OT1OT1

סדרות עתיות

אילון סולן*

9 ביוני 2004

רשימות אלו מתבססות על רשימות של פרופסור דוד שטיינברג ופרופסור יואב בנימיני. תודתי נתונה לשניהם. בהכנת הרשימות השתמשתי גם בספרים הבאים:

- Time Series Analysis, Forecasting and Control. By G.E.P. Box, G.M. Jenkins and G.C. Reinsel, Prentice-Hall International Inc., 1994.
- Time Series Analysis - Univariate and Multivariate Methods. By William W.S. Wei, Addison-Wesley Publishing Company, 1989.
- Fourier Analysis of Time Series: an Introduction. By Peter Bloomfield, John Wiley and Sons, 1976.
- The Analysis of Time Series - an Introduction. By C. Chatfield, Chapman & Hall, 1996.

מעט חומר נלקח גם מהספרים הבאים:

- The Statistical Analysis of Time Series. By T.W. Anderson, John Wiley and Sons, 1958.
- Spectral Analysis and Time Series. By M.B. Priestly, Academic Press, 1981.

ברצוני להודות לרון ארטשטיין, צפריר כהן, רמה פורת ודקל צור על עזרתם בענייני LaTeX עברי.

*המחלקה לסטטיסטיקה וחקר ביצועים, בית הספר למדעי המתמטיקה, אוניברסיטת תל אביב. דואר אלקטרוני: eilons@post.tau.ac.il

תוכן ענינים

1 הקדמה

סדרה עתית היא אוסף של תצפיות שנעשות לאורך זמן. בדרך כלל נסמן סדרה עתית על ידי

$$\{y_t\}, t = 1, \ldots, N.$$

<u>מספר דוגמאות לסדרות עתיות הן:</u>

- המחיר השנתי הממוצע של חיטה במשך 70 השנים האחרונות.
- הטמפרטורה החודשית הממוצעת ברסיפה, ברזיל.
- המכירות החודשיות של חברה מסויימת.
- גובה פני הכנרת, אחת לשבועיים.
- הריכוז של תהליך כימי מסויים, אחת לשעתיים.
- מחיר הסגירה היומי של מניות IBM.
- סייסמוגרם.
- הזמן בו קרו רעידות אדמה גדולות במיוחד.

<u>תכונות של סדרת עתיות:</u>

1. התהליך אותו דוגמים יכול להיות בזמן בדיד או רציף.
2. התצפיות אינן בלתי תלויות: הן קורלטיביות.

<u>מטרת הניתוח:</u>

1. תאור: האם ניתן לזהות מגמות ארוכות טווח (trend)? האם יש גורם עונתי? האם ניתן לתאר את התהליך באופן פשוט על ידי מודל סטטיסטי?
2. חיזוי: מה ניתן לאמר על תצפיות עתידיות?
3. הסבר: האם וכיצד ניתן לקשר את סדרת התצפיות $\{y_t\}$ עם סדרת קלט ידועה $\{x_t\}$? אם התצפיות של שתי סדרות אלה היו בלתי תלויות, היינו ניגשים לבעיה זו בעזרת רגרסיה. גישת הסדרה העתית למידול סדרת הקלט-פלט מאפשרת לנו לקחת בחשבון את השפעת הזמן על התפתחות הסדרה. מודלים דינמיים מאפשרים גם לנבא ערכים עתידיים של הסדרה.

   דוגמא למקרה כזה הוא אם הסדרה $\{x_t\}$ מייצגת את כמות המשקעים היורדים באגן ניקוז של נהר מסויים, ו-$\{y_t\}$ מייצגת את כמות המים הזורמים בנהר.
4. שליטה: אם מבינים כיצד סדרת התצפיות $\{y_t\}$ מושפעת מסדרת הקלט $\{x_t\}$ ניתן "לשלוט" על הסדרה $\{y_t\}$ באמצעות שליטה על הסדרה $\{x_t\}$. למשל, אם הסדרה $\{y_t\}$ מייצגת רמה של תהליך ייצור מסויים או את מיקומו של טיל. שיטות שליטה יעילות תלויות במידה רבה במידול נכון של הקשר בין סדרת הקלט לסדרת הפלט.

גישות לניתוח סדרות עתיות:

- תחום הזמן: כאן ננסה לתאר באופן ישיר את התפתחות התהליך. הניתוח מתרכז בסדרה הנצפית $\{y_t\}$.

כלי מפתח בגישה זו הוא פונקצית המתאם המשותף (autocorrelation) של הסדרה, המתארת כיצד תצפיות בפער קבוע קשורות זו לזו.

- תחום התדר: כאן מתייחסים לערכים הנצפים כבאים מתוך עקום שהוא סכום של סינוסים וקוסינוסים בתדרים שונים. זוהי הצגה טבעית לסדרות עתיות רבות הבאות מתחומי ההנדסה והפיזיקה.

מטרת הניתוח היא לזהות את התדירויות הבסיסיות המגדירות את הסדרה, ולהעריך את הפאזה והאמפליטודה שלהן. להשגת מטרה זו הסדרה מוצגת כסכום של קוסינוסים, בעזרת כלי הנקרא "פריודוגרם". ניתוח זה נקרא גם "ניתוח ספקטרלי" או "ניתוח פורייה".

מודלים לסדרות עתיות:

1. מודלים פיזיקליים: לעתים יש מודל פיזיקלי המתאר היטב כיצד מתפתח התהליך עם הזמן. נקודה זו נכונה במיוחד עבור נתונים פיזיקליים. למשל, הסייסמוגרם ניתן לתאור די טוב בעזרת התאוריה של תנועת הגלים. התאוריה מרמזת כי הנתונים ניתנים לייצוג כסכום של קוסינוסים, ולכן כיוון הניתוח הראשון אמור להיות בתחום התדר. לרוב מודלים פיזיקליים הם בעלי הצורה

$$\frac{dY}{dt} = k(y, t; \theta).$$

פתרונות של משוואה זו מתארים מסלולים של הפונקציה Y כתלות בזמן.

לרוב המודל שאנו נחקור יהיה תלוי בפרמטר לא ידוע θ, שאותו יש לאמוד מסדרת הנתונים. כשנתונות לנו שתי סדרות או יותר, נצטרך לעיתים לפתור מערכת של משוואות דיפרנציאליות.

אנו אומרים כי המודל דטרמיניסטי אם המשוואה נותנת תאור מדוייק של התפתחות הסדרה. לרוב, אפילו אם בידינו תאור פיזיקלי טוב, הסדרה הנצפית תסטה מהעקום האידאלי. במקרה כזה אנו יכולים לחשוב על המודל הפיזיקלי כעל מודל הנותן את התוחלת של Y_t, ועל הסטיה כעל רעש מקרי סביב התוחלת. סטיות מהמודל המוצע (residuals) לרוב אינן רנדומיות, ויש למדל את הקשר ביניהן.

2. פירוק לרכיבים - מגמה, עונתיות ורעש: מודלים מסוג זה רווחים בנתונים כלכליים או אטמוספריים שיש בהם מחזורים שנתיים או יומיים. דוגמאות כלליות הן:

$$\begin{aligned} Y_t &= m_t + S_t + \epsilon_t, \\ Y_t &= m_t S_t + \epsilon_t, \\ Y_t &= m_t S_t \epsilon_t, \end{aligned}$$

באשר m_t מייצג מגמה ארוכת טווח, S_t הוא רכיב עונתי, ו-ϵ_t הוא הרעש.

כל אחד מהרכיבים הללו מעניין לצרכים אחרים. לדוגמא, אם הסדרה $\{Y_t\}$ מייצגת את מדד המניות, משקיע סולידי החוסך כסף לפנסיה מתעניין רק ברכיב המגמה, משקיע מתוחכם יותר המחזיק במניותיו מספר שנים מתעניין ברכיב העונתי, ומשקיע מאוד מתוחכם, המוכר וקונה מניות באופן יומיומי, מתעניין ברכיב הרעש.

דוגמא נוספת היא מזג האויר: התחממות כדור הארץ מיוצגת על ידי רכיב המגמה, המחזוריות השנתית מיוצגת על ידי הרכיב העונתי, ורכיב הרעש חשוב לחיזוי מזג האויר מחר.

לעיתים ניתן לתת צורה סגורה לרכיבי המגמה והעונתיות. למשל, $m_t = a + bt$ יתאר מגמה לינארית בזמן.

הרכיב העונתי יכול להיות קוסינוס או אוסף של אינדיקטורים. למשל, אם ברשותנו נתונים כלכליים לכל רבעון, התרומה של S_t יכולה להיות תוספת של קבוע שונה לכל רבעון.

לעיתים אפשר להשתמש בטכניקות סטטיסטיות סטנדרטיות כמו רגרסיה לינארית כדי להתאים מודל כזה. יש לשים לב כי רכיב המגמה במודל לעיתים לא לוקח בחשבון את כל התלות בזמן, כך שגורמי הרעש ϵ_t יהיו מתואמים. ניתוח רגרסיה נכון יצטרך לקחת בחשבון קורלציה זו.

ייתכן ויהיו מספר אינדיקטורים עונתיים שייצגו התנהגות מחזורית בתדרים שונים. למשל, מודל של הטמפר-טורה (קריאה אחת כל שעה) ייקח בחשבון מחזוריות יומית ושנתית. לעיתים קשה להבדיל בין מגמה ארוכת טווח ובין מחזוריות עונתית. נניח, לדוגמא, שברשותנו קריאת טמפרטורה (אחת לשעה) בין החודשים מרץ ומאי. לולא הידע שלנו על התנהגות הטמפרטורה היינו עלולים לטעות ולחשוב כי העליה בטמפרטורה היא מגמה ארוכת טווח, ולא חלק ממחזוריות עונתית.

כדי לזהות מגמה ארוכת טווח אנו מנסים להחליק את המחזוריות העונתית קצרת-הטווח ואת גורם הרעש. לשם כך משתמשים לרוב ב"פילטרים לינאריים". הפילטר הלינארי הפשוט ביותר הוא מיצוע-נע:

$$X_t = \frac{1}{2q+1}\sum_{u=t-q}^{t+q} Y_u. \tag{1}$$

מגמה ארוכת טווח ניתנת לזיהוי ביתר קלות בסדרה X_t. אפשרות אחרת היא להשתמש בפילטר עם מקדמים שהם הסתברויות בינומיאליות.

שני הפילטרים שציינו כאן הם פילטרים לינאריים: הסדרה $\{X_t\}$ היא פונקציה לינארית של הסדרה $\{Y_t\}$. דוגמא לפילטר לא-לינארי שימושי הוא הפילטר הבא.

$$X_t = \text{median}\{X_{t-q}, X_{t-q+1}, \ldots, X_{t+q}\}. \tag{2}$$

כלומר, X_t הוא החציון של $2q+1$ הערכים שסביב Y_t. פילטר זה לא רגיש לערכים חריגים בסדרה (outliers): אם יש לכל היותר q חריגים בין $2q+1$ הערכים שסביב Y_t, החציון עדיין יהיה אחד מ-$q+1$ הערכים שאינם חריגים.

ישנם פילטרים היכולים להעלים את המגמה ארוכת-הטווח, ולאפשר להתרכז ברכיב העונתי. דוגמא לכך הוא פילטר ההפרש הכפול:

$$X_t = Y_{t+1} - 2Y_t + Y_{t-1}.$$

לזיהוי גורמים עונתיים וגורמים מחזוריים שאינם עונתיים, כמו בדוגמא של הטמפרטורה, נשתמש בניתוח בתחום התדר, אותו נראה לקראת סוף הסימסטר.

בתחום התדר, פילטר המיצוע-הנע מחליש את התדירויות הגבוהות, ומשאיר רק את התדירויות הנמוכות. לכן פילטר זה נקרא פילטר תדירויות נמוכות (low pass filter). פילטרים כאלה רווחים בין מהנדסים. גישה רווחת

היא להפעיל פילטרים במרחב התדר; כך ניתן לשלוט במדויק אילו תדירויות מוחקים ואילו תדירויות משאירים. פילטר תדירויות גבוהות (high pass filter) מעלים תדירויות נמוכות, ומשאיר תדירויות גבוהות. פילטר ההפרש הכפול הוא פילטר מסוג זה.

לאחר שניקינו מהסדרה את קווי המגמה ואת הגורמים המחזוריים אנו נשארים עם סדרת שאריות. רוב הקורס יתמקד בהבנת הקשר שבין השאריות הנ"ל. כלומר, בהינתן סדרת השאריות עד הזמן הנוכחי, מה נוכל לאמר על השארית הבאה. נעשה זאת בעזרת מודלים סטוכסטים, שאותם נסקור כעת.

3. מודלים סטוכסטים: המודלים בהם אנו נשתמש ברובו של הקורס יתארו את Y_t כריאליזציה של תהליך סטוכסטי. משפחה שימושית במיוחד של תהליכים סטוכסטים היא משפחת המודלים האוטו-רגרסיביים-מיצוע-נע (autoregressive-moving average, ARMA). מודלים אלה משתמשים במשוואת הפרשים כדי לתאר את Y_t. אחד המודלים אותו נראה הוא מודל $\mathrm{ARMA}(p,q)$, שהוא מהצורה

$$Y_t - \phi_1 Y_{t-1} - \ldots - \phi_p Y_{t-p} = X_t + \theta_1 X_{t-1} + \ldots + \theta_q X_{t-q},$$

באשר (X_t) היא סדרה של רעש מקרי בלתי תלוי עם תוחלת 0. אגף שמאל הוא אוטו-רגרסיה מסדר p, ואגף ימין הוא ממוצע-נע מסדר q. (נשים לב כי השימוש כאן במונח "מיצוע-נע"שונה מהשימוש שנעשה במונח זה בפילטר תדירויות נמוכות).

ניתן לחשוב על משוואות הפרשים כעל האנלוג הדיסקרטי של משוואות דיפרנציאליות בתאור מודלים פיזיקליים. הצורה המדוייקת של הרעש מגדירה את טבעו של המודל.

