Κεφάλαιο 7

ΑΝΑΛΥΣΗ ΒΙΟΛΟΓΙΚΩΝ ΣΗΜΑΤΩΝ ΣΤΟ ΠΕΔΙΟ ΤΟΥ ΧΡΟΝΟΥ

7.1. Εισαγωγή

Όλοι οι ζώντες οργανισμοί, από τα κύτταρα μέχρι τα διάφορα όργανα, παράγουν σήματα βιολογικής προέλευσης. Τέτοια σήματα μπορεί να είναι ηλεκτρικά(π.χ. η εκπόλωση ενός νευρικού κυττάρου ή του μύ της καρδιάς), μηχανικά (π.χ. ο ήχος που παράγεται από τις βαλβίδες της καρδιάς) ή χημικά (π.χ. η ένωση PCO_2 στο αίμα). Τα βιολογικά σήματα αυτά μπορεί να παρουσιάζουν ενδιαφέρον στη διάγνωση, στην παρακολούθηση των ασθενών και στη βιοϊατρική έρευνα.

Κατά τη διάρκεια της ζωής τους, οι ζώντες οργανισμοί παράγουν μία μεγάλη πληθώρα σημάτων, τα οποία συχνά καλύπτονται από την παρουσία άλλων σημάτων και από συνιστώσες θορύβου. Ο κύριος σκοπός της επεξεργασίας των βιολογικών σημάτων είναι το ξεκαθάρισμα του σήματος που μας ενδιαφέρει από τις συνιστώσες θορύβου και κατόπιν ο εντοπισμός των σημαντικών χαρακτηριστικών που περιέχονται σ' αυτό. Τέτοια χαρακτηριστικά πρέπει να είναι σημαντικά στο σχεδιασμό των ιατρικών αποφάσεων όπως, για παράδειγμα, στην επίλυση ενός ιατρικού προβλήματος ή στην κατανόηση της βασικής βιολογικής διεργασίας. Σύμφωνα με την άποψη αυτή, ο σκοπός της επεξεργασίας των βιολογικών σημάτων είναι η εξαγωγή της υπάρχουσας πληροφορίας και η απόκτηση γνώσης.

Η επεξεργασία των βιολογικών σημάτων αποτελείται συνήθως από τέσσερις τουλάχιστον φάσεις:

1. Ανάκτηση σημάτων χρησιμοποιώντας μετρήσεις ή παρατηρήσεις
2. Μετασχηματισμός και ελάττωση των σημάτων
3. Υπολογισμός των χαρακτηριστικών των σημάτων που είναι σημαντικά από διαγνωστική άποψη
4. Ερμηνεία ή ταξινόμηση των σημάτων

Στην πρώτη φάση της ανάκτησης των σημάτων χρησιμοποιούμε μετατροπείς για να δημιουργήσουμε ηλεκτρικά σήματα των οποίων η επεξεργασία μπορεί να επιτευχθεί με τη βοήθεια των υπολογιστών. Στη φάση αυτή τα χημικά ή τα μηχανικά σήματα μετατρέπονται σε ηλεκτρική μορφή τα οποία μπορούν να καταγραφούν με ηλεκτρόδια. Επίσης είναι πολύ σημαντικό να επιτύχουμε σήματα με χαμηλή διαταραχή, δηλ. ένα υψηλό λόγο σήματος προς θόρυβο. Τέλος, στην ίδια φάση τα σήματα μετατρέπονται σε ψηφιακά για να είναι δυνατή η επεξεργασία τους στον υπολογιστή.

Στη δεύτερη φάση επιθυμούμε να μετασχηματίσουμε τα σήματα με τέτοιο τρόπο ώστε να μπορέσουμε να εξάγουμε σημαντικά χαρακτηριστικά στην τρίτη φάση. Τα σήματα μερικές φορές περιέχουν πολύ περισσότερα δεδομένα από ότι είναι απαραίτητο για να εξάγουμε χαρακτηριστικά που δίνουν σημαντική πληροφορία. Για παράδειγμα, αν θέλουμε να διαγνώσουμε μπλοκάρισμα του κλάδου της αριστερής δέσμης (left bundle-branch block) από ένα ηλεκτροκαρδιογράφημα (ECG), χρειαζόμαστε για την ανάλυση μόνο το ένα από τα τρία επάρματα του ECG που καταγράφονται συνήθως. Για να διαγνώσουμε, όμως, ορισμένους τύπους καρδιακών αρρυθμιών, χρειάζονται μερικές φορές καταγραφές ECG πολλών ωρών. Επίσης μερικές φορές η ελάττωση των δεδομένων χρησιμοποιείται για να απομακρύνουμε τις συνιστώσες θορύβου με κατάλληλο φιλτράρισμα.

Στην τρίτη φάση διακρίνουμε τα σημαντικά χαρακτηριστικά που μπορούν να χρησιμοποιηθούν για τη λήψη περαιτέρω αποφάσεων. Τέτοια χαρακτηριστικά των σημάτων πρέπει να έχουν ισχύ διαχωρισμού για να μπορούμε να εξακριβώσουμε αν ο ασθενής έχει την ασθένεια Α ή την ασθένεια Β ή αν υπάρχει μία μεταβολή στην εξέλιξη της ασθένειας. Τα σημαντικά χαρακτηριστικά των σημάτων ξεχωρίζονται με τεχνικές επεξεργασίας σημάτων. Μόλις εξαχθούν τα χαρακτηριστικά των σημάτων χρησιμοποιούνται για τη λήψη αποφάσεων στη φάση της ερμηνείας.

Η φάση της ερμηνείας σχετικά με την επεξεργασία των βιολογικών σημάτων δεν διαφέρει ουσιαστικά από τις μεθόδους που χρησιμοποιούνται στην αναγνώριση προτύπων ή σε άλλες μεθόδους διάγνωσης. Οι μέθοδοι αυτοί μπορεί να έχουν λογική βάση, να στηρίζονται σε τεχνικές στατιστικής ανάλυσης ή να είναι συνδυασμός κάποιων τεχνικών.

7.1.1. Χαρακτηριστικά Βιολογικών Σημάτων

Τα Βιολογικά Σήματα παράγονται από βιολογικές διεργασίες που παρατηρούνται στην Ιατρική. Τέτοιες διεργασίες είναι πολύ πολύπλοκες και παρουσιάζουν δυναμικά χαρακτηριστικά. Τα βιολογικά σήματα είναι συνήθως μία συνάρτηση του χρόνου και συμβολίζονται με $x(t)$, όπου x είναι η τιμή του σήματος και t ο χρόνος. Μερικά σήματα μπορούν να περιγραφούν με λίγες παραμέτρους μόνο. Για παράδειγμα, μία ημιτονοειδής κυματομορφή ορίζεται ως εξής : $x(t)=A\sin(\omega t+\varphi)$. Μόνο τρεις παράμετροι, το πλάτος A, η συχνότητα ω και η φάση φ αρκούν για να περιγραφεί πλήρως το σήμα $x(t)$. Μόλις υπολογίσουμε τις παραμέτρους η κυματομορφή του σήματος καθορίζεται πλήρως. Όμως, όταν το σήμα $x(t)$ παραμορφώνεται από θόρυβο $\in(t)$, $s(t) = x(t) + \in(t)$, η συμπεριφορά του σήματος μπορεί να περιγραφεί μόνο με στατιστικές μεθόδους. Πλήρης πρόβλεψη της συμπεριφοράς του δεν είναι εφικτή. Τα βιολογικά σήματα μπορούν σπάνια να περιγραφούν από λίγες παραμέτρους μόνο, οι οποίες, εν γένει, επηρεάζονται σε σημαντικό βαθμό από το θόρυβο. Εάν οι βιολογικές διεργασίες που δημιουργούν τα σήματα βρίσκονται σε μία δυναμική κατάσταση, δηλ. αλλάζουν συνεχώς, η συμπεριφορά τους σπάνια μπορεί να προβλεφθεί επαρκώς. Οι παράμετροι που περιγράφουν τα σήματα αλλάζουν συνεχώς. Για παράδειγμα, στην περίπτωση ενός ασθενούς που βρίσκεται στην εντατική, οι παράμετροι που περιγράφουν τη λειτουργία της καρδιάς και το κυκλοφορικό, τη λειτουργία των πνευμόνων και την αναπνοή, τις διάφορες χημικές ουσίες του αίματος και του ορμονικού συστήματος μπορεί να μεταβάλλονται συνεχώς. Για το λόγο αυτό, τα σήματα που παράγονται από τέτοιες διεργασίες αντανακλούν το δυναμικό και μη-στάσιμο χαρακτήρα αυτών των διεργασιών.

Μια ταξινόμηση των βιολογικών σημάτων σύμφωνα με την κυματομορφή και τη στασιμότητα σχετίζεται με τον τύπο της διεργασίας από τον οποίο παράγονται, αν είναι δυναμικός ή όχι και αν παρουσιάζει μία επαναληπτική συμπεριφορά ή όχι.

