Χρονοσειρές Μάθημα 1

Περιεχόμενα

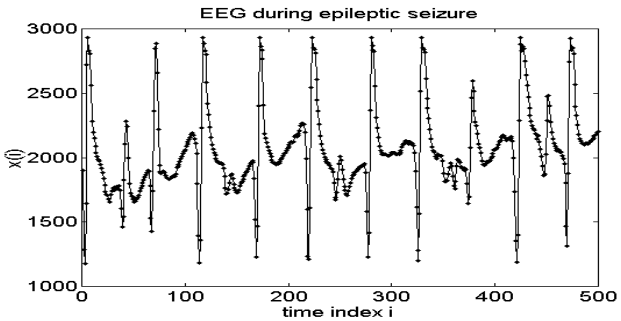
- Στασιμότητα, αυτοσυσχέτιση, μερική αυτοσυσχέτιση, απομάκρυνση στοιχείων μη-στατικότητας, έλεγχος ανεξαρτησίας για χρονικές σειρές
- Γραμμικές στοχαστικές διαδικασίες: αυτοπαλινδρομούμενη (AR), κινούμενου μέσου (MA), μικτή (ARMA)
- Προσαρμογή στάσιμων μοντέλων AR, MA και ARMA σε χρονικές σειρές
- Μη-στάσιμα γραμμικά μοντέλα χρονικών σειρών
- Πρόβλεψη χρονικών σειρών
- Φασματική ανάλυση χρονικών σειρών
- Μη-γραμμική ανάλυση χρονικών σειρών με στοχαστικά μοντέλα
- Μη-γραμμική ανάλυση χρονικών σειρών και δυναμικά συστήματα

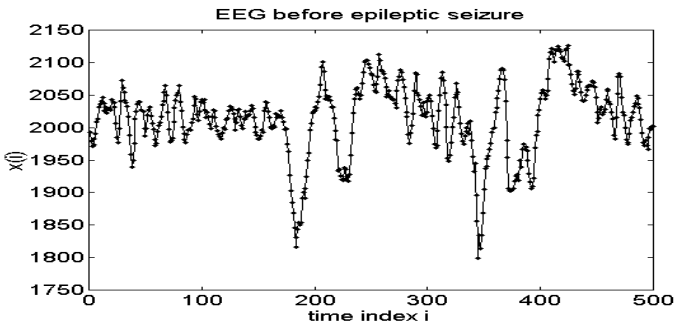
Βιβλία

- “The Analysis of Time Series, An Introduction”, Chatfield C., Sixth edition, Chapman & Hall, 2004
- “Introduction to time series and forecasting”, Brockwell P.J. and Davis R.A., Second edition, Springer, 2002
- “Non-Linear Time Series, A Dynamical System Approach”, Tong H., Oxford University Press, 1993
- “Nonlinear Time Series Analysis”, Kantz H. and Schreiber T., Cambridge University Press, 2004

Πραγματικές χρονοσειρές

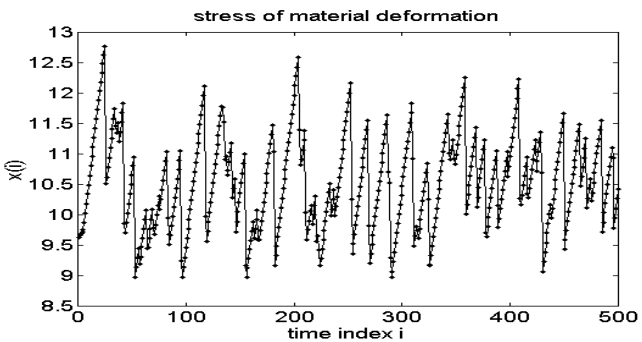
φυσιολογία





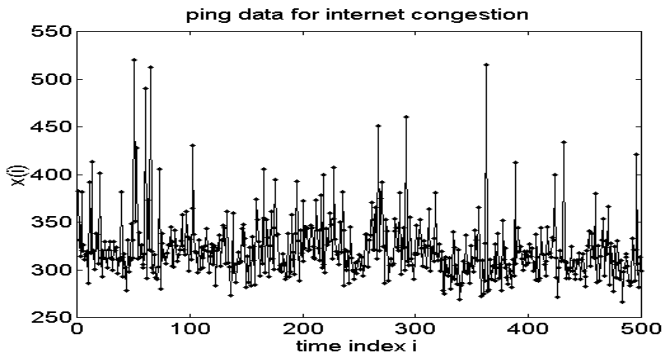
μονοδιάστατη χρονοσειρά

μηχανική



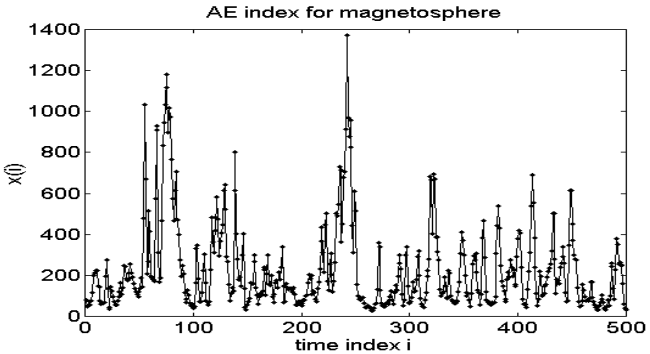
ηλεκτρονική

μόνο μια χρονοσειρά



περιορισμένο μήκος

γεωφυσική



οικονομία

μη-στασιμότητα

θόρυβος

Ορισμοί / συμβολισμοί

Παρατηρούμενο μέγεθος → μεταβλητή [variable] X

Οι τιμές του παρατηρούμενου μεγέθους αλλάζουν με κάποια μικρή ή μεγάλη τυχαιότητα (στοχαστικότητα) → τυχαία μεταβλητή (τ.μ.) [random variable] X

Οι παρατηρήσεις γίνονται συνήθως με συγκεκριμένο χρονικό βήμα
→ χρόνος δειγματοληψίας [sampling time].

Για κάθε χρονική στιγμή t θεωρούμε την τιμή x_t της τυχαίας μεταβλητής X.

