4. ΕΠΕΞΕΡΓΑΣΊΑ ΣΗΜΆΤΩΝ

Συσχέτιση

Εντροπία

Αμοιβαία Πληροφορία

Electrical & Computer Engineering
University of Cyprus

Στατιστική

ΟΡΟΣ	ΟΡΙΣΜΟΣ
Μέση τιμή (expectation)	$E[(X)] = \sum_i x_i p_i = \frac{1}{N}\sum_i x_i$
n^{th} Ροπή (moment), μ: μέση τιμή	$E[(X-\mu)^n]$
2η ροπή (2^{nd} moment): διασπορά (variance)	$E[(X-\mu)^2]$ $\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu_x)^2$
Στατιστική ανεξαρτησία (statistical independence)	$p(x_1,\ldots,x_n) = p(x_1)\ldots p(x_n)$

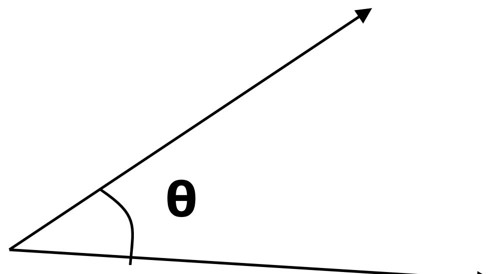
Συσχέτιση (Correlation)

- Δείχνει τη δύναμη και κατεύθυνση της γραμμικής σχέσης μεταξύ δύο τυχαίων μεταβλητών

$$C(\mathrm{X},\mathrm{Y}) = \frac{Cov(\mathrm{X},\mathrm{Y})}{\sigma_{\mathrm{X}}\sigma_{\mathrm{Y}}} = \frac{\mathrm{E}\left[(\mathrm{X}-\mu_{\mathrm{X}})(\mathrm{Y}-\mu_{\mathrm{Y}})\right]}{\sigma_{\mathrm{X}}\sigma_{\mathrm{Y}}}$$

$$c_{xy} = \frac{1}{n-1}\frac{\sum (x_i-\mu_x)(y_i-\mu_y)}{\sigma_x\sigma_y}$$

- Συμμετρική, [-1,1]
- Συσχέτιση=0 αν και μόνο αν X και Y στατιστικά ανεξάρτητες. ΠΡΟΣΟΧΗ! Το αντίθετο δεν ισχύει!
- C(X,X)=1



