



Τμήμα Πολιτικών Μηχανικών
Πανεπιστήμιο Θεσσαλίας

ΥΔΡΟΛΟΓΙΚΗ ΠΡΟΣΟΜΟΙΩΣΗ ΚΑΙ ΠΡΟΓΝΩΣΗ

Ενότητα 3: Υδρολογική πρόγνωση

3.2. Μοντέλα Χρονοσειρών

Καθ. Αθανάσιος Λουκάς

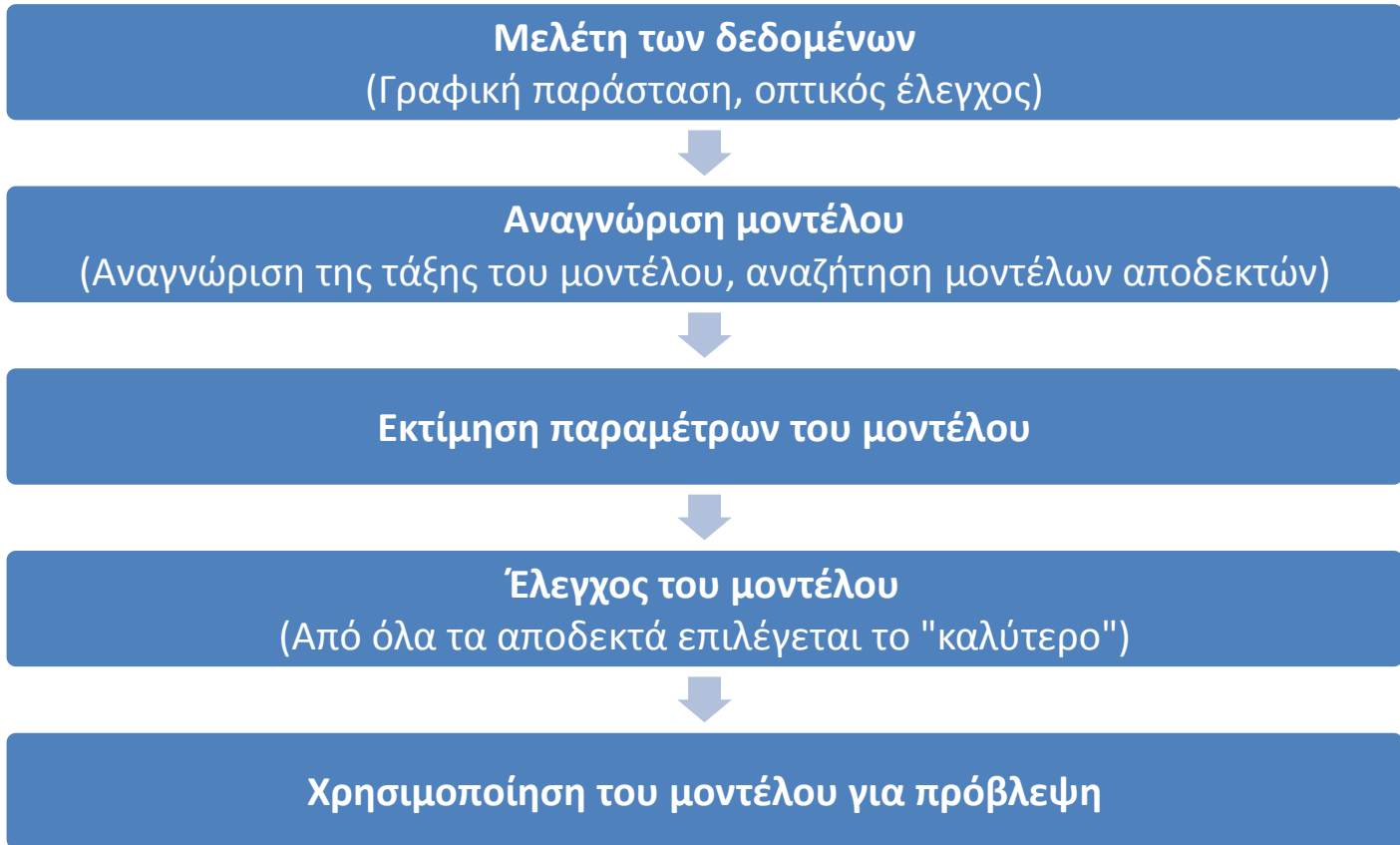
Εργαστήριο Υδρολογίας και Ανάλυσης Υδατικών Συστημάτων

Τμήμα Πολιτικών Μηχανικών

Πολυτεχνική Σχολή

ΑΝΑΛΥΣΗ ΥΔΡΟΛΟΓΙΚΩΝ ΧΡΟΝΟΣΕΙΡΩΝ

Time Series Analysis



Στοχαστική Ανάλυση Χρονολογικής Σειράς

Μοντέλα χρονοσειρών	
Αυτοπαλινδρομούμενα μοντέλα τάξης p (Autoregressive models)	AR(p)
Κινούμενου μέσου μοντέλα τάξης q (Moving Average models)	MA(q)
Μικτά μοντέλα τάξης p, q (Autoregressive Moving Average models)	ARMA(p, q)
Ολοκληρωμένα μικτά μοντέλα τάξης p, q	ARIMA(p, d, q)
Πολαπλασιαστικό ή εποχικό μοντέλο	ARIMA(p, d, q)(P, D, Q)_{s}



Στοχαστικά Μοντέλα - Υποδείγματα

- Γενικά υπάρχουν τρεις βασικές κατηγορίες στοχαστικών υποδειγμάτων χρονολογικών σειρών:
 - Τα αυτοπαλίνδρομα υποδείγματα ή υποδείγματα AR (Autoregressive Models)
 - Τα υποδείγματα κινητών μέσων (Moving Average Models) ή υποδείγματα MA
 - Και τα μεικτά υποδείγματα ή υποδείγματα ARMA (Mixed Autoregressive - Moving Average Models)



Γενικό γραμμικό μοντέλο-General linear model (GLM)

$$\text{Έκφραση: } z_t = a_t + \psi_1 a_{t-1} + \dots = \sum_{i=0}^{\infty} \psi_i a_{t-i} \quad \mu\epsilon \quad \psi_0 = 1$$

a_t λευκός θόρυβος (white noise)

$$E(a_t) = 0 \text{ και } Cov(a_t, a_{t+k}) = \begin{cases} \sigma_a^2 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

Αποδεικνύεται:

$$\alpha) Var(z_t) = \sum_{i=1}^{\infty} \psi_i^2 \sigma_a^2 \text{ και } \exists \text{ όταν } \left(\sum_{i=1}^{\infty} \psi_i^2 \right) \text{ συγκλίνει}$$

$$\beta) \gamma_k = cov(z_t, z_{t+k}) = \left(\sum_{i=0}^{\infty} \psi_i \psi_{i+k} \right) \sigma_a^2$$

για να είναι στάσιμο (το GLM) θα πρέπει $\sum_{i=0}^{\infty} \psi_i \psi_{i+k}$ να συγκλίνει



Τελεστές

$Bz_t = z_{t-1}$	$B^m z_t = z_{t-m}$	Προς τα πίσω μετατόπισης (backward operator)
$Fz_t = z_{t+1}$	$F^m z_t = z_{t+m}$	Προς τα εμπρός μετατόπισης (forward operator)
$\nabla z_t = z_t - z_{t-1}$	$\nabla^m z_t = \nabla^{m-1}(\nabla z_t)$	Προς τα πίσω διαφορών (backward difference operator)



Αυτοπαλίνδρομα μοντέλα τάξης p ($AR(p)$)

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \dots + \varphi_p z_{t-p} + a_t \quad \text{ή} \quad \varphi(B)z_t = a_t$$

όπου $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$ (AR πολυώνυμο)

και $E(a_t) = 0 \quad Var(a_t) = \sigma_a^2$

Θεώρημα: Η ικανή και αναγκαία συνθήκη για να είναι στάσιμο το $AR(p)$ μοντέλο, είναι οι ρίζες της εξίσωσης $\varphi(\omega) = 0 \quad \omega \in \mathbb{C}$ να βρίσκονται έξω από το μοναδιαίο κύκλο.

Συνέπεια του θεωρήματος:

$$\text{για } p = 1 \quad -1 < \varphi_1 < 1 \quad \text{για } p = 2 \quad \begin{array}{l} \varphi_2 - \varphi_1 < 1 \\ \varphi_1 + \varphi_2 < 1 \\ -1 < \varphi_2 < 1 \end{array}$$



Αυτοπαλίνδρομα μοντέλα τάξης p ($AR(p)$)

Εκτίμηση παραμέτρων

- Μέθοδος ροπών ή Yule-Walker
- Μέθοδος ελαχίστων τετραγώνων με συνθήκη
- Μέθοδος ελαχίστων τετραγώνων χωρίς συνθήκη
- Βελτιωμένη μέθοδος ροπών



Αυτοπαλίνδρομα μοντέλα τάξης p (AR(p))

Μέθοδος ροπών ή Yule-Walker

$$Ez_{t-k}z_t = \varphi_1 Ez_{t-k}z_{t-1} + \varphi_2 Ez_{t-k}z_{t-2} + \dots + \varphi_p Ez_{t-k}z_{t-p} + Ez_{t-k}a_t$$

(για στάσιμη χρονοσειρά) $\gamma_k = \varphi_1 \gamma_{k-1} + \varphi_2 \gamma_{k-2} + \dots + \varphi_p \gamma_{k-p}$

ή $\rho_k = \varphi_1 \rho_{k-1} + \varphi_2 \rho_{k-2} + \dots + \varphi_p \rho_{k-p}$

- Για $k = 1, 2, \dots, p$ η λύση του γραμμικού συστήματος εκτιμά τις παραμέτρους $\hat{\varphi}_1, \hat{\varphi}_2, \dots, \hat{\varphi}_p$

- Για $k = 0$ $\gamma_0 = \varphi_1 \gamma_1 + \varphi_2 \gamma_2 + \dots + \varphi_p \gamma_p + \sigma_a^2$

ή $\gamma_0 = \sigma_z^2 = \frac{\sigma_a^2}{1 - \varphi_1 \rho_1 - \dots - \varphi_p \rho_p}$



Αυτοπαλίνδρομα μοντέλα τάξης p ($AR(p)$)

- Ένα **Αυτοπαλίνδρομο μοντέλο** ή **υπόδειγμα p τάξης** ή **$AR(p)$** στη γενική του μορφή είναι:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

$\phi_1, \phi_2, \dots, \phi_p$ οι προς εκτίμηση παράμετροι του μοντέλου.

Η **τάξη p** αναφέρεται στο **μήκος της υστέρησης** και ο όρος **αυτοπαλίνδρομο** οφείλεται στο ότι η παραπάνω σχέση είναι στην ουσία ένα **υπόδειγμα παλινδρόμησης** με ερμηνευτικές μεταβλητές τις τιμές της εξαρτημένης μεταβλητής Y_t με χρονική υστέρηση.

Η μεταβλητή ε_t καλείται **λευκός θόρυβος**.



Αυτοπαλίνδρομο μοντέλο πρώτης τάξης AR(1)

Το **Αυτοπαλίνδρομο υπόδειγμα πρώτης τάξης AR(1)** είναι της μορφής

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \varepsilon_t$$

(αν υποθέσουμε ότι $\mu = 0$) ή αν οι μεταβλητές εκφράζονται ως αποκλίσεις από τους μέσους όρους:

$$y_t = Y_t - \mu, \quad y_{t-1} = Y_{t-1} - \mu$$

θα καταλήξουμε στο τύπο:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t$$



Αυτοπαλίνδρομο μοντέλο πρώτης τάξης AR(1)

Για να είναι **στάσιμη η χρονολογική σειρά** θα πρέπει

$-1 < \phi_1 < 1$ (ο ϕ_1 είναι ο συντελεστής αυτοσυσχέτισης)

Αν $\phi_1 > 0$ η συνάρτηση αυτοσυσχέτισης ACF αρχίζει από τη μονάδα και φθίνει γεωμετρικά προς το μηδέν καθώς αυξάνει το s .

✓ Αν $\phi_1 < 0$ η συνάρτηση αυτοσυσχέτισης ACF τείνει προς το μηδέν αλλά με εναλλασσόμενο πρόσημο καθώς αυξάνει το s .

Αν εφαρμοσθεί ο **τελεστής υστερήσεως (lag operator) B**:

$$BY_t = Y_{t-1}, B_2Y_t = Y_{t-2}, \dots, B_pY_t = Y_{t-p}$$

Έχουμε ότι: $y_t = \phi_1 y_{t-1} + \varepsilon_t = \phi_1 B y_t + \varepsilon_t$

ή

$$(1 - \phi_1 B) y_t = \varepsilon_t$$



Αυτοπαλίνδρομο μοντέλο δεύτερης τάξης AR(2)

Το **Αυτοπαλίνδρομο υπόδειγμα δεύτερης τάξης AR(2)** είναι της μορφής:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

ή

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

Αν εφαρμοσθεί ο **τελεστής υστερήσεως (lag operator) B** η σχέση γράφεται και ως εξής:

$$(1 - \phi_1 B - \phi_2 B^2) y_t = \varepsilon_t$$



Αυτοπαλίνδρομο μοντέλο δεύτερης τάξης AR(2)

- Για να είναι η σειρά **στάσιμη** πρέπει να ισχύουν οι σχέσεις:

$$\phi_1 + \phi_2 < 1$$

$$-\phi_1 + \phi_2 < 1$$

$$-1 < \phi_2 < 1$$

Οι σχέσεις $\rho_1 = \phi_1 / (1 - \phi_2)$ και $\rho_2 = \phi_2 + \phi_1^2 / (1 - \phi_2)$ μας δίνουν την δυνατότητα να υπολογίζουμε τους συντελεστές αυτοσυσχέτισης από τους συντελεστές ϕ_i και αντίστροφα.

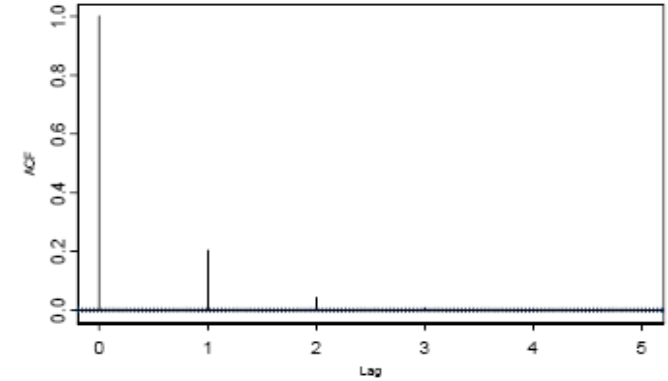
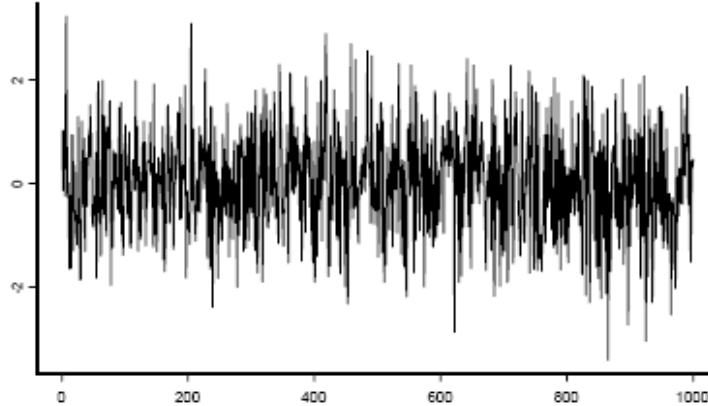
- **Η συνάρτηση αυτοσυσχέτισης AFC μιας AR(2) διαδικασίας τείνει προς το μηδέν καθώς αυξάνεται η υστέρηση s .**



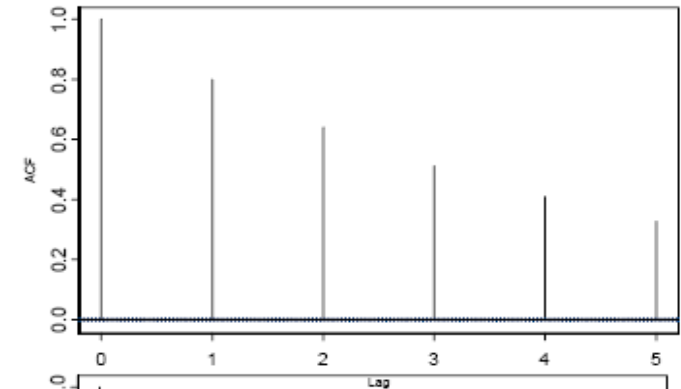
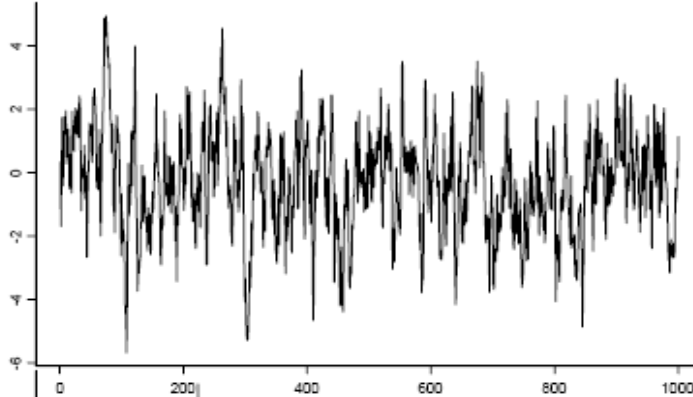
ΓΡΑΦΗΜΑΤΑ ΤΗΣ ACF ΓΙΑ ΔΙΑΦΟΡΕΣ ΤΙΜΕΣ ΤΗΣ Φ ΤΟ Φ ΕΙΝΑΙ Ο ΣΥΝΤΕΛΕΣΤΗΣ ΑΥΤΟΣΥΣΧΕΤΙΣΗΣ

AR(1) processes

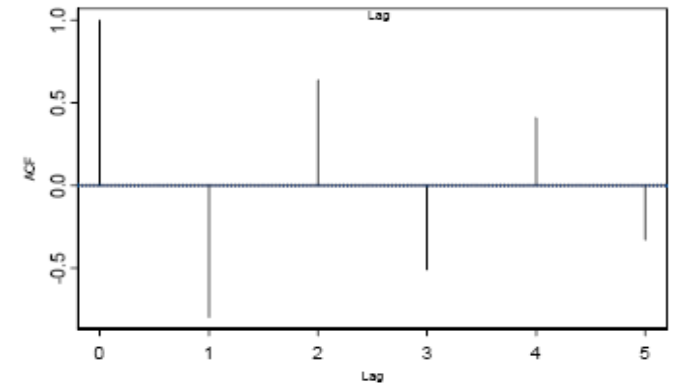
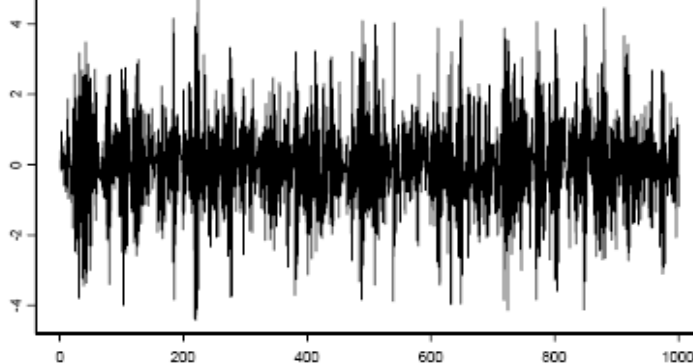
$$\phi_1 = 0.2$$



$$\phi_1 = 0.8$$



$$\phi_1 = -0.8$$



Αυτοπαλίνδρομα μοντέλα τάξης p ($AR(p)$)