<u>שיטת הניתוח:</u>

1. נפתח משפחה שימושית של מודלים.
2. נפעיל כלים סטטיסטיים על הסדרה העתית הנצפית בכדי לזהות מודל אפשרי מהמשפחה.
3. נאמוד את הפרמטרים של המודל.
4. נבדוק האם המודל המועמד מתאים לסדרה הנתונה.
5. אם המודל מתאים, נסיק מסקנות, ננבא ערכים עתידיים וכו'.

אם המודל אינו מתאים, נזהה משפחה אלטרנטיבית של מודלים, ונחזור ל-2.

1.1 הצגה גרפית של הסדרה

לאחר שקיבלנו נתונים כלליים על הסדרה, והגדרנו מה מטרת הניתוח, הצעד הראשון הוא להציג את הסדרה באופן גרפי. הצגה זו חשובה הן כדי לתאר את הנתונים, והן כדי לנסות ולפתח מודל מתאים שיסביר את הסדרה.

כדי להציג את הסדרה נכון, יש לבחור סקאלות, ואת הדרך בה יוצגו הנתונים (למשל, הצגה רציפה או נקודות נפרדות). לבחירות אלו יש לעיתים השפעה גדולה על הבנת הסדרה.

כיום ההצגה נעשית לרוב בעזרת תוכנות סטטיסטיות. יש תוכנות סטטיסטיות בהן ההצגה הגרפית נוחה לשימוש, ויש כאלו בהן הדבר אינו כך.

1.2 טרנספורמציות על הנתונים

ההצגה הגרפית יכולה לרמוז לנו כי יש לעשות טרנספורמציה מסויימת על הנתונים, למשל, לוגריתם או שורש ריבועי. שלוש הסיבות העיקריות להפעלת טרנספורמציה הן:

1. ייצוב השונות.

אם הנתונים מראים קיום מגמה, ואם נראה כי השונות עולה עם התוחלת, ייתכן ויש להפעיל טרנספורמציה על הסדרה. סטית התקן הפרופורציונית לתוחלת מרמזת על הצורך בטרנספורמציה לוגריתמית.

2. להפיכת הרכיב העונתי לאדיטיבי.

אם ניתן לזהות מגמה, והרכיב העונתי עולה עם התוחלת, מרמז הדבר כי יש להפעיל טרנספורמציה על הנתונים שתהפוך את הרכיב העונתי לאדיטיבי. אם הרכיב העונתי פרופורציוני לתוחלת, הוא נקרא "כיפלי", וניתן להפעיל טרנספורמציה לוגריתמית. יש לזכור כי טרנספורמציה זו תייצב את השונות רק אם רכיב הרעש הוא גם כיפלי.

3. להפיכת הנתונים למתפלגים נורמלית.

הניתוח שנפתח בקורס מניח ברובו כי הרעש מתפלג נורמלית. באופן מעשי הנחה זו לעיתים אינה מתקיימת.

טרנספורמציות הלוגריתם והשורש הריבועי הן שתי דוגמאות מתוך משפחה של טרנספורמציות שנקראות טרנספורמציות בוקס-קוקס (Box-Cox transformations). לכל פרמטר λ ממשי, הטרנספורמציה עם פרמטר λ המופעלת על הסדרה $\{y_t\}$ מגדירה סדרה חדשה $\{z_t\}$ באופן הבא:

$$z_t = \begin{cases} \frac{(y_t)^\lambda - 1}{\lambda} & \lambda \neq 0, \\ \\ \log(y_t) & \lambda = 0. \end{cases}$$

לרוב הניסיון ויכולת הניחוש מצביעים על הערך האופימלי של λ שיש לבחור.

לעיתים טרנספורמציה שהופכת את הרכיב העונתי לאדיטיבי לא מייצבת את השונות, ולא ניתן למצוא דרך להשיג את שתי המטרות בו זמנית.

1.3 מציאת רכיב המגמה

באופן כללי, מגמה היא שינוי ארוך טווח בתוחלת של התהליך. אין הגדרה פורמלית מקובלת של רכיב זה.

רכיב המגמה הפשוט ביותר הוא מגמה לינארית:

$$y_t = \alpha + \beta t + x_t,$$

באשר $\{x_t\}$ היא סדרת "רעש", ו-α, β הם קבועים. תרומת רכיב המגמה בזמן t היא, אם כן,

$$m_t = \alpha + \beta t.$$

לרוב רכיב המגמה משתנה עם הזמן, ואנו נתאים קו מגמה שונה בקטעי זמן שונים.

אם ניתן לזהות גם רכיב עונתיות, נחשב תחילה ממוצע של הסדרה בכל מחזור, וננסה לנתח את סדרת הממוצעים כדי לאפיין את רכיב המגמה.

אם לא ניתן לזהות רכיב עונתי, הגישה המסורתית היא למצוא קו מגמה שהוא פונקציה פשוטה, למשל עקום פולינומיאלי (קו לינארי, עקום ריבועי, וכו'). ניתן להתאים את עקום Gompertz הנתון על ידי

$$\log(y_t) = a + br^t,$$

באשר a, b, r הם קבועים ו-$0 < r < 1$, או את ה-logistic curve הנתון על ידי

$$x_t = \frac{a}{1 + be^{-ct}}.$$

דרך אחרת לזיהוי רכיב המגמה היא על ידי פילטר לינארי, לדוגמא פילטר המיצוע-נע

$$z_t = \frac{1}{2q+1} \sum_{j=t-q}^{t+q} y_t.$$

אם q גדול דיו, התרומה של רכיבי הרעש ורכיב העונתיות לסדרה $\{z_t\}$ יהיה קטן, וניתן יהיה לזהות בקלות רבה יותר את תרומת רכיב המגמה.

ישנם פילטרים מקובלים נוספים (כלומר, פילטרים לינארים עם מקדמים השונים מהמקדמים בפילטר מיצוע-נע, או פילטרים לא לינארים). הבחירה בפילטר הספציפי בו צריך להשתמש דורשת נסיון וידע תאורטי על ניתוח בתחום התדר, אותו נראה בהמשך.

2 מודלים סטוכסטיים לסדרות עתיות

הגדרה 2.1 **תהליך סטוכסטי** הוא סדרה אינסופית של משתנים מקריים $\{Y_t\}$.

אנו נראה סדרה עתית כריאליזציה של תהליך סטוכסטי. לדוגמא, אם התהליך הוא ההילוך מקרי, כלומר, $Y_0 = 0$ ו-$Y_{t+1} = Y_t + X_t$, באשר X_t הוא משתנה מקרי ברנולי עם פרמטר חצי, הסדרה העתית היא סדרה ספציפית של מספרים שלמים שיוצרה בעזרת התהליך הנ"ל.

שלוש דוגמאות פשוטות לתהליכים סטוכסטיים הם:

1. $\{Y_t\}$ היא סדרה של משתנים מקריים נורמלים בלתי תלויים שווי התפלגות. תהליך כזה נקרא **תהליך אקראי טהור** או **תהליך רעש לבן**.

2. $\{Y_t\}$ מוגדרת באופן הבא: $Y_0 = 0$ ו-$Y_{t+1} = Y_t + X_t$, באשר X_t היא סדרה של משתנים מקריים בלתי-תלויים, שווי-התפלגות עם תוחלת 0 ושונות סופית. תהליך כזה נקרא **תהליך הילוך מקרי הוגן**.

3. $\{Y_t\}$ מוגדרת באופן הבא: $Y_0 = 0$ ו-$Y_{t+1} = Y_t + X_t$, באשר X_t היא סדרה של משתנים מקריים בלתי-תלויים, שווי-התפלגות עם תוחלת שאינה 0 ושונות סופית. תהליך כזה נקרא **תהליך הילוך מקרי לא-הוגן**.

הגדרה 2.2 תהליך סטוכסטי $\{Y_t\}$ הוא **סטציונרי** אם לכל m חיובי, לכל $t_1, \ldots, t_m$, ולכל τ חיובי, למשתנים המקריים הרב-ממדיים $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m})$ ו-$(Y_{t_1+\tau}, Y_{t_2+\tau}, \ldots, Y_{t_m+\tau})$ יש אותה התפלגות.

בפרט, אם התהליך סטציונרי אזי ההתפלגות של Y_t אינה תלויה ב-t. לכן, התוחלת

$$\mu(t) = \mathbf{E}[Y_t]$$

והשונות

$$\sigma^2(t) = \mathrm{Var}(Y_t) = \mathbf{E}[(Y_t - \mu(t))^2]$$

(אם הן קיימות) אינן תלויות ב-t, וניתן פשוט לסמנן ב-μ ו-σ^2.

סדרה עתית היא ריאליזציה סופית של התהליך הסטוכסטי. נסמן את הסדרה העתית הנתונה לנו על ידי $(y_1, \ldots, y_N)$. כלומר, אנו מקבלים N תצפיות בלבד.

מכיוון שאנו רואים תצפית אחת מכל אחד מהמשתנים המקריים $Y_1, \ldots, Y_N$, אם התהליך הסטוכסטי הוא כללי לא נוכל ללמוד דבר על התהליך מתוך התצפיות.

כדי שהסדרה הנצפית תספק לנו מידע על התהליך, התהליך שיצר את הסדרה חייב לקיים חוקיות כלשהי, למשל, שהוא סטציונרי.

אם התהליך סטציונרי, ניתן לאמוד את התוחלת והשונות על ידי שימוש בתצפיות:

$$\begin{aligned}\widehat{\mu} &= \frac{1}{N}\sum_{t=1}^{N} y_t = \overline{y}, \\ \widehat{\sigma}^2 &= \frac{1}{N}\sum_{t=1}^{N}(y_t - \overline{y})^2.\end{aligned}$$

האומד הלא מוטה לשונות מתקבל על ידי חלוקה ב-$N-1$ במקום חלוקה ב-N. ההגדרה שנתנו כאן לא משנה בהרבה את האומד, ומפשטת את הנוסחאות. קירובים דומים נעשה גם בהמשך.

2.1 פונקצית השונות המשותפת (auto-covariance) ופונקצית המתאם המשותף (auto-correlation)

מאפיין אחד של סדרה עתית הוא כיצד ערכים שכנים תלויים האחד בשני.
פונקצית השונות המשותפת מודדת את השונות המשותפת בין ערכי הסדרה בפער קבוע. היא מוגדרת על ידי:

$$\gamma_k(t) = \mathrm{Cov}(Y_t, Y_{t+k}) = \mathbf{E}[(Y_t - \mu(t))(Y_{t+k} - \mu(t))].$$

נשים לב כי $\gamma_0 = \sigma^2(t)$.
אם $\{Y_t\}$ הוא תהליך סטציונרי, השונות המשותפת γ_k אינה תלויה ב-t.
כאשר התהליך סטציונרי נגדיר את פונקצית המתאם המשותף על ידי

$$\rho_k = \frac{\gamma_k}{\gamma_0}.$$

טענה 2.3 פונקצית המתאם המשותף מקיימת את התכונות הבאות:

1. $\rho_0 = 1$.
2. $\rho_{-k} = \rho_k$.
3. $|\rho_k| \le 1$ לכל k.
4. פונקצית המתאם המשותף אינה מתארת באופן יחיד את התהליך הסטוכסטי. כלומר, יש שני תהליכים סטוכסטים שונים להם יש אותה פונקצית מתאם משותף.

הוכחה. שתי הטענות הראשונות נובעות מתוך הגדרת פונקצית המתאם המשותף.
כדי להוכיח את הטענה השלישית נשתמש בזהות

$$\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\mathrm{Cov}(X, Y).$$

מכיוון שהתהליך סטציונרי,

$$\begin{aligned} 0 &\le \mathrm{Var}(Y_t \pm Y_{t+k}) \\ &= \mathrm{Var}(Y_t) + \mathrm{Var}(Y_{t+k}) \pm 2\mathrm{Cov}(Y_t, Y_{t+k})) \\ &= 2\sigma^2 \pm 2\gamma_k. \end{aligned}$$

ומכאן נובעת הטענה.
כדי להוכיח את הטענה הרביעית, יהא Z משתנה מקרי בעל שונות סופית, ונגדיר תהליך $\{Y_t\}$ באופן הבא. Y_t הוא בלתי תלוי ב-$\{Y_j\}_{j<t}$, והתפלגותו זהה להתפלגות Z. כלומר, התהליך $\{Y_t\}$ מייצר סדרה של תצפיות בלתי תלויות של המשתנה המקרי Z. במקרה זה פונקצית המתאם המשותף של התהליך $\{Y_t\}$ היא $\rho_k = 0$ לכל $k \neq 0$. ■

2.2 סטציונריות חלשה

בסעיף זה נגדיר את מושג הסטציונריות החלשה.

הגדרה 2.4 תהליך סטוכסטי $\{Y_t\}$ נקרא **סטציונרי חלש** *(weak stationary)* אם $\mathbf{E}[Y_t]$ אינו תלוי ב-t, ו-$\mathrm{Cov}(Y_t, Y_{t+k})$ תלוי רק בפער k, ולא ב-t.

מההגדרות נובע כי כל תהליך סטציונרי הוא סטציונרי חלש, אך ההיפך אינו בהכרח נכון.

עבור רוב המאפיינים של תהליכים שנראה, די לדרוש כי התהליך הוא סטציונרי חלש.

דוגמא למשפחה של תהליכים סטציונריים חלשים היא משפחת התהליכים הנורמלים, בהם ההתפלגות המשותפת של $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m})$ היא רב-נורמלית לכל אוסף של אינדקסים שונים $t_1, t_2, \ldots, t_m$. ההתפלגות הרב-נורמלית מאופיינת על ידי המומנטים הראשון והשני, ולכן תהליך נורמלי מאפיין על ידי תוחלתו ופונקצית השונות המשותפת.

כדי לראות כי לא כל תהליך סטציונרי חלש הוא סטציונרי נתבונן בדוגמא הבאה. לכל t יהא Y_t משתנה מקרי ברנולי עם פרמטר $1/2$, כלומר,

$$\mathbf{P}(Y_t = 0) = \mathbf{P}(Y_t = 1) = \frac{1}{2}.$$

כעת נגדיר את ההתפלגות המשותפת של הסדרה $(\{Y_t\})$.

