Κεφάλαιο 6

Χρονοσειρές

Στο προηγούμενο κεφάλαιο μελετήσαμε τη σχέση ενός μεγέθους με άλλα μεγέθη καθώς και την εξάρτηση του μεγέθους (της εξαρτημένης τυχαίας μεταβλητής) από άλλα μεγέθη (τις ανεξάρτητες μεταβλητές). Εδώ θα επικεντρωθούμε σε ένα μέγεθος που αλλάζει τιμές με το χρόνο και θα μελετήσουμε την εξάρτηση του μεγέθους X σε κάποια χρονική στιγμή t, x_t, από το ίδιο μέγεθος σε προηγούμενες χρονικές στιγμές, $x_{t-1}, x_{t-2}, \ldots$. Θα μελετήσουμε λοιπόν τη δυναμική εξέλιξη της διαδικασίας ή συστήματος που παράγει το μέγεθος που παρατηρούμε και για αυτό θα θεωρήσουμε πως παρατηρούμε το μέγεθος στο χρόνο με σταθερό χρονικό βήμα, δηλαδή με σταθερό **χρόνο δειγματοληψίας** (sampling time). Ένα σύνολο τέτοιων παρατηρήσεων λέγεται **χρονική σειρά** ή **χρονοσειρά** (time series). Σε κάποια προβλήματα ο χρόνος δειγματοληψίας μπορεί να μην είναι σταθερός και τότε χρειάζεται ειδικότερη επεξεργασία της χρονοσειράς για να γίνει η ανάλυση. Για παράδειγμα οι ημερήσιες τιμές ενός χρηματιστηριακού δείκτη (π.χ. του Χρηματιστηρίου Αξιών Αθηνών) συνιστούν μια χρονοσειρά με μεταβλητό φυσικό χρόνο δειγματοληψίας, αφού μεσολαβούν Σαββατοκύριακα και γιορτές που είναι κλειστό το χρηματιστήριο. Η πιο απλή προσέγγιση σε αυτήν την περίπτωση είναι να ορίσουμε ως χρόνο αναφοράς όχι το φυσικό χρόνο αλλά τον οικονομικό χρόνο συναλλαγών και να θεωρήσουμε σταθερό χρονικό βήμα μιας (οικονομικής) ημέρας ακόμα και από Παρασκευή σε Δευτέρα. Στη συνέχεια θα θεωρούμε το χρόνο δειγματοληψίας σταθερό που είναι και η τυπική περίπτωση στην ανάλυση μιας χρονοσειράς.

Το πρόβλημα στην ανάλυση χρονοσειρών είναι να εκτιμήσουμε το σύστημα που παράγει τη χρονοσειρά και ενδεχομένως να κάνουμε προβλέψεις μελλοντικών τιμών του μεγέθους που παρατηρούμε. Η πρώτη υπόθεση που θα πρέπει να απορρίψουμε για να έχει νόημα η ανάλυση της χρονοσειράς είναι ότι η μεταβολή των τιμών του μεγέθους που παρατηρούμε είναι εντελώς τυχαία, δηλαδή το σύστημα που παρατηρούμε είναι λευκός θόρυβος. Αν οι παρατηρήσεις της χρονοσειράς δεν είναι ανεξάρτητες, η πληροφορία που υπάρχει στη χρονοσειρά μπορεί να δίνεται με διαφορετικές μορφές και τα κυριότερα χαρακτηριστικά που θα πρέπει να μελετήσουμε πριν προχωρήσουμε να προσαρμόσουμε κάποιο μοντέλο στη χρονοσειρά είναι:

1. **Στασιμότητα** (Stationarity): Απλά αυτό σημαίνει ότι οι διακυμάνσεις των τιμών της χρονοσειράς δε διαφοροποιούνται με το χρόνο. Μια μη-στάσιμη χρονοσειρά μπορεί να έχει **τάσεις** (trends), δηλαδή (αργές) αλλαγές στη μέση τιμή της με το χρόνο, π.χ. η τιμή βενζίνης μπορεί να έχει διακυμάνσεις λόγω της διεθνούς αγοράς αλλά και να παρουσιάζει μια αυξητική τάση σε βάθος χρόνου λόγω πληθωρισμού. Μια μη-στάσιμη χρονοσειρά μπορεί επίσης να παρουσιάζει **περιοδικότητα** (periodicity), που όταν αναφέρεται σε συγκεκριμένες περιόδους που σχετίζονται με φυσικές εποχές του έτους (μήνα, τρίμηνο, τετράμηνο) λέγεται και **εποχικότητα** (seasonality), π.χ. η τιμή του όζοντος στην ατμόσφαιρα υπόκειται σε εποχικές διακυμάνσεις πέρα από τις διακυμάνσεις που μπορεί να οφείλονται στην εξέλιξη του οικοσυστήματος.

2. **Αιτιοκρατία** (determinism) και **στοχαστικότητα** (stochasticity): Όλες οι χρονοσειρές από πραγματικά μεγέθη περιέχουν θόρυβο και με αυτήν την έννοια όλες οι πραγματικές χρονοσειρές είναι στοχαστικές. Η μεγαλύτερη πρόκληση στην ανάλυση πραγματικών χρονοσειρών είναι η διερεύνηση και ταύτιση ή εντοπισμός του αιτιοκρατικού μέρους του συστήματος που παράγει τη χρονοσειρά. Όταν αυτό είναι κρυμμένο μέσα στο θόρυβο ή γενικότερα δεν κυριαρχεί στην εξέλιξη της χρονοσειράς, τότε θεωρούμε πως το σύστημα είναι στοχαστικό και περιοριζόμαστε σε στατιστική περιγραφή του συστήματος, όπως κάναμε και για τις τυχαίες μεταβλητές στην στατική περίπτωση (ανεξάρτητες μετρήσεις ενός μεγέθους, δες Κεφάλαιο 3 και 4).

   Αν για κάποιο λόγο μπορούμε να υποθέσουμε ότι το σύστημα που παράγει τη χρονοσειρά είναι κυρίως αιτιοκρατικό με κάποιες στοχαστικές διαταραχές που όμως δεν κυριαρχούν στην εξέλιξη του συστήματος (και της χρονοσειράς που μελετάμε), τότε μπορούμε να χρησιμοποιήσουμε διαφορετικές προσεγγίσεις που είναι κατάλληλες για αιτιοκρατικά δυναμικά συστήματα, π.χ. ανίχνευση κύριων περιόδων αν το σύστημα φαίνεται να είναι περιοδικό ή διερεύνηση της μη-γραμμικής δυναμικής αν το σύστημα φαίνεται να είναι χαοτικό.