Παράδειγμα 1: Να υπολογισθούν οι συναρτήσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης της χρονοσειράς $z_t = 0.7z_{t-1} + a_t$

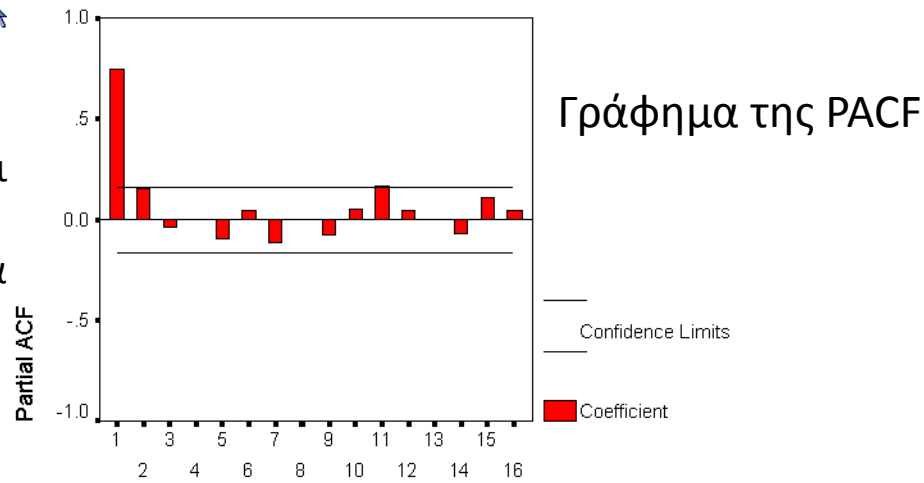
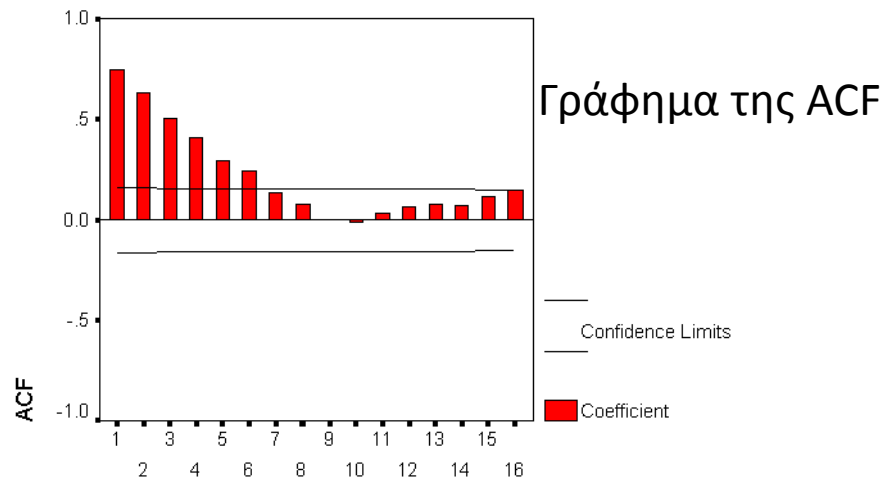
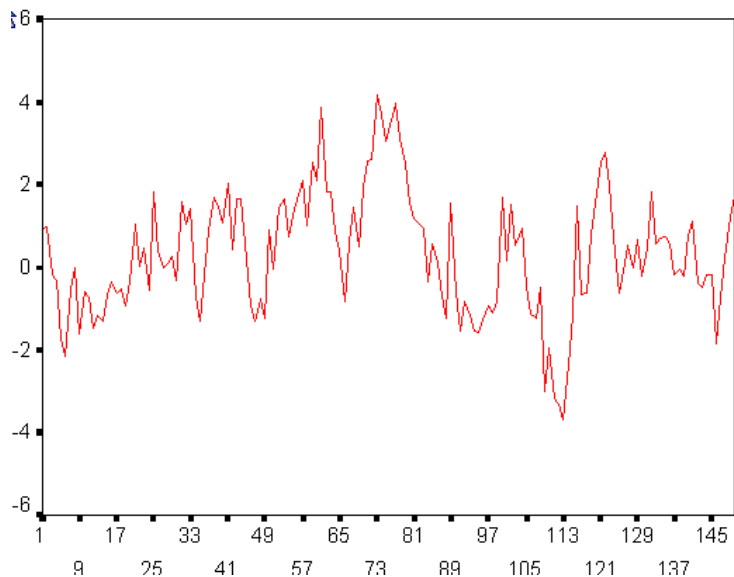
- Από τις εξισώσεις Yule-Walker για $p=1$ προκύπτει $\rho_k = \varphi_1 \rho_{k-1} \Rightarrow$

$$\begin{array}{lll} k=1 & \rho_1 = 0.7\rho_0 = 0.7 & \varphi_{11} = \rho_1 = 0.7 \\ k=2 & \rho_2 = 0.7\rho_1 = 0.7^2 & \varphi_{22} = 0 \\ \text{για} & \vdots & \vdots \\ k=m & \rho_m = 0.7^m & \varphi_{22} = 0 \text{ για } k \geq 2 \end{array}$$



Αυτοπαλίνδρομο μοντέλο πρώτης τάξης AR(1)

Παράδειγμα 2: Με προσομοίωση φτιάξαμε 150 παρατηρήσεις από την AR(1) χ.σ. $(1-0.8B)z_t = a_t$. Η σειρά έχει το παρακάτω γράφημα:



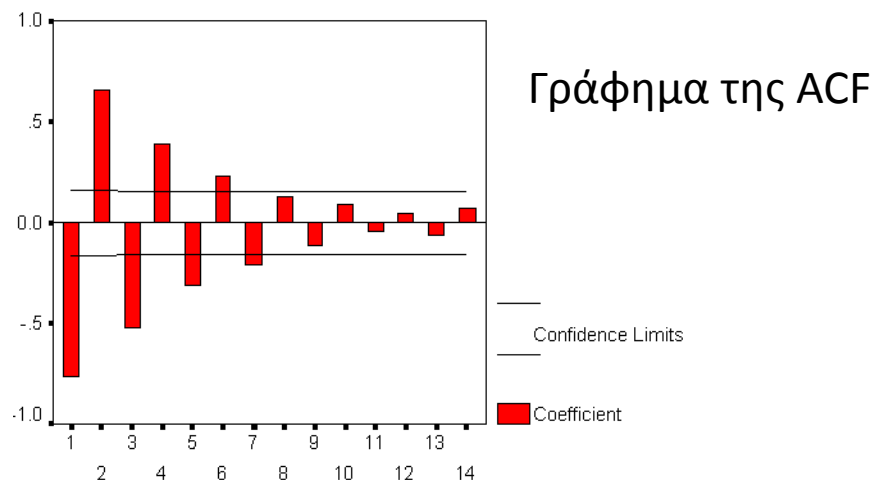
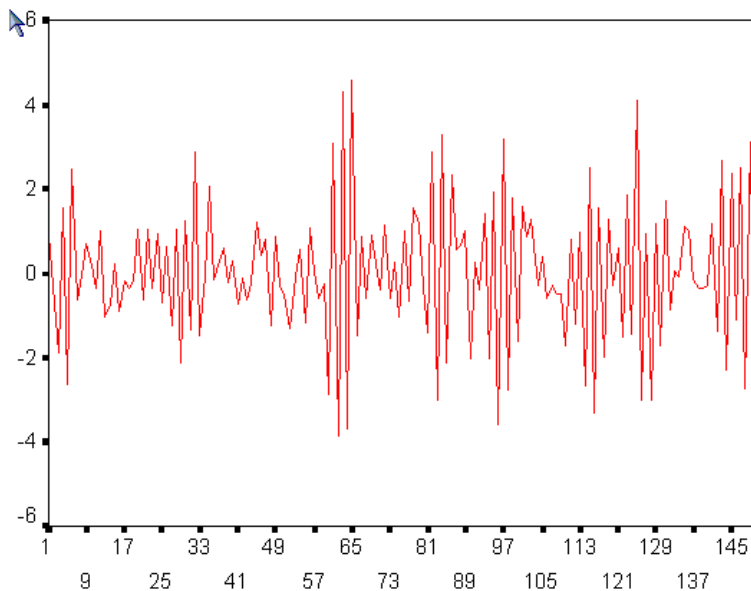
Παρατηρήσεις:

- Οι παρατηρήσεις της χρονοσειράς είναι συσχετισμένες
- Οι αυτοσυσχετίσεις ελαττώνονται εκθετικά ($\rho_k=0.8^k$)
- Η συνάρτηση μερικής αυτοσυσχέτισης υστέρησης $k=1$ είναι πολύ σημαντική, είναι θετική, ενώ οι υπόλοιπες είναι ασήμαντες



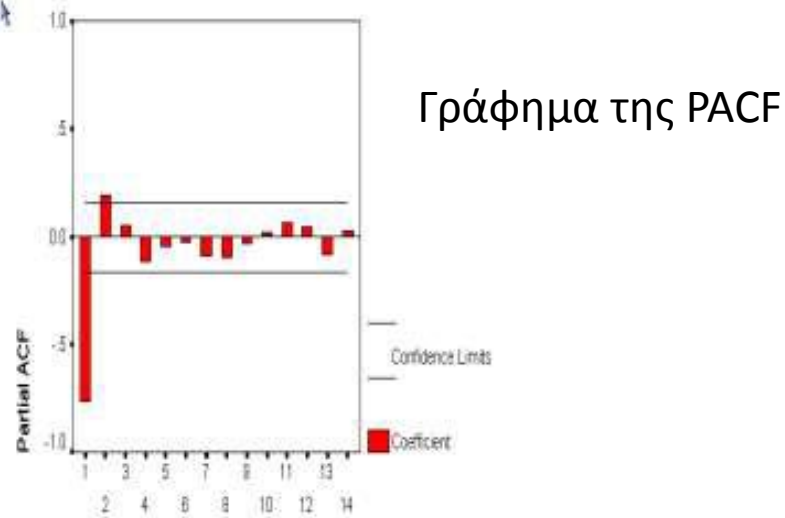
Αυτοπαλίνδρομο μοντέλο πρώτης τάξης AR(1)

Παράδειγμα 3: Με προσομοίωση φτιάξαμε 150 παρατηρήσεις από την AR(1) χ.σ. $(1+0.8B)z_t = a_t$. Η σειρά έχει το παρακάτω γράφημα:



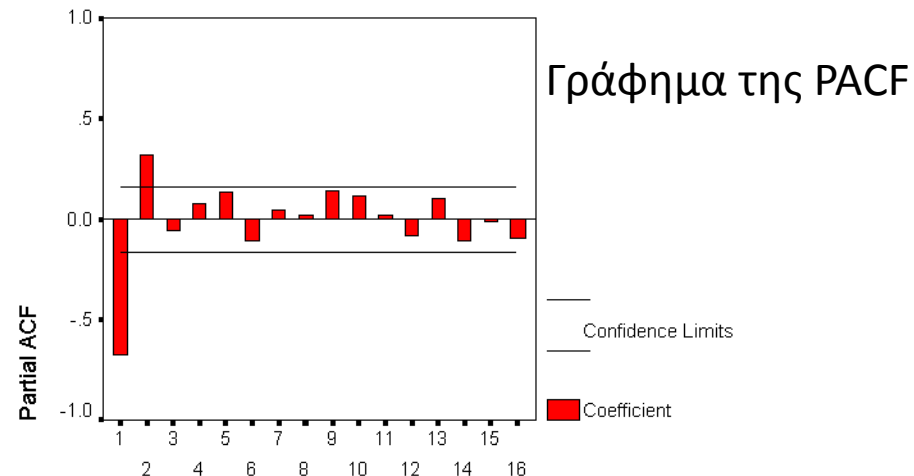
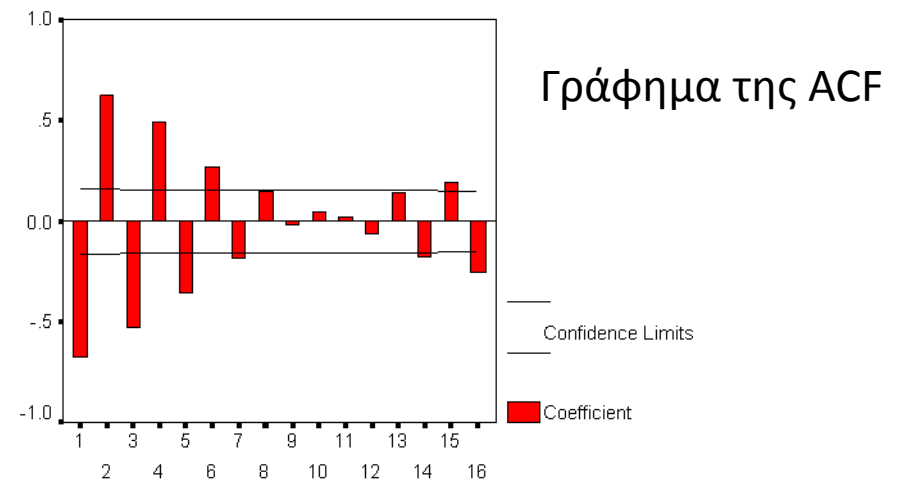
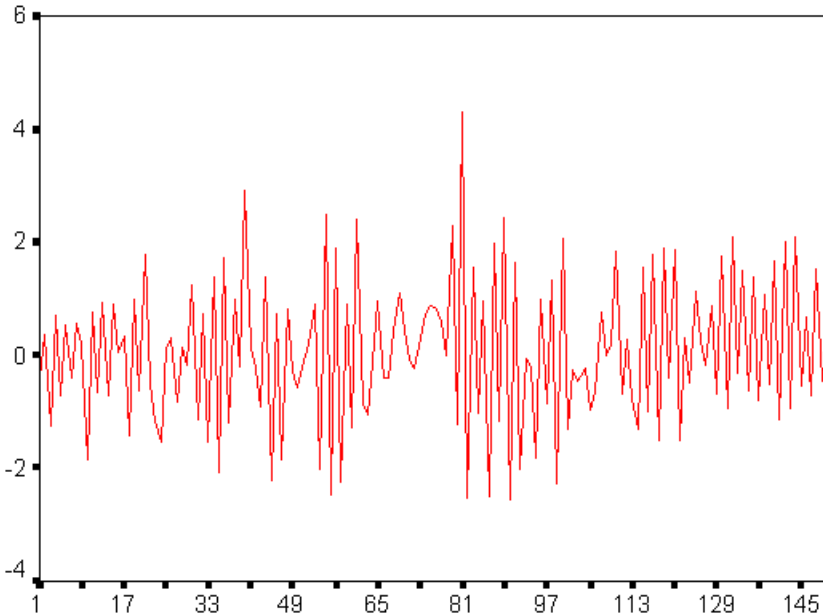
Παρατηρήσεις:

- Οι αυτοσυσχετίσεις εναλλάσσονται ως προς το πρόσημο και ελαττώνονται εκθετικά ($\rho_k = (-0.8)^k$)
- Από τις μερικές αυτοσυσχετίσεις πολύ σημαντική είναι η ϕ_{11} η οποία είναι αρνητική, ενώ η ϕ_{22} είναι οριακά σημαντική και οι υπόλοιπες ασήμαντες



Αυτοπαλίνδρομο μοντέλο δεύτερης τάξης AR(2)

Παράδειγμα 4: Με προσομοίωση φτιάξαμε 150 παρατηρήσεις από μια AR(2) χ.σ. $(1+0.5B-0.3B^2)z_t = a_t$. Η σειρά έχει το παρακάτω γράφημα:



Παρατηρήσεις:

- το correlogram της σειράς παρουσιάζει μείωση εκθετική ενώ τα πρόσημα των αυτοσυσχετίσεων πηγαίνουν εναλλάξ
- στο γράφημα των ϕ_{kk} σημαντικές είναι μόνο οι ϕ_{11} και ϕ_{22}



Κινούμενου μέσου μοντέλα τάξης q (MA(q))

$$z_t = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad \text{ή} \quad z_t = \Theta(B)a_t$$

$$\text{όπου } \Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \text{ (MA πολυώνυμο)}$$

$$Ea_t = 0 \quad \text{Var}(a_t) = \sigma_a^2$$

Παρατήρηση: η MA(q) χ.σ. ισοδυναμεί με ένα GLM πεπερασμένης τάξης

Είναι πάντα στάσιμη

Αποδεικνύεται ότι είναι αντιστρέψιμη, όταν οι ρίζες της $\Theta(\omega)=0$ βρίσκονται έξω από τον μοναδιαίο κύκλο.

Συνθήκες αντιστρεψιμότητας:

για $p=1$

$$-1 < \theta_1 < 1$$

$$\theta_2 - \theta_1 < 1$$

για $p=2$

$$\theta_1 + \theta_2 < 1$$

$$-1 < \theta_2 < 1$$



Κινούμενου μέσου μοντέλα τάξης q ($MA(q)$)

Εκτίμηση παραμέτρων

- Μέθοδος Durbin
- Μέθοδος ελαχίστων τετραγώνων
- Μέθοδος ελαχίστων τετραγώνων σε χώρο συχνοτήτων
- Βελτιωμένη μέθοδος ροπών



Κινούμενου μέσου μοντέλα τάξης q (MA(q))

Μέθοδος ροπών

$$\gamma_k = E z_{t-k} z_t = E(a_{t-k} - \theta_1 a_{t-k-1} - \dots - \theta_q a_{t-k-q})(a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q})$$

$$\gamma_k = \begin{cases} 0 & k > q \\ (-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \dots + \theta_{q-k} \theta_q) \sigma_a^2 & k = 1, 2, \dots, q \\ \gamma_{-k} & k < 0 \end{cases}$$

για $k=0$ $\sigma_z^2 = \gamma_0 = (1 + \theta_1^2 + \dots + \theta_q^2) \sigma_a^2$ οπότε

$$\rho_k = \begin{cases} 0 & k > q \\ \frac{-\theta_k + \theta_k \theta_{k+1} + \theta_2 \theta_{k+2} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & k = 1, 2, \dots, q \\ \rho_{-k} & k < 0 \end{cases} \quad (\text{A})$$



Κινούμενου μέσου μοντέλα τάξης q (MA(q))

Μέθοδος ροπών

Παρατηρήσεις:

- αυτοσυσχετίσεις διάφορες του μηδενός τόσες, όσο το q.
- για γνωστά ρ_k το σύστημα (A) μπορεί να λυθεί, ως προς $\vartheta_1, \vartheta_2, \dots, \vartheta_q$ με μεθόδους αριθμητικής ανάλυσης.