המשתנים המקריים $(Y_t)_{t \neq 3}$ הם בלתי תלויים. לעומת זאת, המשתנה המקרי Y_3 מוגדר באופן הבא:

$$Y_3 = Y_1 + Y_2 \mod 2.$$

קל לראות כי Y_3 בלתי תלוי ב-Y_t, לכל $t \neq 3$. לכן, Y_i ו-Y_j בלתי תלויים לכל $i \neq j$, ולכן הסדרה $\{Y_t\}$ היא סטציונרית חלש.

מכיוון שההתפלגות המשותפת של $\{Y_1, Y_2, Y_3\}$ שונה מההתפלגות המשותפת של $\{Y_4, Y_5, Y_6\}$, הסדרה אינה סטציונרית.

2.3 דוגמאות לתהליכים סטוכסטים

2.3.1 תהליך אקראי טהור (**purely random process**)

הגדרה 2.5 סדרה של משתנים מקריים נקראת **תהליך אקראי טהור** אם המשתנים המקריים הם בלתי תלויים ושווי התפלגות.

תהליך אקראי טהור הוא סטציונרי, ומקיים:

$$\gamma_k = 0, k = \pm 1, \pm 2, \ldots$$

תהליך אקראי טהור נקרא גם רעש לבן, וישמש דוגמא לרעש ללא כל מבנה מיוחד.

2.3.2 הילוך מקרי (**random walk**)

הגדרה 2.6 נניח ש-$\{X_t\}$ הוא תהליך אקראי טהור עם תוחלת μ_X ושונות σ_X^2. תהליך $\{Y_t\}$ יקרא **הילוך מקרי** אם

$$Y_t = Y_{t-1} + X_t.$$

בדרך כלל נתחיל עם $Y_0 = 0$, ואז

$$Y_t = \sum_{j=1}^{t} X_j.$$

תכונות התהליך $\{Y_t\}$ הן:

$$\begin{aligned}\mathbf{E}[Y_t] &= t\mu_X, \\ \mathrm{Var}(Y_t) &= t\sigma_X^2.\end{aligned}$$

בפרט, תהליך זה אינו סטציונרי.

נשים לב כי התהליך $\{\nabla Y_t\}$ המוגדר על ידי

$$\nabla Y_t = Y_t - Y_{t-1} = X_t$$

הוא תהליך אקראי טהור, ובפרט סטציונרי.

2.3.3 תהליך מיצוע-נע (moving average process)

הגדרה 2.7 נניח ש-$\{X_t\}$ הוא תהליך אקראי טהור עם תוחלת 0 ושונות σ_X^2. תהליך $\{Y_t\}$ יקרא **תהליך מיצוע נע מסדר** q אם קיימים קבועים $\theta_0, \theta_1, \ldots, \theta_q$ כך שלכל t מתקיים

$$Y_t = \theta_0 X_t + \theta_1 X_{t-1} + \cdots + \theta_q X_{t-q}.$$

משפחת התהליכים האלה תסומן על ידי $\mathrm{MA}(q)$.

מכיוון שניתן להכפיל את כל הקבועים θ_k בקבוע, ולחלק את $\{X_t\}$ באותו קבוע, ופעולה זו אינה משנה את התהליך $\{Y_t\}$, בדרך כלל נניח כי $\theta_0 = 1$.

כמה תכונות של תהליך מיצוע נע מסדר q הן:

$$\begin{aligned}\mu(t) &= \mathbf{E}[Y_t] = \sum_{j=0}^{q} \theta_j \mathbf{E}[X_{t-j}] = 0, \\ \sigma^2(t) &= \mathrm{Var}(Y_t) = \sigma_X^2 \sum_{j=0}^{q} (\theta_j)^2.\end{aligned}$$

מכיוון ש:

$$\mathrm{Cov}(X_t, X_{t+k}) = \begin{cases} 0 & k \neq 0, \\ \sigma_X^2 & k = 0, \end{cases}$$

ומכיוון ש:

$$\begin{aligned}\gamma_k &= \mathrm{Cov}(Y_t, Y_{t+k}) \\ &= \mathrm{Cov}(\theta_0 X_t + \cdots + \theta_q X_{t-q}, \theta_0 X_{t+k} + \cdots + \theta_q X_{t+k-q})\end{aligned}$$

נובע כי

$$\gamma_k = \begin{cases} 0 & k > q, \\ \sigma_X^2 \sum_{j=0}^{q-k} \theta_j \theta_{j+k} & k = 0, 1, \ldots, q, \\ \gamma_{-k} & k < 0. \end{cases}$$

לכן

$$\rho_k = \begin{cases} 0 & k > q, \\ \frac{\sum_{j=0}^{q-k} \theta_j \theta_{j+k}}{\sum_{j=0}^{q} \theta_j^2} & k = 0, 1, \ldots, q, \\ \rho_{-k} & k < 0. \end{cases} \tag{3}$$

נשים לב כי פונקצית המתאם המשותף נקטמת אחרי פער של q.

לדוגמא, עבור $q = 1$ נקבל:

$$\begin{aligned} \gamma_0 &= \sigma_X^2(1 + \theta_1^2), \\ \gamma_1 &= \theta_1 \sigma_X^2, \\ \gamma_k &= 0, \qquad k > 1. \end{aligned}$$

ולכן פונקצית המתאם המשותף נתונה על ידי:

$$\rho_k = \begin{cases} 1 & k = 0, \\ \frac{\theta_1}{1+\theta_1^2} & k = \pm 1, \\ 0 & k = \pm 2, \pm 3, \ldots \end{cases}$$

2.3.4 תהליך גאוסי (gaussian process)

הגדרה 2.8 תהליך $\{Y_t\}$ יקרא **תהליך גאוסי** אם ההתפלגות המשותפת של $\{Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m}\}$ היא מולטי-נורמלית לכל m ולכל $t_1, \ldots, t_m$ (עם פרמטרים שונים לכל m ולכל $t_1, \ldots, t_m$).

מסתבר כי תהליך כזה נקבע באופן חד ערכי על ידי התוחלת μ, השונות σ^2, ופונקצית המתאם המשותף.

2.4 אמידת γ_k ו-ρ_k

אם ברשותנו N תצפיות $y_1, \ldots, y_N$ יש לנו $N-k$ זוגות של תצפיות בפער k. יש מספר דרכים בהן ניתן להשתמש במידע לשם אמידת γ_k ו-ρ_k.

האומדים המקובלים הם:

$$\begin{aligned} c_k &= \widehat{\gamma}_k = \frac{1}{N} \sum_{t=1}^{N-k} (y_t - \overline{y})(y_{t+k} - \overline{y}), \\ r_k &= \widehat{\rho}_k = \frac{c_k}{c_0}. \end{aligned}$$

בחישוב c_k, החלוקה היא ב-N במקום ב-$N-k$. ראשית, דבר זה אינו משנה את התוצאות באופן ניכר כאשר $N \gg k$. באופן מעשי, אין סיבה לחשב את c_k עבור $k \geq N/4$. שנית, למרות שלאומד בו המכנה הוא $N-k$ יש הטיה קטנה יותר, הרי ש-Jenkins and Watts (1968) טוענים כי לאומד זה יש שגיאה ריבועית ממוצעת (mean square error) גדולה יותר. למרות ש-c_k הוא אומד מוטה, ניתן להראות כי ההטיה היא מסדר גודל של $1/N$, וכן

$$\lim_{N \to \infty} \mathbf{E}[c_k] = \gamma_k,$$

ולכן אסימפטוטית האומד אינו מוטה.

שיטה נוספת לאמידת פונקצית השונות המשותפת פותחה על ידי Quenouille, והיא נקראת **שיטת האולר** (jackknife estimation). בשיטה זו מחלקים את הסדרה העתית לשני חצאים, ואומדים את פונקצית השונות המשותפת בכל חצי ובסדרה כולה. נסמן ב-c_{k1} וב-c_{k2} את האומדים לפונקצית השונות המשותפת בכל אחד מחצאי הסדרה, וב-c_k את פונקצית השונות המשותפת בסדרה כולה. אזי אומד האולר לפונקצית השונות המשותפת נתון על ידי:

$$\widehat{c}_k = 2c_k - \frac{1}{2}(c_{k1} + c_{k2}).$$

ניתן להראות כי אומד זה מוריד את ההטיה של האומד מסדר גודל של $\frac{1}{N}$ לסדר גודל של $\frac{1}{N^2}$. יתרון נוסף של שיטה זו הוא שניתן לראות אם לשני חצאי הסדרה יש תכונות מתאם דומות, ובכך לראות האם הסדרה יוצרה מתהליך סטציונרי. חסרון של שיטה זו הוא שהיא דורשת יותר חישובים.

אומד האולר ל-ρ_k הוא

$$\widehat{r}_k = 2r_k - \frac{1}{2}(r_{k1} + r_{k2}),$$

באשר r_{k1} ו-r_{k2} הם האומדים לפונקצית המתאם המשותף בשני חצאי הסדרה.

קשה מאוד למצוא תכונות של פונקצית המתאם המשותף (למשל, את $\mathbf{E}[r_k]$ או $\mathrm{Cov}(r_k, r_{k+t})$), מכיוון שפונקצית המתאם המשותף היא מנה של שני משתנים מקריים לא פשוטים.

2.5 טעות התקן של r_k (standard error)

אם $\{Y_t\}$ הם משתנים מקריים בלתי תלויים ושווי התפלגות, Kendall, Stuart and Ord הוכיחו כי

$$\mathbf{E}[r_k] \approx -\frac{1}{N}, \quad \mathrm{Var}(r_k) \approx \frac{1}{N},$$

וההתפלגות של r_k שואפת להתפלגות נורמלית כאשר N שואף לאינסוף, תחת הנחות חלשות.

עבור תהליך גאוסיאני סטציונרי, Bartlett הוכיח כי

$$\mathrm{Var}(r_k) \approx \frac{1}{N} \sum_{j=-\infty}^{\infty} \left(\rho_j^2 + \rho_{j+k}\rho_{j-k} - 4\rho_k\rho_j\rho_{j-k} + 2\rho_j^2\rho_{j-k}^2\right).$$

אם פונקצית המתאם המשותף מתאפסת עבור פערים גבוהים, כלומר, אם יש q כך ש-$\rho_j = 0$ כאשר $|j| > q$ אזי השונות של r_k עבור $k > q$ היא בקרוב

$$\mathrm{Var}(r_k) \approx \frac{1}{N}\left(1 + \sum_{j=1}^{q} \rho_j^2\right). \tag{4}$$

השורש של ערך זה הוא טעות התקן עבור פערים גדולים.

באופן דומה, השונות המשותפת בין r_k ו-r_{k+s} עבור פערים גדולים הוא בקירוב:

$$\mathrm{Cov}(r_k, r_{k+s}) \approx \frac{1}{N} \sum_{j=-\infty}^{\infty} \rho_j\rho_{j+s}.$$

אם s גדול, אזי r_k ו-r_{k+s} הם כמעט לא מתואמים (uncorrelated), אבל r_k שכנים יכולים להיות מתואמים.

דוגמא: נתונות לנו 200 תצפיות מתהליך סטוכסטי $\{Y_t\}$, כלומר $N = 200$. נניח עוד כי במציאות $\rho_1 = 0.4$ (כלומר, תצפיות שכנות הן מתואמות) ולכל $k \geq 2$ מתקיים $\rho_k = 0$.

נניח כי האומדים ל-ρ_k שחישבנו מהנתונים הם:

k	1	2	3	4	5	6	7	8	9	10	11	12
r_k	.38	$-.04$	$-.17$	$-.05$	$-.05$	$-.08$	$-.12$	.02	$-.12$	$-.11$	$-.04$	$-.07$

.

אנו משערים כי התצפיות באות מתהליך גאוסיאני סטציונרי וכי הן בלתי תלויות. במקרה כזה $q = 0$, ולכן $\mathrm{Var}(r_k) \approx \frac{1}{200}$ לכל k. לכן טעות התקן היא $1/\sqrt{200} \approx 0.07$. מכיוון ש-$r_1 = 0.38$ הוא 5 פעמים טעות התקן, אנו מקבלים עדות די חזקה לכך שההשערה שלנו אינה נכונה.

אם אנו משערים כי $\rho_k = 0$ לכל $k \geq 2$, נקבל כי

$$\mathrm{Var}(r_k) \approx \frac{1}{200}\left(1 + 2(.38)^2\right) = 0.0064,$$

ולכן טעות התקן היא בערך $\sqrt{0.0064} = 0.08$. מהנתונים r_3 הוא מעט יותר מפעמיים טעות התקן, ולכן יש עדות חלשה לכך שיש קורלציה בפער 3. כל שאר הקורלציות הן בהחלט בטווח של טעות מדידה.

2.6 הקורלוגרם (correlaogram)

כלי שימושי בניתוח סדרה עתית הוא הקורלוגרם: זהו הגרף של הערכים r_k לכל פער k.

מציאת משמעות בקורלוגרם היא משימה לא פשוטה הדורשת נסיון. ניתן כאן כמה קווים מנחים.

2.6.1 סדרה אקראית

אם הסדרה העתית יוצרה על ידי תהליך אקראי טהור, אזי אם הסדרה ארוכה דיה (כלומר, אם N גדול), נקבל כי $r_k \approx 0$ לכל $k \neq 0$. ליצר דיוק, במקרה זה r_k מתפלג $N(0, \frac{1}{N})$, ולכן 95% ממקדמי המתאם יהיו באינטרבל $\pm\frac{2}{\sqrt{N}}$. צידה השני של המטבע הוא ש-5% ממקדמי המתאם יראו משמעותיים.

נקודה זו מדגימה מדוע אין זה קל לפרש את הקורלוגרם: מספר מקדמים עלולים להראות משמעותיים גם כאשר אין כל סיבה הגורמת לכך (ראה איור 4.1).

2.6.2 מתאם קצר-טווח (short-term correlation)

בסדרות עתיות סטציונריות רבות יש מתאם קצר-טווח, המאופיין על ידי מתאם בפער 1 (ערך גבוה של r_1). ככל שהפער גדל, המתאם קטן, עד שמפער מסויים לא קיים מתאם משמעותי. כלומר, הערך של r_k עבור פערים גדולים נוטה להיות קרוב ל-0. דוגמא פונקצית מתאם משותף כזו ניתן לראות באיור 2.1.