7.1.2. Στοχαστικές Διεργασίες

Με τον όρο «στοχαστική διεργασία» θεωρούμε ένα στατιστικό φαινόμενο το οποίο εξελίσσεται στο χρόνο σύμφωνα με πιθανοκρατικούς νόμους. Στη μαθηματική ορολογία, η στοχαστική διεργασία ή ανέλιξη ορίζεται ως μία συλλογή τυχαίων μεταβλητών $\{x(t),\ t \in T\}$, όπου T είναι ένα σύνολο σημείων στα οποία η διεργασία ορίζεται. Θα σημειώνουμε την τυχαία μεταβλητή τη χρονική στιγμή t με $x(t)$ αν T είναι ένα συνεχές σύνολο (συνήθως $-\infty < t < +\infty$) και x_t αν T είναι ένα σύνολο διακριτών τιμών (συνήθως $t=0,\ \pm 1,\ \pm 2, \ldots$). Επειδή το T είναι ένα σύνολο χρονοσημείων η στοχαστική διεργασία είναι γνωστή με την ονομασία χρονοσειρά. Συνήθως μία χρονοσειρά ανήκει σε ένα άπειρο σύνολο χρονοσειρών που θα ήταν δυνατό να είχε παρατηρηθεί. Κάθε μέλος του συνόλου των χρονοσειρών αποτελεί μία δυνατή πραγμάτωση ή δειγματοσυνάρτηση της στοχαστικής διεργασίας. Στις πρακτικές εφαρμογές σημειώνουμε τη δειγματοσυνάρτηση με $x(t)$ για $0 \leq t \leq T$ αν οι παρατηρήσεις είναι συνεχείς και με x_t *για* $t=1, 2, \ldots, N$ αν είναι διακριτές.

Ένας τρόπος περιγραφής μιας χρονοσειράς είναι ο προσδιορισμός της από κοινού συνάρτησης κατανομής των $X(t_1), X(t_2), \ldots, X(t_n)$ για τις χρονικές στιγμές $t_1, t_2, \ldots, t_n$ και για κάθε τιμή του n. Αλλά αυτό είναι πολύπλοκο και αποφεύγεται στην πράξη. Μία απλούστερη, περισσότερο χρήσιμη μέθοδος περιγραφής μιας χρονοσειράς είναι να υπολογίσουμε τις ροπές (*moments*) της χρονοσειράς, ειδικότερα την πρώτη ροπή και τη δεύτερη ροπή, οι οποίες είναι γνωστές ως μέση τιμή, διασπορά και συνδιασπορά της χρονοσειράς αντίστοιχα.

Συγκεκριμένα έχουμε:

(α) Η ***μέση τιμή***, $\mu(t)$, ορίζεται ως εξής

$$\mu(t) = E\{X(t)\}$$

(β) Η ***συνάρτηση διασποράς***, $\sigma^2(t)$, ορίζεται από τη σχέση

$$\sigma^2(t) = Var\,\{X(t)\}$$

(γ) Η ***συνάρτηση αυτοδιασποράς***, $\gamma(t_1, t_2)$, δίνεται από τον τύπο

$$\gamma(t_1,t_2) = cov\{X(t_1), X(t_2)\} = E\{[\,X(t_1) - \mu(t_1)][\,X(t_2) - \mu(t_2)]\}$$

(δ) Ο **συντελεστής αυτοσυσχέτισης**, $\rho(t_1, t_2)$, δίνεται από τη σχέση

$$\rho(t_1,t_2) = \gamma(t_1,t_2) \Big/ \sqrt{Var\{x(t_1\}\ Var\{x(t_2\}}$$

Η συνάρτηση διασποράς είναι ειδική περίπτωση της συνάρτησης αυτοδιασποράς όταν θέσουμε $t_1=t_2$. Ροπές μεγαλύτερης τάξης μπορούν να ορισθούν μ’ έναν ανάλογο τρόπο, αλλά χρησιμοποιούνται αραιά στην πράξη, επειδή ο υπολογισμός των συναρτήσεων $\mu(t)$ και $\gamma(t_1,t_2)$ είναι συχνά επαρκής. θα ασχοληθούμε κυρίως με στάσιμες χρονοσειρές που περιγράφονται κατωτέρω και θα δούμε πως μπορούν να εφαρμοσθούν στη μελέτη των βιολογικών σημάτων.

7.2. Στάσιμες χρονοσειρές

Μία σπουδαία ομάδα χρονοσειρών είναι εκείνες που είναι στάσιμες (*stationary*). Μία χρονοσειρά λέγεται αυστηρώς στάσιμη αν η από κοινού συνάρτηση κατανομής (*joint distribution*) των $X(t_1),...,X(t_n)$ είναι η ίδια με την από κοινού συνάρτηση κατανομής των $X(t_1+\tau),...,X(t_n+\tau)$ για όλα τα $t_1,t_2,\dots,t_n$, τ. Με άλλα λόγια, μετατοπίζοντας την αρχή του χρόνου κατά μία ποσότητα τ δεν μεταβάλλουμε την από κοινού συνάρτηση κατανομής των $X(t_1),...,X(t_n)$, η οποία, γι' αυτό το λόγο, πρέπει να εξαρτάται μόνο από τα διαστήματα μεταξύ των $t_1,...,t_n$. Ο παραπάνω ορισμός ισχύει για κάθε τιμή του n. Ειδικότερα, εάν $n=1$, συνεπάγεται ότι η συνάρτηση κατανομής της $X(t)$ πρέπει να είναι ίδια για όλα τα t, έτσι ώστε

$$\mu(t) = \mu \quad \text{και} \quad \sigma^2(t) = \sigma^2$$

δηλ. η μέση τιμή και η διασπορά είναι σταθερές και δεν εξαρτώνται από την τιμή του t. Περαιτέρω, αν $n=2$, η από κοινού συνάρτηση κατανομής των $X(t_1)$ και $X(t_2)$ εξαρτάται μόνο από τη διαφορά t_2-t_1, η οποία καλείται καθυστέρηση (*lag*). Έτσι η συνάρτηση αυτοδιασποράς εξαρτάται μόνο από τη διαφορά t_2-t_1 και μπορεί να γραφεί ως εξής:

$$\gamma(\tau) = E\{[X(t+\tau)-\mu][X(t)-\mu]\}$$

Η συνάρτηση $\gamma(\tau)$ ονομάζεται ***συνάρτηση αυτοδιασποράς*** με καθυστέρηση τ. Όταν η μέση τιμή της χρονοσειράς είναι μηδέν τότε η συνάρτηση αυτοδιασποράς είναι γνωστή ως συνάρτηση αυτοσυσχέτισης (*autocorrelation function*).

Η κανονικοποιημένη συνάρτηση αυτοδιασποράς μας δίνει το ***συντελεστή αυτοσυσχέτισης*** (*autocorrelation coefficient*), ο οποίος ορίζεται ως εξής :

$$\rho(\tau) = \gamma(\tau) \,/\, \gamma(0)$$

Ο συντελεστής αυτός περιγράφει τη συσχέτιση μεταξύ των σημείων $X(t)$ και $X(t+\tau)$ της χρονοσειράς, που απέχουν μεταξύ τους τ βήματα.

Δευτέρας-τάξης στασιμότητα. Στην πράξη είναι συχνά χρήσιμο να ορίσουμε τη στασιμότητα μ' ένα λιγότερο περιοριστικό τρόπο απ' αυτόν που περιγράψαμε παραπάνω. Μια χρονοσειρά καλείται δευτέρας-τάξης στάσιμη (ή ασθενώς στάσιμη), αν η μέση τιμή είναι σταθερά και αν η αυτοδιασπορά εξαρτάται μόνο από την καθυστέρηση (lag), έτσι ώστε

$$E\{X(t)\}=\mu \quad \text{και} \quad Cov\{X(t+\tau), X(t)\}=\gamma(\tau)$$

Προϋποθέσεις για ροπές μεγαλύτερης τάξης απ' αυτές της δευτέρας τάξης δε γίνονται. Από τον ορισμό αυτό συνεπάγεται ότι η διασπορά είναι σταθερά. Ακόμη πρέπει να σημειωθεί ότι η μέση τιμή και η διασπορά είναι πεπερασμένες.

Η κανονική κατανομή πολλών μεταβλητών είναι τελείως καθορισμένη από την πρώτη και δεύτερη ροπή, και επομένως από τις συναρτήσεις μ και $\gamma(\tau)$. Η παρατήρηση αυτή δείχνει ότι η στασιμότητα δευτέρας τάξης συνεπάγεται αυστηρή στασιμότητα στην περίπτωση χρονοσειρών με κανονικές κατανομές. Όμως για χρονοσειρές που δεν ακολουθούν κανονικές κατανομές, οι συναρτήσεις μ και $\gamma(\tau)$ μπορεί να μη περιγράφουν επαρκώς τη χρονοσειρά.