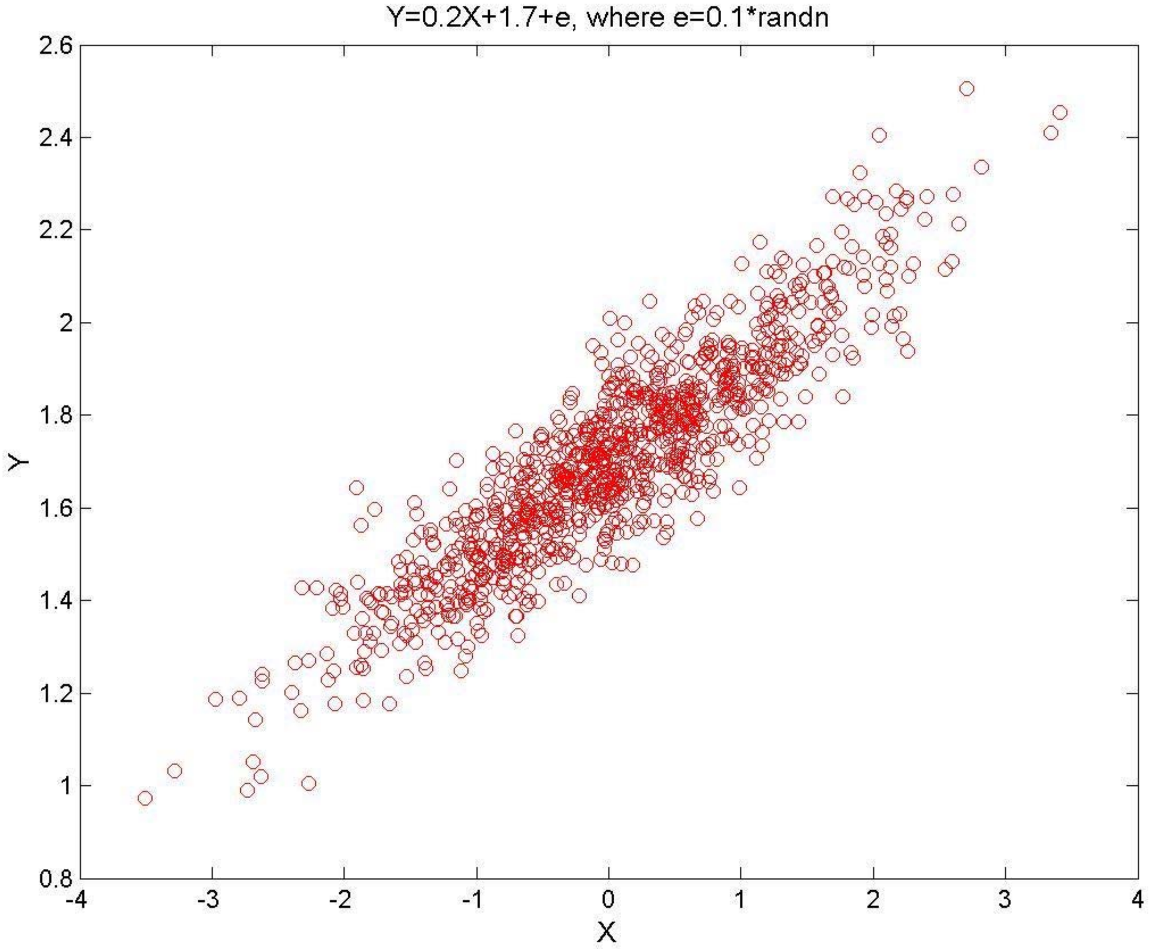
$$C_{XY} = \cos\theta = \frac{X \cdot Y}{\|X\|\|Y\|} = \frac{\sum_i (x_i - y_i)}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}}$$

- Γεωμετρικά:ισχύει μόνο για κεντραρισμένες μεταβλητές
- Π.χ.

  X=(0.5,0.2,1.8,3.2)
  Y=(2.1,1.2,6,10.2)

  Cxy=?

Παράδειγμα



Π'ΙΝΑΚΑΣ ΣΥΣΧ'ΕΤΙΣΗΣ

- Για n τυχαίες μεταβλητές, $X_1,\ldots,X_n$:

$$C(X) = \begin{bmatrix} C(X_1,X_1) & C(X_1,X_2) & \ldots & C(X_1,X_n) \\ C(X_2,X_1) & C(X_2,X_2) & \ldots & C(X_2,X_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(X_n,X_1) & C(X_n,X_2) & \cdots & C(X_n,X_n) \end{bmatrix} \quad \text{και} \quad C(X_2,X_1) = C(X_1,X_2)$$

π.χ. x=[0.5 1.2 4.0 6.0
2.3 7.0 1.0 -2.0
0.3 -0.5 1.2 1.1]

→ C(x)=[1.0000 -0.7942 0.7931
-0.7942 1.0000 -0.9166
0.7931 -0.9166 1.0000]

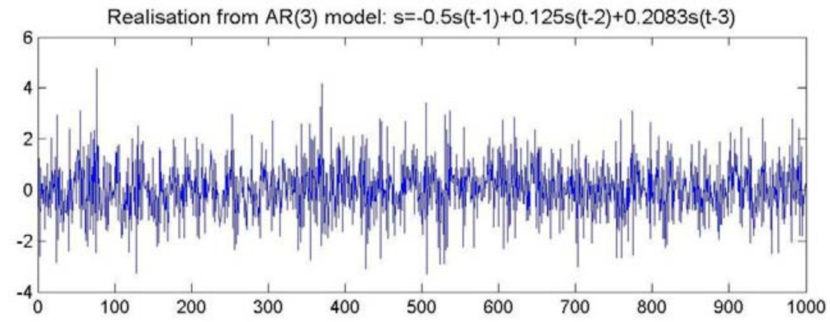
Αυτοσυσχέτιση (autocorrelation)