בסדרה כזו, אם התצפית בזמן t היא מעל לתוחלת, התצפיות הבאות יטו להיות מעל התוחלת, בעוד שאם התצפית בזמן t היא מתחת לתוחלת, התצפיות הבאות יטו להיות גם כן מתחת לתוחלת.

2.6.3 סדרה עתית מתחלפת (alternating series)

אם התצפיות נוטות להתחלף, כלומר אם אחרי תצפית גבוהה נקבל לרוב תצפית נמוכה ולהיפך, אזי גם פונקצית המתאם המשותף תיטה להתחלף: r_1 יהיה שלילי, r_2 יהיה חיובי (מכיוון שתצפיות בפער 2 יטו להיות מאותו כיוון של התוחלת), וכו' (ראה איורים 2.2, 2.7).

2.6.4 סדרה עתית לא-סטציונרית

אם בסדרה העתית יש מגמה (trend) מקדמי המתאם המשותף (r_k) לא ישאפו ל-0, למעט בפערי זמן גדולים מאוד. זאת מכיוון שהתצפיות הבאות אחרי תצפית הנמצאת מעל התוחלת נוטות להיות גם כן מעל התוחלת, וכנ"ל לגבי תצפית הנמצאת מתחת לתוחלת (ראה איור 2.3). לא ניתן להסיק רבות מהקורלוגרם במקרה זה, מכיוון שהשפעת המגמה מאפילה על כל התופעות הסטטיסטיות האחרות. למען האמת, פונקצית המתאם המשותף היא בעלת משמעות רק עבור תהליכים סטציונרים.

2.6.5 השפעה של גורמים עונתיים

אם הסדרה העתית כוללת השפעות עונתיות, הקורלוגרם יגלה זאת. לדוגמא, עבור תצפיות חודשיות בהן יש מחזוריות שנתית, r_6 יטה להיות 'גדול' ושלילי, ו-r_{12} יהיה 'גדול' וחיובי.

באיור $2.4(a)$ ניתן לראות את הקורלוגרם של טמפרטורות חודשיות ברסיפה, ברזיל. לא ניתן לפספס את המחזוריות השנתית. אבל מעט מאוד ניתן ללמוד מעבר לכך: המחזוריות השנתית מאפילה על כל תופעה סטטיסטית אחרת. אם נסיר את ההשפעה העונתית (כלומר, נחסר מכל תצפית את ממוצע התצפיות במרחק $\pm 12n$ ממנה) נקבל סדרה חדשה, ללא ההשפעה העונתית. הקורלוגרם של סדרה זו מראה (ראה איור $2.4(b)$) מראה לנו שיש מתאם בפערים 1, 2, ו-3. מכאן נסיק כי אחרי חודש קר מהממוצע השנתי נקבל לרוב עוד חודש קר מהממוצע השנתי.

2.6.6 Outliers

אם הסדרה הנצפית מכילה outliers, הדבר ישפיע על הקורלוגרם, ויש לתקן את הסדרה לפני שאנו מנתחים אותה. אם יש outlier יחיד, מקדמי הקורלוגרם יהיו קטנים הרבה יותר מהתוצאות שהיינו אמורים לקבל ללא ה-outlier. אם יש שני outliers, בזמנים t_1 ו-t_2, נקבל מתאם גבוה בפער $t_2 - t_1$, ומתאמים נמוכים בכל שאר הפערים.

2.7 מטריצת השונות המשותפת

אם מתבוננים ב-n הרכיבית הראשונים של התהליך, $Y_1, Y_2, \ldots, Y_n$, ניתן לחשב את מטריצת הקובריאנס

$$\Gamma_n = \begin{pmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdot & \cdot & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdot & \cdot & \gamma_{n-2} \\ \gamma_2 & \gamma_1 & \gamma_0 & \cdot & \cdot & \gamma_{n-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \gamma_1 \\ \gamma_{n-1} & \gamma_{n-2} & \gamma_{n-3} & \cdot & \gamma_1 & \gamma_0 \end{pmatrix}$$

האיבר (i,j) במטריצה זו הוא $\mathrm{Cov}(Y_i,Y_j)=\gamma_{i-j}$. מטריצת המתאם המשותף נתונה על ידי:

$$P_n=\left(\begin{array}{cccccc} 1 & \rho_1 & \rho_2 & \cdot & \cdot & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \cdot & \cdot & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \cdot & \cdot & \rho_{n-3} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \rho_1 \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdot & \rho_1 & 1 \end{array}\right)$$

מתוך ההגדרה מקבלים כי $\Gamma_n=\sigma^2 P_n$.

טענה 2.9 המטריצה Γ_n היא חיובית לחלוטין *(positive definite)*, לכל $n\geq 1$. כלומר, לכל וקטור שורה $v\in\mathbf{R}^n$ מתקיים

$$v\Gamma_n v^t\geq 0.$$

אם $\mathrm{Var}(Y_t)>0$, אזי האי-שוויון הוא חזק כאשר v אינו וקטור האפס.

נשים לב שמכיוון ש-$\Gamma_n=\sigma^2 P_n$, נובע כי המטריצה P_n היא גם חיובית לחלוטין.

הוכחה. נקבע וקטור $v\in\mathbf{R}^n$. אזי

$$\begin{aligned} v\Gamma_n v^t &= \sum_{i,j} v_i v_j \mathrm{Cov}(Y_i,Y_j) \\ &= \mathrm{Var}\left(\sum_{i=1}^n v_i Y_i\right)\geq 0. \end{aligned}$$

במשוואה זו יש שוויון אם ורק אם $\sum_{i=1}^n v_i Y_i$ הוא קבוע. אך אם $\mathrm{Var}(Y_t)>0$, דבר כזה קורה רק אם v הוא וקטור האפס. ■

תנאי שקול לכך שמטריצה היא חיובית לחלוטין הוא שהדטרמיננטה של כל תת-מטריצה סימטרית הוא חיובי. בפרט נקבל כי:

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}>0\Rightarrow -1<\rho_1<1.$$

$$\begin{vmatrix} 1 & \rho_2 \\ \rho_2 & 1 \end{vmatrix}>0\Rightarrow -1<\rho_2<1.$$

$$\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}>0\Rightarrow \rho_1^2<\frac{1+\rho_2}{2}.$$

כדי להסיק את האי-שוויון האחרון נשים לב כי התנאי שהדטרמיננטה חיובית שקול ל-

$$2\rho_1^2\rho_2-2\rho_1^2-\rho_2^2+1>0,$$

הגורר את הנדרש.

באופן דומה ניתן לקבל אי-שוויונות נוספים שעל פונקצית המתאם המשותף לקיים.

3 הכנות מתמטיות

בפרק זה נציג מספר שאלות מתמטיות העולות בעת ניתוח מודלים לינאריים, וניתן את הפתרון התאורטי שלהן.

3.1 הפיכות של פולינומים

משפט 3.1 יהא $\Phi(B) = 1 + \sum_{i=1}^{d} \phi_i B^i$ פולינום מדרגה d שכל שורשיו (במישור המרוכב) נמצאים מחוץ לעיגול היחידה. יהא $\{X_t\}$ תהליך רעש אקראי טהור, ויהא $\{Y_t\}$ תהליך המקיים $\Phi(B)Y_t = X_t$. אזי יש פולינום $\Theta(B) = \sum_{j=0}^{\infty} \theta_j B^j$ עם רדיוס התכנסות לפחות 1 המקיים:

- $\Phi(B)\Theta(B) = 1$.
- $\lim_{n\to\infty} \mathbf{E}\left[(Y_t - \sum_{j=1}^{n} \theta_j X_{t-j})^2\right] = 0$.

מכיוון ש-$Y_t = \Phi(B)X_t$ ניתן היה לצפות כי $X_t = \Phi(B)^{-1}Y_t$. מכיוון שבאופן כללי ההופכי Θ של Φ הוא טור אינסופי, לא ברור שהסכום $\sum_{j=1}^{\infty} \theta_j$ מתכנס, ולכן לא ברור שההופכי מוגדר (כלומר, שרדיוס ההתכנסות של הטור האינסופי הוא חיובי). משפט זה נותן תנאי מספיק לכך שההפכי יהיה מוגדר.

הוכחה. ראשית, נשים לב לשוויון הבא:

$$\begin{aligned} &\frac{\alpha_0 z^k + \alpha_1 z^{k+1} + \cdots + \alpha_p z^{k+p}}{1 + \beta_1 z + \cdots + \beta_p z^p} = \\ &= \alpha_0 z^k + \frac{(\alpha_1 - \alpha_0\beta_1)z^{k+1} + (\alpha_2 - \alpha_0\beta_2)z^{k+2} + \cdots + (\alpha_p - \alpha_0\beta_p)z^{k+p}}{1 + \beta_1 z + \cdots + \beta_p z^p}. \end{aligned} \tag{5}$$

מכיוון ש-

$$\Phi(B)Y_t = X_t,$$

הרי ש-

$$Y_t = \Phi(B)^{-1}X_t.$$

כעת נחשב את $\Phi(B)^{-1}$.

נזכור כי

$$\Phi(B) = 1 + \sum_{i=1}^{p} \phi_i B^i.$$

נגדיר באופן רקורסיבי את הסדרות $(\delta_k, \alpha_{k,1}, \ldots, \alpha_{k,p})$ באופן הבא

$$\begin{aligned} \alpha_{0,i} &= -\theta_i, \quad i = 1, \ldots, p, \\ \alpha_{k+1,i} &= \alpha_{k,i+1} - \alpha_{k,1}\phi_i, \quad i = 1, 2, \ldots, p-1, \\ \alpha_{k+1,p} &= -\alpha_{k,1}\phi_p, \\ \delta_{k+1} &= \alpha_{k1}. \end{aligned}$$

טענה 3.2 לכל n מתקיים

$$\frac{1}{1+\phi_1 z+\cdots+\phi_p z^p}=1+\sum_{k=1}^{n}\delta_k z^k+\frac{\alpha_{n,1}z^{n+1}+\cdots+\alpha_{n,p}z^{n+p}}{1+\phi_1 z+\cdots+\phi_p z^p}.$$

הוכחה. נוכיח באינדוקציה על n.

עבור $n=0$ נקבל ממשוואה (5)

$$\begin{aligned}\frac{1}{1+\phi_1 z+\cdots+\phi_p z^p} &= 1-\frac{\phi_1 z+\cdots+\phi_p z^p}{1+\phi_1 z+\cdots+\phi_p z^p}\\ &= 1+\frac{\alpha_{0,1}z^{n+1}+\cdots+\alpha_{0,p}z^{n+p}}{1+\phi_1 z+\cdots+\phi_p z^p}.\end{aligned}$$

לכל $n\geq 0$ מתקיים, לפי הנחת האינדוקציה ומשוואה (5),

$$\begin{aligned}\frac{1}{1+\phi_1 z+\cdots+\phi_p z^p} &= 1+\sum_{k=1}^{n}\delta_k z^k+\frac{\alpha_{n,1}z^{n+1}+\cdots+\alpha_{n,p}z^{n+p}}{1+\phi_1 z+\cdots+\phi_p z^p}\\ &= 1+\sum_{k=1}^{n}\delta_k z^k+\alpha_{n,1}z^{n+1}\\ &\quad+\frac{(\alpha_{n,2}-\alpha_{n,1}\phi_1)z^{n+2}+\cdots+(\alpha_{n,p}-\alpha_{n,1}\phi_{p-1})z^{n+p}-\alpha_{n,1}z^{n+p+1}}{1+\phi_1 z+\cdots+\phi_p z^p}\\ &= 1+\sum_{k=1}^{n}\delta_k z^k+\delta_{n+1}z^{n+1}+\frac{\alpha_{n+1,1}z^{n+2}+\cdots+\alpha_{n+1,p}z^{n+p+1}}{1+\phi_1 z+\cdots+\phi_p z^p},\end{aligned}$$

כנדרש. ■

הנחנו שכל השורשים של הפולינום Φ נמצאים מחוץ לעיגול היחידה. לכן, לכל מספר מרוכב z המקיים $|z|\leq 1$ מתקיים

$$\Phi(z)=\sum_{i=1}^{p}\phi_i z^i\neq 0.$$

לכן $\frac{1}{\Phi(z)}$ מוגדר. נניח כי $\phi_p\neq 0$, ויהיו $z_1,\ldots,z_p$ השורשים של הפולינום Φ. אזי

$$\frac{1}{\Phi(z)}=\frac{1}{\sum_{i=1}^{p}\phi_i z^i}=\frac{C}{\prod_{i=1}^{p}\left(1-\frac{z}{z_i}\right)}=C'\prod_{i=1}^{p}\sum_{j=1}^{\infty}\left(\frac{z}{z_i}\right)^j=\sum_{j=1}^{\infty}\delta_j z^j.$$

מכיוון שאגף ימין סופי כאשר $|z|\leq 1$, גם אגף ימין סופי במקרה זה. זה נכון גם כאשר $z=1$, ולכן $\sum_{j\geq n}\delta_j$ חסום לכל $n\geq 0$. בפרט נקבל כי הגודל $\frac{\alpha_{n+1,1}z^{n+2}+\cdots+\alpha_{n+1,p}z^{n+p+1}}{1+\phi_1 z+\cdots+\phi_p z^p}$ שואף ל-0 לכל z. אך זה גורר כי לכל $i=1,\ldots,p$ מתקיים $\lim_{n\to\infty}\alpha_{n,i}=0$, ולכן

$$\begin{aligned}&\lim_{n\to\infty}\mathbf{E}\left[(Y_t-\sum_{j=1}^{n}\psi_j X_{t-j})^2\right]\\ &=\lim_{n\to\infty}\mathbf{E}\left[(\alpha_{n,1}Y_{t-n-1}+\cdots+\alpha_{n,p}Y_{t-n-p})^2\right]=0.\end{aligned}$$

■

3.2 פתרון משוואות הפרשים הומוגניות

יהא $\Phi(z)=\sum_{i=0}^{p}\phi_i z^i$ פולינום ממעלה p, המקיים $\phi_0=1$. בפרט, 0 אינו שורש של הפולינום.
נתבונן במערכת המשוואות הבאה בנעלמים $(y_t)_{t=1}^{\infty}$:

$$\sum_{i=1}^{p}\phi_i y_{t-i}=0, \qquad \forall t\geq p. \tag{6}$$

מערכת משוואות כזו נקראת מערכת משוואות הפרשים הומוגנית.