Παράδειγμα 5: Οι δύο χ.σ. $z_t=(1-0.4B)\alpha_t$ και $z_t=(1-2.5B)\alpha_t$ έχουν την ίδια ACF

$$\rho_k = \begin{cases} -1 & k = 1 \\ 2.9 & \\ 0 & k > 1 \end{cases}$$

Όμως μόνο η μία είναι αντιστρέψιμη και μάλιστα η $z_t=(1-0.4B)\alpha_t$ διότι μόνο η ρίζα του $1-0.4\omega = 0$ βρίσκεται έξω από το μοναδιαίο κύκλο.



Κινούμενου μέσου μοντέλα τάξης q (MA(q))

Παράδειγμα 6: Η συνάρτηση αυτοδιασποράς για μια MA(1) χ.σ.

$z_t = (1 - \theta_1 B)a_t$ είναι:

$$\gamma_k = \begin{cases} (1 + \theta_1^2)\sigma_a^2 & k = 0 \\ -\theta_1\sigma_a^2 & k = 1 \\ 0 & k > 1 \end{cases} \quad \text{και} \quad \rho_k = \begin{cases} -\theta_1 & k = 1 \\ 0 & k > 1 \end{cases}$$

Η PACF για την ίδια σειρά (επειδή $\rho_k = 0$ για $k > 1$) είναι:

$$\varphi_{11} = \rho_1 = \frac{-\theta_1}{1 + \theta_1^2} = \frac{-\theta_1(1 - \theta_1^2)}{1 - \theta_1^4}$$

$$\varphi_{22} = \frac{-\rho_1^2}{1 - \rho_1^2} = \frac{-\theta_1^2(1 - \theta_1^2)}{1 - \theta_1^6}$$

⋮

$$\varphi_{kk} = \frac{-\theta_1^k(1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}} \quad \text{για} \quad k > 1$$

Παρατηρήσεις: Σε μια MA(1) χ.σ.

- $\rho_1 \neq 0$, $\rho_k = 0$ για $k > 1$
- $2|\rho_k| < 1$ συνεπώς $|\rho_k| < 0.5$
- τα φ_{kk} μειώνονται εκθετικά, διατηρώντας αρνητικό πρόσημο, όταν $\theta_1 > 0$ ή με εναλλασσόμενο πρόσημο και μάλιστα ξεκινώντας από θετική τιμή, όταν $\theta_1 < 0$.



Κινούμενου μέσου μοντέλα τάξης q (MA(q))

Παράδειγμα 7: Δίνεται η MA(2) σειρά $z_t = (1 - \theta_1 B - \theta_2 B^2) a_t$. Να υπολογισθεί η ACF και η PACF. Από τις σχέσεις (A) προκύπτει:

$$\rho_k = \begin{cases} \frac{-\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2} & k = 1 \\ \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2} & k = 2 \\ 0 & k > 2 \end{cases} \quad \text{Οπότε} \quad \begin{aligned} \phi_{11} = \rho_1 &= \frac{-\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2} \\ \phi_{22} &= \frac{\rho_2 - \rho_1}{1 - \rho_1^2} = \dots \\ &\vdots \\ &\vdots \end{aligned}$$

Παρατηρήσεις: Σε μια MA(2) χ.σ.

- ρ_1 και ρ_2 σημαντικά και τα υπόλοιπα ασήμαντα
- η PACF ελαττώνεται εκθετικά ή έχει φθίνουσα ημιτονοειδή συμπεριφορά, ανάλογα με το πρόσημο και το μέγεθος των θ_1 και θ_2
 - ημιτονοειδή συμπεριφορά έχει όταν οι ρίζες της $(1 - \vartheta_1 B - \vartheta_1 B^2) = 0$ είναι μιγαδικές



Υποδείγματα κινητού μέσου όρου MA

Στη γενική του μορφή ένα υπόδειγμα **κινητού μέσου q τάξης ή MA(q)** είναι το ακόλουθο:

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

Η **τάξη q** αναφέρεται στο **μήκος της υστέρησης της μεταβλητής ε_t** για την οποία υποθέτουμε ότι είναι **λευκός θόρυβος** (η **ε_t** έχει μέση τιμή μηδέν, διακύμανση σταθερή και οι συντελεστές αυτοσυσχέτισης είναι μηδέν)

❖ Ο όρος «**κινητός μέσος**» αναφέρεται στο γεγονός ότι η Y_t εμφανίζεται ως ένα **σταθμισμένο άθροισμα των τιμών της ε_t**



Υπόδειγμα κινητού μέσου όρου πρώτης τάξης MA(1)

Η μορφή του υποδείγματος **κινητού μέσου πρώτης τάξης ή MA(1)** είναι η ακόλουθη:

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

ή

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

ή

$$Y_t - \mu = y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Αν χρησιμοποιήσουμε τον **τελεστή υστερήσεως (Lag operator) L** η ανωτέρω σχέση γίνεται:

$$y_t = (1 + \theta_1 L) \varepsilon_t$$



❖ Όλες οι αυτοσυνδιακυμάνσεις και συνεπώς οι αυτοσυσχετίσεις είναι μηδέν εκτός από την πρώτη.

Αυτό σημαίνει ότι η μνήμη της διαδικασίας **δεν ξεπερνά την μια περίοδο.**

[δηλαδή μια οποιαδήποτε παρατήρηση της Y έστω η Y_{11} συσχετίζεται με την Y_{10} ή την Y_9 και με καμιά άλλη παρατήρηση].

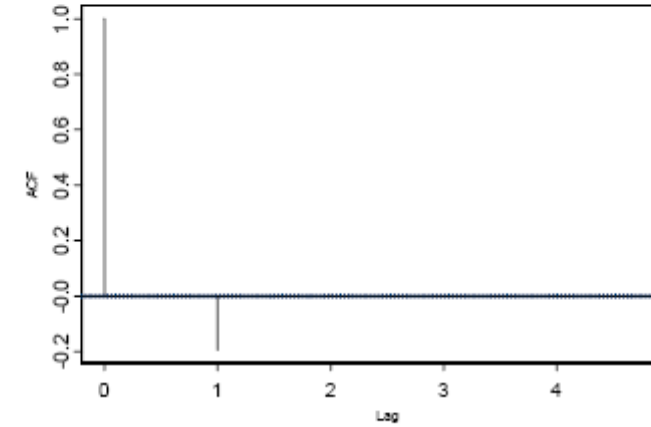
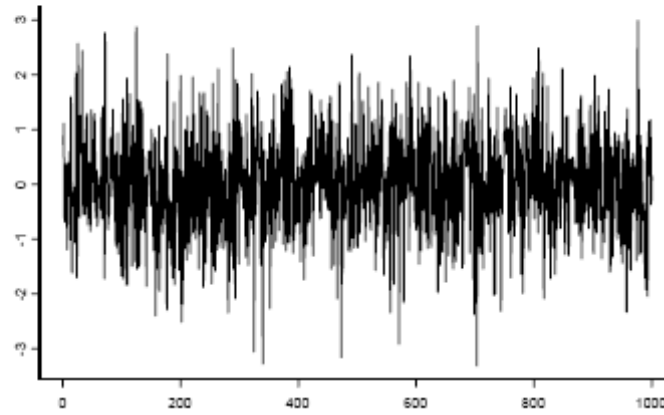
[ΣΤΗΝ ΕΠΟΜΕΝΗ ΔΙΑΦΑΝΕΙΑ ΠΑΡΟΥΣΙΑΖΟΝΤΑΙ ΔΙΑΓΡΑΜΜΑΤΑ ΤΗΣ ACF ΓΙΑ ΔΙΑΦΟΡΕΣ ΤΙΜΕΣ ΤΗΣ θ]



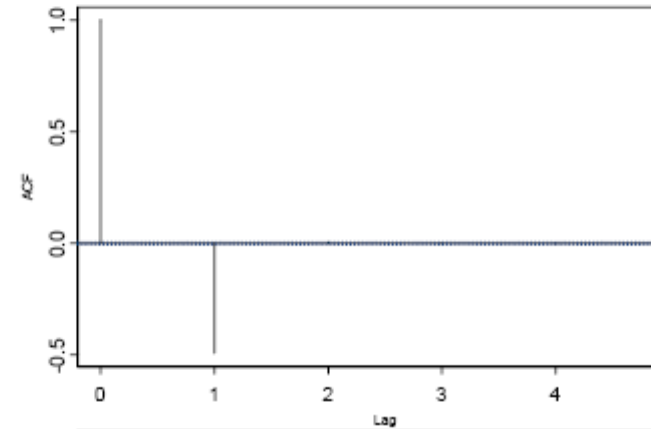
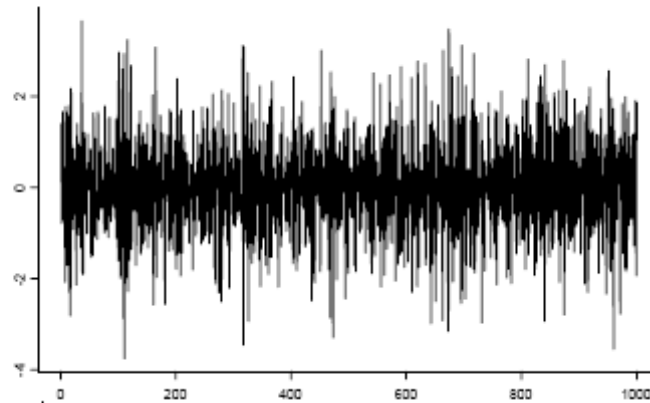
MA(1) processes

ΔΙΑΓΡΑΜΜΑΤΑ ΤΗΣ ACF ΓΙΑ ΔΙΑΦΟΡΕΣ ΤΙΜΕΣ ΤΗΣ θ

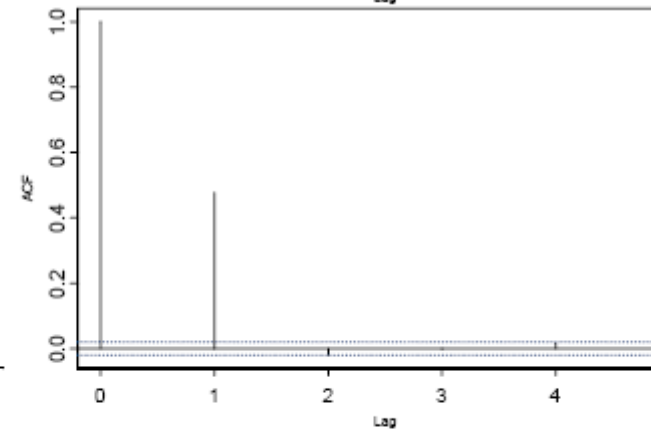
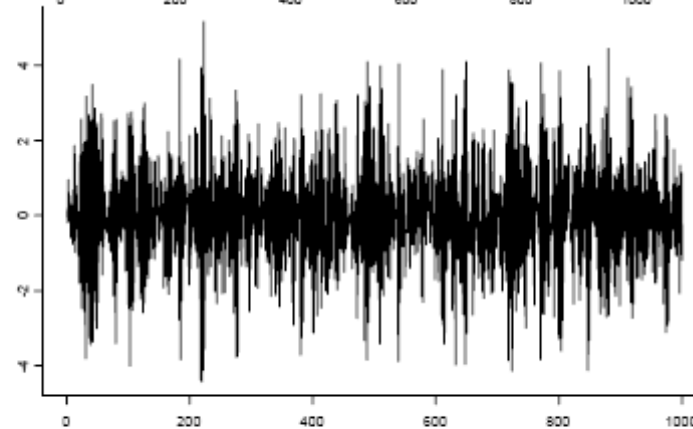
$$\theta_1 = 0.2$$



$$\theta_1 = 0.8$$

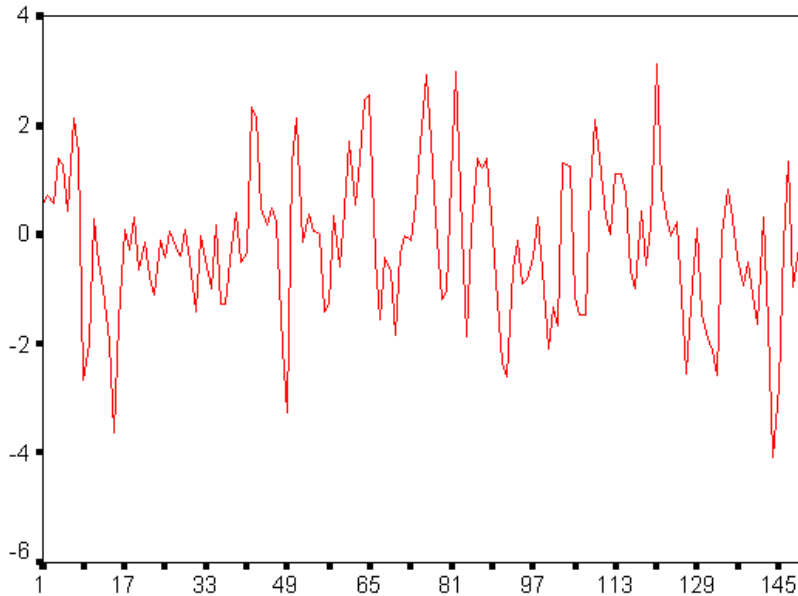


$$\theta_1 = -0.8$$



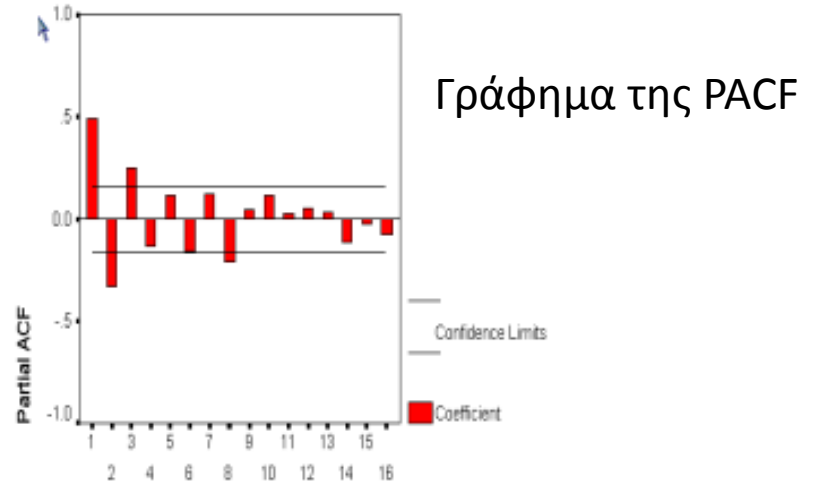
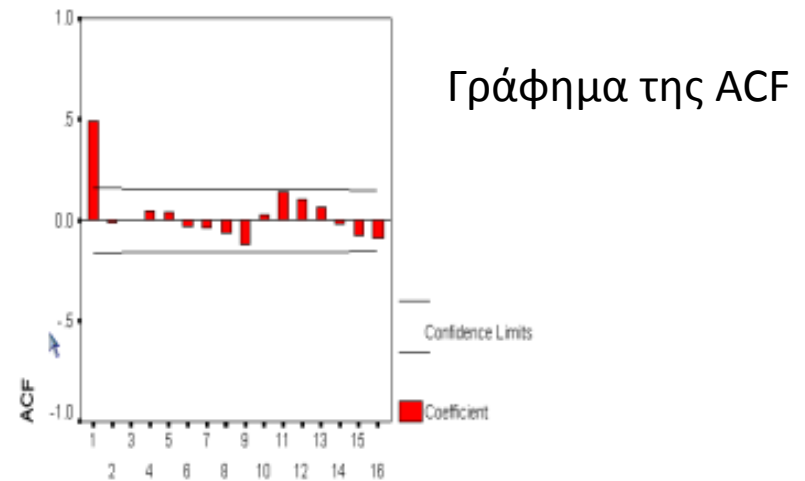
Κινούμενου μέσου μοντέλα τάξης q (MA(q))

Παράδειγμα 8: Με προσομοίωση φτιάξαμε 150 παρατηρήσεις από την MA(1) χ.σ. $z_t = (1 + 0.8B)a_t$. Η σειρά έχει το παρακάτω γράφημα:



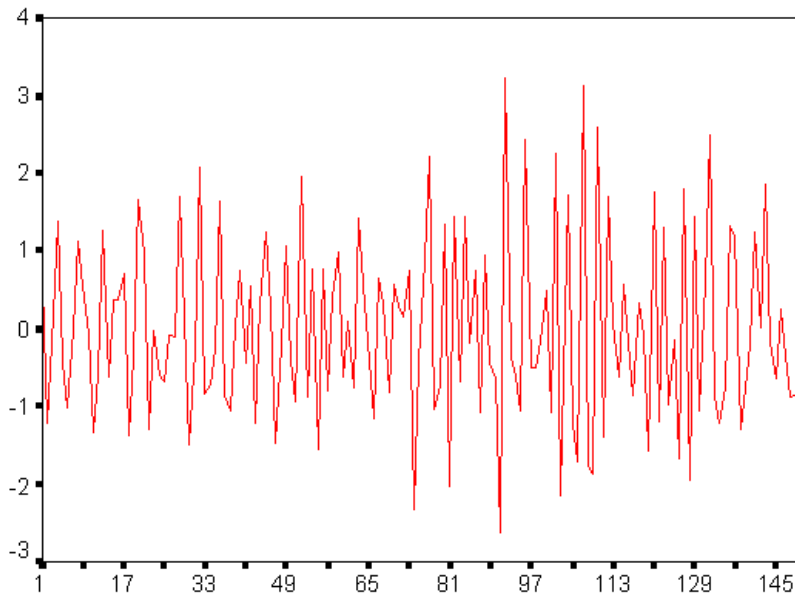
Παρατηρήσεις:

- ρ_1 σημαντική, ρ_k για $k > 1$ ασήμαντες
- φ_{kk} μειώνονται εκθετικά, με εναλλασσόμενο πρόσημο ($\theta_1 = -0.8 < 0$) ξεκινώντας από θετική τιμή