   Για παράδειγμα η μεταβολή τής στάθμης του όζοντος στην ατμόσφαιρα μπορεί να έχει διαφορετικές περιοδικότητες που θέλουμε να εντοπίσουμε με ακρίβεια (περίοδο έτους αλλά ίσως και άλλες περιόδους) και τότε καταφεύγουμε σε μεθόδους της φασματικής ανάλυσης. Μπορεί όμως απαλείφοντας την ετήσια εποχικότητα, να θεωρήσουμε ότι το σύστημα είναι στοχαστικό με ενδεχομένως κάποιες γραμμικές συσχετίσεις βραχείας διάρκειας που θα θέλαμε να εκτιμήσουμε και να προσαρμόσουμε κάποιο κατάλληλο στοχαστικό μοντέλο. Τέλος μπορεί να υποθέσουμε ότι η μη-κανονικότητα της χρονοσειράς (απαλλαγμένης από την ετήσια εποχικότητα) οφείλεται σε κάποιο μη-γραμμικό αιτιοκρατικό δυναμικό σύστημα, ενδεχομένως χαμηλής διάστασης και χαοτικό, που έχει τη δυνατότητα να παρουσιάζει φαινομενικά τυχαία συμπεριφορά.

3. **Γραμμικότητα** (linearity) και **μη-γραμμικότητα** (nonlinearity): Σύμφωνα με τα παραπάνω φαίνεται αυτές οι δύο έννοιες να σχετίζονται με την αιτιοκρατία και στοχαστικότητα αλλά γενικά μπορούν να ορισθούν ανεξάρτητα από αυτές. Η γραμμικότητα του συστήματος σημαίνει πως οι μεταβλητές του συστήματος (που μπορεί να μην έχουμε τη δυνατότητα να τις παρατηρήσουμε) αλληλο-επιδρούν γραμμικά, δηλαδή αν θα εκφράζαμε το σύστημα με αναλυτική μορφή όλοι οι όροι θα ήταν γραμμικοί ως προς τις μεταβλητές του συστήματος. Σε αντίθετη περίπτωση το σύστημα είναι μη-γραμμικό. Για τη χρονοσειρά αυτό σημαίνει πως για ένα γραμμικό σύστημα ορίζουμε την εξέλιξη της χρονοσειράς ως γραμμικό συνδυασμό των προηγούμενων παρατηρήσεων της χρονοσειράς, ενώ για ένα μη-γραμμικό σύστημα μπορούμε να ορίσουμε την εξέλιξη της χρονοσειράς με μεγαλύτερη ακρίβεια αν θεωρήσουμε και τη συνδυασμένη επίδραση των προηγούμενων παρατηρήσεων σε διαφορετικές χρονικές στιγμές ή τις ίδιες.

   Άρα λοιπόν ένα στοχαστικό σύστημα μπορεί να είναι γραμμικό ή μη-γραμμικό και το ίδιο ισχύει για ένα αιτιοκρατικό σύστημα. Βέβαια ένα αιτιοκρατικό γραμμικό σύστημα δεν παρουσιάζει ιδιαίτερο ενδιαφέρον γιατί τα γραμμικά αιτιοκρατικά δυναμικά συστήματα έχουν απλές λύσεις που στην απουσία θορύβου μπορούμε εύκολα να εντοπίσουμε (σταθερό σημείο, περιοδικά σημεία ή τροχιές). Εδώ σημειώνεται ότι κάποια δυσκολία μπορεί να παρουσιαστεί αν το σύστημα είναι πολλών διαστάσεων, υπάρχει κάποια τυχαία διαταραχή και το πλήθος των παρατηρήσεων είναι σχετικά μικρό. Από την άλλη μεριά, είναι ιδιαίτερα δύσκολο να εντοπίσουμε μη-γραμμικότητα σε ένα στοχαστικό σύστημα (ή διαδικασία όπως συνήθως λέγεται) αφού ο θόρυβος στο σύστημα δεν επιτρέπει τον εντοπισμό πολύπλοκων μη-γραμμικών σχέσεων. Σε μια τέτοια περίπτωση θα πρέπει να έχουμε ορίσει μια συγκεκριμένη μη-γραμμική μορφή που θέλουμε να διερευνήσουμε. Συνήθως λοιπόν οι δύο κυρίαρχες κλάσεις συστημάτων που υποθέτουμε για στάσιμες χρονοσειρές είναι η **γραμμική στοχαστική διαδικασία** (linear stochastic process) και το **μη-γραμμικό δυναμικό (πιθανώς χαοτικό) σύστημα** (nonlinear dynamical (possibly chaotic) system).

Στη συνέχεια θα αναλύσουμε τη χρονοσειρά επικεντρώνοντας σε κάθε μια από τις παραπάνω μορφές χρονοσειρών.

Γενικά για κάθε χρονική στιγμή t θεωρούμε την τιμή x_t της μεταβλητής X που παρατηρούμε. Το σύνολο των παρατηρήσεων x_t για κάποια χρονική περίοδο n (σε μονάδες του χρόνου δειγματοληψίας), δηλαδή για χρονικές στιγμές $t = 1, \ldots, n$, αποτελεί τη χρονοσειρά $\{x_t\}_{t=1}^{n} = \{x_1, \ldots, x_n\}$. Όταν έχουμε τη δυνατότητα ταυτόχρονης παρατήρησης πολλών μεγεθών για το ίδιο σύστημα, όπως π.χ. καταγραφές σεισμικών κυμάτων από διαφορετικούς σταθμούς ή καταγραφή θερμοκρασίας και πίεσης, έχουμε πολλαπλές ταυτόχρονες χρονοσειρές ή αλλιώς έχουμε μια **πολυ-διάστατη χρονοσειρά** (multivariate time series). Συνήθως όμως η μελέτη περιορίζεται στη συλλογή τιμών ενός μεγέθους, δηλαδή μιας μονοδιάστατης χρονοσειράς, και εδώ δε θα επεκταθούμε σε πολυ-διάστατες χρονοσειρές.

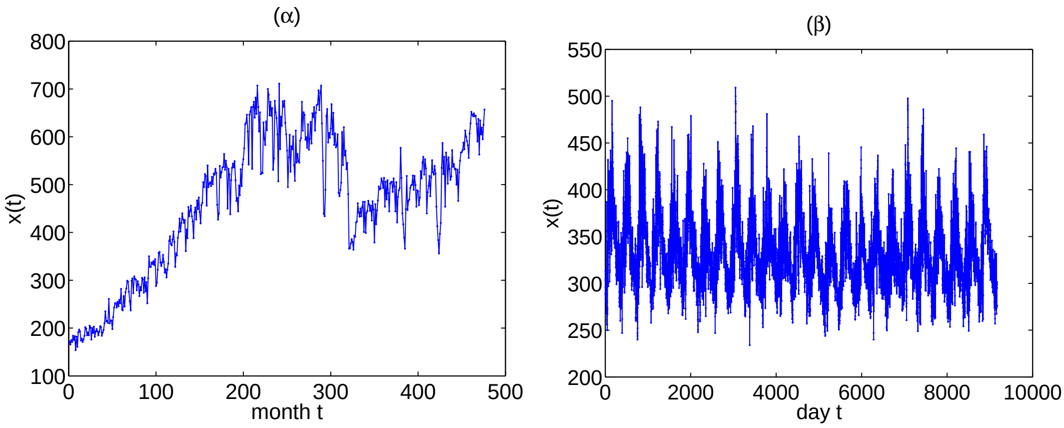
Η μελέτη θα αρχίσει δίνοντας τα πρώτα στάδια της ανάλυσης χρονοσειρών και περιγράφοντας βασικά χαρακτηριστικά χρονοσειρών από στοχαστικά συστήματα, όπως στασιμότητα, τάση, περιοδικότητα και αυτοσυσχέτιση. Στη συνέχεια θα μελετήσουμε γραμμικά μοντέλα για την πρόβλεψη στοχαστικών χρονοσειρών και ειδικότερα τη διαδικασία Box-Jenkins. Η μελέτη θα ολοκληρωθεί με την παρουσίαση κάποιων μη-γραμμικών μεθόδων ανάλυσης χρονοσειρών καθώς και κάποιων απλών μη-γραμμικών μοντέλων πρόβλεψης.