בסעיף זה נראה מהם הפתרונות של מערכת כזו.

יהיו $\frac{1}{G_1},\ldots,\frac{1}{G_p}$ השורשים של הפולינום ϕ. כלומר,

$$\Phi(z)=\prod_{i=1}^{p}(1-G_i z),$$

ולכל $k=1,\ldots,p$ מתקיים

$$0=\Phi(\frac{1}{G_k})=\sum_{i=1}^{p}\phi_i G_k^{-i}.$$

הטענה הבאה נותנת לנו p **סדרות** (y_t) הפותרות את (6).

טענה 3.3 הסדרה (y_t) המוגדרת על ידי

$$y_t=G_k^{t-1}$$

היא פתרון של המערכת (6).

הוכחה. לכל $t\geq p$,

$$\begin{aligned}\sum_{i=1}^{p}\phi_i y_{t-i} &= \sum_{i=1}^{p}\phi_i G_k^{t-i-1}\\ &= G_k^{t-1}\sum_{i=1}^{p}\phi_i G_k^{-i}=0,\end{aligned}$$

כנדרש. ■

מכיוון שהמערכת (6) היא לינארית ב-(y_t), מרחב הפתרונות הוא מרחב לינארי. הוכחת הטענה הבאה מושארת לקורא.

טענה 3.4 יהיו (y_t) ו-(z_t) שני פתרונות של המערכת (6), ויהיו $\alpha,\beta\in\mathbf{R}$ שני מספרים ממשיים. אזי הסדרה (w_t) המוגדרת על ידי

$$w_t=\alpha y_t+\beta z_t$$

היא פתרון של המערכת (6).

כאשר השורשים $\frac{1}{G_1},\ldots,\frac{1}{G_p}$ שונים זה מזה, אלו הם כל הפתרונות של המערכת.

טענה 3.5 אם $\frac{1}{G_1},\dots,\frac{1}{G_p}$ שונים זה מזה, אזי כל פתרון (y_t) של המערכת (6) הוא מהצורה
$$y_t=\sum_{k=1}^{p}A_kG_k^{t-1},$$
באשר $A_1,\dots,A_p$ הם קבועים ממשיים.

הוכחה. יהא (y_t) פתרון של המערכת (6). נתבונן במטריצה בת $p+1$ שורות ו-∞ עמודות, בה בשורה k (עבור $k=1,\dots,p$) מופיע הסדרה $(G_k^0,G_k^1,G_k^2,G_k^3,\dots)$, ובשורה התחתונה מופיעה הסדרה (y_t).

דרגת העמודות של המטריצה היא p: העמודה ה-t ניתנת להצגה כצרוף לינארי של p העמודות הקודמות, מכיוון שכל שורה מקיימת את המשוואה (6). לכן דרגת השורות של המטריצה אף היא p.

מכיוון ש-$\frac{1}{G_1},\dots,\frac{1}{G_p}$, p השורות הראשונות הן בלתי תלויות. אכן, תת-המטריצה מסדר $p\times p$ המכילה את p העמודות הראשונות ב-p השורות הראשונות נראות
$$\begin{pmatrix} 1 & G_1 & G_1^2 & G_1^3 & \cdots & G_1^p\\ 1 & G_2 & G_2^2 & G_2^3 & \cdots & G_2^p\\ & \cdots & & & & \\ 1 & G_p & G_p^2 & G_p^3 & \cdots & G_p^p \end{pmatrix}$$
ומטריצה זו היא מטריצת ון-דר-מונדה, שהיא הפיכה.

לכן השורה התחתונה ניתנת להצגה כצרוף לינארי של p השורות הראשונות, ולכן יש לה הצגה כנדרש. ■

כעת נראה מה קורה כאשר לפולינום Φ יש שורש עם ריבוי.

לפני שנטפל במקרה זה, ניזכר בשתי זהויות קומבינטוריות.

3.2.1 שתי זהויות קומבינטוריות

נזכור כי $\binom{n}{i}=\frac{n!}{(n-i)!i!}$ הוא המקדם הבינומי המתאים. זהו מספר האפשרויות לבחור תת-קבוצה בת i איברים מתוך קבוצה של n איברים.

הזהות הבסיסית שמקדמי הבינום מקיימים היא
$$\binom{n}{i}=\binom{n-1}{i-1}+\binom{n-1}{i}$$
(אנו מפרשים כאן $\binom{n-1}{n}=0$). אכן, מספר האפשרויות לבחור ת-קבוצה בת i איברים מתוך קבוצה של n איברים שווה למספר האפשרויות לבחור קבוצה המכילה את האיבר הראשון (נותר לבחור $i-1$ איברים מתוך שאר $n-1$ האיברים) ועוד מספר האפשרויות לבחור קבוצה שאינה מכילה אותו (נותר לבחור i איברים מתוך שאר $n-1$ האיברים).

נשים לב כי
$$\binom{n}{0}=\binom{n}{n}=1.$$

הסכום המתחלף של המקדמים הבינומים הוא 0:
$$\sum_{i=0}^{n}(-1)^n\binom{n}{i}=0. \tag{7}$$

אכן האיבר $\binom{n}{i}$ הוא של $\binom{n-1}{i-1}$ ו-$\binom{n-1}{i}$. לכן כל איבר מהצורה $\binom{n-1}{i}$ תורם לשני גורמים בטור: ל-$\binom{n}{i-1}$ ול-$\binom{n}{i}$, אך התרומה לאחד היא ב-$+1$ ולשני ב--1 (כי מכפילים את $\binom{n}{i}$ ב-$(-1)^i$). לכן הסכום מתאפס. באופן פורמלי:

$$\sum_{i=0}^{n}(-1)^n(-1)^n\binom{n}{i} = \sum_{i=0}^{n}(-1)^n(-1)^n(\binom{n-1}{i-1}+\binom{n-1}{i}) = \sum_{i=0}^{n-1}\binom{n}{i}(1-1) = 0.$$

הטור הבא אף הוא מתאפס:

$$\sum_{i=0}^{n}(-1)^n i\binom{n}{i} = 0. \tag{8}$$

אכן,

$$\begin{aligned}
\sum_{i=0}^{n}(-1)^n i\binom{n}{i} &= \sum_{i=0}^{n}(-1)^n i(\binom{n-1}{i-1}+\binom{n-1}{i}) \\
&= \sum_{i=1}^{n}(-1)^n i\binom{n-1}{i-1}+\sum_{i=0}^{n-1}(-1)^n i\binom{n-1}{i} \\
&= -\sum_{i=0}^{n-1}(-1)^{n-1}(i+1)\binom{n-1}{i-1}+\sum_{i=0}^{n-1}(-1)^n i\binom{n-1}{i} \\
&= -\sum_{i=0}^{n-1}(-1)^{n-1} i\binom{n-1}{i-1}-\sum_{i=0}^{n-1}(-1)^{n-1}\binom{n-1}{i-1}+\sum_{i=0}^{n-1}(-1)^n i\binom{n-1}{i}.
\end{aligned}$$

האיבר הראשון והשלישי מתאפסים מהנחת האינדוקציה, והאיבר השני מתאפס ממשוואה (7). לסיום ההוכחה יש לבדוק את תנאי ההתחלה של האינדוקציה. דבר זה מושאר לקורא.

כעת ננסה להכליל את המשוואה (8).

טענה 3.6 לכל $j > 1$, לכל $0 \le k \le j$, ולכל $n > j$ מתקיים

$$\sum_{i=k}^{n}(-1)^i\binom{j+i-k}{j}\binom{n}{i} = 0.$$

הוכחה. ההוכחה היא באינדוקציה כפולה: על j ועל n.

נבדוק כי הטענה נכונה כאשר $j = 0$. במקרה זה, $k = 0$. ממשוואה (7) נקבל

$$\sum_{i=0}^{n}(-1)^i\binom{n}{i} = 0.$$

נניח באינדוקציה כי הטענה נכונה עבור $j-1$, ונוכיחה עבור j. תחילה נבדוק מה קורה כאשר $k = j = n-1$. במקרה זה

$$\begin{aligned}
\sum_{i=j}^{j+1}(-1)^i\binom{j+i-k}{j}\binom{j+1}{i} &= (-1)^j\binom{j}{j}\binom{j+1}{j}+(-1)^{j+1}\binom{j+1}{j}\binom{j+1}{j+1} \\
&= (-1)^j(\binom{j+1}{j}-\binom{j+1}{j}) = 0.
\end{aligned}$$

כעת נניח כי $k \leq j$ ו-$n \geq j+1$. במקרה זה

$$\begin{aligned}
\sum_{i=k}^{n}(-1)^i\binom{j+i-k}{j}\binom{n}{i} &= \sum_{i=k}^{n}(-1)^i\binom{j+i-k}{j}(\binom{n-1}{i}+\binom{n-1}{i-1}) \\
&= \sum_{i=k}^{n-1}(-1)^i\binom{j+i-k}{j}\binom{n-1}{i} \\
&\quad ++\sum_{i=k}^{n}(-1)^i\binom{j+i-k}{j}\binom{n-1}{i-1} \\
&= \sum_{i=k}^{n-1}(-1)^i\binom{j+i-k}{j}\binom{n-1}{i} \\
&\quad +\sum_{i=k-1}^{n-1}(-1)^i\binom{j+i+(k-1)}{j}\binom{n-1}{i}.
\end{aligned}$$

הסכום הראשון מתאפס מהנחת האינדוקציה לגבי k ו-$n-1$. הסכום השני מתאפס מהנחת האינדוקציה על $k-1$ ו-$n-1$. ■

3.2.2 פולינום עם שורש מרובה

נניח כי $\frac{1}{G_0}$ הוא שורש מריבוי d של הפולינום Φ, כלומר,

$$\Phi(z) = (1-G_0z)^d \times \prod_{i=1}^{p-d}(1-G_iz).$$

לפני שנפתח את הפתרון הכללי, נתבונן במקרה בו

$$\Phi(z) = (1-G_0z)^2 = 1-2G_0z+G_0^2z^2.$$

כמו קודם, הסדרה (y_t) המוגדרת על ידי

$$y_t = G_0^{t-1}$$

היא פתרון. אבל גם הסדרה (z_t) המוגדרת על ידי

$$\begin{aligned}
y_1 &= 0 \\
y_t &= (t-1)G_0^{t-2}
\end{aligned}$$

היא פתרון. אכן $y_3 = 2G_0$ ו-$y_2 = 1$, ולכן

$$y_3 - 2G_0y_2 + y_1 = 2G_0 - 2G_0 = 0,$$

ולכל $t > 3$

$$y_t - 2G_0y_{t-1} + y_{t-2} = (t-1)G_0^{t-2} - 2G_0(t-2)G_0^{t-3} + G_0^2(t-3)G_0^{t-4} = 0.$$

טענה 3.7 נניח כי Φ הוא פולינום ממעלה p, שיש לו שורש $\frac{1}{G_0}$ מריבוי d. אזי הסדרה (y_t) המוגדרת על ידי:

$$\begin{aligned} y_t &= 0, \qquad t = 1, \ldots, j \\ y_t &= \binom{t-1}{j} G_0^{t-j}, \qquad t > j. \end{aligned}$$

היא פתרון של מערכת המשוואות (6).

הוכחה. תחילה נניח כי $d = p$, כלומר, $\Phi(z) = (1 - G_0 z)^d$. במקרה זה ניתן להציג במפורש את המקדמים של הפולינום Φ:

$$\phi_i = (-1)^i \binom{n}{i} G_0^i.$$

כדי להראות שהסדרה (y_t) המוגדרת הטענה היא פתרון של מערכת המשוואות (6), יש להראות כי

$$\sum_{i=0}^{d} \phi_i y_{t-i} = 0, \qquad \forall t \geq j.$$

נחלק לשני מקרים. אם $t > 2j$, הצבת הצורה המפורשת של ϕ_i ושל y_t מראה כי יש להראות ש-

$$\sum_{i=0}^{d} (-1)^i \binom{n}{i} G_0^i \binom{t-i}{j} G_0^{t-i-j} = 0.$$

הגורם $G_0^i \times G_0^{t-i-j} = G_0^{t-j}$ תורם קבוע, שניתן לצמצם, ולכן יש להוכיח כי

$$\sum_{i=0}^{d} (-1)^i \binom{n}{i} \binom{t-i}{j} = 0.$$

שוויון זה מתקיים מטענה 3.6.

כעת לא נניח כי $d = p$. מכיוון ש-$\frac{1}{G_0}$ הוא שורש מריבוי d, מתקיים

$$\Phi(z) = \Theta \circ \Psi(z),$$

באשר Θ הוא פולינום ממעלה $p - d$, ו-$\Psi(z) = (1 - G_0 z)^d$. לכן, הפעלת הפולינום Φ על סדרה (y_t) מסויימת משמעה תחילה הפעלת הפולינום Ψ על הסדרה, ולאחר מכן הפעלת הפולינום Θ על הסדרה המתקבלת. אבל מהחלק הראשון, אחרי הפעלת הפולינום Ψ על הסדרה (y_t) המוגדרת בטענה מתקבלת סדרת אפסים, ולכן גם אחרי הפעלת הפולינום Θ נישאר עם סדרת אפסים. ■

כעת ניתן לאפיין את כל הפתרונות של מערכת המשוואות (6) כאשר יש שורש מרובה.