Το σύνολο των τιμών της μεταβλητής x_t για κάποια χρονική περίοδο n (σε μονάδες δειγματοληψίας) → (μονοδιάστατη) **χρονοσειρά** [(univariate) time series]

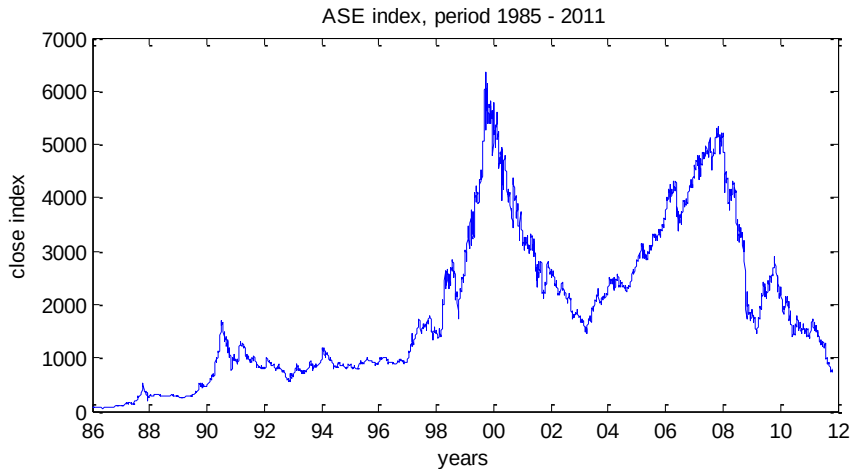
$$\{x_t\}_{t=1}^{n} = \{x_1, x_2, \ldots, x_n\}$$

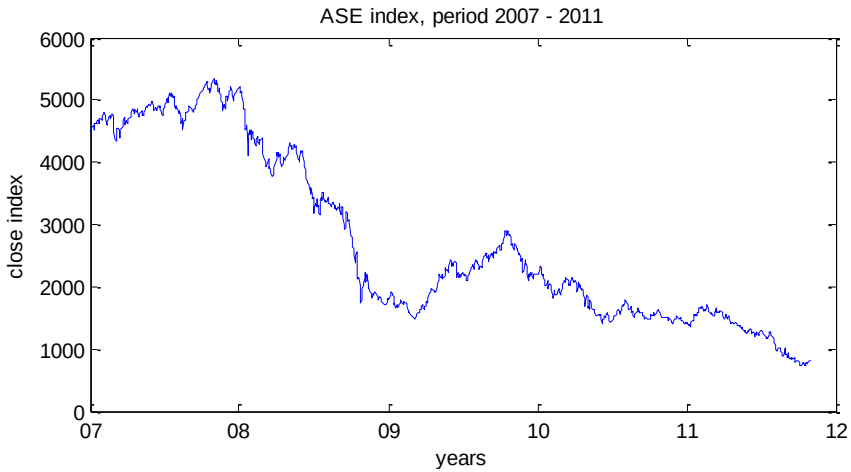
Αν υπάρχουν ταυτόχρονες παρατηρήσεις περισσότερων από μιας μεταβλητής
→ πολυδιάστατη χρονοσειρά [multivariate time series]

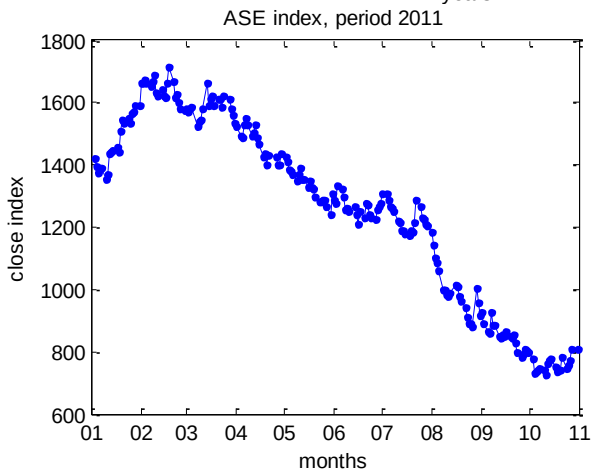
Στη μονοδιάστατη ή πολυδιάστατη χρονοσειρά εφαρμόζουμε μεθόδους και τεχνικές για να αντλήσουμε πληροφορίες για το σύστημα που την παράγει
→ **ανάλυση χρονοσειρών** [time series analysis]

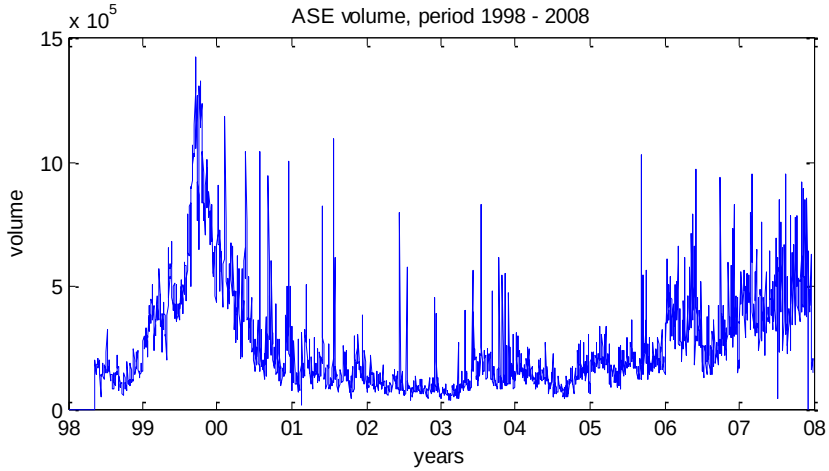
Η χρονοσειρά μπορεί να θεωρηθεί ως πραγματοποίηση μιας στοχαστικής ή καθοριστικής διαδικασίας (δυναμικό σύστημα) $\{X_t\}_{t=-\infty}^{\infty}$

Δείκτης και όγκος συναλλαγών Χρηματιστηρίου Αξιών Αθηνών (ΧΑΑ)









Πρόβλεψη ? Ποια είναι η τιμή του δείκτη αύριο? Μεθαύριο?

Δυναμικό σύστημα
Στοχαστική διαδικασία ? Ποιος είναι ο μηχανισμός της ελληνικής χρηματιστηριακής αγοράς?

Γενικός δείκτης τιμών καταναλωτή (GICP)

Τάση?

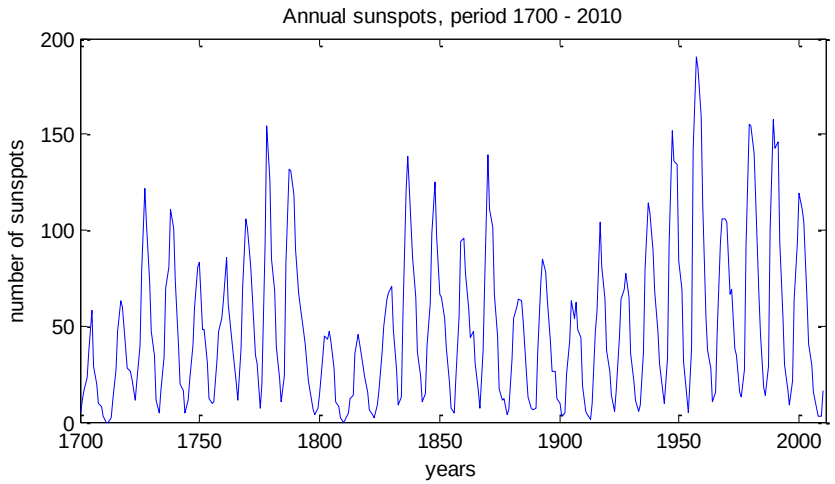
Εποχικότητα / περιοδικότητα?