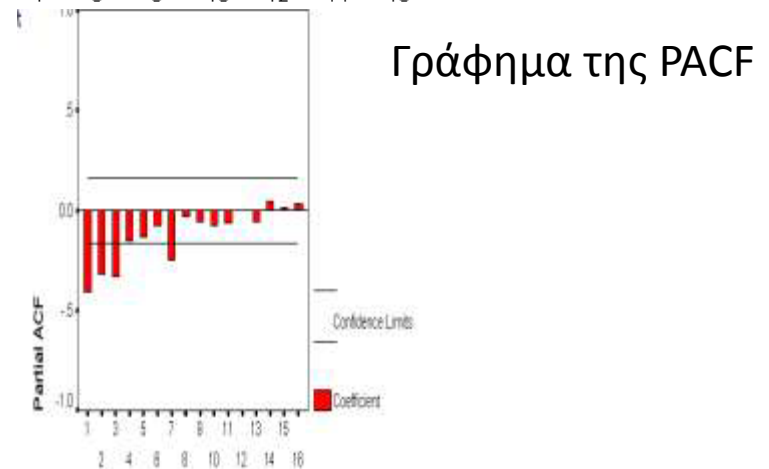
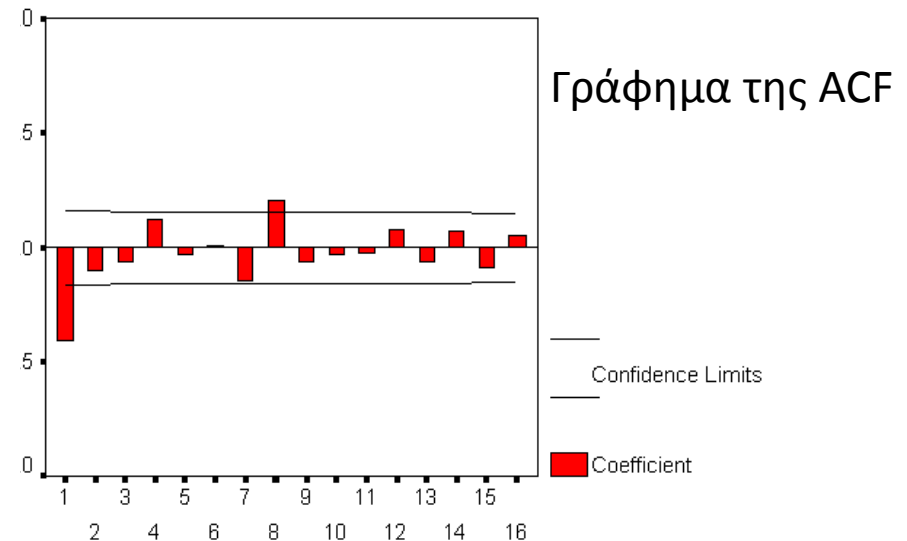
Κινούμενου μέσου μοντέλα τάξης q (MA(q))

Παράδειγμα 9: Με προσομοίωση φτιάξαμε 150 παρατηρήσεις από την MA(1) χ.σ. $z_t = (1 - 0.8B)a_t$. Η σειρά έχει το παρακάτω γράφημα:



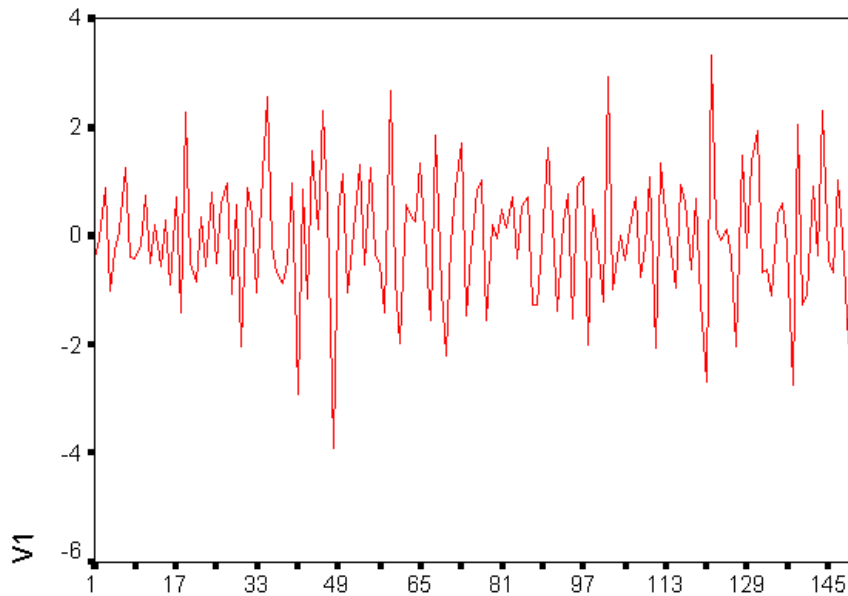
Παρατηρήσεις:

- ρ_1 σημαντική, $\rho_k=0$ για $k>1$
- φ_{kk} μειώνονται εκθετικά και έχουν το ίδιο πρόσημο ($\theta_1=0.8>0$)



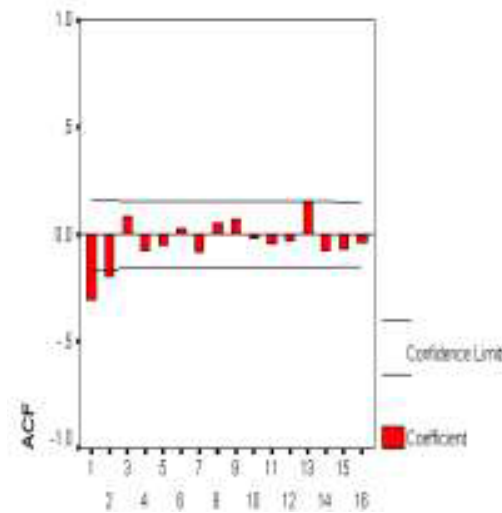
Κινούμενου μέσου μοντέλα τάξης q (MA(q))

Παράδειγμα 10: Με προσομοίωση φτιάξαμε 150 παρατηρήσεις από την MA(2) χ.σ. $z_t = (1 - 0.65B - 0.24B^2)a_t$. Η σειρά έχει το παρακάτω γράφημα:

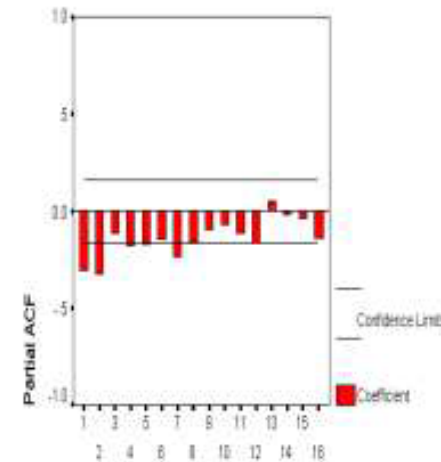


Παρατηρήσεις:

- ρ_1, ρ_2 σημαντικές, $\rho_k=0$ για $k>2$
- φ_{kk} μειώνονται εκθετικά



Γράφημα της ACF



Γράφημα της PACF



Σχέση μεταξύ $AR(p)$ και $MA(q)$ διαδικασίας

$$AR(p) \rightarrow \Phi_p(B)z_t = a_t \Rightarrow z_t = \frac{1}{\Phi_p(B)} a_t = \Psi(B)a_t \text{ δηλ. } MA(\infty)$$

$$\text{με } \Phi_p(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p \text{ και } \Psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

$$\text{Ισχύει } \Phi_p(B) \cdot \Psi(B) = 1$$

Παράδειγμα 11: Εστω μια $AR(1)$ χ.σ. $z_t = \varphi_1 z_{t-1} + a_t \Rightarrow (1 - \varphi_1 B)z_t = a_t \Rightarrow$

$$z_t = \frac{1}{1 - \varphi_1 B} a_t = (1 + \varphi_1 B + \varphi_1^2 B^2 + \dots) a_t \text{ δηλ. } MA(\infty)$$

Αντίστροφα

$$MA(q) \rightarrow z_t = \Theta_q(B)a_t \text{ με } \Theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

$$\pi(B) z_t = \frac{1}{\Theta_q(B)} z_t = a_t \text{ όπου } \pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$$

$$\text{άρα } MA(q) \sim AR(\infty)$$



Μικτό μοντέλο τάξης p,q (ARMA(p,q))

$$z_t = \varphi_1 z_{t-1} + \dots + \varphi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

$$\text{ή } \Phi(B) z_t = \Theta(B) a_t$$

$$\text{όπου } \Phi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p, \quad \Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

$$\text{και } E(a_t) = 0 \quad \text{Var}(a_t) = \sigma_a^2$$

Ένα $ARMA(p,q)$ μοντέλο, ισοδυναμεί με AR μοντέλο τάξης p , με σφάλματα που ακολουθούν $MA(q)$ σειρά, δηλαδή $\Phi(B) z_t = \varepsilon_t$ όπου $\varepsilon_t = \Theta(B) a_t$ ή ισοδυναμεί με $MA(q)$ χ. σ., με σφάλματα που ακολουθούν $AR(p)$ σειρά, δηλαδή $z_t = \Theta(B) b_t$ όπου $b_t = \Phi^{-1}(B) a_t$.

Είναι **στάσιμη** όταν οι ρίζες της $\Phi(\omega) = 0$ είναι έξω από το μοναδιαίο κύκλο και **αντιστρέψιμη** όταν οι ρίζες της $\Theta(\omega) = 0$ είναι έξω από το μοναδιαίο κύκλο



Συνάρτηση αυτοσυσχέτισης μιας ARMA(p,q) χρονοσειράς

$$\gamma_k = E(z_{t-k} z_t) = \varphi_1 \gamma_{k-1} + \dots + \varphi_p \gamma_{k-p} + E(z_{t-k} \alpha_t) - \theta_1 E(z_{t-k} \alpha_{t-1}) - \dots - \theta_q E(z_{t-k} \alpha_{t-q})$$

Επειδή $E(z_{t-k} \alpha_{t-i}) = 0$ για $k > i$ έχουμε :

$$\gamma_k = \varphi_1 \gamma_{k-1} + \dots + \varphi_p \gamma_{k-p} \quad k \geq (q + 1)$$

ή

$$\rho_k = \varphi_1 \rho_{k-1} + \dots + \varphi_p \rho_{k-p} \quad k \geq (q + 1)$$

- η παρατήρηση της ACF βοηθά στον καθορισμό του μοντέλου.
- Η ACF ενός ARMA(p,q) μοντέλου, μετά την υστέρηση k , συμπεριφέρεται όμοια με τη συνάρτηση αυτοσυσχέτισης μιας AR(p) χ.σ.
- Οι πρώτες q αυτοσυσχετίσεις, $\rho_q, \rho_{q-1}, \dots, \rho_1$, εξαρτώνται από τις αυτοπαλινδρομούμενες και από τις κινούμενου μέσου παραμέτρους.



Μικτό μοντέλο τάξης p, q ($ARMA(p, q)$)

Παράδειγμα 12: Έστω το $ARMA(1, 1)$ μοντέλο

$$(1 - \varphi_1 B)z_t = (1 - \theta_1 B)\alpha_t$$

Θα πρέπει $|\varphi_1| < 1$ για να είναι στάσιμο και $|\theta_1| < 1$ για να είναι αντιστρέψιμο.

Έχει συνάρτηση αυτοδιασποράς

$$\gamma_k = E(z_{t-k} z_t) = \varphi_1 \gamma_{k-1} + E(z_{t-k} \alpha_t) - \theta_1 E(z_{t-k} \alpha_{t-1})$$

- για $k=0$ $\gamma_0 = \varphi_1 \gamma_1 + E(z_t \alpha_t) - \theta_1 E(z_t \alpha_{t-1}) = \varphi_1 \gamma_1 + \sigma_a^2 - \theta_1 (\varphi_1 - \theta_1) \sigma_a^2$

$$(E z_t \alpha_t = \sigma_a^2, E(z_t \alpha_{t-1}) = \varphi_1 E(z_{t-1} \alpha_{t-1}) + E(\alpha_t \alpha_{t-1}) - \theta_1 E(\alpha_t^2) = (\varphi_1 - \theta_1) \sigma_a^2)$$

- για $k=1$ $\gamma_1 = \varphi_1 \gamma_0 - \theta_1 \sigma_a^2 = \frac{(\varphi_1 - \theta_1)(1 - \varphi_1 \theta_1)}{(1 - \varphi_1^2)} \sigma_a^2$

- για $k \geq 2$ $\gamma_k = \varphi_1 \gamma_{k-1}$



Υπόδειγμα ARMA (p, q)

Το μοντέλο - υπόδειγμα ARMA (p,q):

Είναι συνδυασμός **p αυτοπαλίνδρομων όρων** και **q όρων κινητού μέσου** και γι' αυτό αποκαλείται **μεικτό αυτοπαλίνδρομο – κινητού μέσου υπόδειγμα τάξης (p, q)**, (Mixed Autoregressive – Moving Average Model) ή ARMA (p, q).

- Κάθε διαδικασία AR(p) είναι διαδικασία ARMA (p,0) και κάθε διαδικασία MA (q) είναι ARMA (0, q).



Υπόδειγμα ARMA (1,1)

Το υπόδειγμα **ARMA (1,1)** έχει τη μορφή:

$$Y_t = \delta + \alpha Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

ή

$$y_t = \alpha y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Για να είναι **στάσιμη η σειρά** πρέπει να είναι

$$|\alpha| < 1$$

➤ Η **συνάρτηση μερικής αυτοσυσχέτισης** συμπεριφέρεται όπως στην περίπτωση **της AM (1) διαδικασίας**, δηλαδή **φθίνει γεωμετρικά**.



ΓΕΝΙΚΑ:

- Η συνάρτηση αυτοσυσχέτισης μιας $ARMA(p, q)$ διαδικασίας συμπεριφέρεται όπως αυτή μιας $AR(p)$ διαδικασίας,
- Η συνάρτηση μερικής αυτοσυσχέτισης μιας $ARMA(p, q)$ διαδικασίας συμπεριφέρεται όπως αυτή μιας $MA(q)$ διαδικασίας για $s > q - p$.

(αυτά συνοψίζονται στον παρακάτω πίνακα)



Πίνακας: Μορφές της συνάρτησης αυτοσυσχέτισης και μερικής αυτοσυσχέτισης

Διαδικασία	Αυτοσυσχετίσεις ACF	Μερικές αυτοσυσχετίσεις PACF
AR(1)	Φθίνουν γεωμετρικά από ρ_1	Μηδέν μετά το ρ_{11}
AR(2)	Φθίνουν γεωμετρικά από ρ_2	Μηδέν μετά το ρ_{22}
AR (p)	Φθίνουν γεωμετρικά ή φθίνουν με ημιτονοειδή συμπεριφορά από ρ_p	Μηδέν μετά το ρ_{pp}



MA(1)	Μηδέν μετά το ρ_1	Φθίνου γεωμετρικά μετά από ρ_{11}
MA(2)	Μηδέν μετά το ρ_2	Φθίνου γεωμετρικά μετά από ρ_{22}
MA(q)	Μηδέν μετά το ρ_q	Φθίνου γεωμετρικά μετά από ρ_{qq}
ARMA (1,1)	Φθίνου γεωμετρικά από ρ_1	Φθίνου γεωμετρικά μετά από ρ_{11}
ARMA (p,q)	Φθίνου γεωμετρικά από ρ_q	Φθίνου γεωμετρικά μετά από ρ_{pp}



Κριτήρια επιλογής τάξης μοντέλου

- Κριτήριο Akaike AIC

$$AIC(k) = N \log \hat{\sigma}_a^2 + k$$

- N πλήθος παρατηρήσεων
- σ_a^2 η διασπορά των σφαλμάτων του τάξης t μοντέλου
- $k=t+1$ όπου t , το πλήθος των παραμέτρων του μοντέλου, συν ένα, αν έχει εκτιμηθεί και η μέση τιμή του μοντέλου.

$$AIC^*(k) = -2 \ln \hat{L} + 2k$$

- k = πλήθος παραμέτρων
- $\ln \hat{L} = -\frac{N}{2} \ln \hat{\sigma}_a^2 - \frac{N}{2} (1 + \ln 2\pi)$



Κριτήρια επιλογής τάξης μοντέλου

- **Κριτήριο Shibata BIC**

$$BIC(k) = n \ln \hat{\sigma}_a^2 - (N - k) \ln \left(1 - \frac{k}{N}\right) + k \ln N + k \ln \left[\frac{\hat{\sigma}_z^2}{\hat{\sigma}_a^2} - \frac{1}{k} \right]$$

- k αριθμός παραμέτρων
- $\hat{\sigma}_z^2$ η διασπορά της σειράς
- $\hat{\sigma}_a^2$ η εκτιμώμενη βάσει του μοντέλου διασπορά των σφαλμάτων

- **Κριτήριο Schwartz SBC**

$$SBC(k) = N \ln \hat{\sigma}_a^2 + k \ln N$$



Κριτήρια επιλογής τάξης μοντέλου

- Κριτήριο Parzen CAT

$$CAT(p) = \begin{cases} -(1 + \frac{1}{N}) & p = 0 \\ \frac{1}{N} \sum_{j=1}^p \frac{1}{\hat{\sigma}_j^2} - \frac{1}{\hat{\sigma}_p^2} & p = 1, 2, \dots \end{cases}$$

- $\hat{\sigma}_j^2$ εκτιμητής του $\hat{\sigma}_a^2$, όταν ένα $AR(j)$ μοντέλο προσαρμόσται στα δεδομένα.

Το καλύτερο μοντέλο, σε όλες τις παραπάνω περιπτώσεις, είναι εκείνο που αντιστοιχεί στη μικρότερη τιμή του κριτηρίου.



Έλεγχος του μοντέλου

- Ελέγχεται εάν τα σφάλματα ακολουθούν κανονική κατανομή, με οποιοδήποτε **goodness of fit test** (Kolmogorov-Smirnov, χ^2 -test, κλπ.)
- Ελέγχεται εάν οι **ACF** και οι **PACF** των σφαλμάτων που προκύπτουν από την προσαρμογή του μοντέλου είναι ασήμαντες για οποιοδήποτε k .
- Η υπόθεση $H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$ ελέγχεται με το στατιστικό

$$Q = N \sum_{s=1}^k \hat{r}_s^2(\hat{a}_t) \quad \text{ή} \quad \text{το} \quad Q^* = N(N+2) \sum_{s=1}^k (N-s)^{-1} \hat{\sigma}_s^2$$

- το οποίο ακολουθεί χ^2 κατανομή με $k-m$ βαθμούς ελευθερίας όπου m το πλήθος των παραμέτρων που εκτιμήθηκαν



Κριτήρια επιλογής τάξης μοντέλου

Παράδειγμα 13: Προσαρμογή διάφορων μοντέλων σε κάποια δεδομένα και οι τιμές του κριτηρίου του Akaike που πήραμε είναι οι παρακάτω:

p	q	0	1	2	3	4
0		142.1200	93.6185	70.0174	57.8203	55.3785
1		94.7730	69.9294	62.2191	56.4811	58.4985
2		31.9081	23.7781	41.5118	49.7238	48.1051
3		24.0529	25.2286	43.7349	31.1166	47.2080
4		25.6708	27.4769	88.6398	60.0407	75.5548

Από τον πίνακα φαίνεται ότι το ARMA(2,1) μοντέλο είναι το καλύτερο για τα δεδομένα μας.



Έλεγχος του μοντέλου

Παράδειγμα 14: Έστω ότι προσαρμόσαμε σε κάποια δεδομένα μια $AR(2)$ χ.σ. Εάν $k=12$, το στατιστικό $Q^*=18.0$ ακολουθεί χ^2 κατανομή με 10 βαθμούς ελευθερίας.

Όμως $\chi^2_{10;0.05}=18.307 > Q^*=18.0 \rightarrow$ η H_0 απορρίπτεται οριακά.