טענה 3.8 הסדרה (y_t) היא פתרון של מערכת המשוואות (6) אם ורק אם קיימים קבועים $A_0, \ldots, A_{d-1}$ וקבועים $C_1, \ldots, C_{p-d}$ כך ש-

$$y_t = \left(A_0 + A_1 \frac{t-1}{1} + \cdots + A_{d-1} \frac{t-1}{d-1} \right) G_0^{t-1} + \sum_{i=1}^{p-d} C_i G_i^{t-1}.$$

הוכחה. העובדה שכל סדרה כזו היא פתרון נובעת מטענה 3.7. העובדה שכל הפתרונות הם כאלו נובעת באופן דומה להוכחת טענה 3.5. ■

על ידי שימוש במשוואה (8) ניתן להסיק כי באופן כללי הסדרה (y_t) המוגדרת על ידי

$$\begin{aligned} y_1 &= 0 \\ y_t &= (t-1)\alpha^{t-2} \end{aligned}$$

היא פתרון עבור הפולינום $\Phi(z) = (1 - \frac{1}{G}z)^d$.

על ידי שימוש בטענה 3.4 נקבל כי כל צרוף לינארי של פתרונות הוא פתרון, וכמו בטענה 3.5 ניתן לקבל כי מימד מרחב הפתרונות הוא p.

ניתן להראות בדרך דומה כי הפתרון הכללי למשוואה (6), כאשר לפולינום Φ יש שורש אחד עם ריבוי d, וכל שאר השורשים עם ריבוי 1, הוא

$$y_t = \left(A_0 + A_1(t-1) + A_2(t-1)^2 + \cdots + A_{d-1}(t-1)^{d-1}\right) G_0^{t-1} + \sum_{k=1}^{p-d} C_i G_k^{t-1}.$$

באופן כללי, נקבל כי כל שורש $\frac{1}{G}$ של Φ מריבוי d מגדיר d פתרונות ספציפיים למערכת המשוואות ההומוגנית, ומרחב כל הפתרונות הוא המרחב הלינארי ה-p מימדי הנפרש על ידי p הפתרונות המוגדרים על ידי כל השורשים.

4 פילטר לינארי כללי

מודל פשוט ושימושי לניתוח סדרות עתיות הוא מודל הפילטר הלינארי. במודל זה אנו מניחים כי הנתונים מיוצרים באמצעות טרנספורמציה לינארית, או פילטר לינארי, מתהליך של רעש אקראי $\{X_t\}$.

נתחיל במתן דוגמא של תהליך לינארי פשוט, נמשיך עם המודל הכללי, ולאחר מכן נגדיר שתי משפחות חשובות של תהליכים לינאריים: תהליכי מיצוע-נע ותהליכים אוטו-רגרסיביים. אחר כך נחקור את המאפיינים של המודל הלינארי.

4.1 דוגמא: מודל לינארי $\mathrm{MA}(1)$

נניח כי $\{X_t\}$ היא סדרה של רעש אקראי טהור, המקיימת

$$\mathbf{E}[X_t] = 0, \quad \mathrm{Var}(X_t) = \sigma_X^2.$$

נניח כי התהליך $\{Y_t\}$ מחושב באמצעות הנוסחא:

$$Y_t = X_t + \theta X_{t-1}. \tag{9}$$

התהליך $\{Y_t\}$ תלוי בפרמטר θ, שיכול להיות כל מספר ממשי.

התהליך $\{Y_t\}$ הוא תהליך מיצוע-נע מסדר 1, כפי שהוגדר בסעיף 2.3.3. ראינו בסעיף 2.3.3 כי פונקצית המתאם המשותף של תהליך $\mathrm{MA}(1)$ היא

$$\rho_k = \rho_k(\theta) \begin{cases} 1 & k = 0, \\ \frac{\theta}{1+\theta^2} & k = \pm 1, \\ 0 & \text{אחרת}. \end{cases}$$

נשים לב כי מכיוון ש-

$$\frac{1/\theta}{1 + (1/\theta)^2} = \frac{\theta}{1+\theta^2},$$

נובע כי $\rho_k(\theta) = \rho_k(1/\theta)$.

כלומר, פונקצית המתאם המשותף עבור הפרמטר θ זהה לפונקצית המתאם המשותף עבור הפרמטר $1/\theta$. בפרט, המומנטים מסדר שני אינם מאפיינים באופן מלא תהליכים סטוכסטים מסוג $\mathrm{MA}(1)$.

כאן בנינו את סדרת הפלט $\{Y_t\}$ מתוך סדרת הקלט $\{X_t\}$. לעיתים נרצה לשחזר את סדרת הקלט $\{X_t\}$ מתוך סדרת הפלט $\{Y_t\}$. סיבה אחת לעשות זאת היא כדי לוודא שאנו יודעים מהו המודל שייצר את סדרת הפלט: אם המודל נכון, הסדרה המשוחזרת צריכה להיות סדרה של רעש אקראי.

קל לראות כי:

$$\begin{aligned} X_t &= Y_t - \theta X_{t-1} \\ &= Y_t - \theta(Y_{t-1} - \theta X_{t-2}) \\ &= Y_t - \theta Y_{t-1} + \theta^2 Y_{t-2} - \theta^3 Y_{t-2} \cdots . \end{aligned}$$

הסדרה המתאימה להצגה זו מתכנסת אם $|\theta| < 1$, ומתבדרת אחרת. בפרט, אם $\theta \geq 1$ לא ניתן לשחזר את הסדרה $\{X_t\}$ מתוך הסדרה $\{Y_t\}$.

הגדרה 4.1 תהליך $\mathrm{MA}(1)$ המוגדר במשוואה (9) הוא **הפיך** *(invertible)* אם $|\theta| < 1$, ואינו **הפיך** אחרת.

4.2 המודל הלינארי הכללי

לפני שנגדיר את המודל, נגדיר את אופרטור **ההזזה לאחור** (backward shift). אופרטור ההזזה לאחור B מקבל סדרה אינסופית, ומחזיר את הסדרה מוזזת באיבר אחד ימינה:

$$B(\ldots, y_{-1}, y_0, y_1, y_2, y_3, \ldots) = (\ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \ldots).$$

באופן שקול,

$$B(X_t) = X_{t-1}.$$

אם נפעיל את האופרטור B על הסדרה j פעמים, נקבל את הסדרה מוזזת j איברים ימינה:

$$B^j(X_t) = B(B^{j-1}(X_t)) = B(\underbrace{B \cdots B}_{j-1 \text{ פעמים}}(X_t)) = X_{t-j}.$$

כעת נחזור להגדרת הפילטר הלינארי.

הגדרה 4.2 נניח כי $\{X_t\}$ הוא תהליך של רעש אקראי טהור, המקיים

$$\mathbf{E}[X_t] = 0, \qquad \mathrm{Var}(X_t) = \sigma_X^2.$$

התהליך $\{Y_t\}$ הוא **תהליך לינארי** אם קיימים קבועים $\theta_1, \theta_2, \ldots$ כך שמתקיים לכל t:

$$\begin{aligned} Y_t &= X_t + \theta_1 X_{t-1} + \theta_2 X_{t-2} + \cdots \\ &= X_t + \sum_{j=1}^{\infty} \theta_j X_{t-j}. \end{aligned}$$

הערה 4.3 מכיוון שהפרמטר j אינו חסום, הסכום הנ"ל כולל גם X_k עם אינדקס שלילי. האינטרפרטציה היא שסדרת הרעש $\{X_t\}$ מוגדרת לכל $-\infty < t < +\infty$, ואנו רואים את הערכים של הסדרה $\{Y_t\}$ רק עבור $t \geq 0$.

אם נשתמש באופרטור ההזזה B נקבל כי התהליך הלינארי מקיים:

$$Y_t = \left(1 + \sum_{j=1}^{\infty} \theta_j B^j\right) X_t.$$

כלומר, הפילטר הלינארי מוגדר באמצעות הטרנספורמציה הלינארית

$$\Theta(B) = 1 + \sum_{j=1}^{\infty} \theta_j B^j.$$

4.3 תהליך מיצוע-נע מסדר q

הגדרה 4.4 תהליך לינארי $\{Y_t\}$ הוא תהליך מיצוע נע מסדר q אם $\theta_{q+1} = \theta_{q+2} = \cdots = 0$. כלומר, אם מתקיים

$$Y_t = X_t + \theta_1 X_{t-1} + \cdots + \theta_q X_{t-q}.$$

אם נשתמש באופרטור

$$\Theta(B) = 1 + \sum_{j=1}^{q} \theta_j B^j,$$

נקבל כי

$$Y_t = \Theta(B) X_t.$$

תהליך המיצוע-נע שהגדרנו כאן תוחלתו 0. לעיתים נתקלים בתהליכי מיצוע-נע שתוחלתם אינה אפס. במקרה כזה ניתן להגדיר תהליך מיצוע-נע מוכלל כתהליך שמקיים:

$$Y_t = \mu + \Theta(B) X_t$$

כך אנו מקבלים תהליך מיצוע-נע מסדר q עם תוחלת $\mathbf{E}[Y_t] = \mu$ שאינה 0.

בעמוד 16 ראינו כי פונקצית המתאם המשותף של תהליך מיצוע-נע נתונה על ידי:

$$\rho_k = \begin{cases} 0 & k > q, \\ \frac{\sum_{j=0}^{q-k} \theta_j \theta_{j+k}}{\sum_{j=0}^{q} \theta_j^2} & k = 0, 1, \ldots, q, \\ \rho(-k) & k < 0. \end{cases}$$

בפרט, פונקציה זו מתאפסת עבור פער $k > q$.

4.4 התאמת מודל מיצוע-נע

נניח שנתונה לנו סדרה שיוצרה על ידי תהליך מיצוע-נע, אך איננו יודעים מהו סדר התהליך q ומהם המקדמים $\theta_1, \ldots, \theta_q$.

את q קל לזהות לפי פונקצית המתאם המשותף. מכיוון שפונקציה זו מתאפסת עבור פער $k > q$, נחשב אומדים לערכי המתאם המשותף, וננסה למצוא q כך שבפערים גדולים מ-q האומדים קרובים ל-0.

נראה תחילה כיצד לאמוד את המקדמים כאשר התהליך הוא מסדר 1, ותוחלת התהליך μ אינה בהכרח 0, כלומר, כאשר התהליך מקיים את המשוואה

$$Y_t = \mu + X_t + \theta X_{t-1}.$$

בהינתן הצעות ל-μ, θ ו-x_0 ניתן לחשב רקורסיבית

$$x_t = y_t - \mu - \theta x_{t-1}. \tag{10}$$

נסמן ב-

$$S = S(\mu, \theta, x_0) = \frac{1}{N}\sum_{t=1}^{N}(x_t)^2.$$

אם הסדרה אכן יוצרה על ידי תהליך מיצוע-נע מסדר 1, אזי עבור הערכים הנכונים של θ, μ ו-x_0 (ורק עבורם) נקבל כי S הינו בקירוב σ_X^2.

נתבונן כעת ב-μ', θ', x'_0 כלשהם. נגדיר רקורסיבית $x'_t = y_t - \mu' - \theta' x'_{t-1}$. אזי

$$S(\mu', \theta', x'_0) = \frac{1}{N}\sum_{t=1}^{N}(x'_t)^2,$$

וכן

$$x'_t - x_t = (\mu - \mu') + (\theta x_{t-1} - \theta' x'_{t-1}) = (\mu - \mu') + (\theta - \theta')x_{t-1} + \theta'(x_{t-1} - x'_{t-1}).$$

נפעיל נוסחא זו אינדוקטיבית ונקבל

$$x'_t - x_t = (\mu - \mu')\sum_{j=1}^{t}\theta'^j + (\theta - \theta')\sum_{j=1}^{t}\theta^j x_{t-j}.$$

לכן

$$x'_t = x_t + (\mu - \mu')\sum_{j=1}^{t-1}\theta'^{j-1} + (\theta - \theta')\sum_{j=1}^{t-1}\theta^j x_{t-j} = x_t + A + \sum_{j=1}^{t-1}C_j x_{t-j}.$$

בפרט

$$\begin{aligned}
S(\mu', \theta', x'_0) &= \frac{1}{N}\sum_{t=1}^{N}(x'_t)^2 \\
&= \frac{1}{N}\sum_{t=1}^{N}(x_t + A + \sum_{j=1}^{t-1}C_j x_{t-j})^2 \\
&= \frac{1}{N}\sum_{t=1}^{N}(x_t)^2 + A^2 + \frac{1}{N}\sum_{t=1}^{N}(\sum_{j=1}^{t-1}C_j x_{t-j})^2 \\
&\quad + \frac{2A}{N}\sum_{t=1}^{N}x_t + \frac{2A}{N}\sum_{t=1}^{N}\sum_{j=1}^{t-1}C_j x_{t-j} + \frac{2}{N}\sum_{t=1}^{N}\sum_{j=1}^{t-1}C_j x_t x_{t-j}.
\end{aligned}$$

האיבר הראשון באגף ימין הוא $S(\mu, \theta, x_0)$, האיבר השני הוא אי-שלילי, האיבר השלישי הוא אי-שלילי, האיבר הרביעי הוא קירוב ל-$2A\mathbf{E}[X_t] = 0$, האיבר החמישי הוא קירוב ל-$C\mathbf{E}[X_t] = 0$, באשר C הוא סכום מתאים של ה-C_j, והאיבר השישי הוא צרוף לינארי של קירובים של $\mathrm{Cov}(X_t, X_t - j) = 0$. לכן נצפה לקבל כי

$$S(\mu, \theta, x_0) \geq S(\mu', \theta', x'_0).$$

לצערנו הנוסחא המקשרת את μ ו-θ לסכום ריבועי ההפרשים אינה ריבועית, אלא ממעלה גבוהה הרבה יותר. לכן לא ניתן לגזור אותה, להשוותה לאפס ולמצוא את ערכי μ, θ ו-x_0 עבורם סכום ריבועי ההפרשים מינימלי.

הדרך המקובלת למצוא מקדמים אלה היא הדרך הבאה.

- נניח כי $x_0 = 0$.
- נחשב את סכום ריבועי ההפרשים עבור רשת צפופה (grid) של μ ו-θ, ונמצא את הערכים בהם סכום זה מינימלי.