6.1 Βασικά Χαρακτηριστικά Χρονοσειράς

Θα μελετήσουμε πρώτα κάποια βασικά χαρακτηριστικά των χρονοσειρών μέσα από πραγματικά παραδείγματα.

6.1.1 Στασιμότητα, τάση και περιοδικότητα

Στο Σχήμα 6.1α δίνεται η μηνιαία παραγωγή σιδήρου στην Αυστραλία για κάποια χρονική περίοδο σε ένα **διάγραμμα ιστορίας** (history diagram). Παρατηρούμε ότι η παραγωγή παρουσιάζει αυξητική τάση, μετά σε κάποια περίοδο η παραγωγή σταθεροποιείται σε ένα επίπεδο αλλά έχει έντονες διακυμάνσεις, στη συνέχεια έχει πτώση και μετά συνεχίζει και πάλι να αυξάνει σταθερά. Τέτοιες αργές μεταβολές λέγονται τάσεις στη χρονοσειρά. Το διάγραμμα ιστορίας των ημερήσιων τιμών όζοντος στην περιοχή Θεσσαλονίκης για 25 έτη που δίνεται στο Σχήμα 6.1β δεν παρουσιάζει τάσεις αλλά φαίνεται να έχει μια ετήσια περιοδικότητα, ή όπως αναφέρεται στη στατιστική βιβλιογραφία, ετήσια εποχικότητα.



Σχήμα 6.1: (α) Μηνιαία παραγωγή (βασικού) σιδήρου στην Αυστραλία την περίοδο Ιανουάριος 1956 - Αύγουστος 1995 (σε χιλιάδες τόνων). (β) Ημερήσιες τιμές όζοντος στην περιοχή Θεσσαλονίκης για 25 έτη.

Η εμφάνιση τάσης ή περιοδικότητας στη χρονοσειρά υποδηλώνει ότι τα στατιστικά χαρακτηριστικά του συστήματος που παράγει τη χρονοσειρά αλλάζουν με το χρόνο και η χρονοσειρά δεν είναι στάσιμη. Η *αυστηρή στασιμότητα* (strict stationarity) ορίζεται μαθηματικά ως η διατήρηση στο χρόνο t της κοινής κατανομής των $\{x_t, x_{t+1}, \ldots, x_{t+\tau}\}$ για κάποιο αυθαίρετο παράθυρο υστερήσεων τ. Η συνθήκη στασιμότητας περιορίζεται συνήθως στη διατήρηση της μέσης τιμής και αυτοδιασποράς (θα ορίσουμε αυτόν τον όρο στη συνέχεια) και αναφέρεται ως *ασθενής στασιμότητα* (weak stationarity). Μάλιστα πρακτικά αντί της αυτοδιασποράς εξετάζουμε μόνο τη σταθερότητα της διασποράς με το χρόνο.

Η μη-στασιμότητα αποτελεί σοβαρό πρόβλημα στην ανάλυση χρονοσειρών και ιδιαίτερα όταν προσπαθούμε να κάνουμε προβλέψεις. Για παράδειγμα, τη χρονοσειρά της μηνιαίας παραγωγής σιδήρου στο Σχήμα 6.1α, θα τη θεωρούσαμε μη-στάσιμη επειδή η μέση τιμή της δεν παραμένει σταθερή. Σε χρονοσειρές με έντονη περιοδικότητα ή εποχικότητα θα θέλαμε πρώτα να ουδετεροποιήσουμε την επίδραση της περιοδικής ή εποχικής συνιστώσας πριν αναλύσουμε τη χρονοσειρά. Υπάρχουν συγκεκριμένες τεχνικές καθώς και στατιστικοί έλεγχοι για να διερευνήσουμε τη στασιμότητα σε μια χρονοσειρά αλλά δε θα μας απασχολήσουν εδώ. Θα δούμε όμως κάποιους βασικούς μετασχηματισμούς που εφαρμόζουμε σε μια φανερά μη-στάσιμη χρονοσειρά για να την κάνουμε στάσιμη και να προχωρήσουμε με την ανάλυση.

6.1.2 Απαλοιφή τάσης και περιοδικότητας

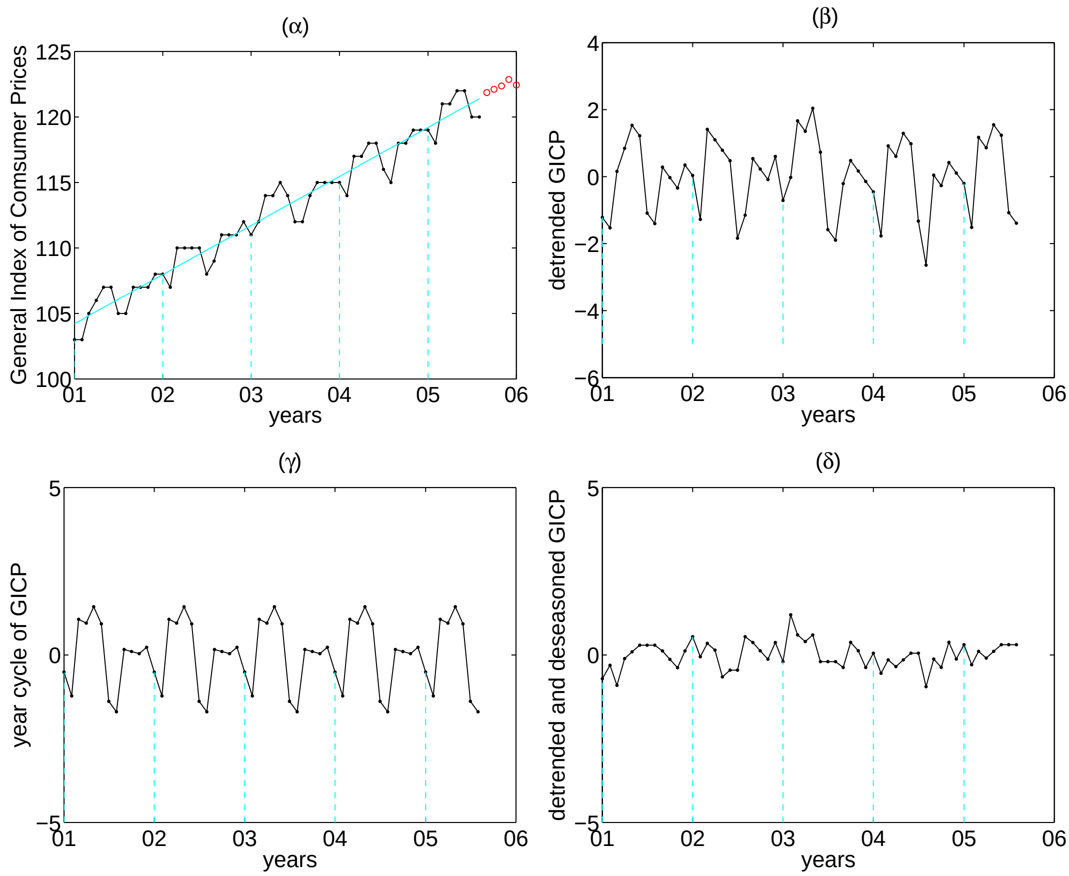
Γενικά μια χρονοσειρά μπορούμε να τη χωρίσουμε σε τρεις συνιστώσες ως εξής