Με $H_0 : \rho_1 = \rho_2 = \dots = \rho_{12} = 0$ και

$H_1 : \rho_i \neq 0$ για κάποιο $i = 1, 2, \dots, 12$

όπου ρ_1, ρ_2, \dots οι αυτοσυσχετίσεις των σφαλμάτων που προέκυψαν από την προσαρμογή του $AR(2)$ μοντέλου.

- Δηλαδή το $AR(2)$ μοντέλο δεν είναι ικανοποιητικό και θα πρέπει να βελτιωθεί.



Γραμμικά μη στάσιμα μοντέλα (Μέθοδος Box&Jenkins)

- Εάν η **χρονοσειρά δεν είναι στάσιμη**, τα στοχαστικά μοντέλα που αναφέρθηκαν, δεν μπορούν να χρησιμοποιηθούν. Με μετασχηματισμούς που αναφέρονται στη «**Θ.Ε. 3.1: Ανάλυση χρονοσειρών**», επιτυγχάνεται η στασιμότητα της σειράς και προσαρμόζεται κατόπιν κάποιο από τα μοντέλα **$AR(p)$, $MA(q)$ ή $ARMA(p,q)$** .
- Οι Box και Jenkins προτείνουν κάποια μοντέλα τα οποία συγχρόνως πετυχαίνουν στασιμότητα και προσαρμογή. Αυτά τα μοντέλα είναι:
 - **Ολοκληρωμένα μικτά μοντέλα $ARIMA(p,d,q)$**
 - **Εποχικά μοντέλα**
 - **Το πολλαπλασιαστικό μοντέλο $ARIMA(p,d,q) \times (P,D,Q)_s$**



Υπόδειγμα ARIMA (p, d, q)

Μέχρι τώρα ασχοληθήκαμε με στάσιμες στοχαστικές διαδικασίες πράγμα που σημαίνει ότι **ο μέσος, η διακύμανση και οι αυτοδιακυμάνσεις** δεν εξαρτώνται από το χρόνο.

Μια απλή μορφή **αυτοπαλίνδρομης διαδικασίας πρώτης τάξης μη στάσιμης** είναι:

$$Y_t = Y_{t-1} + \varepsilon_t$$

όπου η μεταβλητή ε_t είναι **λευκός θόρυβος**.

❖ Η ανωτέρω μορφή είναι γνωστή ως **τυχαίος περίπατος ή τυχαία διαδρομή (Random walk)**.



Υπόδειγμα ARIMA (p, d, q)

- ❖ Όταν υπάρχει σταθερός όρος είναι γνωστή ως **τυχαία διαδρομή με περιπλάνηση (Random walk with drift)**.

$$Y_t = \alpha + Y_{t-1} + \varepsilon_t$$

- Μια στοχαστική διαδικασία που ακολουθεί την τυχαία διαδρομή δεν είναι στάσιμη ως προς τη διακύμανση.

Αν υποθέσουμε ότι ε_t ακολουθεί κατανομή (μ, σ^2) η στοχαστική διαδικασία που ακολουθεί την τυχαία διαδρομή δεν είναι στάσιμη ως προς τη διακύμανση και το μέσο.

- ❖ Αν πάρουμε τις **πρώτες διαφορές της Y_t** η **στοχαστική διαδικασία** που ακολουθεί την **τυχαία διαδρομή είναι στάσιμη**.



Υπόδειγμα ARIMA (p, d, q)

Οι περισσότερες από τις χρονολογικές σειρές (οικονομικού ή φυσικού περιεχομένου) δεν έχουν χαρακτηριστικά στάσιμων διαδικασιών. Μπορούν όμως να μετατραπούν σε στάσιμες παίρνοντας τις πρώτες ή τις δεύτερες κ.λ.π διαφορές.

Αυτό είναι αναγκαίο για να αποφύγουμε το πρόβλημα της **φαινομενικής ή νόθου παλινδόμησης (spurious regression)**.

❖ Όταν μια σειρά μετατρέπεται σε στάσιμη παίρνοντας τις **πρώτες διαφορές** λέμε ότι η σειρά είναι **ολοκληρωμένη πρώτης τάξης (Integrated First Order)** και συμβολίζεται με **I(1)**.

❖ Αν η σειρά μετατρέπεται σε στάσιμη παίρνοντας τις **δεύτερες διαφορές** λέμε ότι η σειρά είναι **ολοκληρωμένη δεύτερης τάξης** και συμβολίζεται με **I(2)**.



Υπόδειγμα ARIMA (p, d, q)

- ❖ Αν η σειρά μετατρέπεται σε στάσιμη παίρνοντας τις **d διαφορές** λέμε ότι η σειρά είναι **ολοκληρωμένη d τάξης** και συμβολίζεται με **I(d)**.
- ❖ Κάθε **στάσιμη σειρά** θεωρείται **ολοκληρωμένη μηδενικής τάξης** δηλαδή **I(0)**.

Αν ορισθεί ο τελεστής των διαφορών Δ ως εξής:

$\Delta = 1-L$ ο τελεστής των πρώτων διαφορών

$\Delta^2 = (1-L)^2$ ο τελεστής των δεύτερων διαφορών

.....

$\Delta^d = (1-L)^d$ ο τελεστής των d διαφορών



Υπόδειγμα ARIMA (p, d, q)

Τότε η μορφή της **Αυτοπαλίνδρομης διαδικασίας πρώτης τάξης**

$$Y_t = Y_{t-1} + \varepsilon_t$$

Γίνεται:

$$Y_t - Y_{t-1} = (1 - L)Y_t = \Delta Y_t = \varepsilon_t$$

ή

$$\Delta Y_t = \varepsilon_t$$



Υπόδειγμα ARIMA (p, d, q)

Ένα υπόδειγμα ARMA(p, q) που εφαρμόζεται σε μια **ολοκληρωμένη σειρά d τάξης** ονομάζεται **αυτοπαλίνδρομο ολοκληρωμένο υπόδειγμα κινητού μέσου τάξης(p,d,q) (Autoregressive Integrated Moving Average)** και **συμβολίζεται ως ARIMA (p,d,q).**

(Όπου p συμβολίζει τους όρους του αυτοπαλίνδρομου υποδείγματος, q τους όρους του υποδείγματος των κινητών μέσων).

Για παράδειγμα **ARIMA (2,1,1)** σημαίνει ότι η σειρά καθίσταται **στάσιμη** με τις **πρώτες διαφορές** και στην προκύπτουσα σειρά των πρώτων διαφορών εφαρμόζεται το **ARMA (2,1).**

❖ Η **τάξη του ARIMA (p,d,q)** προσδιορίζεται με τη **μεθοδολογία των Box – Jenkins**



Ολοκληρωμένα μικτά μοντέλα $ARIMA(p,d,q)$

Μια χρονική σειρά $\{z_t\}$ προσαρμόζεται σε ένα $ARIMA(p,d,q)$ μοντέλο, αν οι διαφορές d τάξης ακολουθούν $ARMA(p,q)$ μοντέλο.

- Προσφέρεται για χρονικές σειρές που παρουσιάζουν τάση.

Έκφραση: $\Phi(B)(1-B)^d z_t = \Theta(B)a_t$ όπου

$$\Phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$



Τυχαίος περίπατος

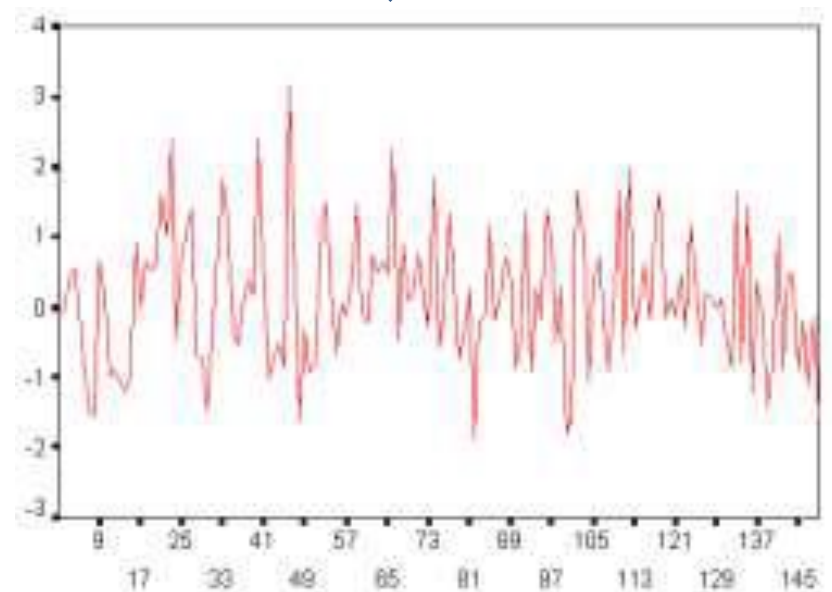
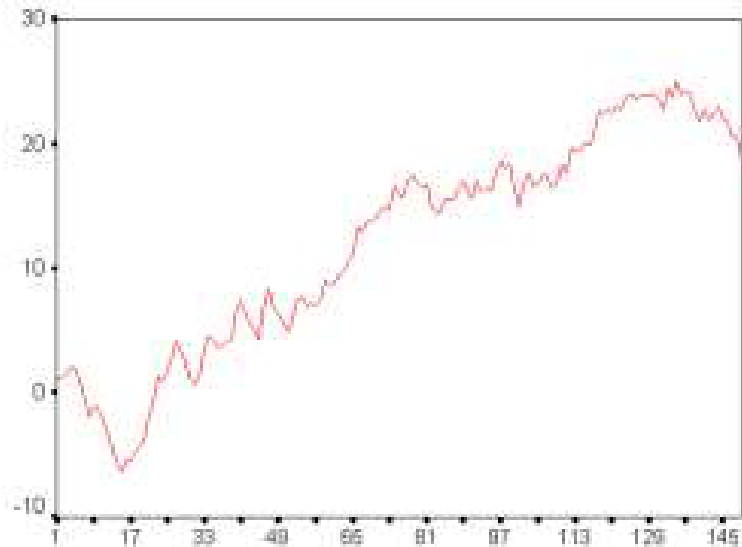
Έκφραση:

$$z_t = z_{t-1} + \alpha_t \quad \text{ή,}$$

$$(1-B)z_t = \alpha_t \quad \text{με } z_1 = \alpha_1 \quad \text{και}$$

α_t = λευκός θόρυβος (white noise)

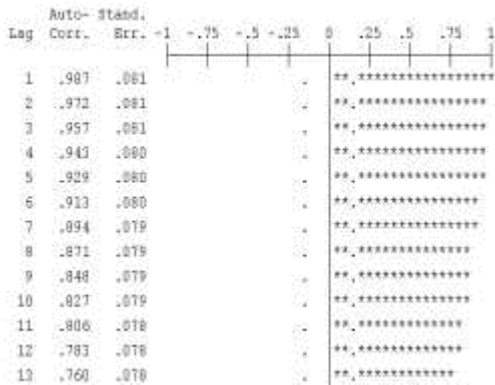
Τα γραφήματα μιας σειράς 150 παρατηρήσεων που φτιάσαμε με προσομοίωση από **μοντέλο τυχαίου περιπάτου** και των **πρώτων διαφορών αυτής** είναι:



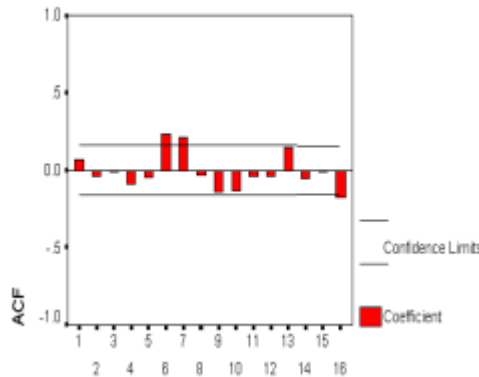
Τυχαίος περίπατος

Τα γραφήματα των ACF και PACF είναι:

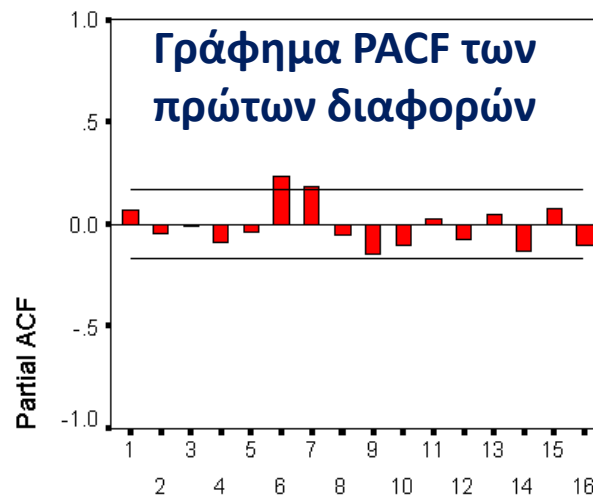
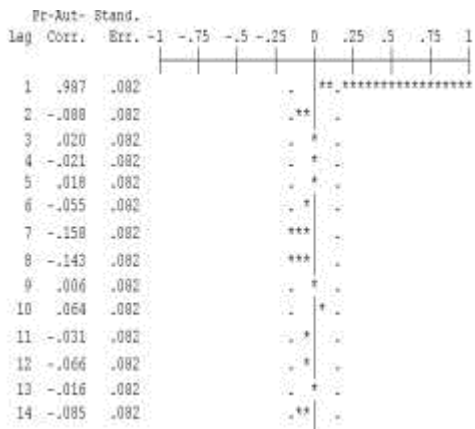
Γράφημα ACF της σειράς



Γράφημα ACF των πρώτων διαφορών



Γράφημα PACF της σειράς



Παρατηρήσεις:

- Εάν θεωρηθεί $AR(1)$ με $\phi_1=1$, επειδή $\rho_k = \phi_1^k$ τα ρ_k είναι πολύ σημαντικά όπως φαίνεται στο πάνω αριστερό γράφημα ACF της σειράς.
- Παρατηρούμε ότι τα ρ_k και τα ϕ_{kk} των πρώτων διαφορών είναι όλα σχεδόν ασήμαντα

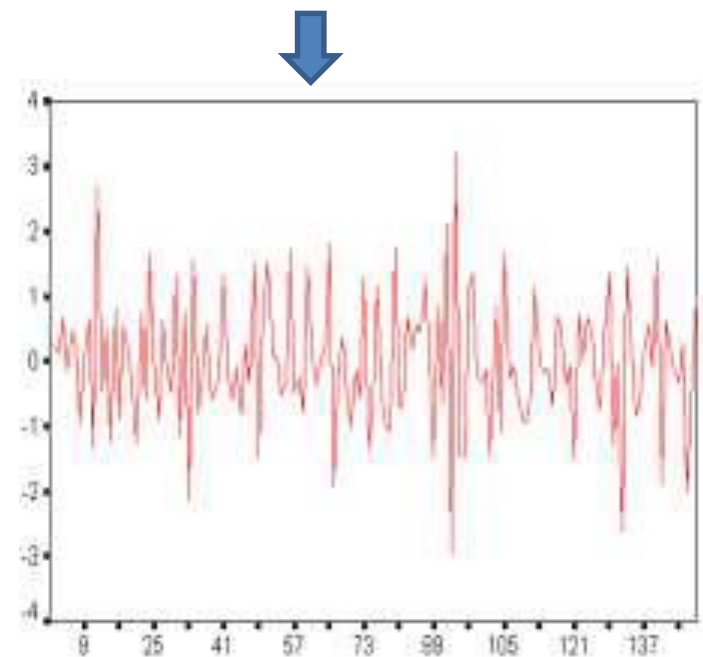
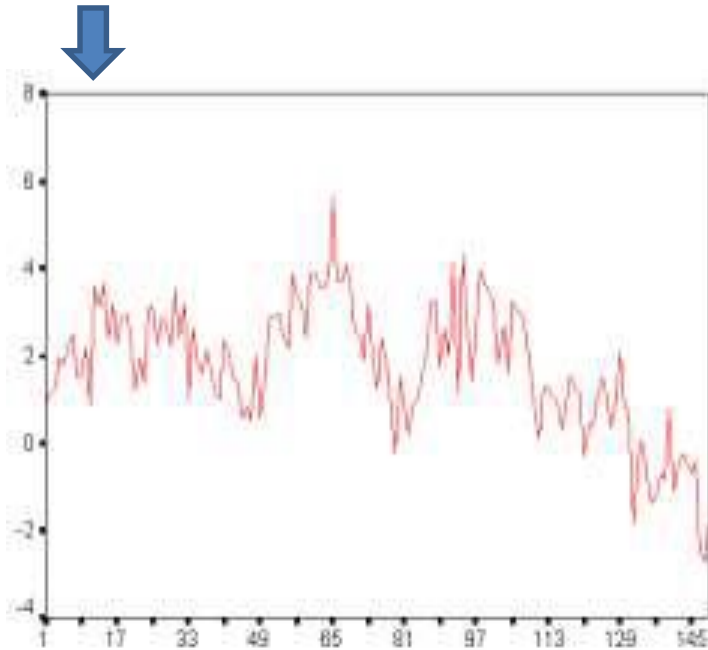


ARIMA(0,1,1) ή IMA(1,1) μοντέλο

Έκφραση: $(1-B)z_t = (1-\theta_1 B)a_t$ ή

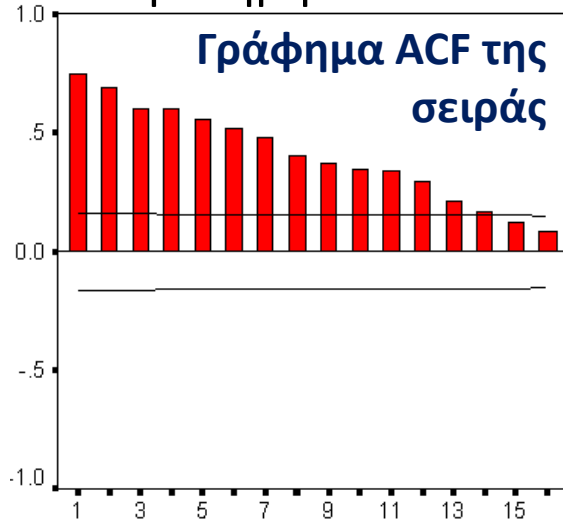
$$z_t = z_{t-1} + a_t - \theta_1 a_{t-1} \quad \text{με} \quad -1 < \theta_1 < 1$$

Τα γραφήματα μιας σειράς 150 παρατηρήσεων που φτιάχτηκαν με προσομοίωση από **IMA(1,1)** μοντέλο με $\theta_1=0.5$ και των **πρώτων διαφορών της σειράς** είναι:



ARIMA(0,1,1) ή IMA(1,1) μοντέλο

Τα γραφήματα των ACF και PACF των παρατηρήσεων και των πρώτων διαφορών των παρατηρήσεων είναι:



Παρατηρήσεις:

- Τα r_k των αρχικών παρατηρήσεων παραμένουν σημαντικά και για μεγάλα lags που σημαίνει ότι η χρονοσειρά παρουσιάζει τάση
- Το γράφημα των πρώτων διαφορών των παρατηρήσεων δεν παρουσιάζει τάση
- Το γράφημα των r_k των πρώτων διαφορών παρουσιάζεται “όμοιο” με μιας $MA(1)$ σειράς με $\theta_1 > 0$
- Στο γράφημα των φ_{kk} των πρώτων διαφορών, το φ_{11} είναι σημαντικό ενώ τα υπόλοιπα φ_{kk} για $k > 2$ είναι ασήμαντα