7.3. Ο συντελεστής αυτοσυσχέτισης

Στην παράγραφο αυτή εξετάζουμε τις ιδιότητες του συντελεστή αυτοσυσχέτισης. Υποθέτουμε ότι η στάσιμη χρονοσειρά $X(t)$ έχει μέση τιμή μ, διασπορά σ^2, και συνάρτηση αυτοδιασποράς $\gamma(\tau)$. Τότε

$$\rho(\tau)=\frac{\gamma(\tau)}{\gamma(0)}=\frac{\gamma(\tau)}{\sigma^2}$$

Ας σημειωθεί ότι $\rho(0)=1$.

Πρώτη ιδιότητα: Ο συντελεστής αυτοσυσχέτισης είναι μία άρτια συνάρτηση ως προς την καθυστέρηση, δηλ.

$$\rho(\tau) = \rho(-\tau)$$

Από τον ορισμό της $\gamma(\tau)$ έχουμε

$$\gamma(\tau) = cov\{X(t+\tau), X(t)\} = cov\ \{X(t'), X(t'-\tau)\} = \gamma(-\tau)\ (\text{θέτουμε } t'=t+\tau)$$

και επομένως συνεπάγεται ότι $\rho(\tau)=\rho(-\tau)$.

Δευτέρα ιδιότητα: Για κάθε τιμή του τ έχουμε ότι

$$|\ \rho(\tau)| \leq 1$$

Η ιδιότητα αυτή αποδεικνύεται αν λάβουμε υπόψη μας ότι

$$Var\ [\lambda_1 X(t)+\lambda_2 X(t+\tau)] \geq 0 \qquad \text{για όλες τις σταθερές } \lambda_1, \lambda_2.$$

Από τον ορισμό της διασποράς προκύπτει ότι :

$$\lambda_1^2\, Var\{X(t)\}+\lambda_2^2\, \{Var\{X(t+\tau)\}+2\lambda_1\lambda_2 cov\{X(t), X(t+\tau)\}=$$

$$(\lambda_1^2+\lambda_2^2)\sigma^2+2\lambda_1\lambda_2\gamma(\tau) \geq 0 \Rightarrow \lambda_1^2+\lambda_2^2+\lambda_1\lambda_2[\gamma(\tau)/\sigma^2] \geq 0 \Rightarrow \lambda_1^2+\lambda_2^2+\lambda_1\lambda_2\rho(\tau) \geq 0$$

Αν θεωρήσουμε ότι έχουμε ένα διώνυμο ως προς λ_1, για να ισχύει η παραπάνω ανισότητα πρέπει η διακρίνουσα του διωνύμου να είναι μικρότερη ή ίση με το μηδέν, δηλ.

$$4\rho^2(\tau)-4 \leq 0 \Rightarrow \rho^2(\tau) \leq 1 \Rightarrow |\ \rho(\tau)| \leq 1$$

Τρίτη ιδιότητα *(Μη-μοναδικότητα)*: Για ένα συγκεκριμένο συντελεστή αυτοσυσχέτισης αντιστοιχεί μόνο μία δυνατή στάσιμη χρονοσειρά με κανονική κατανομή, επειδή η στάσιμη κανονική χρονοσειρά είναι τελείως καθορισμένη από τη μέση τιμή, τη διασπορά και το συντελεστή αυτοσυσχέτισης. Όμως, είναι πάντοτε δυνατό να βρεθούν πολλές χρονοσειρές που δεν είναι κανονικές με τον ίδιο συντελεστή αυτοσυσχέτισης και αυτό δημιουργεί μια περαιτέρω δυσκολία στην ερμηνεία του συντελεστή αυτοσυσχέτισης μιας χρονοσειράς. Οι *Jenkins and Watts (1968, σελ. 170)* δίνουν ένα παράδειγμα δύο διαφορετικών χρονοσειρών που έχουν τον ίδιο συντελεστή αυτοσυσχέτισης.

7.4. Μερικές χρήσιμες χρονοσειρές

Στην παράγραφο αυτή θα εξετάσουμε διάφορα μοντέλα που είναι κατάλληλα για την περιγραφή μιας χρονοσειράς. Αρχίζουμε με την περιγραφή μίας καθαρώς τυχαίας διεργασίας.

7.4.1. Η καθαρώς τυχαία διεργασία

Η διακριτή διεργασία $\{ Z_t \}$ καλείται μία καθαρώς τυχαία διεργασία αν οι τυχαίες μεταβλητές Z_t αποτελούν μία ακολουθία αμοιβαία ανεξαρτήτων τυχαίων μεταβλητών με την ίδια κατανομή. Από τον ορισμό προκύπτει ότι η μέση τιμή και η διασπορά είναι σταθερές και ότι

$$\gamma(k) = cov\{ Z_t, Z_{t+k}\} = 0 \qquad \text{για } k=\pm 1, \pm 2,\ldots$$

Καθώς η μέση τιμή και η αυτοδιασπορά είναι ανεξάρτητες από το χρόνο, η διεργασία είναι δεύτερης τάξης στάσιμη. Στην πραγματικότητα η διεργασία είναι επίσης και αυστηρώς στάσιμη. Ο συντελεστής αυτοσυσχέτισης δίνεται από τον τύπο

$$\rho(k) = \begin{cases} 1 & k = 0 \\ 0 & k = \pm 1, \ \pm 2,\ldots \end{cases}$$

Μία καθαρώς τυχαία διεργασία καλείται μερικές φορές λευκός θόρυβος (*white noise*) ιδιαίτερα από τους μηχανικούς. Διεργασίες του τύπου αυτού είναι χρήσιμες σε πολλές περιπτώσεις, όπως, για παράδειγμα, στα μοντέλα κινητού μέσου όρου (*moving average*) που θα εξετάσουμε παρακάτω.

Υπάρχει μια αμφιβολία όσον αφορά στη δυνατότητα ύπαρξης μιας καθαρώς τυχαίας διεργασίας σε συνεχή χρόνο. Μια τέτοια διεργασία θα είχε ένα συντελεστή αυτοσυσχέτισης που δίνεται από τον τύπο

$$\rho(k) = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

και η οποία προφανώς δεν είναι συνεχής συνάρτηση. Μπορεί να αποδειχτεί ότι μια τέτοια διεργασία θα έχει άπειρη διασπορά, και επομένως δεν είναι πραγματικό φαινόμενο. Όμως, αποδεικνύεται ότι αποτελεί ένα χρήσιμο μαθηματικό εργαλείο για θεωρητικούς σκοπούς, και επίσης είναι μια προσέγγιση ορισμένων διεργασιών που συμβαίνουν στην πράξη, όπως οι μεταβολές δυναμικού σ' ένα αγωγό που οφείλονται σε θερμικό θόρυβο, και οι ωθήσεις που εξασκούνται σ' ένα σωματίδιο μετεωρούμενο σε υγρό κατά μήκος ενός δεδομένου άξονα για την παραγωγή της κίνησης *Brown*.
Ένας λευκός θόρυβος μπορεί να προσεγγιστεί από μία διεργασία σε συνεχή χρόνο με συντελεστή αυτοσυσχέτισης $\rho(\tau) = e^{-\lambda|\tau|}$. Η συνάρτηση αυτή γίνεται μικρή πολύ γρήγορα καθώς $\lambda \to \infty$ (Βλέπε *Cox and Miller, 1968*).

7.4.2. **Τυχαίος δρόμος (*random walk*)**

Υποθέτουμε ότι $\{ Z_t \}$ είναι μία διακριτή, καθαρώς τυχαία διεργασία με μέση τιμή μ και διασπορά σ_z^2. Μία χρονοσειρά $\{X_t\}$ λέγεται ότι είναι τυχαίος δρόμος αν

$$X_t = X_{t-1}+Z_t$$

Επειδή η χρονοσειρά X_t συνήθως αρχίζει στο μηδέν θα έχουμε ότι

$$X_1 = Z_1 \; και \; X_t = \sum_{i=1}^{t} Z_t$$

Από τις σχέσεις αυτές βρίσκουμε ότι

$$E\{X_t\} = t\mu \qquad και \qquad Var\{X_t\} = t\sigma_z^2$$

Καθώς η μέση τιμή και η διασπορά μεταβάλλονται με το χρόνο *t*, η χρονοσειρά δεν είναι στάσιμη. Όμως είναι ενδιαφέρον να σημειώσουμε ότι οι πρώτες διαφορές ενός τυχαίου δρόμου, που δίνονται από τη σχέση

$$Z_t = X_t - X_{t-1}$$

σχηματίζουν μία καθαρώς τυχαία διεργασία, η οποία ως γνωστό είναι στάσιμη.

Τα πιο γνωστά παραδείγματα χρονοσειρών που συμπεριφέρονται ως τυχαίοι δρόμοι είναι ορισμένα προβλήματα στην κατεύθυνση της βιοπληροφορικής καθώς και οι τιμές των μετοχών.