ניתן להשתמש בשיטות נומריות סטנדרטיות למציאת המינימום. אם יש לאתחל את האלגוריתם בהצעה התחלתית, נשתמש ב-

$$\mu := \frac{1}{N}\sum_{t=1}^{N} y_t,$$

ו-θ הוא הפתרון של המשוואה

$$r_1 = \frac{\theta}{1+\theta^2}.$$

אם סדרת הרעש $\{X_t\}$ מתפלגת נורמלית, הערכים הממזערים את סכום ריבועי ההפרשים ממקסמים את יחס הנראות, בהינתן ש-$x_0 = 0$.

ההנחה ש-$x_0 = 0$ אינה משנה הרבה אם $|\theta| < 1$. במקרה זה תרומת x_0 ל-x_t, עבור t גדול, היא קטנה. אכן, אם נציב רקורסיבית את משוואה (10) בעצמה (ונניח לשם פשטות כי $\mu = 0$) נקבל

$$x_t = y_t - \theta x_{t-1} = y_t - \theta y_{t-1} + \theta^2 x_{t-2} = \cdots = \sum_{j=0}^{t-1}(-1)^j\theta^j y_{t-j} + (-1)^t\theta^t x_0,$$

ולכן תרומת x_0 שואפת ל-0 בקצב גאומטרי.

אם סדר התהליך q גדול מ-1, נשתמש בטכניקה דומה (אם כי היא תדרוש יותר זמן חישובים, מכיוון שצריך למנות על הערכים של $\theta_1, \ldots, \theta_q$).

4.5 תהליכים אוטו-רגרסיביים (auto-regressive processes)

הגדרה 4.5 יהא $\{X_t\}$ תהליך רעש אקראי טהור, המקיים

$$\mathbf{E}[X_t] = 0, \qquad \mathrm{Var}(X_t) = \sigma_X^2.$$

תהליך $\{Y_t\}$ הוא תהליך **תהליך אוטו–רגרסיבי מסדר** p אם קיימים קבועים $\phi_1, \ldots, \phi_p$ כך שמתקיים לכל t:

$$Y_t = X_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p}. \tag{11}$$

ניתן להציג תהליך כזה בעזרת אופרטור ההזזה לאחור:

$$(1 - \phi_1 B - \cdots - \phi_p B^p)Y_t = X_t,$$

ואם נציב $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ נקבל את ההצגה:

$$\Phi(B)Y_t = X_t.$$

אם ניקח במשוואה (11) תוחלת בשני האגפים, נניח כי $\{Y_t\}$ הוא תהליך סטציונרי, ונסמן $\mu = \mathbf{E}[Y_t]$, נקבל כי μ מקיים את משוואת הרקורסיה:

$$\begin{aligned}\mu = \mathbf{E}[Y_t] &= \sum_{i=1}^{p} \phi_i \mathbf{E}[Y_{t-i}] + \mathbf{E}[X_t] \\ &= \mu \times \sum_{i=1}^{p} \phi_i.\end{aligned}$$

ובפרט, נקבל כי $\mu = 0$ (אם $\sum_{i=1}^{p} \phi_i = 0$, כל μ יפתור את המשוואה).

לכן, הצגה זו מתאימה רק לסדרות שהן יציבות סביב 0. כדי להתאים את המודל לסדרות עם תוחלת כלשהי, נגדיר תהליך אוטו-רגרסיבי מסדר p כללי על ידי:

$$Y_t - \mu = X_t + \sum_{i=1}^{p} \phi_i (Y_{t-i} - \mu).$$

מכיוון שהתוחלת אינה משפיעה על השונות המשותפת, נניח כרגע כי $\mu = 0$.

כעת נחשב את השונות המשותפת.

$$\gamma_k = \mathrm{Cov}(Y_t, Y_{t-k}) = \mathbf{E}[Y_{t-k} Y_t].$$

מכיוון ש-

$$Y_{t-k} Y_t = Y_{t-k} X_t + \sum_{i=1}^{p} \phi_i Y_{t-k} Y_{t-i},$$

נקבל כי

$$\gamma_k = \begin{cases} \sum_{i=1}^{p} \phi_i \gamma_{k-i} & k > 0 \\ \sum_{i=1}^{p} \phi_i \gamma_i + \sigma_X^2 & k = 0. \end{cases} \tag{12}$$

נשתמש בזהות $\rho_i = \frac{\gamma_i}{\gamma_0}$ כדי לקבל עבור $k = 0$ כי

$$\begin{aligned}\mathrm{Var}(Y_t) = \gamma_0 &= \sum_{i=1}^{p} \phi_i \gamma_i + \sigma_X^2 \\ &= \gamma_0 \sum_{i=1}^{p} \phi_i \rho_i + \sigma_X^2.\end{aligned}$$

מכיוון ש-$\Phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$, נקבל כי

$$\sigma_X^2 = \gamma_0 (1 - \sum_{i=1}^{p} \phi_i \rho_i) = \gamma_0 \times \Phi(B) \rho_i.$$

לכן קיבלנו כי

$$\gamma_0 = \frac{\sigma_X^2}{\Phi(B)\rho_i} = \frac{\sigma_X^2}{1 - \phi_1 \rho_1 - \cdots - \phi_p \rho_p}. \tag{13}$$

עבור $k > 0$ נקבל כי $\gamma_k = \sum_{i=1}^{p} \phi_i \gamma_{k-i}$, השקול ל-$\Phi(B)\gamma_k = 0$. נחלק את שני האגפים ב-$\gamma_0$ ונקבל

$$\Phi(B)\rho_k = 0.$$

אנו מקבלים כי פונקצית המתאם המשותף במודל אוטו-רגרסיבי מסדר p מקיימת את משוואת ההפרשים

$$\rho_k - \phi_1\rho_{k-1} - \cdots - \phi_p\rho_{k-p} = 0. \tag{14}$$

כפי שראינו בטענה 3.5, אם השורשים של הפולינום האופייני Φ שונים זה מזה אזי הפתרון הכללי של משוואה זו הוא

$$\rho_k = A_1G_1^k + A_2G_2^k + \cdots + A_pG_p^k, \tag{15}$$

באשר $\frac{1}{G_1}, \frac{1}{G_2}, \ldots, \frac{1}{G_p}$ הם השורשים של

$$\Phi(B) = 1 - \phi_1B - \cdots - \phi_pB^p.$$

בטענה 3.8 מצאנו את הפתרון הכללי כאשר השורשים אינם שונים זה מזה.

נראה בסעיף הבא שכדי שהתהליך יהיה סטציונרי נצטרך לדרוש כי $|G_i| < 1$ לכל i. כדי למצוא את השורשים של הפולינום $\Phi(B)$ אפשר להציג פולינום זה כמכפלה של מונומים $(1 - G_iB)$ וגורמים ריבועיים הנותנים זוגות של פתרונות מרוכבים צמודים.

כל שורש G_i ממשי תורם לפונקצית המתאם המשותף (15) גורם אקספוננציאלי $A_iG_i^k$. כל זוג של שורשים מרוכבים G_i, G_j תורם גורם מהצורה $d^k\sin(2\pi fk + F)$, כאשר התדר f והפאזה F קשורים לשורשים G_i, G_j. גורם מהסוג השני נקרא גל סינוס דועך (המונח "דועך" מגיע מכיוון שהמקדם d^k גורם לשאיפה לאפס בקצב גיאומטרי).

לכן, באופן כללי, פונקצית המתאם המשותף היא צירוף של גורמים אקספוננציאליים וגלי סינוס דועכים.

ממשוואה (14) מקבלים את מערכת המשוואות

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 + \cdots + \phi_p\rho_{p-1} \\
\rho_2 &= \phi_1\rho_1 + \phi_2 + \cdots + \phi_p\rho_{p-2} \\
&\;\;\vdots \\
\rho_p &= \phi_1\rho_{p-1} + \phi_2\rho_{p-2} + \cdots + \phi_p.
\end{aligned}$$

או בכתיב מטריציוני

$$P_p\vec{\phi} = \vec{\rho}.$$

פתרון מערכת המשוואות הלינארית הזו הוא

$$\vec{\phi} = P_p^{-1}\vec{\rho}.$$

מערכת משוואות זו נקראת **משוואות יול-ווקר** (Yule-Walker equations).

אם אנו משערים שהסדרה הגיעה מתהליך אוטו-רגרסיבי מסדר p אנו יכולים לאמוד בקלות את הפרמטרים על ידי פתרון משוואות Yule-Walker:

$$\widehat{\vec{\phi}} = \widehat{P}_p^{-1}\widehat{\vec{\rho}}.$$

כאשר $\widehat{\vec{\rho}}$ הם האומדים לפונקצית המתאם המשותף.

4.5.1 הפיכות של תהליכי מיצוע-נע, וסטציונריות של תהליכים אוטו-רגרסיביים

בסעיף 3.1 הוכחנו את המשפט הבא:

משפט 4.6 יהא $\Phi(B) = 1 + \sum_{i=1}^{p} \phi_i B^i$ פולינום מדרגה d שכל שורשיו (במישור המרוכב) נמצאים מחוץ לעיגול היחידה. יהא $\{X_t\}$ תהליך רעש אקראי טהור, ויהא $\{Y_t\}$ תהליך המקיים $\Phi(B)Y_t = X_t$. אזי יש פולינום $\Theta(B) = \sum_{j=0}^{\infty} \theta_j B^j$ עם רדיוס התכנסות לפחות 1 המקיים:

- $\Phi(B)\Theta(B) = 1$.
- $\lim_{n\to\infty} \mathbf{E}\left[(Y_t - \sum_{j=1}^{n} \theta_j X_{t-j})^2\right] = 0$.

כעת נראה כי משפט זה מאפשר לנו לדעת מתי תהליך מיצוע-נע הוא הפיך, ומתי תהליך אוטו-רגרסיבי הוא סטציונרי.

הגדרה 4.7 תהליך מיצוע-נע $Y_t = \Theta(B)X_t$ נקרא **הפיך** *(invertible)* אם כל השורשים של הפולינום $\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ (במישור המרוכב) הם מחוץ לעיגול היחידה.

מסקנה 4.8 יהא $Y_t = \Theta(B)X_t$ תהליך מיצוע-נע הפיך. אזי ניתן לשחזר את הסדרה $\{X_t\}$ מתוך הסדרה $\{Y_t\}$ באמצעות הנוסחא $X_t = (\Theta(B))^{-1}Y_t$.

מסקנה 4.9 יהא $\Phi(B)Y_t = X_t$ תהליך אוטו-רגרסיבי המקיים שכל השורשים של הפולינום $\Phi(B)$ נמצאים מחוץ לעיגול היחידה. אזי $\{Y_t\}$ הוא תהליך סטציונרי.

הוכחה. ממשפט 3.1, הפולינום $\Theta(B) = (\Phi(B))^{-1}$ הוא בעל רדיוס התכנסות חיובי (רדיוס ההתכנסות הוא אף לפחות 1). כמו כן

$$Y_t = (\Phi(B))^{-1}X_t = \Theta(B)X_t.$$

מכיוון שלפולינום $\Phi(B)$ רדיוס התכנסות חיובי, $\sum_{j=1}^{\infty} \theta_j < +\infty$. מכיוון שהתהליך $\{X_t\}$ הוא תהליך רעש אקראי טהור, $\mathbf{E}[Y_t] = \sum_{j=0}^{\infty} \theta_j \mathbf{E}[X_t] = 0$, וכן התהליך $\{Y_t\}$ הוא סטציונרי. ■

4.6 דוגמאות

4.6.1 דוגמא: תהליך $\mathrm{AR}(1)$

תהליך $\mathrm{AR}(1)$, הנקרא גם תהליך מרקוב, נתון על ידי המשוואה

$$Y_t = \phi_1 Y_{t-1} + X_t, \qquad |\phi_1| < 1,$$

התהליך נקרא "תהליך מרקוב" מכיוון שהחלק היחידי מההיסטוריה של התהליך המשפיע על הערך הנוכחי Y_t הוא הערך האחרון Y_{t-1}.

אם נעביר אגפים נקבל:

$$Y_t - \phi_1 Y_{t-1} = X_t.$$

נשתמש באופרטור ההזזה לאחור כדי להציג את התהליך:

$$(1-\phi_1 B)Y_t = X_t.$$

על ידי שימוש בנוסחאות הכלליות (12) מעמוד 35 ו-(13) מעמוד 35 נקבל

$$\begin{aligned}\gamma_1 &= \phi_1\\ \gamma_2 &= \phi_1\gamma_1 = \phi_1^2\\ \gamma_3 &= \phi_1\gamma_2 = \phi_1^3\\ &\cdots\\ \gamma_k &= \phi_1^k, \quad k \geq 2.\end{aligned}$$

קיבלנו כי פונקצית המתאם המשותף דועכת אקספוננציאלית ל-0. אם $\phi_1 < 0$ הוא שלילי, היא מתנדנדת סביב ה-0.
נשים לב כי ממשוואה (13) מעמוד 35

$$\gamma_0 = \mathrm{Var}(Y_t) = \frac{\sigma_X^2}{1-\phi_1\rho_1} = \frac{\sigma_X^2}{1-\phi_1^2}.$$

מכיוון ש-$(1-\phi_1 B)Y_t = X_t$ נקבל כי

$$\begin{aligned}Y_t &= (1-\phi_1 B)^{-1}X_t\\ &= (1+\phi_1 B+\phi_1^2 B^2+\cdots)X_t\\ &= X_t+\phi_1 X_{t-1}+\phi_1^2 X_{t-2}+\cdots.\end{aligned}$$

אם $|\phi_1| < 1$ הטור הנ"ל מתכנס בהחלט, וניתן להביע את $\{Y_t\}$ כפונקציה של $\{X_t\}$.