ARIMA(1,1,0) ή ARI(1,1) μοντέλο

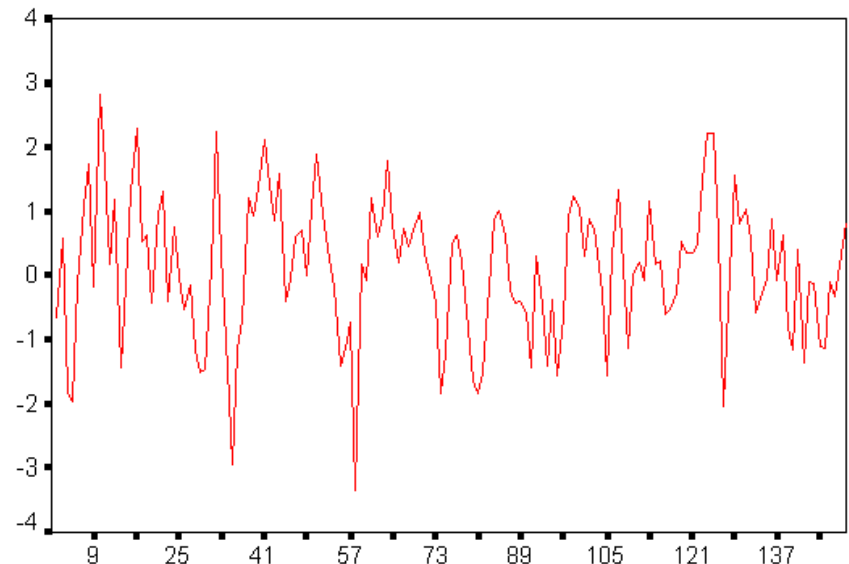
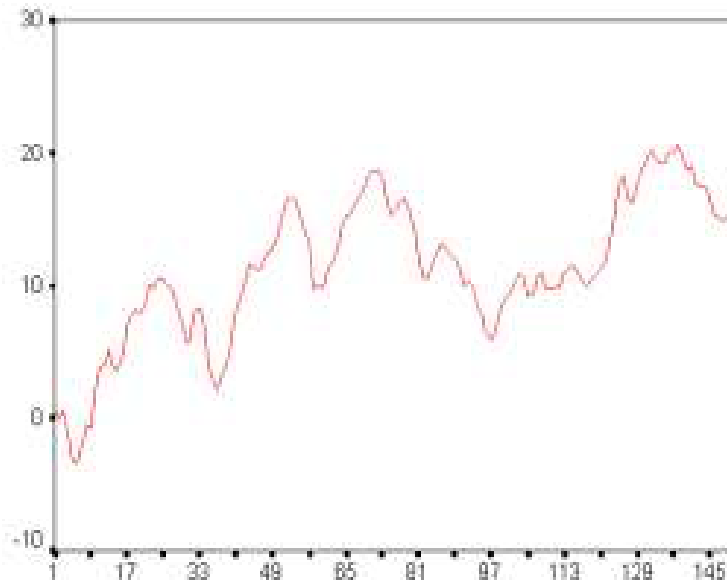
Έκφραση:

$$(1 - B)(1 - \varphi B)z_t = a_t$$

$$(1 - B - \varphi B + \varphi B^2)z_t = a_t \quad \text{όπου } |\varphi| < 1$$

$$z_t = (1 + \varphi)z_{t-1} - \varphi z_{t-2} + a_t$$

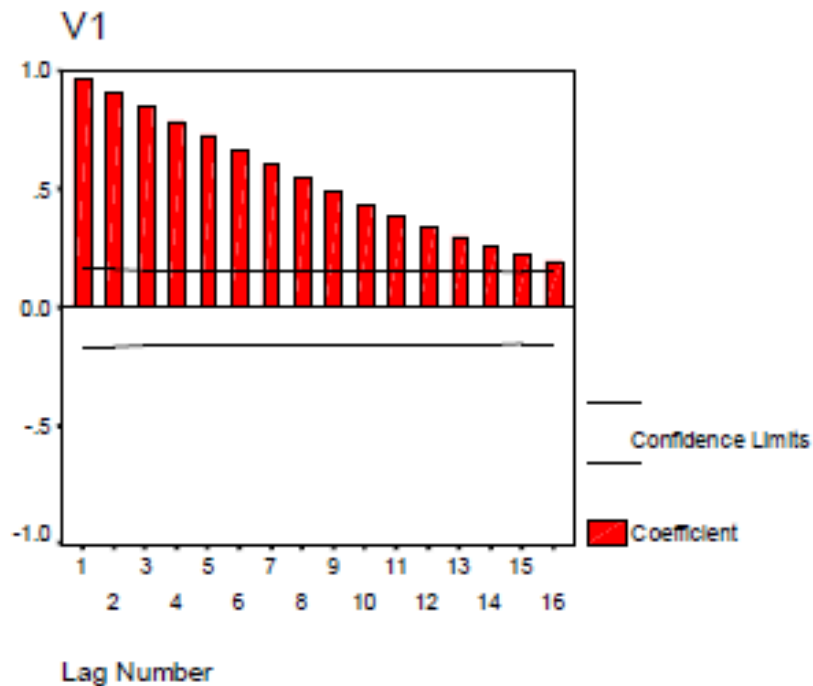
Τα γραφήματα μιας σειράς 150 παρατηρήσεων που φτιάχτηκαν με προσομοίωση από **ARI(1,1)** μοντέλο με $\varphi=0.4$ και των **πρώτων διαφορών της σειράς** είναι:



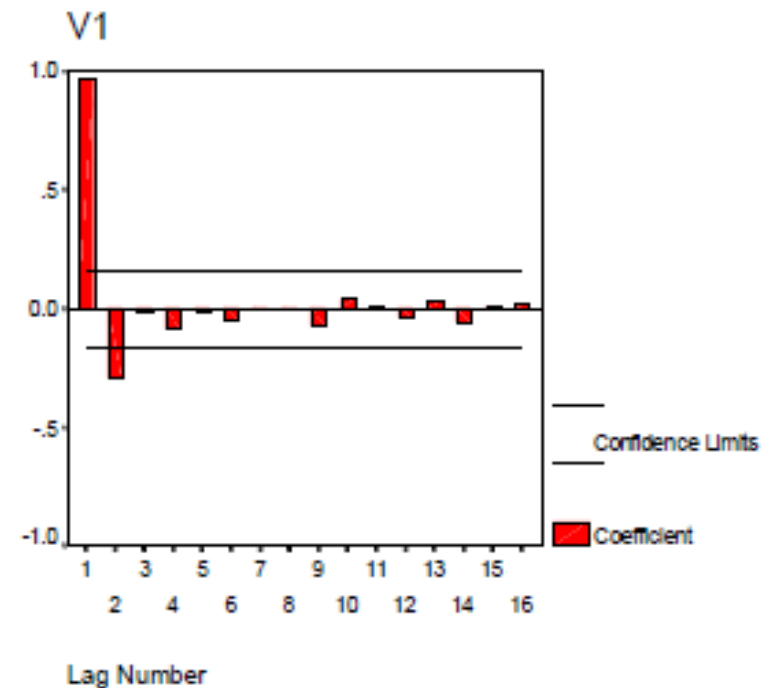
ARIMA(1,1,0) ή ARI(1,1) μοντέλο

Τα γραφήματα των ACF και PACF των παρατηρήσεων είναι:

Γράφημα ACF της σειράς



Γράφημα PACF της σειράς



Εποχικό Υπόδειγμα ARMA (p, q) x (P, Q)_s

- Εποχικό υπόδειγμα ARMA (p, q) x (P, Q)_s έχουμε όταν υπάρχουν **σημαντικές αυτοσυσχετίσεις στις εποχικές συχνότητες** πράγμα που παρουσιάζεται στο **γράφημα της χρονοσειράς**, αλλά και στις **συναρτήσεις ACF και PACF**.

Το **εποχικό υπόδειγμα ARMA** συμβολίζεται όπως αναφέρθηκε ως εξής:

ARMA (p, q) x (P, Q)_s

- Ομοίως έχουμε και το **εποχικό υπόδειγμα:**
ARIMA (p, d, q) x (P, D, Q)_s



Όπου **P, D, Q** είναι τα αντίστοιχα μεγέθη των **p, d, q** όταν αναφερόμαστε στην **εποχικότητα s**.

Άρα χρειάζεται να υπολογισθούν τα αντίστοιχα **SAR(P)** και **SMA(Q)** για να προσδιορισθεί η **τάξη του υποδείγματος που περιγράφει την εποχικότητα**.

Όπου **SAR(P)** η αντίστοιχη παράμετρος της $AR(p)$ αναφερόμενη στην **τάξη P της αυτοπαλινδρόμησης** που εμφανίζεται στην **εποχικότητα**

Ομοίως **SMA(Q)** η αντίστοιχη παράμετρος της $MA(q)$ αναφερόμενη στην **τάξη Q της διαδικασίας κινητού μέσου** που εμφανίζεται στην **εποχικότητα**



Εποχικά μοντέλα

Προσφέρονται για χρονικές σειρές που παρουσιάζουν εποχικότητα s , δηλαδή πολύ υψηλή συσχέτιση μεταξύ των παρατηρήσεων που απέχουν s χρονικά διαστήματα. Τα κυριότερα από αυτά είναι:

- **Εποχικά αυτοπαλίδρομα μοντέλα τάξης P [SAR(P,D)_s]**
- **Εποχικά μοντέλα κινούμενου μέσου τάξης Q [SIMA(D,Q)_s]**
- **Μικτά εποχικά μοντέλα τάξης P,Q [SARIMA(P,D,Q)_s]**



Εποχικά αυτοπαλίδρομα μοντέλα τάξης P [SAR(P,D)_s]

Έκφραση: $\Phi(B^s)w_t = a_t$

όπου $\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_p B^{ps}$

$$w_t = \nabla_s^D \nabla^d z_t$$

και ∇_s^D και ∇^d είναι αντίστοιχα ο τελεστής εποχικών διαφορών τάξης D και ο τελεστής απλών διαφορών τάξης d και απάλειψη μη στασιμότητας.



Εποχικά μοντέλα κινούμενου μέσου τάξης Q [SIMA(D,Q)_s]

Έκφραση: $w_t = \Theta(B^S)\alpha_t$ όπου

$$\Theta(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS}$$

$$w_t = \nabla_s^D \nabla^d z_t$$

και ∇_s^D και ∇^d όπως ορίστηκαν προηγουμένως



Μικτά εποχικά μοντέλα τάξης P,Q [SARIMA(P,D,Q)_s]

Έκφραση: $\Phi(B^s)w_t = \Theta(B^s)a_t$

όπου $\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps}$

$$\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs}$$

$$w_t = \nabla_s^D \nabla^d z_t$$



Γενικό πολλαπλασιαστικό μοντέλο

$ARIMA(p,d,q)(P,D,Q)_S$

Το $ARIMA(p,d,q)(P,D,Q)_S$ είναι συνδυασμός όλων των μοντέλων που αναφέρθηκαν

Έκφραση: $\varphi(B) \Phi(B^S) w_t = \theta(B) \Theta(B^S) a_t$ όπου

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$$

$$\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_P B^{PS}$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$$

$$\Theta(B^S) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS} \text{ και}$$

$$w_t = \nabla^d \nabla_s^D z_t$$

∇^d και ∇_s^D είναι αντίστοιχα ο τελεστής προς τα πίσω διαφορών τάξης d και προς τα πίσω εποχικών διαφορών τάξης D .

p = η τάξη του απλού αυτοπαλινδρομούμενου μοντέλου

q = η τάξη του απλού μοντέλου κινούμενου μέσου

P = η τάξη του εποχικού αυτοπαλινδρομούμενου μοντέλου

Q = η τάξη του εποχικού κινούμενου μέσου μοντέλου

S = η εποχικότητα.



Μεθοδολογία Box – Jenkins

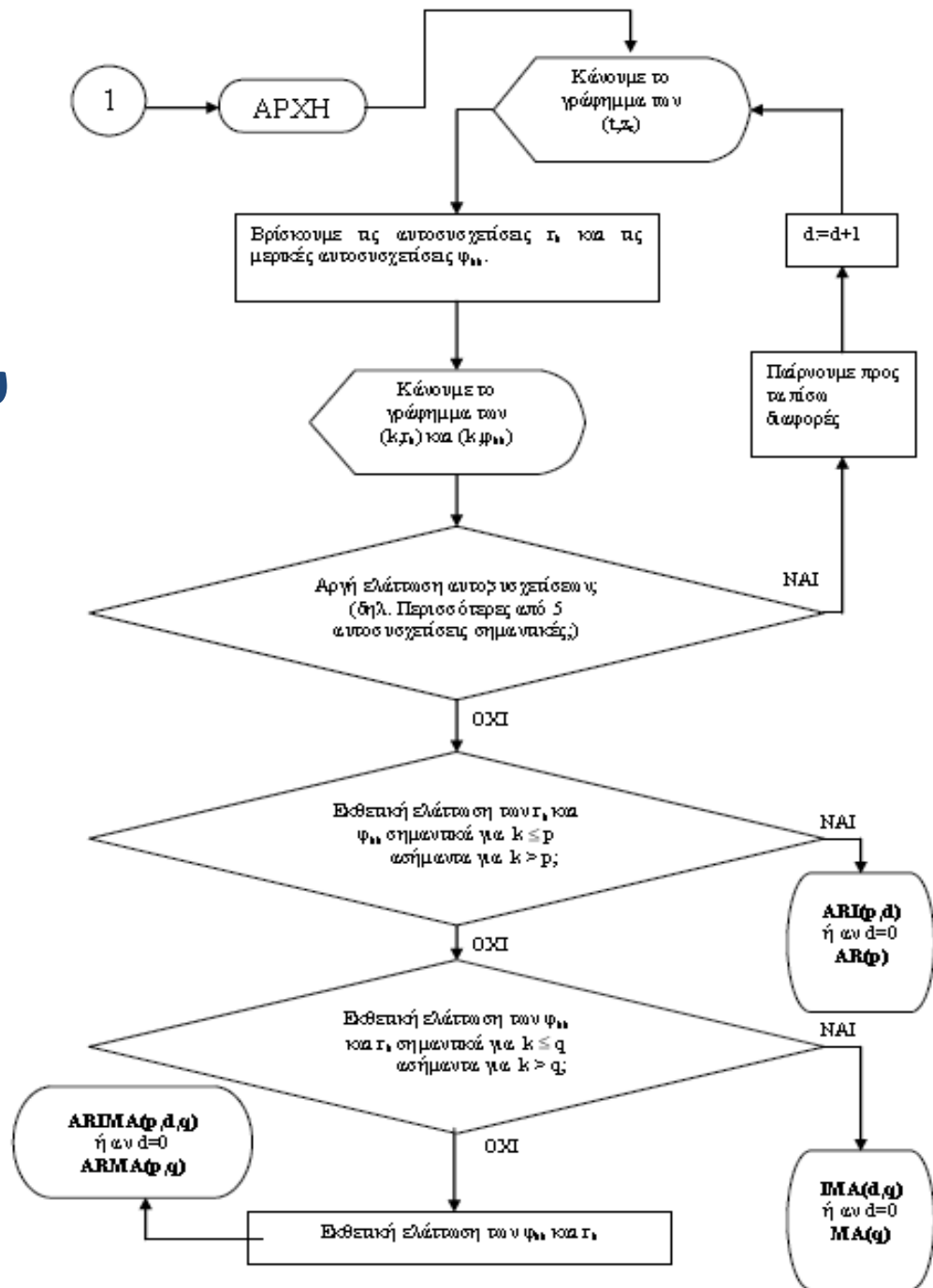
Είναι μια μέθοδος εξεύρεσης ενός στατιστικού υποδείγματος ARIMA που να παριστάνει ικανοποιητικά τη στοχαστική διαδικασία που παρήγαγε τα δεδομένα (το δείγμα).

Η μέθοδος περιλαμβάνει **τρία στάδια**:

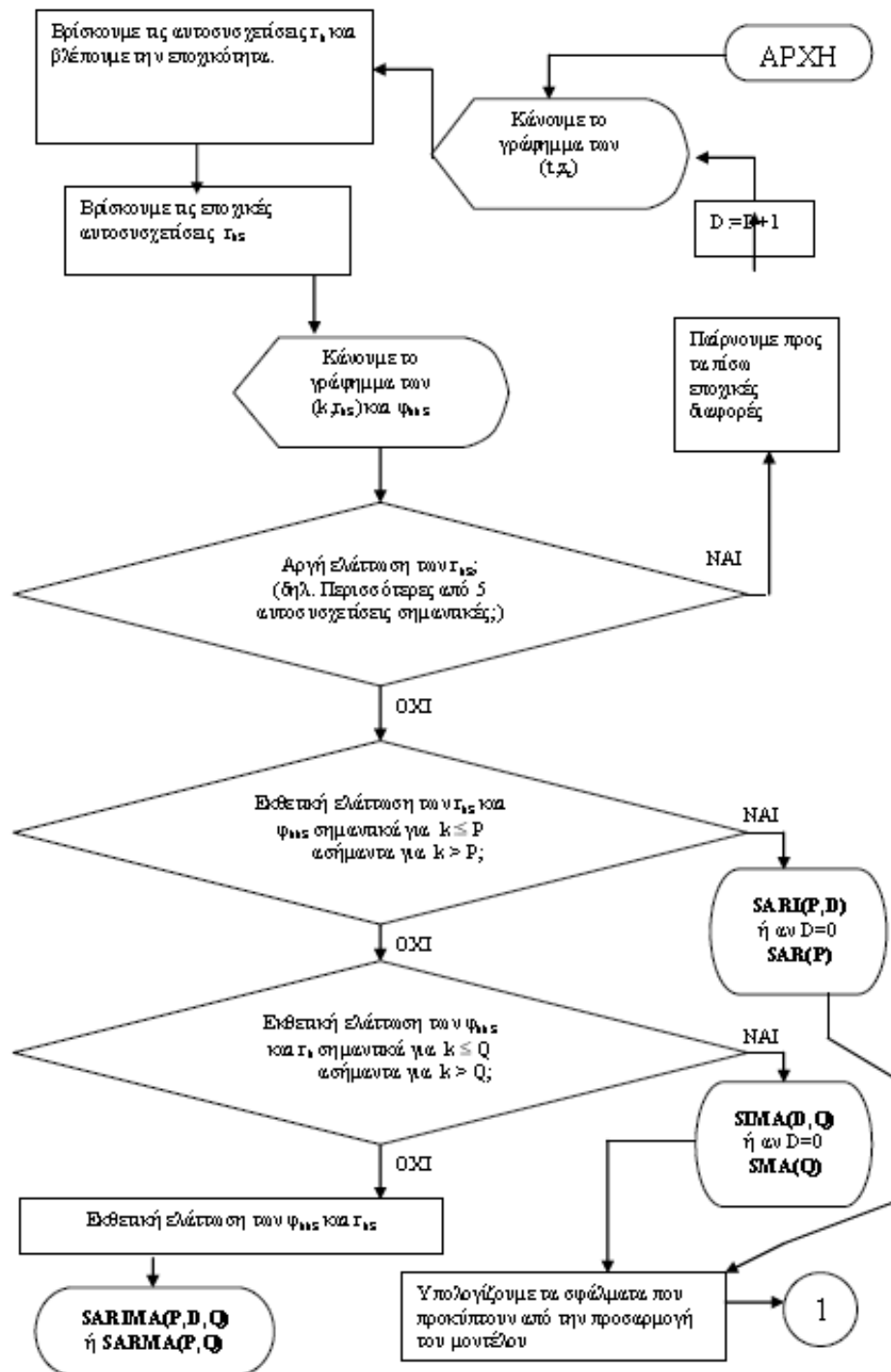
- **Την ταυτοποίηση (Identification)** του υποδείγματος
 - Στη γραφική παράσταση της σειράς
 - Στον έλεγχο της (δειγματικής) συνάρτησης αυτοσυσχέτισης
 - Στην χρησιμοποίηση διάφορων ελέγχων στασιμότητας
- **Την εκτίμηση (Estimation)** του υποδείγματος
- **Την αξιολόγηση (Evaluation)** του υποδείγματος



Διάγραμμα ροής εκτίμησης ενός ARIMA(p,d,q) μοντέλου



Διάγραμμα ροής εκτίμησης ενός $ARIMA(p,d,q)$ (P,D,Q)_s μοντέλου



- **Ταυτοποίηση (Identification) του υποδείγματος**

Περιλαμβάνει τον προσδιορισμό των καταλλήλων τιμών των d , p και q του υποδείγματος.