$$x_t = \mu_t + s_t + y_t, \tag{6.1}$$

όπου μ_t είναι ή συνιστώσα της τάσης, s_t η συνιστώσα της περιοδικότητας για κάποια περίοδο d ($s_{t-d} = s_t$) και y_t είναι η χρονοσειρά των υπολοίπων αν αφαιρέσουμε από την παρατηρούμενη χρονοσειρά την τάση και την περιοδικότητα. Η τάση και η περιοδικότητα είναι και οι δυό συναρτήσεις του χρόνου και δεν περιέχουν πληροφορία για τη δυναμική του συστήματος, δηλαδή την εξάρτηση της παρατήρησης x_t από τις προηγούμενες παρατηρήσεις. Σε κάποια προβλήματα όλη η πληροφορία που μας ενδιαφέρει μπορεί να εντοπίζεται στη συνιστώσα της τάσης και της περιοδικότητας, οπότε το πρόβλημα περιορίζεται στην εκτίμηση των μ_t και s_t. Σε άλλες περιπτώσεις θέλουμε να εξαλείψουμε την τάση και την περιοδικότητα για να διερευνήσουμε τη δυναμική του συστήματος και θα πρέπει αφού εκτιμήσουμε τις συναρτήσεις μ_t και s_t να τις αφαιρέσουμε από της x_t για να πάρουμε τη χρονοσειρά των υπολοίπων y_t. Ας δούμε την απαλοιφή τάσης και περιοδικότητας σε ένα πραγματικό παράδειγμα.

Παράδειγμα 6.1. Στο Σχήμα 6.2α, δίνεται ο γενικός δείκτης τιμών καταναλωτή (general index for consumer price, GICP) σε μηνιαίες τιμές από τον Ιανουάριο 2001 ως τον Αύγουστο 2005 (οι ανοιχτοί κύκλοι στο σχήμα είναι προβλέψεις που θα δούμε παρακάτω). Η χρονοσειρά έχει λοιπόν μήκος $n = 56$. Φαίνεται καθαρά, πως ο δείκτης GICP παρουσιάζει σταθερή πληθωριστική τάση για όλη την περίοδο παρατήρησης αλλά και ασθενή ετήσια περιοδικότητα. Η πληθωριστική τάση μπορεί να περιγραφεί ικανοποιητικά ως γραμμική συνάρτηση μ_t του χρόνου t (μήνα). Η προσαρμογή απλού γραμμικού μοντέλου παλινδρόμησης του μ_t ως προς το t, θεωρώντας τις παρατηρήσεις $\{x_t\}_{t=1}^{56}$ ως τιμές του μ_t, έδωσε $\mu_t = 103.9 + 0.31t$ και φαίνεται με γκρίζα γραμμή στο Σχήμα 6.2α.

Αφαιρώντας αυτή τη γραμμική τάση από τη χρονοσειρά GICP βρίσκουμε τη χρονοσειρά απαλλαγμένη από τάση (detrended) $x'_t = x_t - \mu_t$ που δίνεται στο Σχήμα 6.2β. Αυτή είναι η κατάλληλη χρονοσειρά για ανάλυση αν θέλουμε να μελετήσουμε την μεταβολή του GICP απαλλαγμένη από τον πληθωρισμό. Εδώ φαίνεται καλύτερα η ετήσια περιοδικότητα του GICP αλλά δεν είναι προφανές με ποια περιοδική συνάρτηση του χρόνου μπορούμε να εκτιμήσουμε τον ετήσιο κύκλο. Μπορούμε να εκτιμήσουμε τον ετήσιο κύκλο (annual cycle) του GICP από τις μέσες τιμές του κάθε μήνα για τα έτη 2001 - 2005 (ως το 2004 για τους μήνες Σεπτέμβριο - Δεκέμβριο). Υπολογίζουμε για τον Ιανουάριο το μέσο όρο από τις τιμές Ιανουαρίου για τα 5 χρόνια και



Σχήμα 6.2: (α) Μηνιαίες τιμές γενικού δείκτη τιμών καταναλωτή (GICP) την περίοδο Ιανουάριος 2001 - Αύγουστος 2005 (η κάθετη κατακόρυφη διακεκομμένη γραμμή δηλώνει την αρχή του έτους). Στο διάγραμμα φαίνεται η προσαρμογή γραμμικού μοντέλου τάσης. (β) Η χρονοσειρά που προκύπτει από την αφαίρεση της γραμμικής τάσης στη χρονοσειρά GICP. (γ) Ο εκτιμούμενος ετήσιος κύκλος για την GICP. (δ) Η χρονοσειρά που προκύπτει από την αφαίρεση του εκτιμούμενου ετήσιου κύκλου από τη χρονοσειρά στο (β).