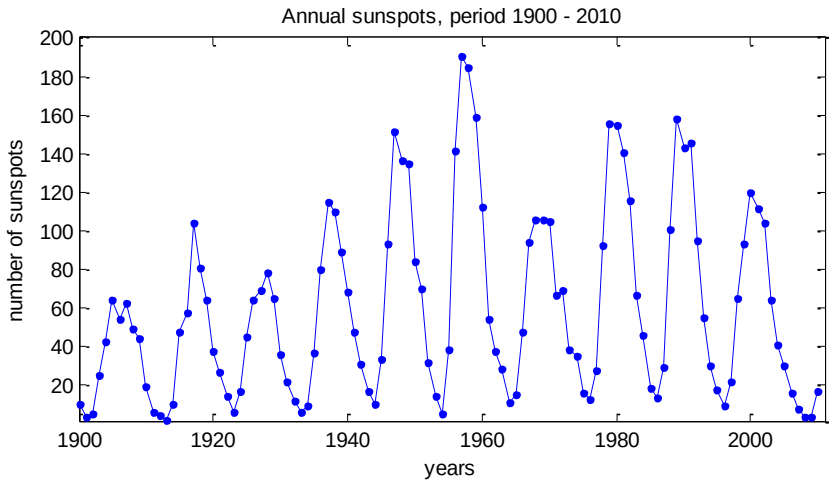
Αυτοσυσχέτιση?

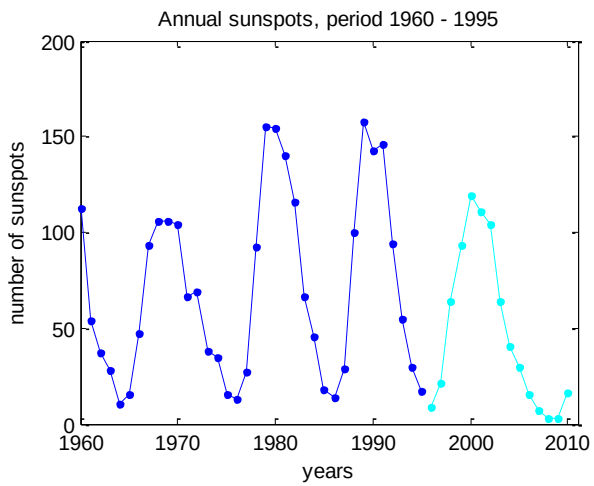
Αυτοπαλινδρόμηση?

Πρόβλεψη ?

Ετήσιες ηλιακές κηλίδες







Ποιος είναι ο μηχανισμός / σύστημα / διαδικασία που δημιουργεί τις ηλιακές κηλίδες?

Είναι **περιοδικό** σύστημα + **θόρυβος** ?

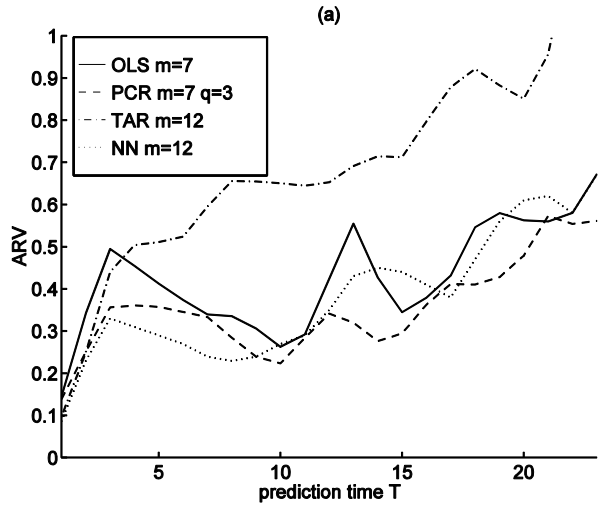
Είναι **στοχαστικό** σύστημα?

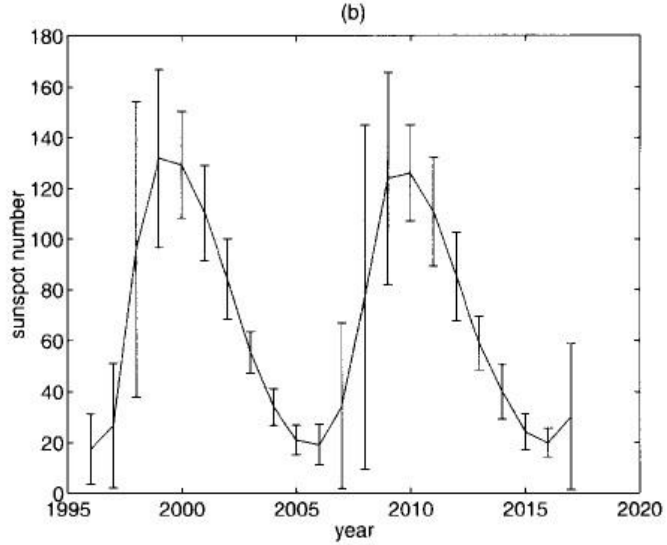
Είναι **χαοτικό** σύστημα?

Γνωρίζοντας τον αριθμό ηλιακών κηλίδων ως το 1995, ποιος θα είναι ο αριθμός τους το 1996 και μετά?

Ποιος θα είναι ο αριθμός των ηλιακών κηλίδων το 2011? 2012? …

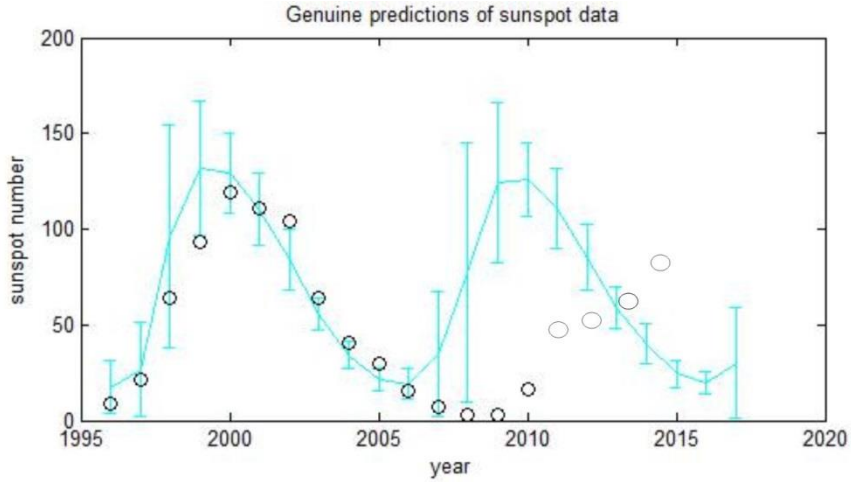
Σύγκριση μοντέλων





Αυθεντική πρόβλεψη

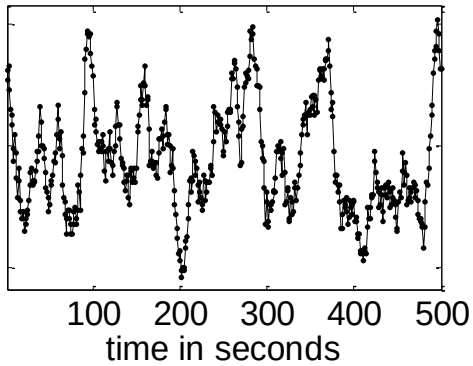
Fig. 9. (a) ARV for iterative multi-step prediction with different models. The test set is the period 1921–1955. The first prediction year of the test set increases with T, e.g. for $T = 1$ the first prediction starts in 1921 and for $T = 23$ in 1944. Thus the number of predictions in the test set decreases with T. The models used in each plot are shown in the legend. (b) The genuine out-of-sample iterative prediction of the sunspot numbers up to the year 2017 based on data up to 1995 using the PCR model with $q = 3$ and $m = 7$. The bars show two times the standard deviation of the point estimate for each step.