Καταρχήν η διαπίστωση αν η σειρά **είναι στάσιμη ή μη στάσιμη** μπορεί να στηριχθεί στα εξής:

- ❖ **Στη γραφική παράσταση της σειράς**

Αν φανεί από τη γραφική παράσταση της σειράς ότι η **διακύμανσή της μεταβάλλεται με το χρόνο**, όπως και **ο μέσος όρος της**, αυτό είναι **σοβαρή ένδειξη** ότι η σειρά είναι **μη στάσιμη**.



Ταυτοποίηση (Identification) του υποδείγματος

- ❖ Στον έλεγχο της (δειγματικής) συνάρτησης αυτοσυσχέτισης
 - Αν οι αυτοσυσχετίσεις συγκλίνουν ταχύτατα προς το μηδέν, αυτό είναι σοβαρή ένδειξη ότι η σειρά είναι στάσιμη.
 - Αντίθετα αν οι αυτοσυσχετίσεις φθίνουν με αργό ρυθμό είναι ένδειξη ότι η σειρά είναι μη στάσιμη.



Ταυτοποίηση (Identification) του υποδείγματος

❖ Στην χρησιμοποίηση διάφορων ελέγχων στασιμότητας

- ✓ Έλεγχος του Bartlett (Bartlett Test),
- ✓ Q Στατιστική (Box – Pierce Test),
- ✓ Στατιστική των Ljung – Box (Ljung – Box Statistic)
- ✓ Έλεγχοι μοναδιαίας ρίζας (unit root test)
(οι τρεις πρώτοι έλεγχοι έχουν αναφερθεί)



Έλεγχοι μοναδιαίας ρίζας (unit root test) - κριτήριο των Dickey – Fuller

Ο έλεγχος στασιμότητας μιας σειράς βασίζεται στις στατιστικές t και F και εξαρτάται από τη μορφή του υποδείγματος.

Παίρνουμε τις παρακάτω τρεις περιπτώσεις:

$$Y_t = \alpha + \beta T + \rho Y_{t-1} + \varepsilon_t \quad (\text{με σταθερό όρο και τάση})$$

$$Y_t = \rho Y_{t-1} + \varepsilon_t \quad (\text{χωρίς σταθερό όρο και χωρίς τάση})$$

$$Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t \quad (\text{με σταθερό όρο και χωρίς τάση})$$

όπου α = σταθερός όρος, T = η τάση, $\rho = 0$
συντελεστής αυτοσυσχέτισης και ε_t = λευκός θόρυβος.



- Η διαδικασία του **ελέγχου μοναδιαίας ρίζας (unit root test)** έχει ως εξής:

1. Το αντίστοιχο υπόδειγμα εκτιμάται με την **OLS**.

Για παράδειγμα η εκτίμηση του υποδείγματος $Y_t = \rho Y_{t-1} + \varepsilon_t$ οδηγεί στο εκτιμώμενο υπόδειγμα:

$$Y_t = \hat{\rho} Y_{t-1}$$

Ελέγχουμε με την **t – στατιστική** την υπόθεση:

$$H_0: \rho = 1$$

$$H_1: \rho < 1$$

$$\text{(με t – στατιστική: } t_\rho = \frac{\hat{\rho}}{s_\rho} \text{)}$$



Έλεγχος μοναδιαίας ρίζας (unit root test)

Αν $t_\rho > t(\alpha, N-2)$, η υπόθεση H_0 **απορρίπτεται**, που σημαίνει ότι η χρονολογική σειρά είναι **στάσιμη**

Αν $t_\rho < t(\alpha, N-2)$, η υπόθεση H_0 **δεν απορρίπτεται**, που σημαίνει ότι η χρονολογική σειρά είναι **μη στάσιμη**

Ο παραπάνω έλεγχος με τη t – στατιστική **δεν ισχύει** όταν η πραγματική τιμή του ρ είναι η μονάδα.

Οι *Dickey – Fuller* ανέπτυξαν μια κατανομή για τον εκτιμητή του ρ ($\hat{\rho}$) που ισχύει ακόμα και όταν η πραγματική τιμή του ρ είναι η μονάδα.

- Στην περίπτωση που η $H_0: \rho = 1$ γίνει δεκτή, δηλαδή η σειρά δεν είναι **στάσιμη**, το υπόδειγμα μετασχηματίζεται έτσι ώστε να αποφευχθούν οι συνέπειες της μη στασιμότητας.

Ένας τέτοιος μετασχηματισμός επιτυγχάνεται αν εκφρασθεί το υπόδειγμα σε **πρώτες διαφορές**.

- Το υπόδειγμα $Y_t = \rho Y_{t-1} + \varepsilon_t$ μπορεί να γραφεί ως εξής:



Έλεγχος μοναδιαίας ρίζας (unit root test)

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + \varepsilon_t$$

ή

$$\Delta Y_t = (\rho - 1) Y_{t-1} + \varepsilon_t$$

ή

$$\Delta Y_t = \gamma Y_{t-1} + \varepsilon_t \quad (\gamma = \rho - 1)$$

Αυτή η μετατροπή αποφεύγει την περίπτωση του $\rho = 1$ και οδηγεί στον έλεγχο της **μηδενικής υπόθεσης**:

$$H_0: \gamma = 0 \quad \text{ή} \quad H_0: \rho - 1 = 0 \quad (\gamma = \rho - 1)$$

$$\text{με } t\text{-στατιστική: } t_\gamma = \frac{\gamma}{s_\gamma}$$

- Ο έλεγχος της H_0 μπορεί να γίνει με βάση τους πίνακες κατανομής του ρ που κατασκεύασαν οι **Dickey – Fuller**.



Έλεγχος μοναδιαίας ρίζας (unit root test)

Έτσι η H_0 γίνεται δεκτή, δηλαδή η χρονολογική είναι μη στάσιμη όταν

$$|\tau_1| > t_\gamma, \text{ όπου } t_1 \text{ από τους πίνακες } D - F$$

Τελικά έχουμε:

- Αν $\gamma = 0$ τότε $\rho = 1$ και η υπόθεση H_0 γίνεται δεκτή, που σημαίνει ότι η χρονολογική σειρά είναι μη στάσιμη
- Αν $\gamma < 0$ τότε $\rho < 1$ και η υπόθεση H_0 απορρίπτεται, που σημαίνει ότι η χρονολογική σειρά είναι στάσιμη
- Αν $\gamma > 0$ τότε $\rho > 1$ και η υπόθεση H_0 απορρίπτεται, όμως στην περίπτωση αυτή εφόσον $\rho > 1$ η χρονολογική σειρά είναι μη στάσιμη



Έλεγχος μοναδιαίας ρίζας (unit root test)

Ομοίως μπορεί να εφαρμοσθεί η ίδια ακριβώς διαδικασία για τον έλεγχο της μοναδιαίας ρίζας και στα άλλα δύο υποδείγματα.

Υπάρχουν οι **κριτικές στους πίνακες D-F**

$$\left| \tau_2 \right| \quad \text{για το υπόδειγμα } Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t \text{ και}$$

$$\left| \tau_3 \right| \quad \text{για το υπόδειγμα } Y_t = \alpha + \beta T + \rho Y_{t-1} + \varepsilon_t$$



Έλεγχος μοναδιαίας ρίζας (unit root test)

Η **F – στατιστική** εφαρμόζεται στην περίπτωση που ο έλεγχος αφορά τον από κοινού έλεγχο όλων των συντελεστών του υποδείγματος.

Ενδεικτικά αν αναφερθούμε στο υπόδειγμα:

$$Y_t = \alpha + \beta T + \rho Y_{t-1} + \varepsilon_t$$

- Η αποδοχή της H_0 σημαίνει μη στάσιμη σειρά, η στασιμότητα της σειράς μπορεί να επιτευχθεί αν το υπόδειγμα εκφρασθεί σε μορφή πρώτων διαφορών:

$$Y_t - Y_{t-1} = \alpha + \beta T + \rho Y_{t-1} - Y_{t-1} + \varepsilon_t = \alpha + \beta T + (\rho - 1) Y_{t-1} + \varepsilon_t$$

ή

$$\Delta Y_t = \alpha + \beta T + (\rho - 1) Y_{t-1} + \varepsilon_t$$

ή

$$\Delta Y_t = \alpha + \beta T + \gamma Y_{t-1} + \varepsilon_t$$



Έλεγχος μοναδιαίας ρίζας (unit root test)

Η μηδενική υπόθεση είναι:

$$H_0: \alpha = \beta = \gamma = 0$$

- Η H_0 απορρίπτεται άρα η σειρά είναι **στάσιμη** όταν:

$$F = \frac{AT\Pi / (k - 1)}{ATK / (N - k)} = \frac{r^2 / (k - 1)}{(1 - r^2) / (N - k)} > \Phi_2$$



Έλεγχος μοναδιαίας ρίζας (unit root test)

Παρατίθεται πίνακας των **Dickey – Fuller** για τους στατιστικούς ελέγχους που αφορούν την στασιμότητα μιας χρονολογικής σειράς

Υπόδειγμα	Υπόθεση	Στατιστική
$\Delta Y_t = \beta Y_{t-1} + \varepsilon_t$	$H_0: \beta = 0$	τ_1
$\Delta Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t$	$H_0: \beta_1 = 0$	τ_2
$\Delta Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t$	$H_0: \beta_0 = \beta_1 = 0$	Φ_1
$\Delta Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 T + \varepsilon_t$	$H_0: \beta_1 = 0$	τ_3
$\Delta Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 T + \varepsilon_t$	$H_0: \beta_0 = \beta_1 = \beta_2 = 0$	Φ_2
$\Delta Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 T + \varepsilon_t$	$H_0: \beta_0 = \beta_1 = 0$	Φ_3



Ταυτοποίηση (Identification) του υποδείγματος

- Εφόσον η σειρά είναι στάσιμη ή γίνει στάσιμη παίρνοντας τις πρώτες ή δεύτερες κ.λ.π διαφορές στη συνέχεια πρέπει να καθορισθεί η τάξη p της αυτοπαλίνδρομης διαδικασίας και η τάξη q της διαδικασίας κινητού μέσου ώστε να προσδιορισθεί το υπόδειγμα ARIMA
- Ο προσδιορισμός των p και q βασίζεται στις (δειγματικές) απλές και μερικές αυτοσυσχετίσεις με βάση τον πίνακα περι των μορφών της συνάρτησης αυτοσυσχέτισης και μερικής αυτοσυσχέτισης που αναφέρθηκε πιο πάνω



Εκτίμηση (Estimation) του υποδείγματος

- Εκτιμούμε τις p παραμέτρους $\alpha_1, \alpha_2, \dots, \alpha_p$, της AR διαδικασίας και τις q παραμέτρους $\theta_1, \theta_2, \dots, \theta_q$ της AM διαδικασίας.
- ❖ Αν έχουμε να εκτιμήσουμε μόνο το AR υπόδειγμα εφαρμόζουμε την μέθοδο των ελαχίστων τετραγώνων OLS.
- ❖ Αν η σειρά περιέχει και όρους κινητού μέσου (MA) τότε για την εκτίμηση των αντιστοιχών παραμέτρων χρησιμοποιούνται μη γραμμικές μέθοδοι.



Αξιολόγηση (Evaluation) του υποδείγματος

Ελέγχεται η καταλληλότητα του υποδείγματος ARIMA δηλαδή πόσο καλά ταιριάζει στα δεδομένα μας.

Ο έλεγχος της καταλληλότητας του υποδείγματος ARIMA περιλαμβάνει:

- ❖ τους στατιστικούς ελέγχους της σημαντικότητας των συντελεστών,
- ❖ την συμπεριφορά των υπολοίπων
- ❖ την τάξη του υποδείγματος
- ❖ την προβλεπτική του ικανότητα



Αξιολόγηση (Evaluation) του υποδείγματος

❖ Στατιστικοί έλεγχοι της σημαντικότητας των συντελεστών

Για τους στατιστικούς ελέγχους της σημαντικότητας των συντελεστών έγινε λόγος.

❖ Η συμπεριφορά των υπολοίπων

Αν το υπόδειγμα ARIMA ταιριάζει στα δεδομένα μας θα πρέπει **τα υπόλοιπα να συμπεριφέρονται** ως μια διαδικασία **λευκού θορύβου**. Αυτό σημαίνει ότι **τα υπόλοιπα δεν θα πρέπει να αυτοσυσχετίζονται**.

Ο έλεγχος των υπολοίπων γίνεται με τη **στατιστική Q των Box – Pierce** με την οποία ελέγχεται η σημαντικότητα από κοινού ενός αριθμού συντελεστών αυτοσυσχέτισης έστω m .



Στατιστική Q των Box – Pierce

Θεωρούμε την **μηδενική υπόθεση**

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0 \quad (\text{τα υπόλοιπα δεν αυτοσυσχετίζονται})$$

Η στατιστική Q των Box – Pierce (Portmanteau lack of fit test) που χρησιμοποιείται για τον έλεγχο της H_0 είναι:

$$Q = T \sum_{s=1}^m P_s^*{}^2$$

όπου m = το μήκος της χρονικής υστέρησης, ρ_s = οι δειγματικές αυτοσυσχετίσεις των υπολοίπων, T = ο αριθμός των παρατηρήσεων (υπολοίπων).

Συνήθως λαμβάνουμε το $m = \sqrt{T}$ (ή $m=0.25T$)



Στατιστική Q των Box – Pierce

Η στατιστική Q των Box – Pierce ακολουθεί την χ^2 κατανομή με $m - p - q$ βαθμούς ελευθερίας $\chi^2(\alpha, m - p - q)$.

Για δεδομένο επίπεδο σημαντικότητας α έχουμε ότι:

Αν $Q > \chi^2(\alpha, m - p - q)$ η μηδέν υπόθεση H_0 **απορρίπτεται** άρα τα **υπόλοιπα αυτοσυσχετίζονται**

Αν $Q < \chi^2(\alpha, m - p - q)$ γίνεται **δεκτή η H_0** άρα τα **υπόλοιπα δεν αυτοσυσχετίζονται**



Στατιστική των Ljung – Box

Μια τροποποιημένη μορφή της στατιστική Q των Box – Pierce είναι η **στατιστική των Ljung – Box (Ljung – Box Statistic)**

Η στατιστική των **Ljung – Box**, μολονότι ακολουθεί την χ^2 κατανομή με $m - p - q$ βαθμούς ελευθερίας $\chi^2 (\alpha, m - p - q)$ δίνει καλύτερα αποτελέσματα από την Q στατιστική όταν εφαρμόζεται σε μικρά δείγματα $T < 30$.

Θεωρούμε την **μηδενική υπόθεση**:

$H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0$ (τα υπόλοιπα δεν αυτοσυσχετίζονται)

Η **στατιστική των Ljung – Box** που χρησιμοποιείται για τον έλεγχο της H_0 ορίζεται ως ακολούθως:

$$LB = T(T+2) \sum_{s=1}^m \frac{p_s^{*2}}{T-s}$$

Για δεδομένο **επίπεδο σημαντικότητας α** έχουμε:

- Αν $LB > \chi^2_{(\alpha, m-p-q)}$ **απορρίπτεται η H_0** άρα τα υπόλοιπα **αυτοσυσχετίζονται**
- Αν $LB \leq \chi^2_{(\alpha, m-p-q)}$ γίνεται **δεκτή η H_0** άρα τα **υπόλοιπα δεν αυτοσυσχετίζονται**



Αξιολόγηση (Evaluation) του υποδείγματος

❖ Η τάξη του υποδείγματος

Η καταλληλότητα του εκτιμώμενου υποδείγματος ελέγχεται με ένα άλλο υπόδειγμα μεγαλύτερης τάξης.

Δηλαδή το εκτιμώμενο υπόδειγμα $ARIMA(p, d, q)$ **συγκρίνεται** με τα υποδείγματα $ARIMA(p+1, d, q)$ και $ARIMA(p, d, q+1)$.

Αν το υπόδειγμα που εκτιμήθηκε περιγράφει τη διαδικασία που παρήγαγε τα δεδομένα, οι επιπλέον συντελεστές στα μεγαλύτερα υποδείγματα **δεν θα πρέπει να είναι στατιστικά σημαντικοί (διάφοροι του μηδενός)**.