την τιμή αυτή την αφαιρούμε από τις 5 παρατηρήσεις του Ιανουαρίου (για τα έτη 2001 - 2005). Το ίδιο κάνουμε και για τους άλλους μήνες. Η περιοδική χρονοσειρά $\{s_t\}_{t=1}^{56}$ (του επαναλαμβανόμενου ετήσιου κύκλου) δίνεται στο Σχήμα 6.2γ όπου η περίοδος είναι $d = 12$. Αν μας ενδιαφέρει να εξετάσουμε αν μπορούμε να αντλήσουμε κάποια επιπλέον πληροφορία από τη χρονοσειρά GICP πέρα από τη γραμμική τάση, που απαλείψαμε, και τον ετήσιο κύκλο που εκτιμήσαμε, τότε από τη χρονοσειρά χωρίς τάση αφαιρούμε την περιοδική χρονοσειρά $\{s_t\}_{t=1}^{56}$. Η χρονοσειρά των υπολοίπων $\{y_t\}_{t=1}^{56}$ δίνεται ως (δες επίσης τη σχέση (6.1))

$$y_t = x'_t - s_t = x_t - \mu_t - s_t,$$

και παρουσιάζεται γραφικά στο Σχήμα 6.2δ. Παρατηρούμε ότι αυτή η χρονοσειρά είναι απαλλαγμένη από τάση και περιοδικότητα και δε φαίνεται να έχει κάποια κανονικότητα ή δομή. Αν η χρονοσειρά $\{y_t\}_{t=1}^{56}$ είναι εντελώς τυχαία, δηλαδή δεν έχει να μας δώσει καμιά πληροφορία, τότε η ανάλυση σταματάει εδώ και περιοριζόμαστε να περιγράψουμε τη χρονική μεταβολή του GICP ως μια πληθωριστική γραμμική τάση σε συνδυασμό με έναν ετήσιο κύκλο. Έτσι αν θέλουμε να προβλέψουμε το GICP για το μήνα Σεπτέμβριο θα επεκτείνουμε (extrapolate) τη συνάρτηση τάσης για το μήνα Σεπτέμβριο και θα προσθέσουμε σε αυτήν την τιμή του ετήσιου κύκλου για το μήνα Σεπτέμβριο. Αναλυτικά, έχοντας τις τιμές GICP από τον Ιανουάριο 2001 ως τον Αύγουστο 2005 ($n = 56$), η τάση επεκτείνεται για το Σεπτέμβριο 2005 ως

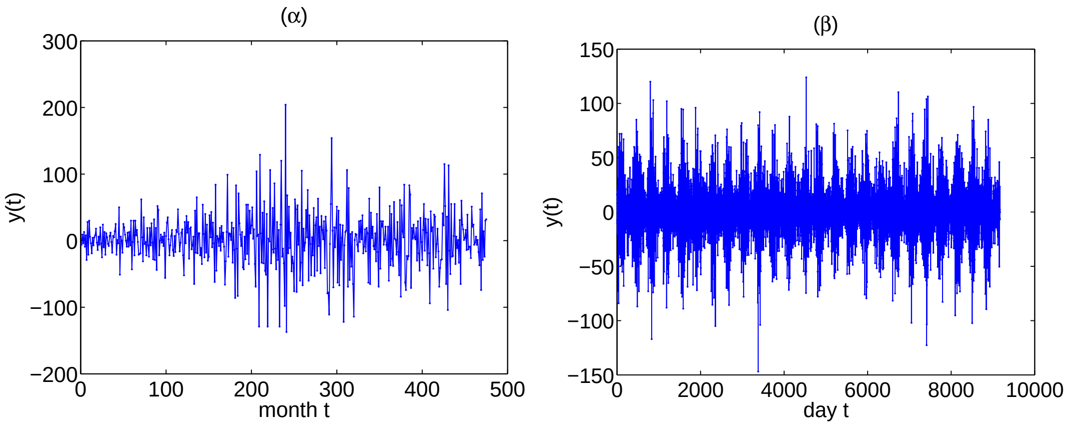
$$\mu_{n+1} = 103.9 + 0.31(n + 1) \quad \Rightarrow \quad \mu_{57} = 103.9 + 0.31 \cdot 57 = 121.7.$$

Η τιμή του ετήσιου κύκλου για το Σεπτέμβριο βρέθηκε ως η μέση τιμή Σεπτεμβρίου για τα έτη 2001 - 2004, $s_9 = 0.16$, και προστίθεται στην τιμή του μ_{57} για να δώσει την πρόβλεψη του GICP για το μήνα Σεπτέμβριο $\hat{x}_{57} = x_{56}(1) = 121.86$. [Η πρόβλεψη συμβολίζεται είτε με το καπελάκι και χρονικό δείκτη τη χρονική στιγμή για την οποία γίνεται η πρόβλεψη ή σημειώνοντας ως δείκτη τη χρονική στιγμή ως την οποία γνωρίζουμε τη μεταβλητή και σημειώνοντας σε παρένθεση τα χρονικά βήματα μπροστά που προβλέπουμε.] Όμοια μπορούμε να κάνουμε προβλέψεις και για τους άλλους μήνες. Στο Σχήμα 6.2α δίνονται με ανοιχτούς κύκλους οι προβλέψεις για τους 5 επόμενους μήνες.

Στο παραπάνω παράδειγμα απαλείψαμε την τάση προσαρμόζοντας στις παρατηρήσεις ένα γραμμικό πολυώνυμο του χρόνου. Αυτήν την προσέγγιση μπορούμε να τη γενικεύσουμε σε πολυώνυμο μεγαλύτερου βαθμού. Πολλές φορές όμως η τάση δε μπορεί να περιγραφεί ικανοποιητικά ως κάποια γνωστή συνάρτηση του χρόνου, αλλά είναι τυχαία και την αναφέρουμε ως **στοχαστική τάση** (stochastic trend). Σε αυτήν την περίπτωση για την απαλοιφή της τάσης είναι καλύτερα να πάρουμε τις **πρώτες διαφορές** (first differences)

$$y_t = \nabla x_t = x_t - x_{t-1}. \tag{6.2}$$

Για το παράδειγμα της μηνιαίας παραγωγής σιδήρου στην Αυστραλία (δες Σχήμα 6.1α), παίρνοντας πρώτες διαφορές η στοχαστική τάση της χρονοσειράς εξαλείφεται, όπως φαίνεται στο Σχήμα 6.3α. Εφαρμόζοντας πρώτες



Σχήμα 6.3: (α) Χρονοσειρά των πρώτων διαφορών της μηνιαίας παραγωγής σιδήρου στην Αυστραλία την περίοδο Ιανουάριος 1956 - Αύγουστος 1995 (σε χιλιάδες τόνων). (β) Χρονοσειρά των πρώτων διαφορών των ημερήσιων τιμών όζοντος στην περιοχή Θεσσαλονίκης για 25 έτη.

διαφορές στη χρονοσειρά των ημερήσιων τιμών όζοντος στην περιοχή Θεσσαλονίκης που έχει περιοδικότητα αλλά όχι τάση, η περιοδικότητα δεν απαλείφεται (δες Σχήμα 6.1β και Σχήμα 6.3β). Για να εξαλείψουμε την περιοδικότητα μπορούμε να χρησιμοποιήσουμε την τεχνική που περιγράφτηκε στο Παράδειγμα 6.1. Υπάρχουν και άλλες μέθοδοι, όπως το **φίλτρο κινούμενου μέσου** (moving average (MA) filter), αλλά δε θα επεκταθούμε σε αυτές. Συχνά αντί να προσπαθήσουμε να εξαλείψουμε την περιοδική συνιστώσα της χρονοσειράς, προσπαθούμε να την εξηγήσουμε μέσα από το μοντέλο που προσαρμόζουμε στη χρονοσειρά.