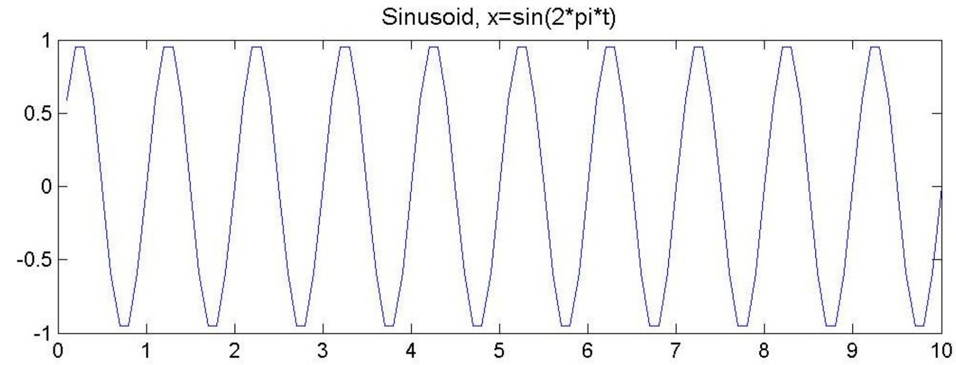
- Συσχέτιση μίας μεταβλητής με τον εαυτό της μεταφερόμενο στο χρόνο

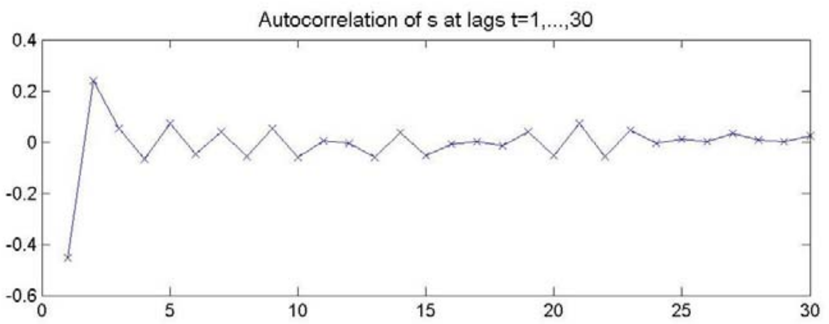
$$AC_X = C\left(X(t), X(t+\tau)\right)$$

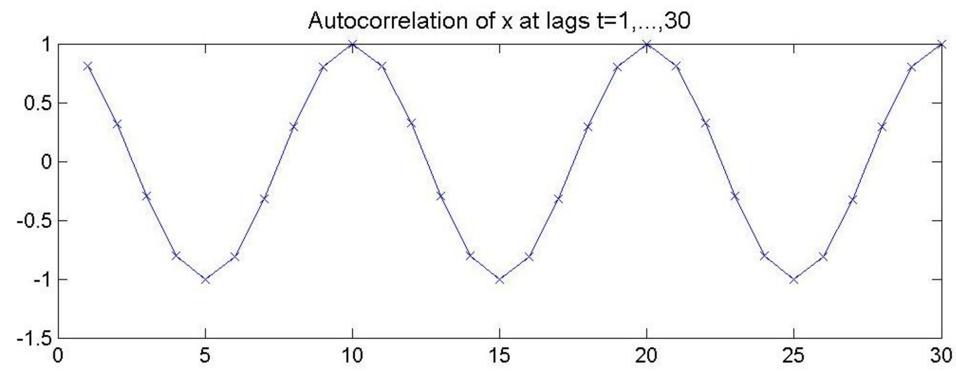
- Δηλ. η σχέση ενός σήματος με τον εαυτό του σε συγκεκριμένη υστέρηση (lag)

ΠΑΡΑΔΕΊΓΜΑΤΑ









Εντροπία (entropy)