Για τον **έλεγχο της τάξης** του υποδείγματος χρησιμοποιούνται δύο **κριτήρια**:

- Το **κριτήριο πληροφοριών Akaike (Akaike Information Criterion) ή AIC**

Ο τύπος που δίνει το κριτήριο AIC είναι:

$$AIC = \ln \frac{\sum u^{*2}}{T} + \frac{2k}{T}$$

όπου

$\sum u^{*2}$ το άθροισμα τετραγώνων των υπολοίπων

T ο αριθμός των παρατηρήσεων

k = ο αριθμός παραμέτρων που εκτιμώνται = $(p + q + 1)$

- Το **Μπαισιανό κριτήριο Schwartz (Schwartz Bayesian Criterion) ή SBC**

Ο τύπος που δίνει το κριτήριο AIC είναι:

$$SBC = \ln \frac{\sum u^{*2}}{T} + \frac{k}{T} \ln T$$

Με βάση τα παραπάνω κριτήρια **επιλέγεται το υπόδειγμα με τη μικρότερη τιμή**

- [Τα **δύο κριτήρια** μπορούν να παίρνουν και **αρνητικές τιμές**]



Αξιολόγηση (Evaluation) του υποδείγματος

❖ Η προβλεπτική ικανότητα του υποδείγματος

■ Προβλεπτική ικανότητα του υποδείγματος AR(1)

Έστω το υπόδειγμα AR(1) το οποίο είναι προς εκτίμηση από T παρατηρήσεις:

$$Y_t = \delta + \alpha_1 Y_{t-1} + \varepsilon_t$$

Η πρόβλεψη την περίοδο $T+1$ είναι:

$$Y_{T+1} = \delta + \alpha_1 Y_T$$

Το **σφάλμα πρόβλεψης** για την περίοδο $T + 1$ είναι:

$$\varepsilon_{T+1} = Y_{T+1} - Y_{T+1}$$



Προβλεπτική ικανότητα του υποδείγματος AR(1)

Για προβλέψεις h περιόδους μπροστά από την τελευταία T παρατήρηση έχουμε:

$$Y_{T+h} = \delta + \alpha_1 Y_{T+h-1}$$

Με τη διακύμανση του σφάλματος πρόβλεψης να αυξάνει μη γραμμικά καθώς αυξάνει η περίοδος πρόβλεψης.

$$V(\varepsilon_{T+h}) = \sigma^2 (1 + \alpha_1^2 + \alpha_1^4 + \dots + \alpha_1^{2(h-1)})$$

Καθώς όμως η περίοδος πρόβλεψης μεγαλώνει το όριό της συγκλίνει προς τον μέσο:

$$\lim Y_{T+h} = \frac{\delta}{1 - \alpha_1} = \mu$$



Προβλεπτική ικανότητα του υποδείγματος MA(1)

- Έστω το υπόδειγμα MA(1) το οποίο είναι προς εκτίμηση από T παρατηρήσεις:

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Για **πρόβλεψη $h = 1$ περίοδο** μπροστά από την **τελευταία T παρατήρηση** έχουμε:

$$Y_{T+1} = E_T(\mu + \theta_1 \varepsilon_t)$$

Με **διακύμανση του σφάλματος πρόβλεψης**: $V(\varepsilon_{T+1}) = \sigma^2$

Για **προβλέψεις $h > 1$ περιόδους** μπροστά από την **τελευταία T παρατήρηση** έχουμε:

$$Y_{T+h} = \mu$$

Με **διακύμανση του σφάλματος πρόβλεψης**: $V(\varepsilon_{T+h}) = \sigma^2 (1 + \theta_1^2)$

- Είναι φανερό ότι ένα υπόδειγμα **κινητού μέσου πρώτης τάξης είναι κατάλληλο για προβλέψεις μόνο μιας περιόδου μπροστά**, αφού για $h > 1$ περιόδους η πρόβλεψη είναι πάντα ο μέσος.



Προβλεπτική ικανότητα του υποδείγματος ARMA(1,1)

- Έστω το υπόδειγμα ARMA(1,1) το οποίο είναι προς εκτίμηση από T παρατηρήσεις:

$$Y_t = \delta + \alpha_1 Y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Για **πρόβλεψη $h=1$ περίοδο** μπροστά από την **τελευταία T παρατήρηση** έχουμε:

$$Y_{T+1} = \delta + \alpha_1 Y_T + \theta_1 \varepsilon_T$$

Με **διακύμανση του σφάλματος πρόβλεψης**: $V(\varepsilon_{T+1}) = \sigma^2$

Ακολουθώντας την ίδια διαδικασία βρίσκουμε τις προβλέψεις για **δύο, τρεις ή για h περιόδους** μπροστά από την **τελευταία T παρατήρηση**.

- Καθώς όμως η περίοδος πρόβλεψης μεγαλώνει το όριο της πρόβλεψης συγκλίνει προς τον μέσο



■ Προβλεπτική ικανότητα του υποδείγματος ARIMA

Έστω το αυτοπαλίνδρομο ολοκληρωμένο υπόδειγμα πρώτης τάξης:

$$w_t = \delta + \alpha_1 w_{t-1} + \varepsilon_t$$

όπου $w_t = \Delta Y_t = Y_t - Y_{t-1}$

Η μη στάσιμη σειρά Y_t γίνεται στάσιμη παίρνοντας τις **πρώτες διαφορές** οι οποίες είναι AR(1).

Μπορούμε να κάνουμε πρόβλεψη με το υπόδειγμα $w_t = \delta + \alpha_1 w_{t-1} + \varepsilon_t$ και στη συνέχεια να κάνουμε πρόβλεψη της Y .

Για $h = 1$ από την **τελευταία T παρατήρηση** θα έχουμε:

$$w_{T+1} = \delta + \alpha_1 w_T$$

οπότε η **πρόβλεψη της Y_{T+1}** θα είναι:

$$Y_{T+1} = \delta + (1 + \alpha_1)Y_T - \alpha_1 Y_{T-1}$$

Γενικά για **προβλέψεις h περιόδους** μπροστά από την **τελευταία T παρατήρηση** έχουμε:

$$Y_{T+h} = Y_T + w_{T+1} + w_{T+2} + \dots + w_{T+h}$$



□ Διάστημα εμπιστοσύνης για την τιμή Y_{T+h}

Με βάση την **προβλεπόμενη τιμή** Y_{T+h} και τη διακύμανση του σφάλματος πρόβλεψης σ_h^2 μπορούμε να ορίσουμε το **1- α διάστημα εμπιστοσύνης** για την **αληθινή τιμή** Y_{T+h} ως εξής:

$$Y_{T+h} - z_{\frac{\alpha}{2}} \sigma_h \leq Y_{T+h} \leq Y_{T+h} + z_{\frac{\alpha}{2}} \sigma_h$$

Η σ_h^2 είναι άγνωστη γι' αυτό την εκτιμάμε από το **άθροισμα των τετραγώνων των υπολοίπων σ^2** και την αντικαθιστούμε στην παραπάνω σχέση στη θέση της σ_h :

$$\sigma^2 = \frac{\sum_{t=1}^T \varepsilon_t^2}{T - p - q}$$



ΑΝΕΛΙΞΕΙΣ MARKOV

- Τα προηγούμενα μοντέλα είναι για συνεχείς μεταβλητές. Στην υδρολογία υπάρχουν και μη-συνεχείς παράμετροι ή συνεχείς παράμετροι μετατρέπονται σε μη-συνεχείς για υπολογιστική ευκολία
- Στις περιπτώσεις μη-συνεχών υδρολογικών παραμέτρων χρησιμοποιείται η θεωρία των **ανερίξεων Markov** (Markov chains)
 - Βροχόπτωση,
 - Απορροή σε ξηρές και ημίξηρες περιοχές,
 - Εδαφική υγρασία,
 - Αποθήκευση σε ταμιευτήρες



Ιδιότητες Διακριτών Χρονοσειρών

- Διακριτή τυχαία μεταβλητή είναι μια μεταβλητή που παίρνει διακριτές τιμές από ένα ορισμένο σετ δεδομένων
- Η πιο κοινή ομάδα μοντέλων ή στοχαστικών διαδικασιών που χρησιμοποιούνται για διακριτές χρονοσειρές είναι οι **ανελίξεις Markov** (Markov chains)
- Οι **ανελίξεις Markov** μπορούν να θεωρηθούν ως συλλογή καταστάσεων ενός συστήματος
- Κάθε κατάσταση ανταποκρίνεται στα στοιχεία του δειγματικού χώρου της τυχαίας χρονοσειράς
- Για κάθε χρονική περίοδο, ίση με το μήκος δειγματοληψίας της χρονοσειράς, οι ανελίξεις Markov μπορούν να παραμείνουν στην ίδια κατάσταση ή να αλλάξουν (μεταβούν) σε άλλες καταστάσεις



ΑΝΕΛΙΞΕΙΣ MARKOV

Ορισμός και Ιδιότητες:

- Αν η $X(t)$ είναι ασυνεχής παράμετρος που αρχίζει στο χρόνο 0 και αναπτύσσεται στο χρόνο t , οι τιμές που παίρνει η $X(t)$ ορίζοντας σαν x_t τότε:

$$P[X(t)=x_t \mid X(0)=x_0, X(1)=x_1, \dots, X(t-1)=x_{t-1}]$$

είναι η πιθανότητα η διαδικασία να είναι ίση με x_t στο χρόνο t .

- Η διαδικασία απλοποιείται αν λάβουμε υπόψη μας μόνο δύο διαδοχικά βήματα $t, t-1$.
 - Η διαδικασία αυτή ονομάζεται **απλή ανέλιξη Markov** ή **ανέλιξη Markov 1^{ης} τάξης** (1st Order Markov chain)

Για ανελίξεις Markov 1^{ης} τάξης

$$\Pr\{X_{t+1} \mid X_t, X_{t-1}, X_{t-2}, \dots, X_1\} = \Pr\{X_{t+1} \mid X_t\}$$

- **Μητρώο Πιθανότητας Μετάβασης** (Transition Probability Matrix)
 - Τετραγωνικός πίνακας: $P_{ij}(t) = P[X(t)=j \mid X(t-1)=i]$ για όλα τα ζευγάρια i, j



ΑΝΕΛΙΞΕΙΣ ΜΑΡΚΟΝ

- Αν μια χρονοσειρά π.χ. βροχόπτωσης θέσουμε:

- 0 για τις ξηρές ημέρες και
- 1 για τις υγρές ημέρες

τότε οι πιθανότητες μετάβασης έχουν τις πιθανότητες

0-0, 0-1, 1-0, 1-1

Μητρώο 2 x 2

$$\begin{matrix} & 0 & 1 \\ 0 & \begin{bmatrix} P_{00} & P_{01} \end{bmatrix} \\ 1 & \begin{bmatrix} P_{10} & P_{11} \end{bmatrix} \end{matrix}$$

- Για Markov-2^{ης} τάξης (για τρεις συνεχόμενες ημέρες)

τότε οι πιθανότητες μετάβασης έχουν πιθανότητες

000, 001, 010, 011, 100, 110, 101, 111

- Και είναι : $\sum_{j=1}^r P_{ij}(t) = 1, i=1, \dots, r.$

$$\frac{1}{r} \sum_{i=1, j=1}^{r, r} P_{ij}(t) = 1, i=1, \dots, r.$$

- Αν το μητρώο πιθανότητας μετάβασης $P(t)$ δεν εξαρτάται από το χρόνο τότε η ανέλιξη Markov είναι **ομοιογενής** (homogeneous) ή **σταθερή** (stationary) ανέλιξη



ΑΝΕΛΙΞΕΙΣ MARKOV

n-step Probability

- Αν η ανέλιξη είναι σε καθεστώς i και έπειτα από n -βήματα χρονικά είναι σε καθεστώς j , η πιθανότητα μετάβασης από i σε j σε n χρονικά βήματα ($P_{ij}^{(n)}$) δίνεται από:

$$P_{ij}^{(n)} = \sum_{k=1}^r p_{ik}^{(n-1)} p_{kj} \quad n > 1$$

- Μπορεί να δειχθεί ότι ο πίνακας $P^{(n)} = P \times P \times P \dots n$; τιμές

Marginal Distribution

- Η πιθανότητα της ανέλιξης να είναι σε οποιοδήποτε καθεστώς j σε χρόνο t ορίζεται ως:

$$q_j(t) = P[X(t) = j], \quad j=1, \dots, r.$$

Marginal distribution της διαδικασίας

- Αν $q_j(0)$ είναι η πιθανότητα στο αρχικό στάδιο (αρχή). Η marginal πιθανότητα είναι: $q_j(t) = \sum_{i=1}^r q_i(0) p_{ij}(t)$



ΑΝΕΛΙΞΕΙΣ MARKOV

Steady State Probability

- Αν υπάρχουν δεδομένα τότε βρίσκονται οι πιθανότητες για ξηρή ή υγρή ημέρα για κάθε ημέρα. Αν όχι βρίσκεται για μακριά περίοδο t το μητρώο $p(t)$ και μια γραμμή του μητρώου είναι για ξηρή και υγρή περίοδο.

Ανελίξεις Markov 1ης τάξης, δύο καταστάσεων

- Πίνακας μίας χρονοσειράς βροχόπτωσης δύο καταστάσεων (Ιανουάριος 1987, Ημέρες με βροχή δίνουν $x_t = 1$, και ημέρες χωρίς βροχόπτωση δίνουν $x_t = 0$).

Ημέρες, t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
x_t	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	1	1

$$X_1 = 0, x_2 = 1, x_3 = 1, x_4 = 0, \dots, \text{ και } x_{31} = 1.$$



Ανελίξεις Markov 1ης τάξης, δύο καταστάσεων

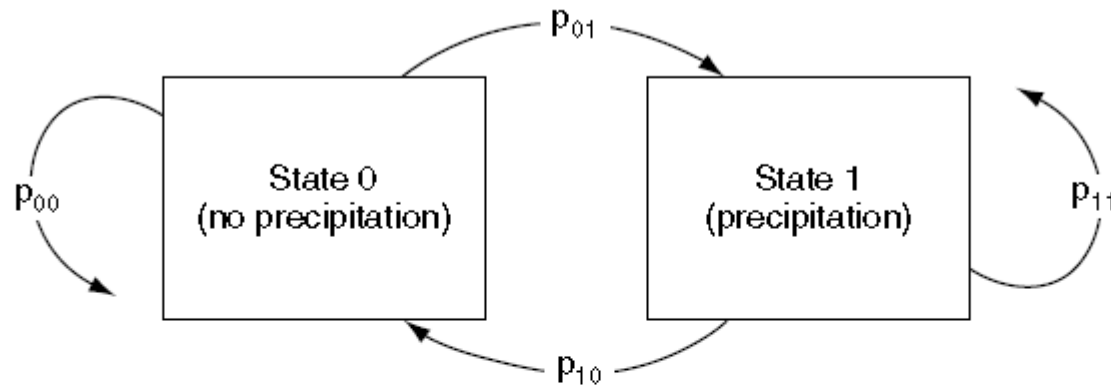


FIGURE 8.1 Schematic representation of a two-state, first-order Markov chain, illustrated in terms of daily precipitation occurrence or nonoccurrence. The two states are labeled 0 for no precipitation, and 1 for precipitation occurrence. For a first-order Markov chain, there are four transition probabilities controlling the state of the system in the next time period. Since these four probabilities are pairs of conditional probabilities, $p_{00} + p_{01} = 1$ and $p_{10} + p_{11} = 1$. For quantities like day-to-day precipitation occurrence that exhibit positive serial correlation, $p_{01} < p_{00}$, and $p_{01} < p_{11}$.

$$p_{00} = \Pr\{X_{t+1} = 0 | X_t = 0\}$$

$$p_{01} = \Pr\{X_{t+1} = 1 | X_t = 0\}$$

$$p_{10} = \Pr\{X_{t+1} = 0 | X_t = 1\}$$

$$p_{11} = \Pr\{X_{t+1} = 1 | X_t = 1\}.$$



Ανεπίξεις Markov 1ης τάξης, δύο καταστάσεων

$$p_{00} = \Pr\{X_{t+1} = 0 | X_t = 0\}$$

$$p_{01} = \Pr\{X_{t+1} = 1 | X_t = 0\}$$

$$p_{10} = \Pr\{X_{t+1} = 0 | X_t = 1\}$$

$$p_{11} = \Pr\{X_{t+1} = 1 | X_t = 1\}.$$

Επειδή, $p_{00} + p_{01} = 1$ και $p_{10} + p_{11} = 1$, ο σκοπός είναι στην εκτίμηση του μητρώου πιθανότητας μετάβασης των p_{01} και p_{11} .

$$\hat{p}_{01} = \frac{\text{\# of 1's following 0's}}{\text{Total \# of 0's}} = \frac{n_{01}}{n_{0\bullet}}$$

$$\hat{p}_{11} = \frac{\text{\# of 1's following 1's}}{\text{Total \# of 1's}} = \frac{n_{11}}{n_{1\bullet}}.$$

$$n_{1\bullet} = n_{10} + n_{11} \text{ και } n_{0\bullet} = n_{00} + n_{01}$$



Ανελίξεις Markov 1ης τάξης, δύο καταστάσεων

- Για μία ανέλιξη Markov που περιγράφει την ημερήσια βροχόπτωση ή μη, η στάσιμη πιθανότητα της βροχόπτωσης, π_1 , είναι η κλιματολογική πιθανότητα της βροχόπτωσης
- Με βάση τις πιθανότητες μετάβασης p_{01} και p_{11}

$$\pi_1 = \frac{p_{01}}{1 + p_{01} - p_{11}}, \quad \pi_0 = 1 - \pi_1$$

- Συνήθως για θετική χρονική συσχέτιση ή εμμονή βρίσκουμε $p_{01} < \pi_1 < p_{11}$.
- Συντελεστής αυτοσυσχέτισης : $r_1 = p_{11} - p_{01}$



ΠΡΟΣΟΜΟΙΩΣΗ ΣΗΜΕΙΑΚΩΝ ΔΙΑΔΙΚΑΣΙΩΝ

Πιο συχνά για τη βροχόπτωση

- Οι Todorigo and Yevjevich (1967) πρότειναν ότι η εμφάνιση καταιγίδων καθορίζεται από μια διαδικασία Poisson. Υπόθεσαν ότι **ο αριθμός αφίξεων καταιγίδων είναι μία διαδικασία Poisson.**
- Έτσι ο αριθμός καταιγίδων $N(t)$ στο χρονικό διάστημα $(0,t)$ κατανέμεται κατά Poisson

$$P[N(t) = n] = \frac{(\lambda t)^n}{n!} \exp(-\lambda t) \quad n=0,1,\dots$$

όπου λ ο μέσος ρυθμός άφιξης καταιγίδων

- Η δεύτερη υπόθεση είναι ότι **ο λευκός θόρυβος του ποσού βροχόπτωσης R είναι συνάρτηση του χρόνου άφιξης.**
 - $N(t)$ και R θεωρούνται ανεξάρτητα και το R κατανέμεται με gamma κατανομή. Η διαδικασία αυτή ονομάζεται **Poisson white noise.**
- Όμοια το **συνολικό ύψος βροχόπτωσης** στο διάστημα $(0,t)$ δίνεται από:
 $z(t) = \sum_{j=1}^{N(t)} R_j$ και ονομάζεται **Compound Poisson process.**



Βιβλιογραφία

Μιμίκου, Μ.Α., 2006. «Τεχνολογία Υδατικών Πόρων», 3^η Έκδοση, Α. Παπασωτηρίου & Σία ΟΕ.



Χρηματοδότηση

- Το παρόν εκπαιδευτικό υλικό έχει αναπτυχθεί στο πλαίσιο του εκπαιδευτικού έργου του διδάσκοντα.
- Το έργο «**Ανοικτά Ακαδημαϊκά Μαθήματα στο Πανεπιστήμιο Θεσσαλίας**» έχει χρηματοδοτήσει μόνο την αναδιαμόρφωση του εκπαιδευτικού υλικού.
- Το έργο υλοποιείται στο πλαίσιο του Επιχειρησιακού Προγράμματος «Εκπαίδευση και Δια Βίου Μάθηση» και συγχρηματοδοτείται από την Ευρωπαϊκή Ένωση (Ευρωπαϊκό Κοινωνικό Ταμείο) και από εθνικούς πόρους.