Στο Παράδειγμα 6.1 για το GICP κάναμε προβλέψεις με βάση την εκτίμηση της τάσης και περιοδικότητας, θεωρώντας ότι η χρονοσειρά των υπολοίπων είναι εντελώς τυχαία. Επίσης στο παράδειγμα για τη μηνιαία παραγωγή σιδήρου στην Αυστραλία, η χρονοσειρά μετά την απαλοιφή της τάσης (από τις πρώτες διαφορές) φαίνεται επίσης να είναι τυχαία. Αν όμως περιέχει συσχετίσεις θα θέλαμε να τις περιγράψουμε μαθηματικά για να κάνουμε καλύτερες προβλέψεις. Στη συνέχεια θα χαρακτηρίσουμε και θα μελετήσουμε τις συσχετίσεις σε μια στάσιμη χρονοσειρά (απαλλαγμένη από τυχόν τάσεις και περιοδικότητες).

6.2 Συσχέτιση σε χρονοσειρά

Πριν να διερευνήσουμε συσχετίσεις σε μια στάσιμη χρονοσειρά $\{x_t\}_{t=1}^{n}$, ας δούμε πρώτα δύο συστήματα χρονοσειρών με μηδενικές συσχετίσεις, το ένα στάσιμο, το άλλο μη-στάσιμο.

6.2.1 Λευκός θόρυβος

Θεωρώντας διαδοχικά στοιχεία της χρονοσειράς ως τυχαίες μεταβλητές, η χρονοσειρά λέγεται ότι αποτελείται από **ανεξάρτητες τυχαίες μεταβλητές με ίδια κατανομή** (independent and identically distributed, iid) όταν οι $x_t, x_{t+1}, \ldots, x_{t+\tau}$ τυχαίες μεταβλητές για $\tau > 1$ έχουν την ίδια κατανομή και είναι ανεξάρτητες μεταξύ τους (για την ανεξαρτησία δες τη σχέση (2.10)). Μια iid χρονοσειρά είναι εντελώς τυχαία και δεν περιέχει αυτοσυσχετίσεις (γραμμικές ή μη-γραμμικές), δηλαδή συσχετίσεις μεταξύ στοιχείων της χρονοσειράς. Μια iid χρονοσειρά λέγεται και **λευκός θόρυβος** (white noise) και θα συμβολίζουμε την κατανομή της ως $\mathrm{WN}(0, \sigma_\epsilon^2)$, με μέση τιμή 0 και διασπορά σ_ϵ^2. Αν επιπλέον τα στοιχεία της χρονοσειράς λευκού θορύβου ακολουθούν κανονική (Γκαουσιανή) κατανομή, τότε η χρονοσειρά λέγεται **Γκαουσιανός λευκός θόρυβος** (Gaussian white noise).

6.2.2 Τυχαίος περίπατος

Ο **τυχαίος περίπατος** (random walk) είναι μια μη-στάσιμη χρονοσειρά, όπου το κάθε στοιχείο της προκύπτει από το προηγούμενο με την πρόσθεση μιας τυχαίας τιμής, δηλαδή η χρονοσειρά είναι τυχαίος περίπατος αν

$$x_t = x_{t-1} + \epsilon_t, \tag{6.3}$$

όπου $\{\epsilon_t\}$ είναι χρονοσειρά λευκού θορύβου, $\epsilon_t \sim \mathrm{WN}(0, \sigma_\epsilon^2)$. Το όνομα υποδηλώνει ακριβώς ότι η χρονοσειρά παράγεται από την κίνηση κάποιου πάνω σε μια ευθεία γραμμή (στο $\mathbf{R}$), που σε κάθε χρονική στιγμή κάνει ένα τυχαίο βήμα μπρος ή πίσω (ϵ_t) από το σημείο που βρίσκεται (x_{t-1}) στο επόμενο (x_t). Σημειώνεται ότι αρχίζοντας από κάποια τιμή μ (για $t = 0$) και αντικαθιστώντας επαναληπτικά τον ορισμό (6.3) του τυχαίου περιπάτου για χρόνους ως t ο ορισμός του τυχαίου περιπάτου μπορεί να γραφεί ως

$$x_t = \mu + \sum_{j=1}^{t} \epsilon_j,$$

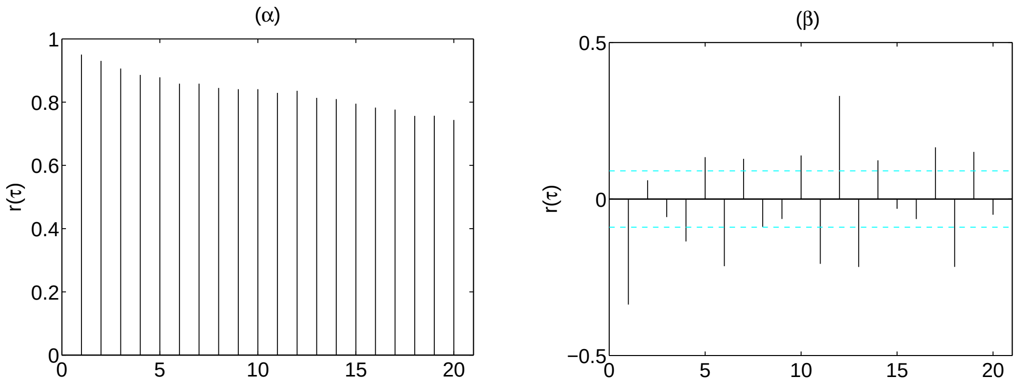
δηλαδή ως άθροισμα όλων των τυχαίων βημάτων ως τη στιγμή t. Οι φαινομενικές συσχετίσεις που φαίνονται από ένα διάγραμμα ιστορίας τυχαίου περιπάτου οφείλονται στο ότι το τυχαίο βήμα σε κάθε χρονική στιγμή, έχει γνωστό σημείο εκκίνησης. Αυτό δημιουργεί στοχαστικές τάσεις για αυτό και η χρονοσειρά του τυχαίου περιπάτου είναι μη-στάσιμη. Παίρνοντας τις πρώτες διαφορές προκύπτει η στάσιμη χρονοσειρά του λευκού θορύβου.