- Μέτρο μέσης αβεβαιότητας μιας τυχαίας μεταβλητής **X**, ή η μέση πληροφορία ανά σύμβολο πληροφορίας, ή ο μέσος αριθμός bits που απαιτούνται για να περιγράψουν την τυχαία μεταβλητή

$$H(X) = -\sum_{i=1}^{M} p(x_i) \log p(x_i)$$

όπου **X**: τυχαία μεταβλητή με τιμές $\{x_1,\ldots,x_M\}$, και $p(x_i)$: συνάρτηση κατανομής πιθανότητας (probability density function)

Αμοιβαία Πληροφορία (Mutual Information)

- Μείωση της αβεβαιότητας μιας τυχαίας μεταβλητής X_1 λόγω ύπαρξης μιας άλλης τυχαίας μεταβλητής X_2.
- Μέτρο γραμμικής και μη-γραμμικής συσχέτισης μεταξύ τυχαίων μεταβλητών X_1 και X_2.

$$MI(\mathbf{X}) = \sum_{i}^{N} H(X_i) - H(X_1,\ldots,X_N)$$

$$H(X_1,\ldots,X_n) = -\sum_{X} p(X_1,\ldots,X_n)\log p(X_1,\ldots,X_n)$$

όπου **X**:{X_1,..., X_n}, H(X1,..., Xn): κοινή εντροπία (joint entropy)

- Kullback-Leibler Divergence

$$KL\big(p(\mathbf{X}) \,\|\, p(X_1)\ldots p(X_N)\big) = \sum_{\mathbf{X}} p(X_1,\ldots,X_N)\log\frac{p(X_1,\ldots,X_N))}{p(X_1)\ldots p(X_N)} = MI(\mathbf{X})$$

- K-L>0, K-L=0 αν και μόνο αν οι μεταβλητές είναι στατιστικά ανεξάρτητες
- Ιδιότητες MI:
  - Αμεταβλητότητα: $MI(X_1,\ldots,X_N) = MI(U_1,\ldots,U_N)$ όταν $U_i = f_i(X_i)$

    και f_i είναι μετατροπή 1-προς-1
  - MI(X,Y)≥0, και MI(X,Y)=0 αν και μόνο αν X και Y στατιστικά ανεξάρτητες

Υπολογισμός της Συνάρτησης Κατανομής Πιθανότητας (ΣΚΠ)

- Για τον υπολογισμό της ΑΠ πρέπει να γνωρίζουμε τη ΣΚΠ:
  - Ιστόγραμμα (histogram) – χωρίζουμε τα δεδομένα σε N ισομερείς κατανομές (bins) πλάτους d_x. Τότε, ΣΚΠ:

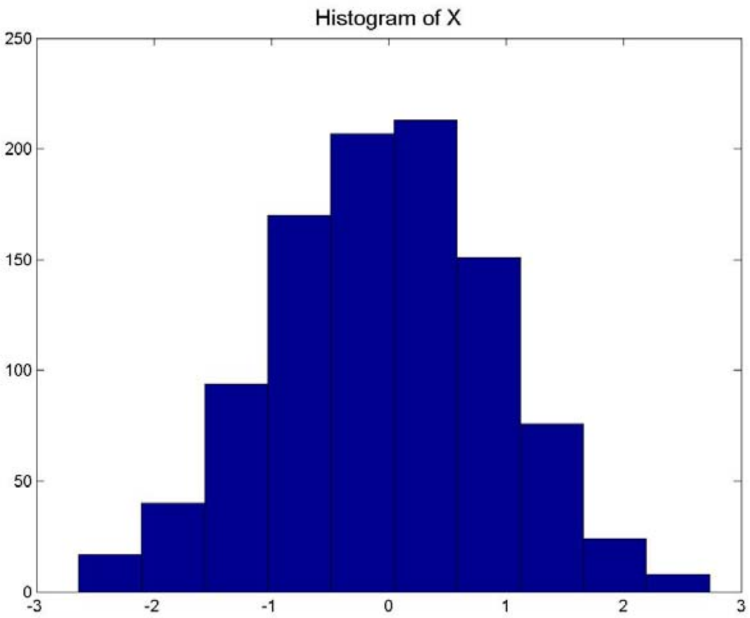
$$\hat{p}(x) = \frac{n_j}{\sum_j^N n_j dx}$$