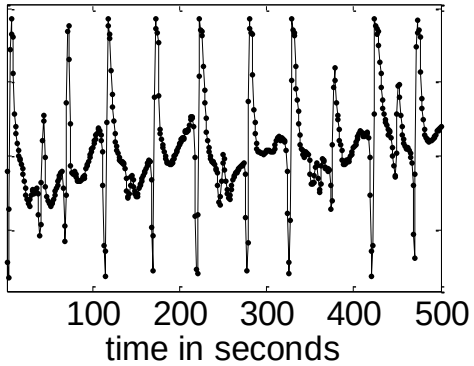
Τι σύστημα παράγει μια πραγματική χρονοσειρά;

Πραγματική χρονοσειρά

preictal EEG

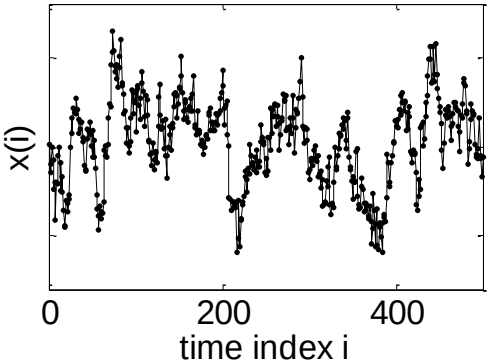


ictal EEG



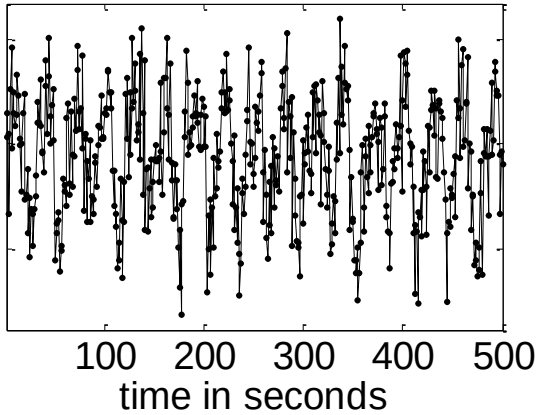
Πιθανά στοχαστικά μοντέλα

stochastic

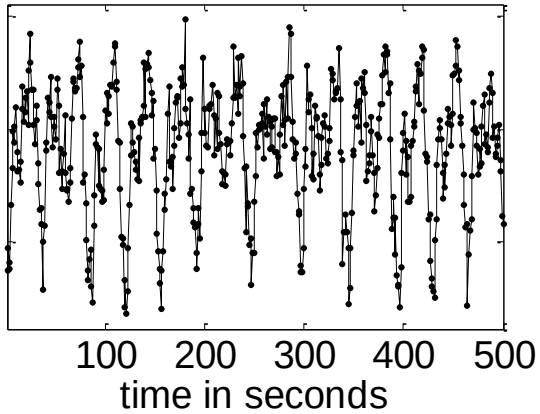


Πιθανά καθοριστικά μοντέλα

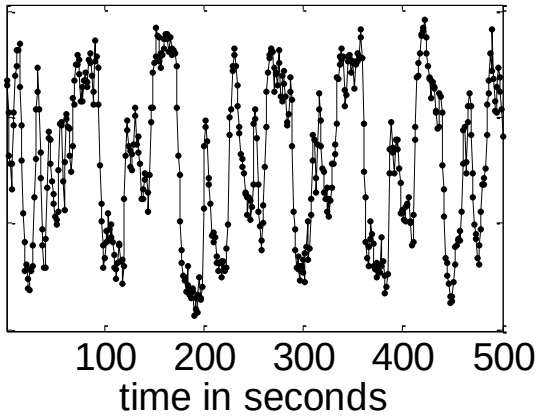
periodic + noise



low dimensional chaos

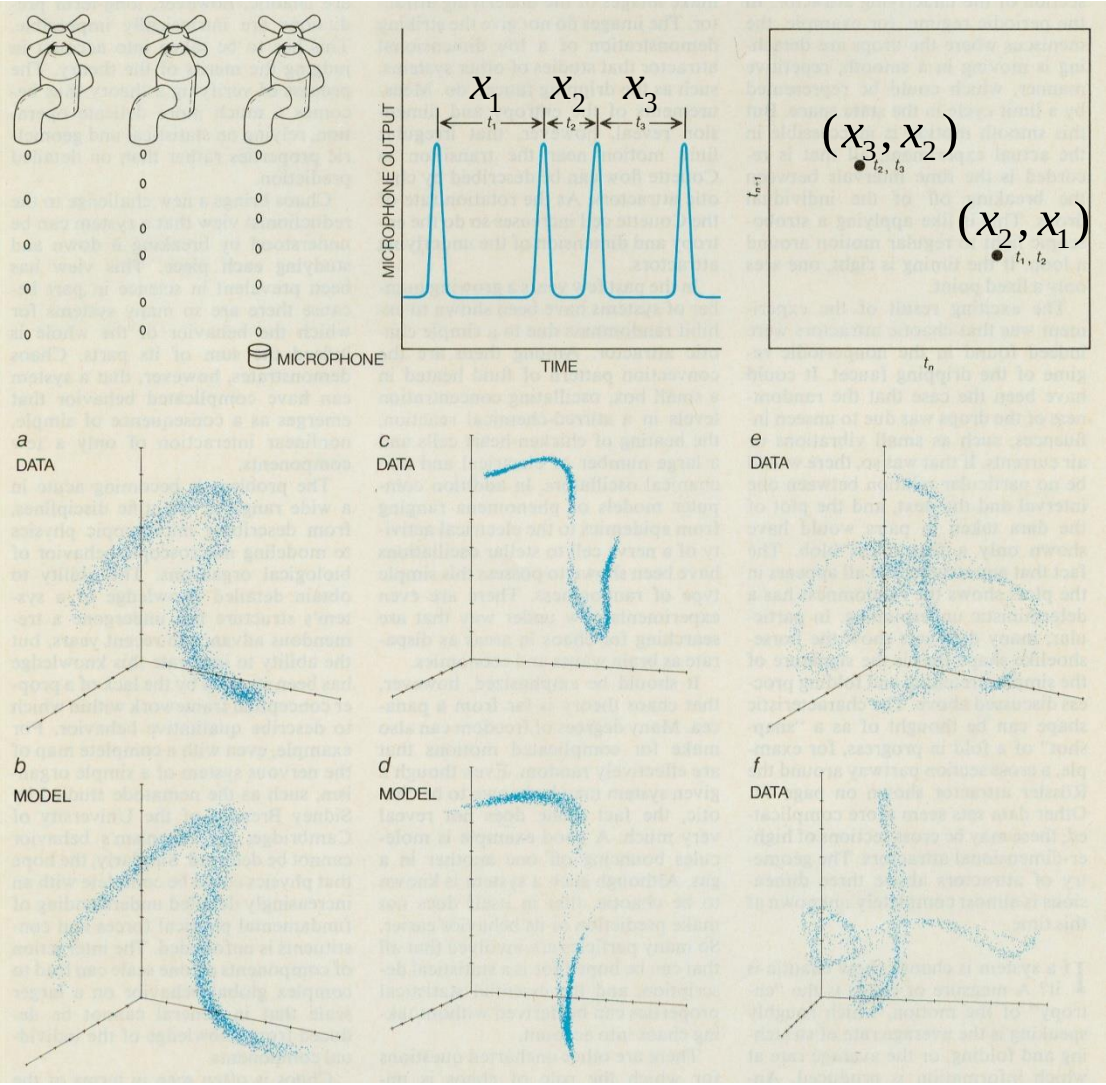


high dimensional chaos



Βρύση που στάζει

(dripping water faucet, UC Santa Cruz).



Crutchfield et al, Scientific American, 1986

Η παρατήρηση της βρύσης που στάζει έδειξε πως για κάποια ταχύτητα ροής, οι σταγόνες δεν τρέχουν σε σταθερά χρονικά διαστήματα.