7.4.3. Μοντέλα κινητού μέσου όρου

Υποθέτουμε ότι $\{ Z_t \}$ είναι μία καθαρώς τυχαία διεργασία με μέση τιμή 0 και διασπορά σ_z^2. Η χρονοσειρά $\{ X_t \}$ θα λέμε ότι περιγράφεται από ένα μοντέλο κινητού μέσου όρου τάξης q, $MA(q)$ αν

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q} \tag{7.1}$$

όπου $\{ \beta_i \}$ είναι σταθερές. Οι μεταβλητές Z_t συνήθως λαμβάνονται με τέτοιο τρόπο έτσι ώστε $\beta_0 = 1$. Προκύπτει αμέσως από την *(7.1)* ότι

$$E\{X_t\} = 0 \quad \kappa\alpha\iota \quad Var\{X_t\} = \sigma_z^2 \sum_{j=0}^{q} \beta_\iota^2$$

επειδή οι μεταβλητές Z_t είναι ανεξάρτητες μεταξύ τους. Επίσης έχουμε:

$$\gamma(\kappa) = cov\{ X_t \ X_{t+k}\} = cov(\beta_0 Z_t + \ldots + \beta_q Z_{t-q} ,\ \beta_0 Z_{t+k} + \ldots + \beta_q Z_{t+k-q})$$

$$\Rightarrow \gamma(k) = \begin{cases} 0 & k > q \\ \sigma_z^2 \sum_{j=0}^{q-k} \beta_j \beta_{j+k} & k = 0, 1, \ldots, q \\ \gamma(-k) & k < 0 \end{cases}$$

Καθώς η *γ(k)* δεν εξαρτάται από το χρόνο *t*, και η μέση τιμή είναι σταθερά, η χρονοσειρά είναι δεύτερης – τάξης στάσιμη για όλες τις τιμές των $\{ \beta_i \}$. Επιπλέον αν οι μεταβλητές Z_t ακολουθούν μία κανονική κατανομή, την ίδια κατανομή ακολουθούν και οι μεταβλητές X_t, και επομένως έχουμε μία κανονική κατανομή τελείως καθορισμένη από τη μέση τιμή και τη συνάρτηση διασποράς. Η χρονοσειρά είναι τότε αυστηρώς στάσιμη.

Ο συντελεστής αυτοσυσχέτισης για ένα μοντέλο *MA(q)* δίνεται από τη σχέση:

$$\rho(k) = \begin{cases} 0 & k > q \\ \sum_{j=0}^{q-k} \beta_j \beta_{j+k} \Big/ \sum_{j=0}^{q} \beta_j^2 & k = 1, \ldots, q \\ 0 & k > q \\ \rho(-k) & k < 0 \end{cases}$$

Ας σημειωθεί ότι ο συντελεστής αυτοσυσχέτισης κόβεται απότομα στην καθυστέρηση *q*, το οποίο είναι ιδιαίτερο χαρακτηριστικό των *MA* μοντέλων. Ειδικότερα για το *MA(1)* μοντέλο με $\beta_0 = 1$ ο συντελεστής αυτοσυσχέτισης δίνεται από τον τύπο:

$$\rho(k) = \begin{cases} 1 & k = 0 \\ \beta_1 / (1 + \beta_1^2) & k = \pm 1 \\ 0 & \text{διαφορετικα} \end{cases}$$

Παρότι δεν υπάρχουν περιορισμοί στις παραμέτρους $\{ \beta_i \}$ για να είναι το *MA* μοντέλο στάσιμο, οι *Box and Jenkins (1970, σελ. 50)* προτείνουν περιορισμούς στις παραμέτρους $\{ \beta_i \}$ για να εξασφαλίσουν αυτό που αποκαλούν αντιστρεψιμότητα. Θεωρούμε τα εξής *MA* μοντέλα πρώτης-τάξης:

$$\text{I. } X_t = Z_t + \theta Z_{t-1} \qquad \text{και} \qquad \text{II. } X_t = Z_t + (1/\theta)Z_{t-1}$$

Μπορεί να δειχτεί εύκολα ότι αυτά τα δύο διαφορετικά μοντέλα έχουν ακριβώς τον ίδιο συντελεστή αυτοσυσχέτισης. Έτσι δεν μπορούμε να προσδιορίσουμε ένα *MA* μοντέλο μοναδικά από ένα δεδομένο συντελεστή αυτοσυσχέτισης. Αν όμως εκφράσουμε τα μοντέλα I και II θέτοντας τις μεταβλητές Z_t σαν συναρτήσεις των X_t, $X_{t-1}, \ldots,$ τότε βρίσκουμε με διαδοχικές αντικαταστάσεις ότι:

$$\text{I. } Z_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} - \ldots$$

$$\text{II. } Z_t = X_t - (1/\theta)X_{t-1} + (1/\theta^2)X_{t-2} - \ldots$$

Αν $|\theta| < 1,$ η σειρά I συγκλίνει ενώ η σειρά II δεν συγκλίνει. Έτσι το σωστό μοντέλο είναι το μοντέλο I. Επίσης αν $|\theta| < 1,$ το μοντέλο I λέμε ότι είναι αντιστρέψιμο, ενώ το μοντέλο II δεν είναι. Η επιβολή της συνθήκης αντιστρεψιμότητας εξασφαλίζει ότι υπάρχει ένα μοναδικό μοντέλο για ένα δεδομένο συντελεστή αυτοσυσχέτισης.

Η εξίσωση (7.1) μπορεί να γραφεί ως εξής:

$$X_t = (\beta_0 + \beta_1 B + \ldots + \beta_q B^q)Z_t = \Gamma(B)Z_t$$

όπου $\Gamma(B)$ είναι ένα πολυώνυμο τάξης q ως προς B, και B είναι ο τελεστής που μεταθέτει προς τα πίσω, δηλ. $B^j X_t = X_{t-j}$ για όλα τα j. Ένα μοντέλο τάξης q είναι αντιστρεπτό αν οι ρίζες της εξισώσεως (θεωρούμε το B ως μία μιγαδική μεταβλητή και όχι ως ένα τελεστή):

$$\Gamma(B) = \beta_0 + \beta_1 B + \ldots + \beta_q B^q = 0$$

όλες βρίσκονται έξω από το μοναδιαίο κύκλο.

Σημειώνουμε επίσης ότι μία αυθαίρετη σταθερά μ μπορεί να προστεθεί στη δεξιά μεριά της εξισώσεως (7.1) στην περίπτωση που η χρονοσειρά μας έχει μέση τιμή μ. Αυτό δεν επηρεάζει το συντελεστή αυτοσυσχέτισης και έχει παραλειφθεί για απλούστευση.

7.4.4. Αυτοπαλινδρομικά μοντέλα (*Autoregressive models*)

Υποθέτουμε ότι $\{ Z_t \}$ είναι μία καθαρώς τυχαία διεργασία με μέση τιμή 0 και διασπορά σ_z^2. Τότε μία χρονοσειρά $\{ X_t \}$ περιγράφεται από ένα αυτοπαλινδρομικό μοντέλο τάξης p αν

$$X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + Z_t \tag{7.2}$$

Αυτό το μοντέλο ομοιάζει με ένα πολλαπλό παλινδρομικό μοντέλο, αλλά η μεταβλητή X_t δεν εξαρτάται από άλλες ανεξάρτητες μεταβλητές αλλά από προηγούμενες τιμές του X_t . Ένα αυτοπαλινδρομικό μοντέλο τάξης p μπορεί να γραφεί με σύντομο τρόπο σαν ένα $AR(p)$ μοντέλο.

7.4.4.1. Το αυτοπαλινδρομικό μοντέλο πρώτης-τάξης

Για απλούστευση, αρχίζουμε με την εξέταση του $AR\ (1)$ μοντέλου. Η εξίσωση του μοντέλου δίνεται από τη σχέση:

$$X_t = \alpha X_{t-1} + Z_t \tag{7.3}$$

Μετά από διαδοχικές αντικαταστάσεις στην (7.3) έχουμε

$$X_t = \alpha[\alpha X_{t-2} + Z_{t-1}] + Z_t = \alpha^2[\alpha X_{t-3} + Z_{t-2}] + \alpha Z_{t-1} + Z_t = \ldots$$
$$= Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \ldots \qquad (\text{αρκεί } -1 < \alpha < 1)$$

δηλ. η χρονοσειρά X_t μπορεί να εκφραστεί ως ένα MA μοντέλο άπειρης τάξης.