6.2.3 Γραμμική συσχέτιση

Στα βιβλία κλασικής ανάλυσης χρονοσειρών με τον όρο συνάρτηση αυτοσυσχέτισης ή απλά αυτοσυσχέτιση εννοείται η **γραμμική αυτοσυσχέτιση** (linear autocorrelation) (το ίδιο ισχύει γενικά για τον όρο συσχέτιση). Η αυτοσυσχέτιση ρ_τ για κάποια υστέρηση τ είναι ο συντελεστής συσχέτισης δύο στοιχείων της χρονοσειράς που απέχουν χρονικά τ βήματα και εκτιμάται από τη χρονοσειρά ως

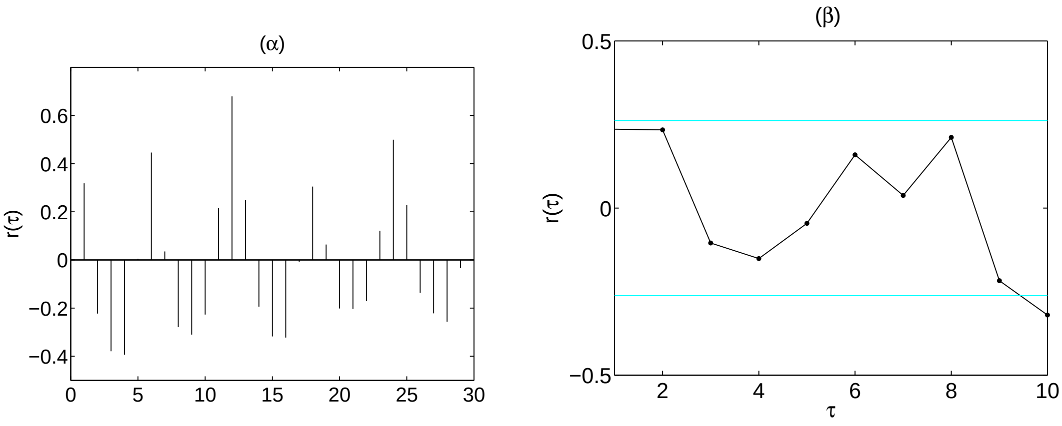
$$\hat{\rho}_\tau = r_\tau = \mathrm{Corr}(x_t, x_{t-\tau}) = \frac{\sum_{t=\tau+1}^{n}(x_t - \bar{x})(x_{t-\tau} - \bar{x})}{\sum_{t=\tau+1}^{n}(x_t - \bar{x})^2}. \qquad (6.4)$$

Οι ορισμοί της αυτοδιασποράς και αυτοσυσχέτισης έχουν νόημα όταν η χρονοσειρά είναι στάσιμη. Όταν δεν είναι στάσιμη δε μπορεί η αυτοσυσχέτιση (και η αυτοδιασπορά) να οριστούν ως συνάρτηση της υστέρησης αλλά ορίζονται για κάθε χρονική στιγμή t. Αν επιχειρήσουμε να υπολογίσουμε τη συνάρτηση αυτοσυσχέτισης ως προς την υστέρηση σε μια μη-στάσιμη χρονοσειρά με τάσεις, παρατηρούμε ότι έχει πολύ υψηλές τιμές και φθίνει πολύ αργά με την υστέρηση. Αυτό σημαίνει ότι υπάρχουν ισχυρές συσχετίσεις μεταξύ κοντινών χρονικά σημείων που είναι λόγο της τάσης. Αυτή η χαρακτηριστική μορφή της αυτοσυσχέτισης φαίνεται στο Σχήμα 6.4α.



Σχήμα 6.4: (α) Διάγραμμα αυτοσυσχέτισης για τη χρονοσειρά της μηνιαίας παραγωγής σιδήρου στην Αυστραλία την περίοδο Ιανουάριος 1956 - Αύγουστος 1995 (σε χιλιάδες τόνων). (β) Διάγραμμα αυτοσυσχέτισης για τη χρονοσειρά των πρώτων διαφορών της χρονοσειράς στο (α). Δίνονται με οριζόντιες διακεκομμένες γραμμές τα όρια της αυτοσυσχέτισης χρονοσειράς iid.

Αντίστοιχα η αυτοσυσχέτιση μιας (μη-στάσιμης) χρονοσειράς με έντονη περιοδικότητα ή εποχικότητα θα παρουσιάσει ταλαντώσεις με κορυφές σε υστερήσεις που είναι πολλαπλάσια της περιοδικότητας. Για παράδειγμα, παρατηρούμε στο Σχήμα 6.5α, πως η αυτοσυσχέτιση της χρονοσειράς των μηνιαίων τιμών GICP την περίοδο Ιανουαρίου 2001 - Αυγούστου 2005 (απαλλαγμένη από την πληθωριστική τάση) έχει χαρακτηριστικές κορυφές για υστέρηση 12 και 24. Επίσης παρατηρούμε πως εμφανίζει κορυφές για υ-



Σχήμα 6.5: (α) Διάγραμμα αυτοσυσχέτισης για τη χρονοσειρά μηνιαίων τιμών GICP την περίοδο Ιανουαρίου 2001 - Αυγούστου 2005 (απαλλαγμένη από την πληθωριστική τάση). (β) Διάγραμμα αυτοσυσχέτισης για τη χρονοσειρά υπολοίπων, όπου έχουμε αφαιρέσει από τη χρονοσειρά στο (α) την περιοδική συνιστώσα. Δίνονται με οριζόντιες διακεκομμένες γραμμές τα όρια της αυτοσυσχέτισης χρονοσειράς iid.

στερήσεις 6 και 18, το οποίο δηλώνει την ύπαρξη και εξαμηνιαίου κύκλου, μικρότερης όμως ισχύος από τον ετήσιο.

Όταν έχουμε μια χρονοσειρά με τάσεις ή περιοδικότητα η μορφή της αυτοσυσχέτισης δεν περιέχει πληροφορία για τις συσχετίσεις σε κάποιες υστερήσεις. Αυτήν την πληροφορία μπορεί ενδεχομένως να μας τη δώσει η αυτοσυσχέτιση της χρονοσειράς των υπολοίπων. Ιδιαίτερα μας ενδιαφέρει αν η χρονοσειρά των υπολοίπων έχει μηδενικές (γραμμικές) αυτοσυσχετίσεις.

Θεωρητικά η αυτοσυσχέτιση της χρονοσειράς λευκού θορύβου είναι μηδενική για $\tau > 0$. Πρακτικά όμως η αυτοσυσχέτιση εκτιμάται από μια πεπερασμένη χρονοσειρά (κάποιου μήκους n) και άρα θα έχει διακυμάνσεις γύρω από το 0. Αποδεικνύεται πως η εκτιμούμενη αυτοσυσχέτιση της χρονοσειράς λευκού θορύβου ακολουθεί κανονική κατανομή, $r_\tau \sim \mathrm{N}(0, 1/n)$. Για αυτό θεωρούμε ότι η αυτοσυσχέτιση για κάποιο τ είναι 'στατιστικά μηδενική' αν $r_\tau \in [-2/\sqrt{n}, 2/\sqrt{n}]$. [Ακριβέστερα αυτό είναι το 95% διάστημα εμπιστοσύνης για την r_τ υποθέτοντας ότι η χρονοσειρά είναι λευκός θόρυβος, όπου το 2 είναι η στρογγυλοποίηση του 1.96 και αντιστοιχεί στην κρίσιμη τιμή της τυπικής κανονικής κατανομής για πιθανότητα 0.975].