Το διάγραμμα διασποράς των δεδομένων έδειξε ότι το στάξιμο των σταγόνων δεν είναι τυχαίο.

διάγραμμα διασποράς (x_i, x_{i-1}) (x_i, x_{i-1}, x_{i-2})

απεικόνιση Henon

$$s_i = 1 - 1.4 s_{i-1}^2 + 0.3 s_{i-2}$$ **χάος**

παρατηρούμενη μεταβλητή

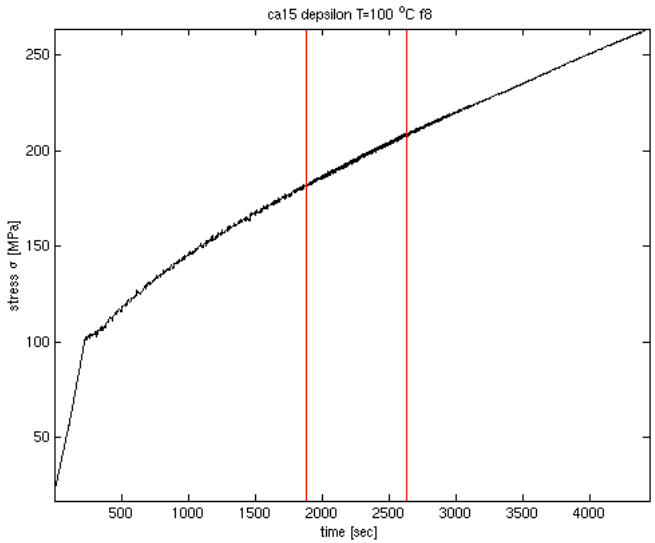
$x_i = s_i + w_i$ w_i θόρυβος

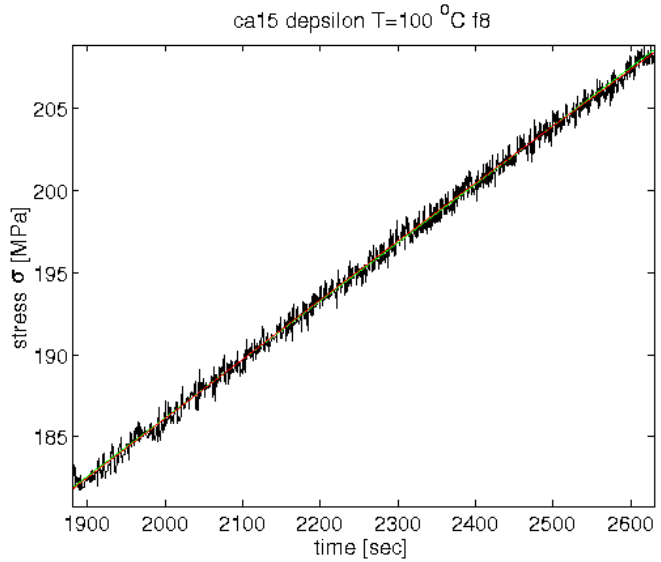
1 **Παρουσίαση του θέματος: Πείραμα / μετρήσεις**

Στασιμότητα - τάση

Τάση [trend]: αργή μεταβολή των τιμών x_t

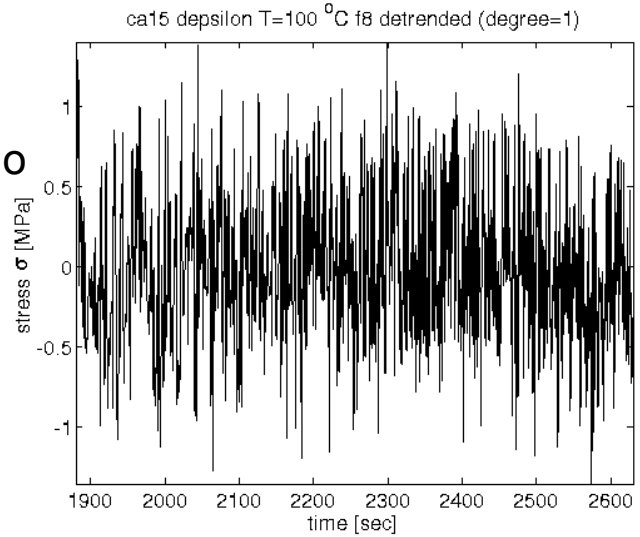
Πλαστική παραμόρφωση

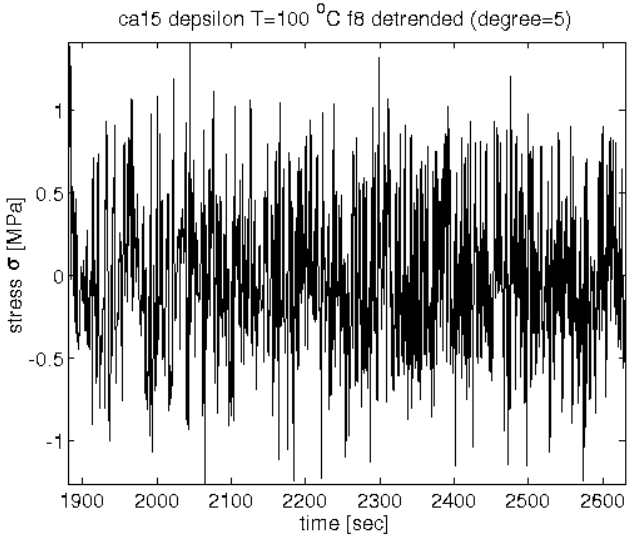




Καθοριστική τάση [deterministic trend]: κάποια συνάρτηση του χρόνου $\mu_t = f(t)$