Αντί να χρησιμοποιήσουμε διαδοχικές αντικαταστάσεις στον παραπάνω υπολογισμό, είναι απλούστερο να χρησιμοποιήσουμε τη μέθοδο με τον τελεστή B που μεταθέτει προς τα πίσω. Στην περίπτωση αυτή η σχέση (7.3) παίρνει την εξής μορφή:

$$(1-\alpha B)X_t = Z_t, \quad \text{έτσι ώστε}$$

$$X_t = (1\text{-}\alpha B)^{-1}\, Z_t = (1+\alpha B+\alpha^2 B^2+\ldots)Z_t = Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \ldots$$

Από τη σχέση αυτή είναι εύκολο να βρούμε ότι

$$E\{X_t\} = 0 \quad \text{και} \quad Var\{X_t\} = \sigma_z^2 (1+\alpha^2 + \alpha^4 + \ldots)$$

Η διασπορά θα είναι πεπερασμένη αρκεί $|\,a\,| < 1$ οπότε

$$Var\{X_t\} = \sigma_z^2 / \left(1 - \alpha^2\right) = \sigma_x^2$$

Η συνάρτηση αυτοδιασποράς υπολογίζεται από την εξής σχέση

$$\gamma(k) = E\{X_t X_{t+k}\} = E\left\{\left[\sum_{i=0}^{\infty} \alpha^i Z_{t-i}\right]\left[\sum_{j=0}^{\infty} \alpha^j Z_{t+k-j}\right]\right\} = \sigma_z^2 \sum_{i=0}^{\infty} \alpha^i \alpha^{k+i} \qquad \forall\ k \geq 0$$

Η συνάρτηση αυτή συγκλίνει για $|\,a\,| < 1$ οπότε έχουμε

$$\gamma(k) = \alpha^k [\sigma_z^2 / (1 - a^2)] = a^k \sigma_x^2$$

Για αρνητικές τιμές του k θα ισχύει ότι $\gamma(k) = \gamma(-k)$.

Επειδή η συνάρτηση $\gamma(k)$ δεν εξαρτάται από το χρόνο t, μία χρονοσειρά X_t που περιγράφεται από ένα *AR(1)* μοντέλο θα είναι δεύτερης - τάξης στάσιμη, αρκεί $|\,a\,| < 1$.

Ο συντελεστής αυτοσυσχέτισης δίνεται από την εξής σχέση:

$$\rho(k) = a^{|k|} \qquad (k = 0, \pm 1, \pm 2, \ldots)$$

Η απόλυτη τιμή στον εκθέτη δικαιολογείται από το γεγονός ότι η $\rho(k)$ είναι άρτια συνάρτηση.

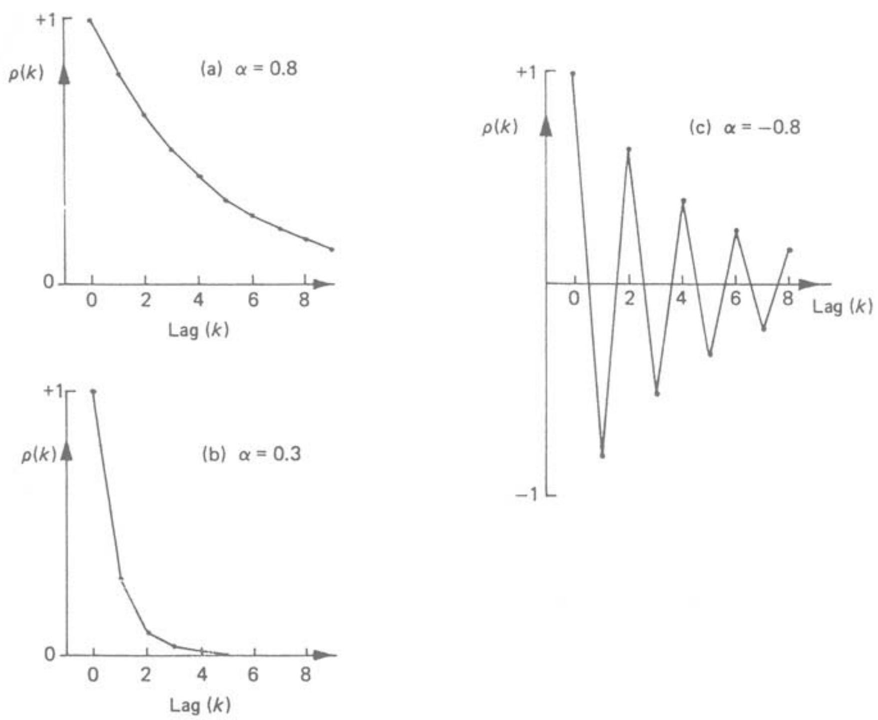
Ο υπολογισμός του συντελεστή αυτοσυσχέτισης είναι δυνατό να γίνει με ένα διαφορετικό τρόπο αν θεωρήσουμε εκ των προτέρων ότι η χρονοσειρά είναι στάσιμη με μέση τιμή 0. Πολλαπλασιάζοντας την (7.3) με X_{t-k} και παίρνοντας μέσες τιμές βρίσκουμε για κάθε $k > 0$, ότι

$$\gamma(-k)= a\gamma(-k+1)$$

Θεωρήσαμε ότι $E\{ Z_t X_{t-k}\}= 0$ για κάθε $k > 0$.
Επειδή η $\gamma(k)$ είναι μία άρτια συνάρτηση, πρέπει επίσης να ισχύει ότι

$$\gamma(k)= a\gamma(k-1) \quad \text{για κάθε } k > 0$$

Για k=0 έχουμε $\gamma(0)= \sigma_x^2$ και επομένως $\gamma(k)= a^k\ \sigma_x^2\ \ (k \geq 0)$. Ο συντελεστής αυτοσυσχέτισης θα δίνεται από τη σχέση $\rho(k)= a^k$ για κάθε $k\geq 0$. Επιπλέον επειδή $|\rho(k)|\leq 1$, έχουμε ότι $|a|\leq 1$. Αλλά αν $|a|=1$, τότε $|\rho(k)|=1$ για όλα τα k, η οποία είναι μία εκφυλισμένη περίπτωση. Επομένως για μία κατάλληλη στάσιμη χρονοσειρά πρέπει να ισχύει ότι $|a|<1$.



Σχήμα 7.1. Τρία παραδείγματα συντελεστών αυτοσυσχέτισης ενός αυτοπαλινδρομικού AR (1) μοντέλλου (a) a=0.8 (b) a=0.3 και (c) a= –0.8.

Η παραπάνω μέθοδος επίτευξης του συντελεστή αυτοσυσχέτισης χρησιμοποιείται συχνά, παρ' ότι χρειάζεται την προϋπόθεση της συνθήκης στασιμότητας.

Τρία παραδείγματα για το συντελεστή αυτοσυσχέτισης μιας σειράς που περιγράφεται από ένα *AR(1)* μοντέλο δίνονται στο Σχ.7.1 για *a= 0.8, 0.3, -0.8.* Είναι φανερό ότι ο συντελεστής αυτοσυσχέτισης μικραίνει γρήγορα για $a = 0.3$ και εναλλάσσεται όταν το a είναι αρνητικό.

7.4.4.2. Το αυτοπαλινδρομικό μοντέλο δεύτερης-τάξης

Το μοντέλο αυτό δίνεται από την εξής εξίσωση

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + Z_t \qquad (7.4)$$

Χρησιμοποιώντας τον τελεστή που μεταθέτει προς τα πίσω βρίσκουμε

$$(1-\alpha_1\beta-\alpha_2\chi^2)X_t = Z_t \quad \text{ή} \quad X_t = \left(1\text{-}\alpha_1 B - \alpha_2 \mathrm{B}^2\right)^{-1} Z_t = f(B)\,Z_t,$$

όπου $f(B)=(1-\alpha_1 B-\alpha_2 B^2)^{-1}=1+c_1 B+c_2 B^2+\ldots$

Η σχέση μεταξύ των παραμέτρων α και c είναι εύκολο να υπολογιστεί

Έχοντας εκφράσει την X_t σαν ένα *MA* μοντέλο βρίσκουμε ότι $E\{X_t\}=0$.

Η διασπορά είναι πεπερασμένη αρκεί το άθροισμα $\sum_{i=0}^{\infty} c_i^2$ να συγκλίνει και αυτό είναι μία αναγκαία συνθήκη η χρονοσειρά να είναι στάσιμη. Η συνάρτηση αυτοδιασποράς δίνεται από τη σχέση:

$$\gamma(k) = \sigma_z^2 \sum_{i=0}^{\infty} c_i c_{i+k} \quad (\text{οπου } c_0 = 1)$$

Μία ικανή συνθήκη γι' αυτή τη συνάρτηση να συγκλίνει, και επομένως η χρονοσειρά να είναι στάσιμη, είναι η σειρά $\sum_{i=0}^{\infty}|c_i|$ να συγκλίνει.