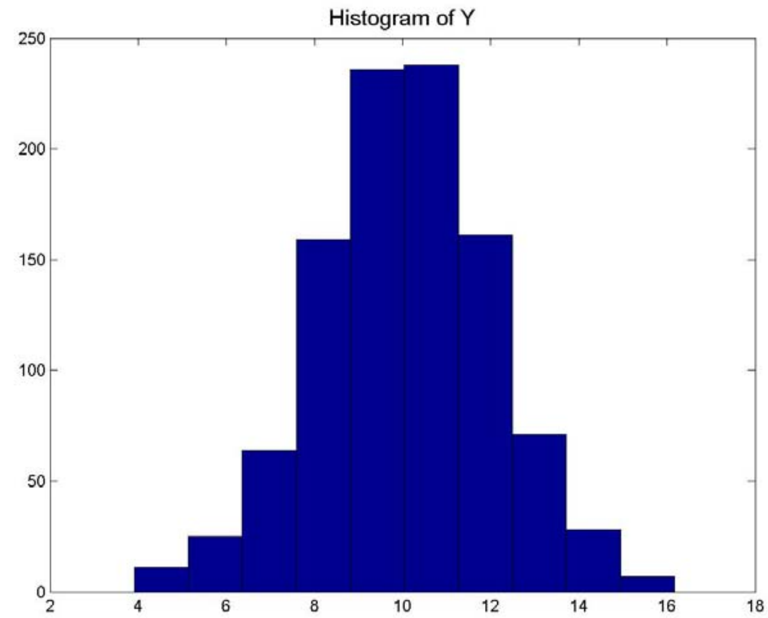
    όπου n_j: αριθμός των δεδομένων που αντιστοιχούν στην κατανομή j πλάτους d_x.
  - Όσο πιο μικρό το d_x, τόσο πιο λεπτομερές είναι το ιστόγραμμα. Προσοχή – πολύ μικρό → πολλές αιχμές, πολύ μεγάλο → πολύ λείο, χάνει το σχήμα του

Παράδειγμα ιστογράμματος

Τυχαίες γκαουσιανές τιμές:

$$X = \frac{1}{\sqrt{2\pi}} e^{-(x)^2/2}, \quad Y = \frac{1}{2\sqrt{2\pi}} e^{-(y-10)^2/8}$$



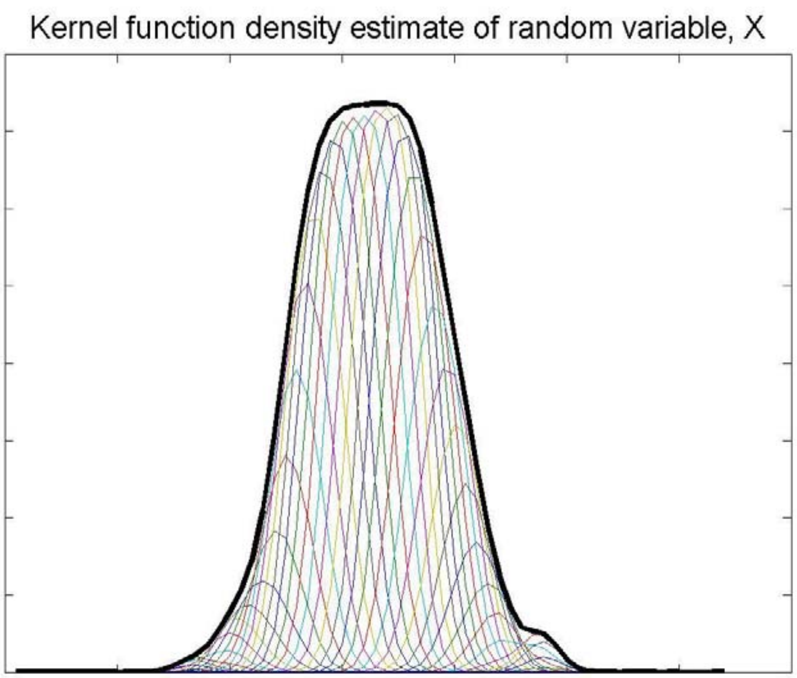


Υπολογισμός ΣΚΠ…

- Μέσω πυρήνα (kernel density estimate):

$$\hat{p}(X)=\frac{1}{Nh}\sum_{i=1}^{N}K\left(\frac{x-x_i}{h}\right)$$

όπου N: αριθμός πυρήνων (kernel functions), h: μέγεθος "κουτιού" με κέντρο x_i ("bandwidth").