Προσαρμογή με πολυώνυμο βαθμού 1



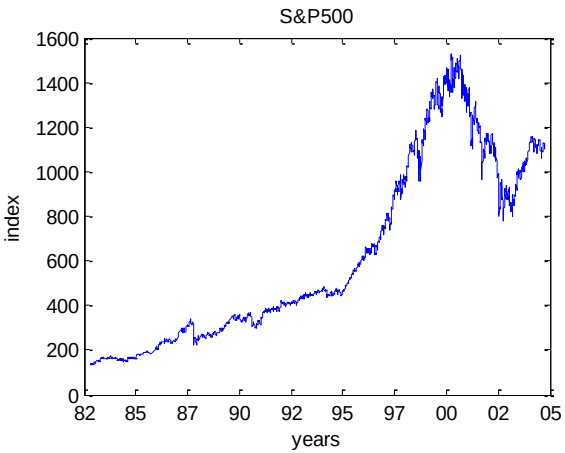


Προσαρμογή με πολυώνυμο βαθμού 5

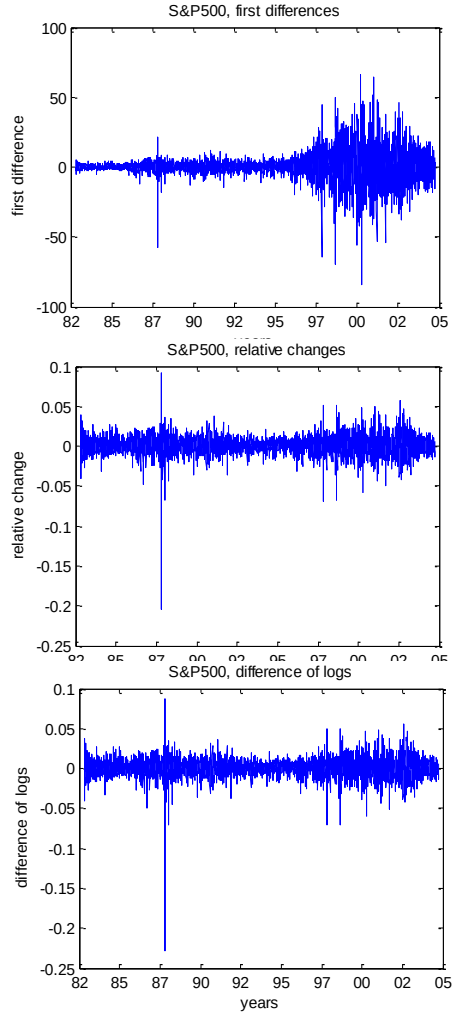
Στοχαστική τάση [stochastic trend]: τυχαία αργή μεταβολή μ_t