Ένας εναλλακτικός τρόπος είναι να θεωρήσουμε ότι η χρονοσειρά είναι στάσιμη και να πολλαπλασιάσουμε την (7.4) με X_{t-k}, να πάρουμε μέσες τιμές και να διαιρέσουμε με σ_x^2 υποθέτοντας ότι η διασπορά της X_t είναι πεπερασμένη. Στην περίπτωση αυτή, επειδή $\rho(k)=\rho(-k)$ για όλα τα k, βρίσκουμε ότι

$$\rho(k)=\alpha_1\rho(k-1)+\alpha_2\rho(k-2) \qquad \text{για κάθε} \quad k>0$$

Η εξίσωση αυτή καλείται εξίσωση *Yule - Walker*. Η γενική λύση της εξίσωσης που είναι μία εξίσωση διαφορών δίνεται από τη σχέση

$$\rho(k)=A_1\pi_1^{|k|}+A_2\pi_2^{|k|} \tag{7.5}$$

όπου $\{\pi_i\}$, i=1, 2 είναι οι ρίζες της εξίσωσης

$$y^2-a_1y-a_2=0$$

Οι σταθερές $\{A_i\}$ λαμβάνονται κατά τέτοιο τρόπο ώστε να ικανοποιούν την αρχική συνθήκη $\rho(0)=1$.

Από την (7.5) είναι φανερό ότι η $\rho(k)$ τείνει στο μηδέν καθώς το k αυξάνει, αρκεί τα $|\pi_i|<1$ $(i=1, 2)$, και αυτό είναι μία αναγκαία και ικανή συνθήκη για να είναι η χρονοσειρά στάσιμη.

Από την εξίσωση $y^2-a_1y-a_2=0$, αν $|\pi_i|<1$, $i=1, 2$ έχουμε:

$$\left|\frac{\alpha_1\pm\sqrt{\alpha_1^2+4\alpha_2}}{2}\right|<1$$

Μπορεί να δειχτεί ότι η περιοχή στην οποία η χρονοσειρά είναι στάσιμη είναι μία τριγωνική περιοχή που ικανοποιεί τις σχέσεις:

$$a_1 + a_2 < 1$$
$$a_2 - a_1 < 1$$
$$a_2 > -1$$

Οι ρίζες π_1 , π_2 είναι πραγματικές αν $\alpha_1^2 + 4\alpha_2 \geq 0$, οπότε ο συντελεστής αυτοσυσχέτισης ελαττώνεται εκθετικά με το k, ενώ οι ρίζες είναι μιγαδικές αν $\alpha_1^2 + 4\alpha_2 < 0$ οπότε βρίσκουμε ότι ο συντελεστής αυτοσυσχέτισης είναι ένα αποσβεννύμενο συνημιτονοειδές κύμα. Όταν οι ρίζες είναι πραγματικές ή μιγαδικές οι σταθερές A_1, A_2 υπολογίζουμε ως εξής:

Επειδή $\rho(0) = 1$ έχουμε $A_1 + A_2 = 1$.

Επιπλέον από την εξίσωση *Yule-Walker* έχουμε

$$\rho(1)=\alpha_1 \rho(0) + \alpha_2 \rho(-1)= \alpha_1 + \alpha_2 \rho(1) \Rightarrow \rho(1)=\alpha_1/(1-\alpha_2)=A_1\pi_1 + A_2\pi_2$$

$$\Rightarrow \rho(1) = A_1\pi_1 + (1-A_1)\pi_2 = \alpha_1/(1-\alpha_2)$$

Επομένως

$$A_1 = \frac{\left[\frac{\alpha_1}{1-\alpha_2} - \pi_2\right]}{\pi_1 - \pi_2} \quad \kappa\alpha\iota \quad A_2 = 1 - A_1$$

Ένας ισοδύναμος τρόπος έκφρασης της συνθήκης στασιμότητας είναι να θεωρήσουμε ότι οι ρίζες της εξίσωσης

$$H(B)=1-\alpha_1 B - \alpha_2 B^2=0$$

βρίσκονται έξω από το μοναδιαίο κύκλο.

7.4.4.3. Το αυτοπαλινδρομικό μοντέλο τάξης p

Στην περίπτωση αυτή η εξίσωση του μοντέλου δίνεται από τη σχέση:

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \ldots + \alpha_p X_{t-p} + Z_t$$

Χρησιμοποιώντας τα επιχειρήματα που παραθέσαμε για το μοντέλο *AR(2)* βρίσκουμε ότι

$$\rho(k)=\alpha_1\,\rho(k-1)+\ldots+\alpha_p\,\rho(k-p) \quad \text{για κάθε} \quad k > 0$$

Η λύση της εξίσωσης αυτής έχει τη μορφή

$$\rho(k)=A_1\,\pi_1^{|k|}+\ldots+A_p\,\pi_p^{|k|}$$

όπου $\{\pi_i\}$ είναι οι ρίζες της εξίσωσης

$$y^p - a_1 y^{p-1} - \ldots - a_p = 0$$

Οι σταθερές $\{ A_i \}$ εκλέγονται για να ικανοποιούν τις αρχικές συνθήκες που εξαρτώνται από τη σχέση $\rho(0) = 1$. Έτσι έχουμε $\sum_{i=1}^{p} A_i = 1$ και οι πρώτες *(p-1) Yule - Walker* εξισώσεις μπορούν να χρησιμοποιηθούν για τον υπολογισμό των σταθερών $\{ A_i \}$ χρησιμοποιώντας ότι $\rho(0) = 1$ και $\rho(k) = \rho(-k)$. Για να είναι η χρονοσειρά μας στάσιμη, πρέπει οι ρίζες της εξίσωσης

$$H(B) = 1-a_1 B-\ldots-a_p B^p = 0$$

να βρίσκονται έξω από το μοναδιαίο κύκλο. Αυτή είναι μία αναγκαία και ικανή συνθήκη.

Για απλούστευση έχουμε θεωρήσει μόνο χρονοσειρές με μέση τιμή *0*, αλλά χρονοσειρές με μη - μηδενικές μέσες τιμές μπορούν να εξετασθούν αν γράψουμε την (7.2) στη μορφή

$$X_t-\mu = \alpha_1 (X_{t-1}-\mu)+\ldots+ \alpha_p (X_{t-p}-\mu)+Z_t$$

Αυτό δεν επηρεάζει το συντελεστή αυτοσυσχέτισης

7.4.5. **Μικτά μοντέλα (*mixed models*)**

Μία χρήσιμη ομάδα μοντέλων για χρονοσειρές είναι επίσης εκείνη που σχηματίζεται με το συνδυασμό των *MA* και *AR* μοντέλων. Ένα μικτό αυτοπαλινδρομικό μοντέλο κινητού μέσου όρου που περιέχει *p* *AR* όρους και *q* *MA* όρους (συντομογραφία *ARMA (p, q)* μοντέλο) λέγεται ότι είναι τάξης (*p, q*). Το μοντέλο αυτό δίνεται από τη σχέση:

$$X_t = \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \ldots + \beta_q Z_{t-q} \qquad (7.6)$$

Χρησιμοποιώντας το τελεστή που μεταθέτει προς τα πίσω μπορούμε να γράψουμε τη σχέση (7.6) στη μορφή

$$H(B)X_t = \Gamma(B)Z_t$$

όπου $H(B)$, $\Gamma(B)$ είναι πολυώνυμα p, q τάξης αντίστοιχα, έτσι ώστε

$$H(B) = 1 - a_1 B - \ldots - a_p B^p \quad \text{και} \quad \Gamma(B) = 1 + \beta_1 B + \ldots + \beta^q B^q$$

Οι τιμές των $\{\alpha_i\}$ που κάνουν τη διεργασία στάσιμη είναι τέτοιες ώστε οι ρίζες του $H(B)$=0 να βρίσκονται έξω από το μοναδιαίο κύκλο. Οι τιμές των $\{\beta_i\}$ που κάνουν τη διεργασία αντιστρεπτή είναι τέτοιες ώστε οι ρίζες της εξίσωσης $\Gamma(B)$=0 να βρίσκονται έξω από το μοναδιαίο κύκλο.

Είναι αλγεβρικά πολύ κοπιαστικό να υπολογίσουμε το συντελεστή αυτοσυσχέτισης ενός *ARMA* μοντέλου και γι’ αυτό δεν θα συζητηθεί εδώ.

Η σπουδαιότητα των *ARMA* μοντέλων βρίσκεται στο γεγονός ότι οι στάσιμες χρονοσειρές μπορούν συχνά να περιγραφούν με ένα *ARMA* μοντέλο που περιέχει λιγότερες παραμέτρους από ότι ένα *MA* ή *AR* μοντέλο.

7.5. Εκτίμηση του συντελεστή αυτοσυσχέτισης

Εξετάζουμε πρώτα έναν εκτιμητή της συνάρτησης αυτοδιασποράς. Η συνάρτηση αυτοδιασποράς με καθυστέρηση k για ένα δείγμα μεγέθους N δίνεται από τη σχέση

$$c_k = \frac{1}{N}\sum_{t=1}^{N-k}\left(x_t - \overline{x}\right)\left(x_{t+k} - \overline{x}\right) \tag{7.7}$$

Η συνάρτηση αυτή είναι ο συνήθης εκτιμητής της θεωρητικής συνάρτησης αυτοδιασποράς με καθυστέρηση *k, γ(k)*.