- Συνήθεις πυρήνες:

$$K\left(\frac{x-x_i}{h}\right)=\frac{1}{(2\pi)^{d/2}h^d\sqrt{\det(S)}}e^{-1/2\left(\frac{x-x_i}{h}S^{-1}\frac{x-x_i}{h}\right)}$$

Υπολογισμός MI χωρίς ΣΚΠ

- Μέθοδος k-πλησιέστερης γειτνίασης (k-nearest neighbor):

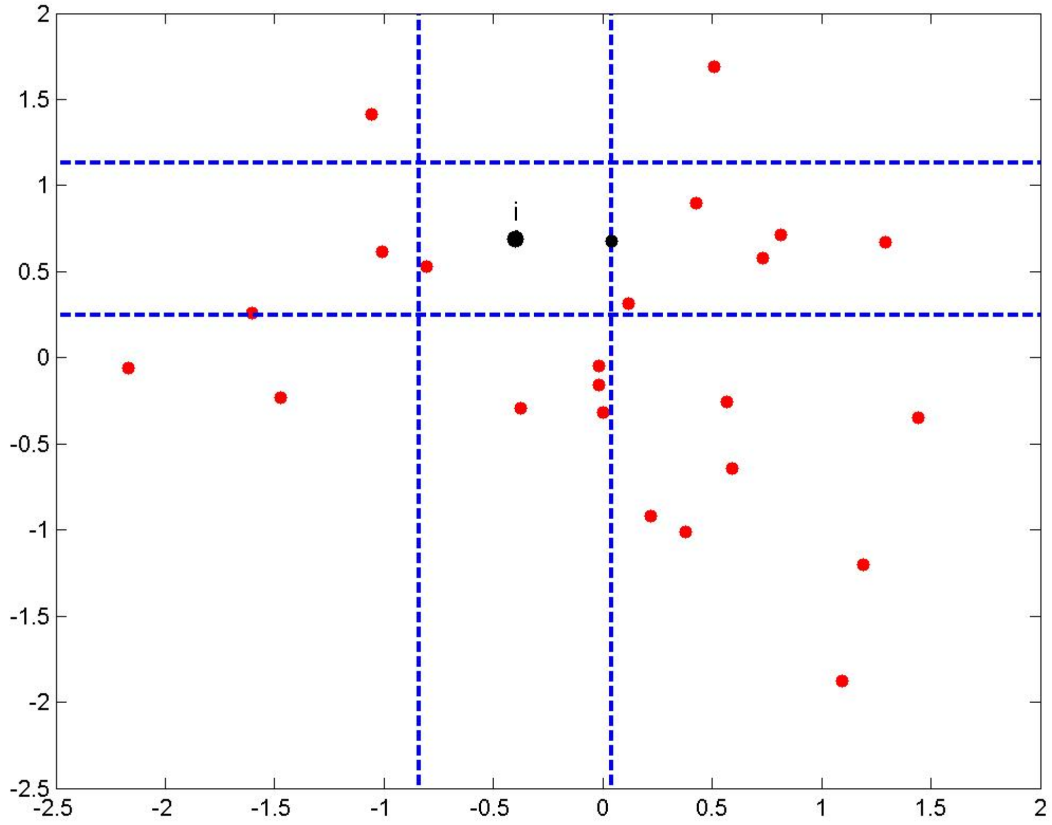
$$MI^{1}(x_1,\ldots,x_m) = \psi(k) + (m-1)\psi(N) - \left\langle \psi(n_{x_1}) + \psi(n_{x_2}) + \ldots + \psi(n_{x_m}) \right\rangle$$

  όπου k: αριθμός κοντινότερου γείτονα, ψ(.): συνάρτηση digamma, $n_{x1},\ldots,n_{xm}$: αριθμός κοντινότερων γειτόνων στις m διαστάσεις, και N: μέγεθος τυχαίας μεταβλητής.