Y_t : η παρατήρηση ενός μεγέθους σε χρόνο t

$\{y_1, y_2, \ldots, y_n\}$ χρονοσειρά



μετασχηματισμός



μεταβολή τιμής

$$x_t = y_t - y_{t-1}$$

σχετική μεταβολή τιμής

$$x_t = \frac{y_t - y_{t-1}}{y_t}$$

μεταβολή λογαριθμού τιμής

$$x_t = \ln y_t - \ln y_{t-1}$$

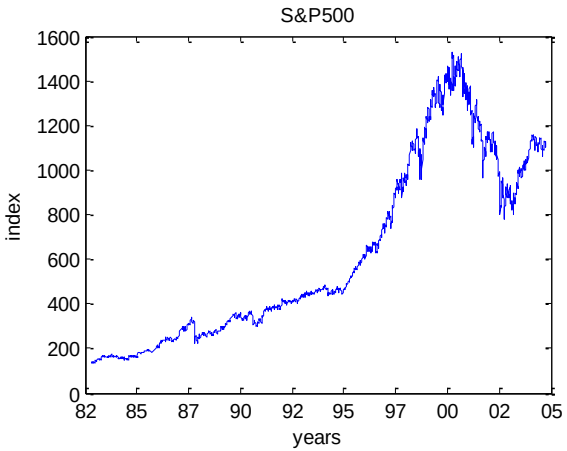
Χρονική συσχέτιση

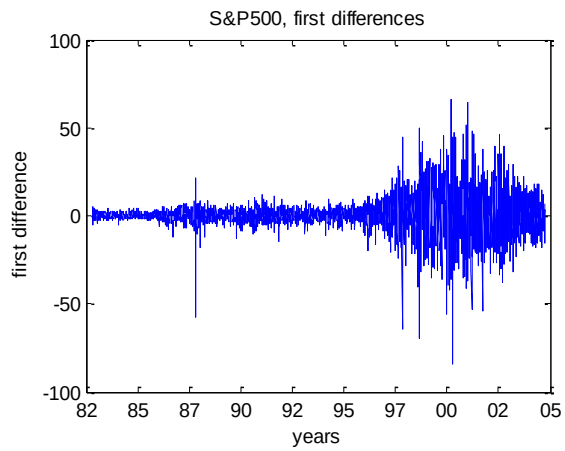
Στοχαστική διαδικασία

Y_t : η τιμή ενός μεγέθους

$\{y_1, y_2, \ldots, y_n\}$ χρονοσειρά

$\{Y_t\}_{t=-\infty}^{\infty}$ $\{X_t\}_{t=-\infty}^{\infty}$



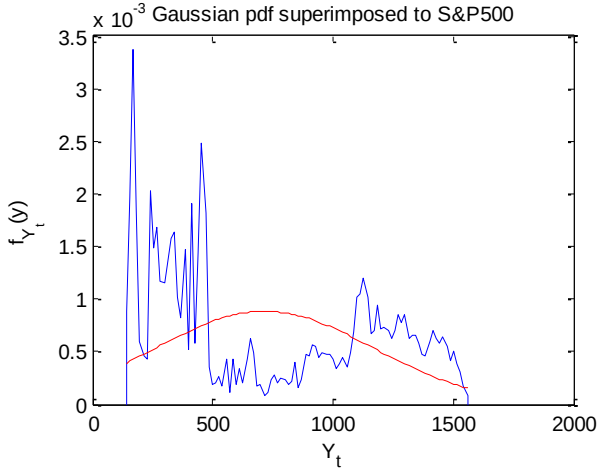


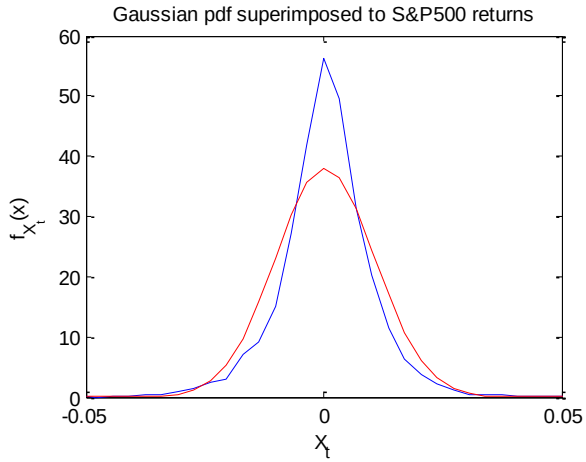
μεταβολή τιμής

$x_t = y_t - y_{t-1}$

$f_{Y_t}(y)$

$f_{X_t}(x)$





Στατική περιγραφή …
→ περιθώρια κατανομή

Δυναμική περιγραφή?
→ Χρονική συσχέτιση

Κατανομές και ροπές στοχαστικής διαδικασίας

Η στοχαστική διαδικασία περιγράφεται από την περιθώρια και τις κοινές κατανομές

$$\left.\begin{array}{l} \forall t \in Z \\ \forall t_1, t_2 \in Z \\ \forall t_1, t_2, t_3 \in Z \end{array}\right\} \begin{array}{l} f_{Y_t}(y) = f_Y(y,t) \\ f_{Y_{t_1},Y_{t_2}}(y_1,y_2) = f_Y(y_1,y_2,t_1,t_2) \\ f_{Y_{t_1},Y_{t_2},Y_{t_3}}(y_1,y_2,y_3) = f_Y(y_1,y_2,y_3,t_1,t_2,t_3) \end{array}$$

περιθώρια κατανομή

κοινή κατανομή 2 μεταβλητών

κοινή κατανομή 3 μεταβλητών

…

Ροπή πρώτης τάξης (μέση τιμή) $\mathrm{E}\left[Y_t\right] = \int_{-\infty}^{\infty} y f_Y(y,t)\mathrm{d}y = \mu_t$

Ροπή δεύτερης τάξης $\mathrm{E}\left[Y_{t_1} Y_{t_2}\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y_1 y_2 f_Y(y_1,y_2,t_1,t_2)\mathrm{d}y_1\mathrm{d}y_2 = \kappa(t_1,t_2)$

Κεντρική ροπή δεύτερης τάξης $\mathrm{E}\left[(Y_{t_1}-\mu_{t_1})(Y_{t_2}-\mu_{t_2})\right] = \kappa(t_1,t_2) - \mu_{t_1}\mu_{t_2} = \gamma(t_1,t_2)$

αυτοδιασπορά [autocovariance]

Ροπές μεγαλύτερης τάξης …

Γενικά η κατανομή και οι ροπές μπορεί να αλλάζουν σε κάθε χρονικό βήμα

Στασιμότητα

Αυστηρή στασιμότητα [strict-sense stationarity]

Οι κατανομές είναι σταθερές στο χρόνο (ισοδύναμα όλες οι ροπές είναι σταθερές)

$$\left.\begin{array}{l} \forall t \in Z \\ \forall t_1, t_2 \in Z \\ \forall t_1, t_2, t_3 \in Z \end{array}\right\} \begin{array}{l} f_{Y_t}(y) = f_Y(y,t) = f_{Y_t}(y) \\ f_{Y_{t_1},Y_{t_2}}(y_1,y_2) = f_{Y_t,Y_{t-\tau}}(y_1,y_2) \\ f_{Y_{t_1},Y_{t_2},Y_{t_3}}(y_1,y_2,y_3) = f_{Y_t,Y_{t-\tau_1},Y_{t-\tau_2}}(y_1,y_2,y_3) \end{array} \left\{ \begin{array}{l} \text{σταθερές} \\ \forall t \in Z \end{array} \right.$$

Ασθενή στασιμότητα [wide-sense stationarity]

Οι δύο πρώτες ροπές είναι σταθερές στο χρόνο

$$\left.\begin{array}{l} \mathrm{E}\left[Y_t\right] = \mu \\ \mathrm{E}\left[Y_{t_1}Y_{t_2}\right] = \mathrm{E}\left[Y_t, Y_{t-\tau}\right] = \kappa(t, t-\tau) = \kappa(\tau) \\ \gamma(t_1,t_2) = \gamma(t, t-\tau) = \gamma(\tau) \end{array}\right\} \begin{array}{l} \text{σταθερές} \\ \forall t \in Z \end{array}$$

σταθερή μέση τιμή και αυτοδιασπορά

και για $\tau=0$ $\quad \mathrm{E}\left[Y_t^2\right] = \kappa(0)$

$$\sigma^2 = \gamma(0) = \mathrm{E}\left[Y_t^2\right] - \left(\mathrm{E}\left[Y_t\right]\right)^2 = \kappa(0) - \mu^2$$

σταθερή διασπορά

Αυτοσυσχέτιση

Στάσιμη χρονοσειρά $\{X_t\}_{t=-\infty}^{\infty}$

Αυτοδιασπορά $\gamma(\tau) = \mathrm{E}\left[(X_t - \mu)(X_{t-\tau} - \mu)\right] = \mathrm{E}\left[X_t X_{t-\tau}\right] - \left(\mathrm{E}\left[X_t\right]\right)^2 = \kappa(\tau) - \mu^2$

Διασπορά $\sigma^2 = \gamma(0) = \mathrm{E}\left[X_t^2\right] - \left(\mathrm{E}\left[X_t\right]\right)^2 = \kappa(0) - \mu^2$

Αυτοσυσχέτιση $\rho(\tau) = \dfrac{\gamma(\tau)}{\gamma(0)}$

$\rho(0) = 1$

Χρονική συσχέτιση μεταβλητών της $\{X_t\}_{t=-\infty}^{\infty}$ σε υστέρηση τ.
Μετράει τη «μνήμη» της $\{X_t\}_{t=-\infty}^{\infty}$

Συμβολισμός: $\gamma(\tau) = \gamma_\tau \qquad \rho(\tau) = \rho_\tau$

Παρατηρήσεις:

- $|\gamma_k| \le \gamma_0$ και $|\rho_k| \le 1$
- $\gamma_k = \gamma_{-k}$ και $\rho_k = \rho_{-k}$
- Πίνακας αυτοδιασπορών

$$\Gamma_n = \begin{vmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \cdots & \vdots \\ \gamma_{n-1} & \cdots & \cdots & \gamma_0 \end{vmatrix}$$

Πίνακας αυτοσυσχετίσεων

$$\mathrm{P}_n = \begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{n-1} & \cdots & \cdots & 1 \end{vmatrix}$$

Κάποιες βασικές στοχαστικές διαδικασίες

1

$\{X_t\}_{t=-\infty}^{\infty}$ **ανεξάρτητες ισόνομες τ.μ.** [independent and identically distributed, **iid**]

$$P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n) = P(X_1 \le x_1)P(X_2 \le x_2)\ldots P(X_n \le x_n)$$

$$\mathrm{E}[X_t] = 0$$

2

$\{X_t\}_{t=-\infty}^{\infty}$ **λευκός θόρυβος** [white noise, **WN**], ασυσχέτιστες τ.μ.