Έχοντας εκτιμήσει τη συνάρτηση αυτοδιασποράς, ορίζουμε ως εκτιμητή της συνάρτησης αυτοσυσχέτισης *ρ(k)* το εξής πηλίκο

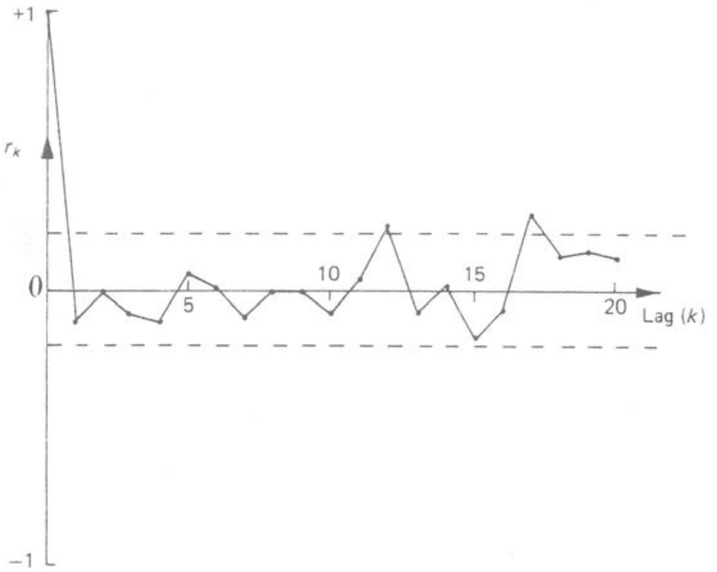
$$r_k = \frac{c_k}{c_0} \tag{7.8}$$

Θα εξετάσουμε τις ιδιότητες του εκτιμητού r_k όταν λαμβάνουμε ένα δείγμα από μία καθαρώς τυχαία διεργασία. Αυτό σημαίνει ότι η θεωρητική συνάρτηση αυτοδιασποράς *γ(k)* είναι μηδέν, εκτός εάν η καθυστέρηση είναι μηδέν *(k=0)*. Το αποτέλεσμα αυτό είναι πολύ χρήσιμο επειδή μας δίνει τη δυνατότητα να αποφασίσουμε αν οι παρατηρούμενες τιμές του r_k από μία δοθείσα χρονοσειρά είναι σημαντικά διάφορες του μηδενός.

Υποθέτουμε ότι οι μεταβλητές x_1 , ..., x_N είναι ανεξάρτητες και έχουν την ίδια κατανομή με αυθαίρετη μέση τιμή. Τότε μπορεί να δειχθεί ότι

$$E\{r_k\} = -\frac{1}{N} \quad \text{και} \quad Var\{r_k\} \cong \frac{1}{N} \tag{7.9}$$

και ότι οι εκτιμητές r_k ακολουθούν ασυμπτωτικά μία κανονική κατανομή κάτω από ορισμένες προϋποθέσεις. Έτσι, έχοντας σχεδιάσει το συσχετόγραμμα, μπορούμε να σχεδιάσουμε προσεγγιστικά τα 95% όρια εμπιστοσύνης στα σημεία $-\frac{1}{N} \pm \frac{2}{\sqrt{N}}$, τα οποία επιπλέον προσεγγίζονται στα σημεία $\pm\frac{2}{\sqrt{N}}$ για μεγάλο N. Παρατηρούμενες τιμές των εκτιμητών r_k που βρίσκονται έξω από τα όρια αυτά είναι σημαντικά διάφορες του μηδενός σε επίπεδο 5%. Παρ' όλα αυτά όταν ερμηνεύουμε το συσχετόγραμμα, πρέπει να έχουμε υπόψη μας ότι η ολική πιθανότητα να πάρουμε ένα συντελεστή έξω από τα όρια εμπιστοσύνης, όταν δίνεται ότι τα δεδομένα είναι πραγματικώς τυχαία, αυξάνει με τον αριθμό των συντελεστών συσχέτισης r_k που παραστάθηκαν γραφικώς. Για παράδειγμα, όταν οι πρώτες 20 τιμές των r_k παριστάνονται γραφικά τότε κανείς αναμένει μία σημαντική τιμή ακόμη κι αν τα δεδομένα είναι τυχαία. Έτσι, αν μόνο ένας ή δύο συντελεστές είναι σημαντικοί, το μέγεθος και η καθυστέρηση των συντελεστών πρέπει να ληφθεί υπόψη όταν αποφασίζουμε ότι ένα σύνολο δεδομένων είναι τυχαίο. Τιμές αρκετά έξω από τα όρια εμπιστοσύνης δείχνουν ότι τα δεδομένα δεν είναι τυχαία.



Σχήμα 7.2. Το συσχετόγραμμα 100 ανεξαρτήτων παρατηρήσεων που ακολουθούν μία κανονική κατανομή.

Το Σχήμα 7.2. δείχνει το συσχετόγραμμα 100 παρατηρήσεων, οι οποίες υποτίθεται ότι είναι ανεξάρτητες και έχουν την ίδια κανονική κατανομή. Τα όρια εμπιστοσύνης δίνονται προσεγγιστικά από τις τιμές $\pm\frac{2}{\sqrt{100}} = \pm 0.2$. Διακρίνουμε ότι 2 από τις 20 τιμές των συντελεστών συσχέτισης r_k βρίσκονται μόλις έξω των ορίων εμπιστοσύνης. Καθώς αυτές οι τιμές συμβαίνουν σε αυθαίρετες καθυστερήσεις (12 και 17) καταλήγουμε ότι δεν υπάρχει ένδειξη να απορρίψουμε την υπόθεση ότι οι παρατηρήσεις είναι ανεξάρτητες.

7.6. Παραδείγματα

1. Να δειχθεί ότι η χρονοσειρά $X_t = acos(\omega t+\theta)$ είναι δεύτερης - τάξης στάσιμη όταν θεωρήσουμε ότι a είναι μία σταθερά, ω είναι επίσης μία σταθερά στο διάστημα (0, π) και θ είναι μία τυχαία μεταβλητή με μία ομοιόμορφη κατανομή στο διάστημα (0, 2π). (Η μεταβλητή θ θεωρείται σταθερή για μία απλή πραγμάτωση)

Λύση:

Η συνάρτηση πυκνότητας πιθανότητας της μεταβλητής θ δίνεται από τον τύπο:

$$p(\theta) = 1 / 2\pi\ , \quad 0 < \theta < 2\pi$$

Η μέση τιμή της μεταβλητής X_t μπορεί να υπολογισθεί τώρα ως εξής:

$$E\{X_t\} = \frac{a}{2\pi}\int_0^{2\pi} cos(\omega t + \theta)\ d\theta = \frac{a}{2\pi}\int_{\omega t}^{2\pi+\omega t} cos\phi\ d\phi = \frac{a}{2\pi}\left[sin(2\pi + \omega t) - sin(\omega t)\right] = 0$$

(θέσαμε $\phi = \omega t+\theta$)

Επίσης βρίσκουμε ότι

$$E\{X_t X_s\} = \frac{a^2}{2\pi}\int_0^{2\pi} cos(\omega t + \theta)\ cos(\omega s + \theta)\ d\theta =$$

$$\frac{a^2}{2\pi}\int_0^{2\pi} (cos\omega t\ \ cos\theta - sin\omega t\ sin\theta)\ (cos\omega s\ \ cos\theta - sin\omega s\ sin\theta)\ d\theta =$$

$$= \frac{a^2}{2}(cos\omega t\ \ cos\omega s + sin\omega t\ sin\omega s) = \frac{a^2}{2} cos\omega(t - s)$$

Από τις σχέσεις αυτές συνεπάγεται ότι η χρονοσειρά είναι δεύτερης - τάξης στάσιμη. Επειδή η θ θεωρείται σταθερά για μία απλή πραγμάτωση, αν εκτιμηθεί μία φορά από υπάρχουσες τιμές της X_t , τότε όλες οι επόμενες τιμές της X_t θα είναι τελείως καθορισμένες. Είναι λοιπόν φανερό ότι το παραπάνω μοντέλο ορίζει μία καθαρώς καθορισμένη διεργασία. Αναφέρουμε, στο σημείο αυτό, ότι όλες οι στάσιμες διεργασίες που ορίζονται από τα *ARMA* μοντέλα είναι καθαρώς μη - καθορισμένες (*non - deterministic* ή *probabilistic*) διεργασίες.

Η χρονοσειρά X_t , όπως ορίζεται παραπάνω, γίνεται πιο ενδιαφέρουσα και δεν είναι πλέον καθαρώς καθορισμένη αν προσθέσουμε ένα τυχαίο σφάλμα (μία καθαρώς τυχαία διεργασία) ως εξής:

$$X_t = acos(\omega t+\theta)+Z_t \tag{7.10}$$

Το μοντέλο αυτό περιγράφει μία συνημιτονοειδή ταλάντωση με περιοδικότητα ω.