- Απόσταση:

$$\|z - z'\| = \max\{\|x - x'\|, \|y - y'\|\}$$

Παράδειγμα πλησιέστερων γειτόνων



Σε αυτό το παράδειγμα:

$n_x(i)=5$

και

$n_y(i)=8$

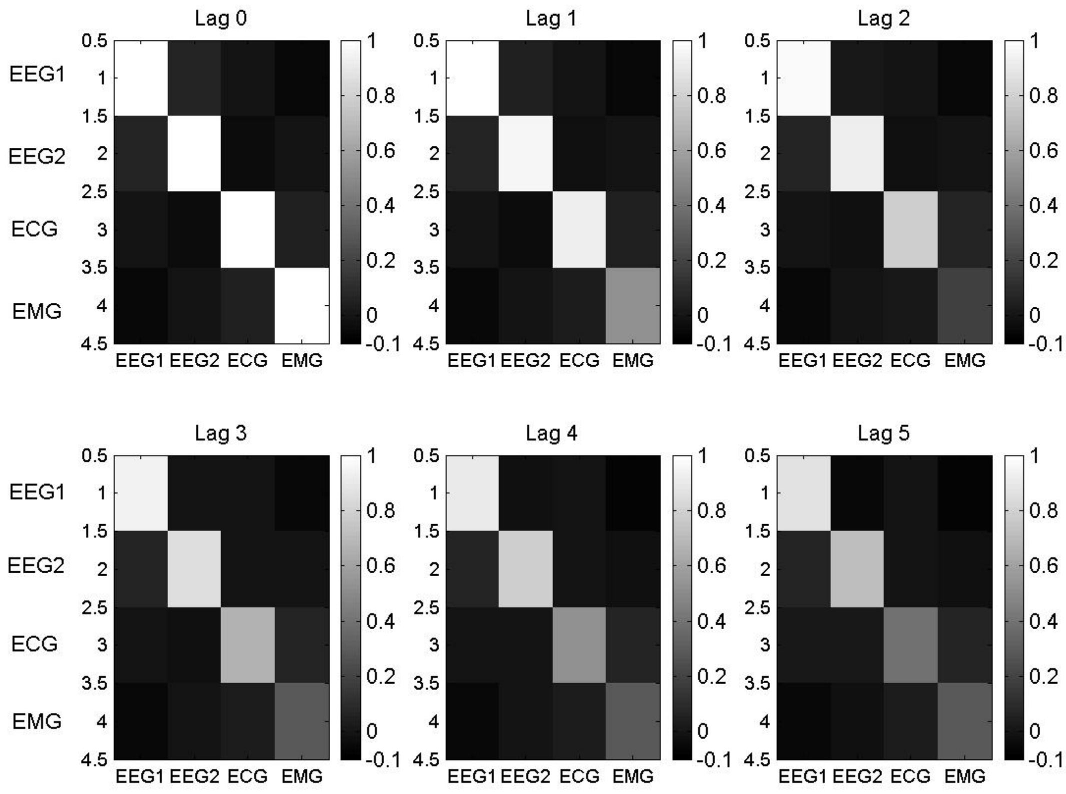
Αμοιβαία αυτοπληροφορία (Auto-Mutual Information - AMI)

- Lagged auto-mutual information: παρόμοια με αυτοσυσχέτιση για τ χρονικές στιγμές, δηλ. χαρακτηρίζει το πώς μια μεταβλητή εξελίσσεται στο χρόνο.
- Διαφορά με αυτοσυσχέτιση: AMI δίνει γραμμικές και μη γραμμικές πληροφορίες!