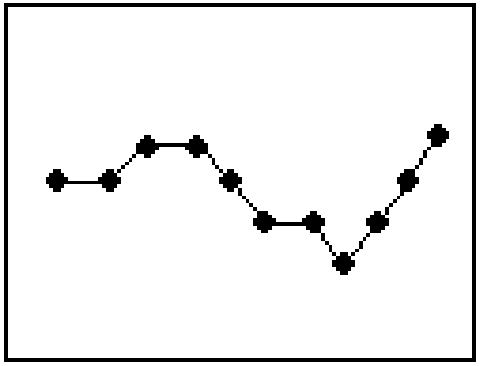
$$\mathrm{E}[X_i X_j] = \delta_{ij}\sigma^2$$

3

$\{Y_t\}_{t=1}^{\infty}$ **τυχαίος περίπατος** [random walk, RW]

$$Y_t = Y_{t-1} + X_t = X_1 + X_2 + \cdots X_t \qquad \{X_t\}_{t=-\infty}^{\infty} \text{ iid} \qquad \mathrm{E}[X_t] = 0 \qquad \mathrm{E}[X_t^2] = \sigma^2$$

$$\mathrm{E}[Y_{t+1} \mid Y_0, Y_1, \ldots, Y_t] = Y_t$$



$$\mathrm{E}[Y_t] = 0 \qquad \mathrm{E}[Y_t^2] = t\sigma^2 \qquad \boxed{?}$$

Η διασπορά αυξάνει γραμμικά με το χρόνο!

$\{X_t\}_{t=-\infty}^{\infty}$ Γκαουσιανή (κανονική) στοχαστική διαδικασία

Για κάθε τάξη p: $f_{X_t, X_{t-\tau_1}, \ldots, X_{t-\tau_{p-1}}}(x_1, x_2, \ldots, x_p)$ είναι p-διάστατη Γκαουσιανή κατανομή

Η κανονική κατανομή καθορίζεται πλήρως από τις δύο πρώτες ροπές

⇓

αυστηρή στασιμότητα ≡ ασθενής στασιμότητα

Παράδειγμα

A τ.μ. $\quad \mathrm{E}[A]=0 \quad \mathrm{Var}[A]=1$

Στοχαστική διαδικασία: $X_t = A\sin(\omega t+\theta)$ $\quad \theta \sim U[-\pi,\pi]$ $\quad \theta$ και A ανεξάρτητα

Είναι ασθενώς στάσιμη;

- $\mathrm{E}[X_t] = \mathrm{E}[A]\mathrm{E}[\sin(\omega t+\theta)] = 0$?
- $\mathrm{E}[X_t X_{t+\tau}] = \mathrm{E}\left[A^2 \sin(\omega t+\theta)\sin(\omega(t+\tau)+\theta)\right] = \ldots = \frac{1}{2}\cos(\omega\tau)$

Οι ροπές πρώτης και δεύτερης τάξης δεν εξαρτώνται από το χρόνο t.

Δειγματική αυτοδιασπορά / αυτοσυσχέτιση

$\{x_1, x_2, \ldots, x_n\}$ χρονοσειρά

Δειγματική μέση τιμή $\bar{x} = \frac{1}{n}\sum_{t=1}^{n} x_t$

- αμερόληπτος εκτιμητής της μέσης τιμής μ της χρονοσειράς ?

Δειγματική αυτοδιασπορά

$$c(\tau) = \frac{1}{n}\sum_{t=\tau+1}^{n}(x_t x_{t-\tau} - \bar{x}^2) \qquad \tau = 0,1,\ldots,n-1$$

$$c(0) = \frac{1}{n}\sum_{t=1}^{n}(x_t^2 - \bar{x}^2)$$

Συμβολισμός

$c(\tau) \equiv c_\tau$

Άλλη εκτίμηση αυτοδιασποράς $c'(\tau) = \frac{1}{n-\tau}\sum_{t=\tau+1}^{n}(x_t x_{t-\tau} - \bar{x}^2)$

Μεροληπτικοί εκτιμητές:

$$\mathrm{E}[c_\tau] \cong \gamma_\tau - \frac{\tau}{n}\gamma_\tau - \left(\frac{n-\tau}{n}\right)\mathrm{Var}[\bar{x}]$$

η μεροληψία αυξάνει με την υστέρηση τ

$$\mathrm{E}[c'_\tau] \cong \gamma_\tau - \mathrm{Var}[\bar{x}]$$

Δειγματική αυτοσυσχέτιση $r(\tau) = \frac{c(\tau)}{c(0)}$ $\qquad r(0) = 1$

Συμβολισμός

$r(\tau) \equiv r_\tau$

Για μεγάλο n: $r_\tau \sim \mathrm{N}(\rho_\tau, \mathrm{Var}[r_\tau])$

$$\mathrm{Var}[r_\tau] \cong \frac{1}{n}\sum_{m=-\infty}^{\infty}(\rho_m^2 + \rho_{m+\tau}\rho_{m-\tau} + 2\rho_\tau^2\rho_m^2 - 4\rho_\tau\rho_m\rho_{m-\tau})$$

τύπος Bartlet

$$\mathrm{Var}[r_\tau] \cong \frac{1}{n}\sum_{m=-\infty}^{\infty}\rho_m^2$$

πολύ μεγάλο n

Αυτοσυσχέτιση λευκού θορύβου

$\{x_1, x_2, \ldots, x_n\}$ χρονοσειρά λευκού θορύβου $\rho_\tau = 0, \ \forall \tau \neq 0$

$r_\tau \sim \mathrm{N}(0, \frac{1}{n})$ **?**

2 **Έλεγχος τυχαιοποίησης για τη σημαντικότητα αυτοσυσχέτισης**

Έλεγχος σημαντικότητας αυτοσυσχέτισης

$\mathrm{H}_0 : \rho_\tau = 0 \qquad \mathrm{H}_1 : \rho_\tau \neq 0$

Απορριπτική περιοχή: $R = \left\{ r_t \ \middle| \ \left| \frac{r_\tau}{\sqrt{1/n}} \right| > z_{1-\alpha/2} \right\}$ για στάθμη σημαντικότητας α

Ζώνη μη-σημαντικής αυτοσυσχέτισης: $\pm z_{1-a/2} \frac{1}{\sqrt{n}}$ για α=0.05 $\simeq \pm \frac{2}{\sqrt{n}}$

Παράδειγμα

Για μια χρονοσειρά 200 παρατηρήσεων δίνονται οι 10 πρώτες αυτοσυσχετίσεις

1	2	3	4	5	6	7	8	9	10
-0.38	-0.28	0.11	-0.08	0.02	0.00	0.01	0.07	-0.08	0.05

Υποθέτοντας ότι η χρονοσειρά είναι τυχαία (H_0:ρ=0): $\mathrm{Var}[r_\tau] \cong \frac{1}{200} = 0.005$

για α=0.05, το 95% των αυτοσυσχετίσεων αναμένουμε να βρίσκεται στο διάστημα

$$\pm 1.96 * \frac{1}{\sqrt{200}} = \pm 1.96 * 0.07 = \pm 0.139$$

$\rho_1 \neq 0$, $\rho_2 \neq 0$ και $\rho_\tau \neq 0$ για τ=3,4,…