testing for nonlinearity in time series: The method of surrogate data, *Physica* D58, 77-94