2. Να δειχθεί ότι για μία *MA(1)* ανέλιξη $\rho(1)\leq 1/2$.

Ο συντελεστής αυτοσυσχέτισης *ρ(1)* μιας *MA(1)* ανέλιξης δίνεται από τη σχέση $\rho(1) = \dfrac{\theta}{1+\theta^2}$. Επειδή $1+\theta^2 \geq 2\theta$ συνεπάγεται ότι

$$\rho(1) \leq \frac{\theta}{2\theta} = \frac{1}{2}$$

3. Δίνεται το μοντέλο $x_t = z_t + \theta_1 z_{t-1} + \theta_2 z_{t-2}$.

(α) Αν οι ρίζες της εξίσωσης $1+\theta_1 B + \theta_2 B^2 = 0$ είναι μιγαδικές, δηλ. της μορφής $\alpha e^{-i\phi}$ και $\alpha e^{i\phi}$, να βρεθεί για ποια τιμή του α το μοντέλο είναι αντιστρέψιμο.

(β) Να δειχθεί ότι οι συντελεστές αυτοσυσχέτισης *ρ(1)* και *ρ(2)* δίνονται από τις σχέσεις :

$$\rho(1) = -2\alpha\left(1+\alpha^2\right)\cos\phi \Big/ \left(1+4\alpha^2\cos^2\phi + \alpha^4\right)$$

$$\rho(2) = \alpha^2 \Big/ \left(1+4\alpha^2\cos\phi + \alpha^4\right)$$

Το πρώτο μέρος αποδεικνύεται ως εξής :

(α) Για να είναι το μοντέλο αντιστρέψιμο πρέπει οι ρίζες του διωνύμου $1+\theta_1 B + \theta_2 B^2 = 0$ να είναι απολύτως μεγαλύτερες από την μονάδα. Στην περίπτωση μας, επειδή οι ρίζες είναι συζυγείς μιγαδικοί αριθμοί, θα πρέπει το μέτρο του μιγαδικού αριθμού να είναι μεγαλύτερο από τη μονάδα, δηλαδή $\left|\alpha e^{-i\phi}\right| = \sqrt{\alpha^2\cos^2\phi + \alpha^2\sin^2\phi} = \sqrt{\alpha^2} = |\alpha| > 1$.

Οι ρίζες ενός διωνύμου δίνονται από τη σχέση

$$\frac{-\theta_1 \pm \sqrt{\theta_1^2 - 4\theta_2}}{2\theta_2} = \frac{-\theta_1 \pm i\sqrt{-\theta_1^2 - 4\theta_2}}{2\theta_2}$$ (επειδή οι ρίζες είναι μιγαδικές)

Θα ισχύει λοιπό

$$\alpha \cos\phi = -\frac{\theta_1}{2\theta_2} \quad \text{και} \quad \alpha \sin\phi = \frac{\sqrt{-\theta_1^2 + 4\theta_2}}{2\theta}$$

Από τις σχέσεις αυτές βρίσκουμε

$$\alpha^2 = \frac{1}{\theta_2} \quad \text{και} \quad \theta_1 = -2\alpha^{-1}\cos\phi$$

(β) Είναι εύκολο να αποδείξουμε ότι οι συντελεστές *ρ(1)* και *ρ(2)* δίνονται από τις σχέσεις

$$\rho(1) = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2} \quad \text{και} \quad \rho(2) = \frac{\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

Αν στις σχέσεις αυτές θέσουμε $\theta_2 = \alpha^{-2}$ και $\theta_1 = -2\ \alpha^{-1}\cos\varphi$, τότε βρίσκουμε εύκολα το ζητούμενο αποτέλεσμα.

4. Η εξίσωση του *AR(2)* μοντέλου δίνεται από την τύπο:

$$X_t = \frac{1}{5}X_{t-1} + \frac{6}{25}X_{t-2} + Z_t$$

Να υπολογισθεί ο συντελεστής αυτοσυσχέτισης *ρ(k)* για $k = 0, \pm 1, \pm 2, \ldots$

Λύση:

Επειδή ο συντελεστής αυτοσυσχέτισης δίνεται από τον τύπο

$$\rho(k) = A_1 \pi_1^{|k|} + A_2 \pi_2^{|k|}$$

θα πρέπει να υπολογίσουμε τις ποσότητες π_1, π_2 και A_1, A_2.

Ως γνωστό π_1 , π_2 είναι οι ρίζες του διωνύμου $y^2–(1/5)y–6/25 = 0$. Άρα βρίσκουμε $\pi_1 =3/5$ και $\pi_2 = -2/5$. Επίσης από τη θεωρία είναι γνωστό ότι:

$$A_1 = \frac{\frac{a_1}{1-a_2} - \pi_2}{\pi_1 - \pi_2} \qquad \text{και} \quad A_2 = 1 - A_1,$$

οπότε λαμβάνουμε $A_1 = 63/95$ και $A_2 = 32/95$.

Επομένως ο συντελεστής αυτοσυσχέτισης δίνεται από τον σχέση:

$$\rho(k) = \frac{63}{95}\left(\frac{3}{5}\right)^{|k|} + \frac{32}{95}\left(-\frac{2}{5}\right)^{|k|} \qquad \text{για} \quad k = 0, \pm 1, \pm 2, \ldots$$

5. Η συνάρτηση αυτοσυσχέτισης ενός *ARMA(1,1)* μοντέλου δίνεται από τς ακόλουθες σχέσεις

$$\rho(1) = \frac{(1+\varphi\theta)(\varphi+\theta)}{1+\theta^2+2\varphi\theta}$$

$$\rho(k) = \varphi\, \rho(k–1) \quad (k=2, 3,\ldots)$$

Λύση:

Το μοντέλο *ARMA(1,1)* δίνεται από τον τύπο

$$X_t – \varphi X_{t-1} = Z_t + \theta Z_{t-1} \tag{5.11}$$

Αν πολλαπλασιάσουμε την (7.11) με X_t και πάρουμε τις μέσες τιμές, βρίσκουμε ότι:

$$\gamma(0) - \varphi\gamma(1) = \sigma_z^2 + (\theta + \varphi)\theta\sigma_z^2$$

επειδή $$E\{X_t Z_t\} = \sigma_z^2 \quad \kappa\alpha\iota \quad E\{X_t Z_{t-1}\} = (\theta+\varphi)\sigma_z^2 \ .$$

Ομοίως βρίσκουμε, αν πολλαπλασιάσουμε την (711) με X_{t-1} και πάρουμε εκ νέου τις μέσες τιμές, ότι

$$\gamma(-1) - \varphi\gamma(0) = \theta\sigma_z^2$$

Λύνοντας αυτές τις εξισώσεις ως προς $\gamma(0)$ και $\gamma(1)=[\gamma(-1)]$ λαμβάνουμε

$$\gamma(0) = \sigma_z^2(1 + 2\varphi\theta + \theta^2) \,/\, (1-\varphi^2)$$

$$\gamma(1) = \sigma_z^2(1+\varphi\theta)\,(1+\varphi\theta) \,/\, (1-\varphi^2)$$

Άρα

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = \frac{(1+\theta\varphi)(\theta+\varphi)}{1+2\varphi\theta+\theta^2}$$

Αν τώρα πολλαπλασιάσουμε την (7.11) με X_{t-k} $(k>1)$ και πάρουμε μέσες τιμές, βρίσκουμε ότι

$$\gamma(k) = \varphi\gamma(k-1)$$

επειδή $E\{Z_t X_{t-k}\} = E\{Z_{t-1} X_{t-k}\} = 0$ για $k>1$.

Επομένως $\rho(k) = \varphi\rho(k-1)$ για k=2, 3,…

6. Ας υποθέσουμε ότι το συσχετόγραμμα μιας χρονοσειράς που αποτελείται από 100 παρατηρήσεις έχει r_1=0.31, r_2=0.37, r_3=-0.05, r_4=0.06, r_5=-0.21, r_6=0.11, r_7=0.08, r_8=0.05, r_9=0.12, r_{10}=-0.01.
Να βρεθεί ένα ARMA μοντέλο που είναι κατάλληλο για την περίπτωση αυτή.

Λύση:

Τιμές του συντελεστή αυτοσυσχέτισης έξω από το διάστημα $\frac{\pm 2}{\sqrt{100}} = \pm 0.2$ είναι σημαντικά διάφορες του μηδενός. Άρα οι τιμές r_1=0.31 και r_2=0.37 είναι σημαντικές. Αυτό σημαίνει ότι ο συντελεστής αυτοσυσχέτισης κόβεται απότομα μετά από καθυστέρηση 2. Στην περίπτωση αυτή ένα δεύτερης τάξης MA μοντέλο είναι κατάλληλο.