Για τα δύο παραπάνω παραδείγματα, οι συναρτήσεις αυτοσυσχέτισης των χρονοσειρών των υπολοίπων δίνονται στο Σχήμα 6.4β και στο Σχήμα 6.5β. Για την παραγωγή του σιδήρου παρατηρούμε ότι η αυτοσυσχέτιση είναι στατιστικά σημαντική για διάφορες υστερήσεις τ και άρα μπορούμε να συμπεράνουμε πως πέρα από τη στοχαστική τάση, η τιμή της παραγωγής σιδήρου σε κάποιο μήνα επηρεάζεται σε κάποιο (μικρό) βαθμό από την παραγωγή του σιδήρου του προηγούμενους μήνες. Για το δείκτη GICP, παρατηρούμε ότι οι αυτοσυσχετίσεις δεν είναι σημαντικές, δηλαδή πέρα από την πληθωριστική τάση και την εποχικότητα οι διακυμάνσεις του δείκτη φαίνεται να είναι τυχαίες.

Το ερώτημα αν μια στάσιμη χρονοσειρά, απαλλαγμένη από τάσεις και περιοδικότητα, είναι ανεξάρτητη είναι πολύ σημαντικό στην ανάλυση χρονοσειρών και αποτελεί την πρωταρχική μηδενική υπόθεση που πρέπει να απορρίψει κάποιος για να προχωρήσει στην ανάλυση της στάσιμης χρονοσειράς και να προσπαθήσει με στοχαστικά ή αιτιοκρατικά μοντέλα να εξηγήσει συσχετίσεις στη χρονοσειρά. Για τη μηδενική υπόθεση της ανεξαρτησίας έχουν αναπτυχθεί διάφοροι έλεγχοι, όπως οι έλεγχοι Portmanteau που βασίζονται στην αυτοσυσχέτιση, αλλά εδώ θα περιοριστούμε στη γραφική παράσταση της αυτοσυσχέτισης.

Ασκήσεις Κεφαλαίου 6

1. Μια από τις πιο πολυ-μελετημένες πραγματικές χρονοσειρές είναι η χρονοσειρά των ετήσιων ηλιακών κηλίδων που άρχισε να μετριέται από το 1700. Τα δεδομένα των ετήσιων ηλιακών κηλίδων από το 1700 ως το 2007 δίνονται στο αρχείο `sunspots.dat` στην ιστοσελίδα του μαθήματος, όπου στην πρώτη στήλη είναι το έτος και στη δεύτερη ο αριθμός ηλιακών κηλίδων.

   (α') Σχηματίστε το διάγραμμα ιστορίας της χρονοσειράς καθώς και το διάγραμμα της συνάρτησης αυτοσυσχέτισης.

   (β') Εντοπίστε την κύρια περίοδο στη χρονοσειρά των ετήσιων ηλιακών κηλίδων και εκτιμείστε τον κύκλο αυτής της περιόδου. Στη συνέχεια αφαιρέστε αυτήν την περιοδική συνάρτηση από την αρχική χρονοσειρά. Σχηματίστε το διάγραμμα ιστορίας της χρονοσειράς των υπολοίπων καθώς και το διάγραμμα της συνάρτησης αυτοσυσχέτισης αυτής της χρονοσειράς (μαζί με τα όρια της αυτοσυσχέτισης λευκού θορύβου). Μπορούμε να υποθέσουμε πως η χρονοσειρά των ετήσιων ηλιακών κηλίδων είναι περιοδική με κάποιο θόρυβο ή φαίνεται να υπάρχει κάποια μορφή δυναμικής;

2. Μια από τις χρονοσειρές που έχουν μελετηθεί ιδιαίτερα για το φαινόμενο του θερμοκηπίου είναι και ο δείκτης της ανωμαλίας θερμοκρασίας εδάφους στο Βόρειο Ημισφαίριο (αναφορικά με το 1961). Οι τιμές του δείκτη για την περίοδο από το 1850 ως το 2006 δίνονται στο αρχείο `crutem3nh.dat`, όπου στην πρώτη στήλη είναι το έτος και στη δεύτερη ο δείκτης θερμοκρασίας.

   (α') Σχηματίστε το διάγραμμα ιστορίας της χρονοσειράς καθώς και το διάγραμμα της συνάρτησης αυτοσυσχέτισης.

   (β') Ο δείκτης θερμοκρασίας φαίνεται να έχει αυξητική τάση (σύμφωνα και με τη θεωρία του φαινομένου του θερμοκηπίου). Εκτιμείστε την τάση αυτή με πολύωνυμο πρώτου, δεύτερου και τρίτου βαθμού. Αφαιρέστε από την αρχική χρονοσειρά τις εκτιμήσεις από κάθε μια από τις τρεις προσαρμογές της τάσης. Σχηματίστε το διάγραμμα ιστορίας για κάθε μια από τις τρεις χρονοσειρές των υπολοίπων, καθώς και τα αντίστοιχα διαγράμματα της συνάρτησης αυτοσυσχέτισης των χρονοσειρών αυτών (μαζί με τα όρια της αυτοσυσχέτισης λευκού θορύβου). Μπορούμε να υποθέσουμε πως η χρονοσειρά του δείκτη παγκόσμιας θερμοκρασίας έχει μόνο κάποια τάση ή φαίνεται να υπάρχει κάποια μορφή δυναμικής;

(γ′) Για την απαλοιφή της τάσης υπολογίστε τις πρώτες διαφορές της αρχικής χρονοσειράς. Σχηματίστε το διάγραμμα ιστορίας για τη χρονοσειρά των πρώτων διαφορών καθώς και το διάγραμμα της συνάρτησης αυτοσυσχέτισης της χρονοσειράς αυτής (μαζί με τα όρια της αυτοσυσχέτισης λευκού θορύβου). Αλλάζει το συμπέρασμα σας για το σύστημα της χρονοσειράς του δείκτη παγκόσμιας θερμοκρασίας για αυτή τη μέθοδο απαλοιφής τάσης;

3. Θεωρείστε τη λογιστική απεικόνιση που δίνεται ως $x_t = 4x_{t-1}(1 - x_{t-1})$. Δημιουργείστε μια χρονοσειρά 100 τιμών από τη λογιστική απεικόνιση. Σχηματίστε το διάγραμμα ιστορίας της χρονοσειράς καθώς και το διάγραμμα της συνάρτησης αυτοσυσχέτισης (μαζί με τα όρια της αυτοσυσχέτισης λευκού θορύβου). Τι συμπέρασμα μπορούμε να βγάλουμε για το σύστημα της λογιστικής απεικόνισης;