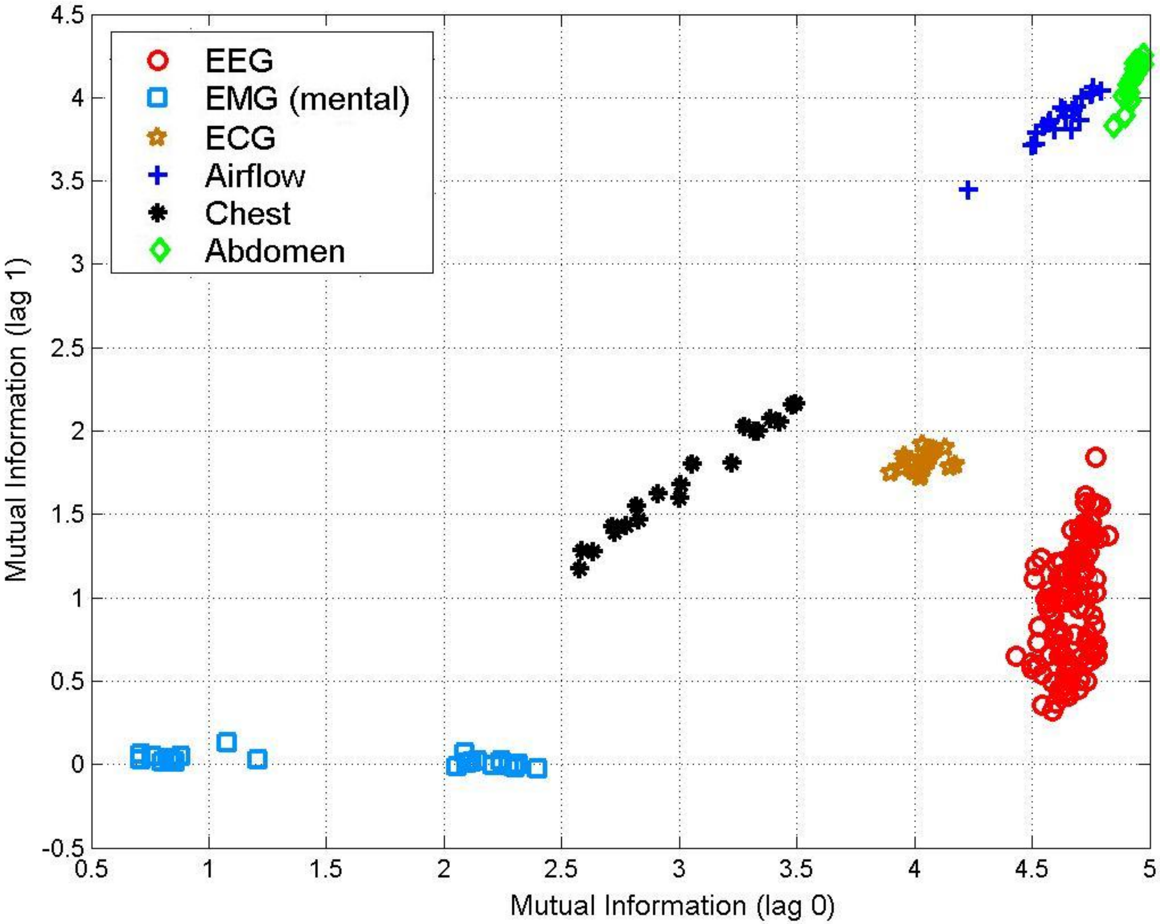
$$AMI_{\tau}(X) = MI(x(T), x(T+\tau))$$

Εφαρμογές σε ΔΕΥ

(1) Αναγνώριση μεταξύ ΗΕΓ και θορύβου



Auto-mutual information για ΗΕΓ και θόρυβο



(2) Συσχέτιση μεταξύ υπολογισμένων independent/principal components και σήματα θορύβου για αναγνώριση θορύβου

(3) Επιλογή κατάλληλων συνδιασμών χαρακτηριστικών για ΔΕΥ

(4) Ομαδοποίηση independent/principal components

(5) Βασικά κομμάτια άλλων μεθόδων επεξεργασίας που χρησιμοποιούνται σε ΔΕΥ

Επόμενο μάθημα:

Μέθοδοι επεξεργασίας σημάτων:

Παλινδρόμηση

Αυτοπαλινδρομικά Μοντέλα