4.6.2 דוגמא: תהליך $\mathrm{AR}(2)$

תהליך $\mathrm{AR}(2)$, נתון על ידי הנוסחא

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + X_t.$$

נשתמש באופרטור ההזזה לאחור ונקבל

$$(1-\phi_1 B-\phi_2 B^2)Y_2 = X_t,$$

או

$$Y_t = (1-\phi_1 B-\phi_2 B^2)^{-1}X_t.$$

נסמן ב-$\frac{1}{G_1}$ ו-$\frac{1}{G_2}$ את השורשים של $\Phi(B)$, כלומר

$$1-\phi_1 B-\phi_2 B^2 = (1-G_1 B)(1-G_2 B).$$

ממסקנה 4.9, כדי שהתהליך יהיה סטציונרי צריך שהשורשים יהיו מחוץ לעיגול היחידה. השורשים של $\Phi(B) = 0$ הם

$$\frac{-\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}.$$

אם הדיסקרימיננטה $\phi_1^2 + 4\phi_2 > 0$ נקבל שני שורשים ממשיים. אם הדיסקרימיננטה $\phi_1^2 + 4\phi_2 = 0$ נקבל שורש ממשי אחד (עם ריבוי 2). כדי שהתהליך יהיה סטציונרי נצטרך שהשורשים יהיו גדולים ממש מ-1 (בערך מוחלט). ניתן להראות שזה קורה כאשר

- $\phi_1 \geq 0$ ו-$\phi_1 + \phi_2 < 1$.
- $\phi_1 < 0$ ו-$\phi_2 - \phi_1 < 1$.

אם הדיסקרימיננטה $\phi_1^2 + 4\phi_2 < 0$ נקבל שני שורשים מרוכבים צמודים:

$$\frac{1}{G_1} = \frac{\phi_1}{2\phi_2} + \frac{\sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} = a + bi = re^{i\theta},$$
$$\frac{1}{G_2} = \frac{\phi_1}{2\phi_2} - \frac{\sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2} = a - bi = re^{-i\theta},$$

כאשר

$$r^2 = a^2 + b^2 = \frac{\phi_1^2}{4\phi_2^2} + \frac{-(\phi_1^2 + 4\phi_2)}{4\phi_2^2} = -\frac{1}{\phi_2}.$$

כדי ששני השורשים המרוכבים יהיו מחוץ לעיגול היחידה צריך ש-$1 < r^2 = -\frac{1}{\theta_2}$, השקול ל-$\phi_2 > -1$. אם נסכם את מה שקיבלנו, כאשר שני השורשים הם מרוכבים, כדי שהתהליך יהיה סטציונרי צריך ש-$\phi_1^2 + 4\phi_1 < 0$ ו-$\phi_2 > -1$, כלומר (ϕ_1, ϕ_2) צריכים להיות במשולש שקדקודיו $(0, 1), (2, -1), (-2, -1)$.

נשים לב כי כאשר שני השורשים מרוכבים

$$G_1 + G_2 = \frac{1}{r}\left(e^{i\theta} + e^{-i\theta}\right) = \frac{2\cos\theta}{r}.$$

אם השורשים שונים, כלומר $G_1 \neq G_2$, אזי

$$\begin{aligned}
\frac{1}{1 - \phi_1 B - \phi_2 B^2} &= \frac{1}{1 - G_1 B} \times \frac{1}{1 - G_2 B} && (16)\\
&= \frac{G_1}{G_1 - G_2} \times \frac{1}{1 - G_1 B} - \frac{G_2}{G_1 - G_2} \times \frac{1}{1 - G_2 B}\\
&= \frac{G_1}{G_1 - G_2}\left(1 + G_1 B + G_1^2 B^2 + \cdots\right) && (17)\\
&\quad - \frac{G_2}{G_1 - G_2}\left(1 + G_2 B + G_2^2 B^2 + \cdots\right). && (18)
\end{aligned}$$

המקדם של B^k בהצגה (18) הוא

$$\begin{aligned}
\frac{1}{G_1 + G_2}\left(G_1^{k+1} + G_2^{k+1}\right) &= \frac{r}{2\cos\theta}\left(\frac{1}{r^{k+1}} e^{-i(k+1)\theta} + \frac{1}{r^{k+1}} e^{i(k+1)\theta}\right)\\
&= \frac{1}{2r^k \cos\theta} \times 2\cos(k+1)\theta\\
&= \frac{\cos(k+1)\theta}{r^k \cos\theta}.
\end{aligned}$$

יתר על כן,

$$\theta = \arctan(\frac{b}{a}) = \arctan \frac{-(\phi_1^2 + 4\theta_2)}{|\theta_1|} = \arccos \frac{|\phi_1|}{2\sqrt{-\theta_2}}.$$

מכאן ניתן להסיק כי במקרה זה תהליך $\mathrm{AR}(2)$ יראה התנהגות פסבדו-מחזורית עם מחזור $2\pi/\theta$ (תדירות של $\theta/2\pi$).

משוואות Yule-Walker עבור מודל זה הינן

$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2\rho_1 \\
\rho_2 &= \phi_1\rho_1 + \phi_2,
\end{aligned}$$

שפתרונן הוא

$$\begin{aligned}
\phi_1 &= \frac{\rho_1(1-\rho_2)}{1-\rho_1^2} \\
\phi_2 &= \frac{\rho_2-\rho_1^2}{1-\rho_1^2}.
\end{aligned}$$

בפרט נקבל כי

$$\rho_1 = \frac{\phi_1}{1-\phi_2},$$

ולכן הסימן של ρ_1 זהה לסימן של ϕ_1, ו:

$$\rho_2 = \phi_2 + \frac{\phi_1^2}{1-\phi_2}.$$

פונקצית המתאם המשותף מקיימת

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}.$$

מכיוון ש-$\rho_0 = 1$ ו-$\rho_1 = \frac{\phi_1}{1-\phi_2}$, ניתן להביע את ρ_k כפתרון של משוואת הפרשים (ראה טענות 3.5 ו-3.8). הפתרון הכללי הוא מהצורה

$$\rho_k = A_1G_1^k + A_2G_2^k.$$

מכיוון ש-$\rho_0 = 1$ נקבל $A_1 + A_2 = 1$. מכיוון ש-$\rho_1 = \frac{\phi_1}{1-\phi_2}$ נקבל כי $A_1G_1 + A_2G_2 = \frac{\phi_1}{1-\phi_2}$. אחרי משחקים אלגבריים קלים נקבל כי:

$$\rho_k = \frac{G_1(1-G_2^2)G_1^k - G_2(1-G_1^2)G_2^k}{(G_1-G_2)(1+G_1G_2)}.$$

מכאן נקבל כי הגרף של פונקצית המתאם המשותף של תהליך $\mathrm{AR}(2)$ תלויה בפרמטרים ϕ_1, ϕ_2. ראה שרטוט מצורף.

עבור המקרה של שורשים מרוכבים,

$$\rho_k = \frac{(\mathrm{sgn}(\phi_1))^k d^k \sin(2\pi f_0 k + F)}{\sin F}, \qquad (19)$$

באשר $d = \sqrt{-\phi_2}$, $\cos(2\pi f_0) = \frac{|\phi_1|}{2\sqrt{-\phi_2}}$, ו-$\tan(F) = \frac{1+d^2}{1-d^2} \times \tan(2\pi f_0)$. במקרה זה לספקטרום יהיה שיא בתדירות f_0. הפונקציה באגף ימין במשוואה (19) נקראת "גל סינוס דועך".

4.7 פונקצית המתאם המשותף החלקית (partial autocorrelation function)

נניח כי אנו משערים שמודל אוטו-רגרסיבי יכול להתאים לסדרה שאנו לומדים, אך איננו יודעים את הסדר של המודל.

משוואות Yule-Walker עבור המודל מסדר k הן:

$$\begin{aligned}
\rho_1 &= \phi_{k1} + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1} \\
\rho_2 &= \phi_{k1}\rho_1 + \phi_{k2} + \cdots + \phi_{kk}\rho_{k-2} \\
&\cdots \\
\rho_k &= \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}.
\end{aligned}$$

כאן $\phi_{k1}, \phi_{k2}, \ldots, \phi_{kk}$ הם הפרמטרים $\phi_1, \ldots, \phi_k$ עבור תהליך אוטו-רגרסיבי מסדר $p = k$. נפתור את המשוואות הללו לכל k לפי הסדר. נקבל בפרט

$$\phi_{11} = \rho_1,$$

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2},$$

ובאופן דומה נקבל פתרון עבור ϕ_{kk} לכל k.

הגודל ϕ_{kk} נקרא המתאם המשותף החלקי בפער k. נשים לב כי עבור תהליך אוטו-רגרסיבי מסדר p, המתאם המשותף החלקי בפער k יתאפס עבור $k > p$. אכן, נקבל עמודה "מיותרת" במערכת משוואות Yule-Walker, ולכן המקדם המתאים יתאפס. למען האמת, לכל $k, l > p$ נקבל כי $\phi_{kl} = 0$.

בדרך זו ניתן לזהות את הסדר של המודל.

מהי טעות התקן של המתאם המשותף החלקי? אם התהליך $\{Y_t\}$ הוא תהליך אוטו-רגרסיבי מסדר p, Quenouille הראה כי עבור $k \geq p+1$ האומדים $\widehat{\phi}_{kk}$ הם בקירוב בלתי תלויים עם תוחלת שהיא בקירוב 0 ושונות $\mathrm{Var}(\widehat{\phi}_{kk}) \approx \frac{1}{N}$, באשר N הוא מספר התצפיות שנתונות לנו.

דוגמא 4.10

נתונות לנו $N = 70$ תצפיות מתהליך כימי מסוים, ואנו חושדים כי הסדרה מגיעה מתהליך אוטו-רגרסיבי. נניח שחישבנו אומדים עבור פונקצית המתאם המשותף החלקי כדלהלן:

k	1	2	3	4	5	6	7	8	9	10
$\widehat{\phi}_{kk}$	$-.4$	$.19$	$.01$	$-.07$	$-.07$	$-.15$	$.05$	0	$-.1$	$.05$

טעות התקן של $\widehat{\phi}_{kk}$ עבור k הגדול מסדר המודל היא בקירוב $.12 \approx \frac{1}{\sqrt{70}}$, ולכן, למעט $\widehat{\phi}_{11}$, כל שאר הערכים הם בתחום של שתי סטיות תקן מתוחלת 0. ניתן לכן להמשיך עם ההשערה כי המודל הוא $\mathrm{AR}(1)$.

עבור מודל מיצוע-נע מסדר 1, $\mathrm{MA}(1)$, הנתון על ידי הנוסחא $Y_t = X_t + \theta_1 X_{t-1}$ מתקיים

$$\phi_{kk} = \theta_1^k \frac{1 - \theta_1^2}{1 - \theta_1^{2(k+1)}},$$

ולכן ϕ_{kk} הוא בקרוב θ_1^k.

עבור מודל מיצוע-נע מסדר 2, $\mathrm{MA}(2)$, הביטוי המדויק למתאם המשותף החלקי הוא מסובך. אם השורשים של הפולינום $\Phi(B)$ הם ממשיים, המתאם המשותף החלקי הוא סכום של גורמים אקספוננציאליים. אם השורשים הם מרוכבים, המתאם המשותף החלקי נתון על ידי גל סינוס דועך (damped sine wave). מבנה דומה ראינו בפונקצית המתאם המשותף של תהליך $\mathrm{AR}(2)$.

4.8 התאמת מודל אוטו-רגרסיבי

נניח שנתונה לנו סדרה שיוצרה על ידי תהליך אוטו-רגרסיבי, אך אנו איננו יודעים את סדר התהליך ואת מקדמיו. את הסדר של התהליך ניתן לזהות בעזרת פונקצית המתאם המשותף החלקית. כפי שראינו במשוואה (14) מעמוד 36, $\phi_{kk} = 0$ עבור $k > p$.

כעת נראה כיצד לאמוד את המקדמים של התהליך. נניח שהתהליך נתון על ידי המשוואה

$$Y_t - \mu = \sum_{i=1}^{p} \phi_i (Y_{t-i} - \mu) + X_t.$$

אזי

$$X_t = (Y_t - \mu) - \sum_{i=1}^{p} \phi_i (Y_{t-i} - \mu).$$

לכן, אם נחשב את הסדרה

$$x_t = (y_t - \mu) - \sum_{i=1}^{p} \phi_i (y_{t-i} - \mu),$$

עבור המקדמים הנכונים סדרה זו תתפלג נורמלית עם תוחלת 0 ושונות σ_X^2, בעוד שעבור המקדמים הלא נכונים שונותה תהיה גדולה יותר. לכן, אם נסמן

$$S = S(\phi_1, \ldots, \phi_p) = \sum_{t=1}^{N} (x_t)^2,$$

נסיק כי ברצוננו למצוא מקדמים $\phi_1, \ldots, \phi_p$ שימזערו את $S(\phi_1, \ldots, \phi_p)$.

מסתבר כי את $\{X_t\}$ מתפלגים נורמלית, אומד זה הוא אומד יחס נראות מקסימלי, אם מתנים כי p הערכים הראשונים של הסדרה הם קבועים.

עבור $p = 1$ נקבל על ידי גזירה כי

$$\widehat{\mu} = \frac{\overline{y}_{(2)} - \widehat{\phi}_1 \overline{y}_{(1)}}{1 - \widehat{\phi}_1},$$

ו-

$$\widehat{\phi}_1 = \frac{\sum_{t=1}^{N-1} (y_t - \widehat{\mu})(y_{t+1} - \widehat{\mu})}{\sum_{t=1}^{N-1} (y_t - \widehat{\mu})^2}$$

הם האומדים של μ ו-ϕ_1, באשר $\overline{y}_{(1)}$ ו-$\overline{y}_{(2)}$ הם ממוצע ערכי הסדרה ללא האיבר האחרון והראשון בהתאמה. מכיוון ש-$\overline{y}_{(1)} \approx \overline{y}_{(2)} \approx \overline{y}$ נקבל כי

$$\widehat{\mu} \approx \overline{y},$$

ו-

$$\widehat{\phi}_1 \approx \frac{\sum_{t=1}^{N-1}(y_t-\overline{y})(y_{t+1}-\overline{y})}{\sum_{t=1}^{N-1}(y_t-\overline{y})^2}. \tag{20}$$

נשים לב כי אומדים אלו היו האומדים שהיינו מקבלים לו היינו מתייחסים למשוואת המודל האוטו-רגרסיבי

$$Y_t-\overline{y}=\phi_1(y_{t-1}-\overline{y})+X_t$$

כאל משוואת רגרסיה כאשר $y_{t-1}-\overline{y}$ הוא המשתנה הבלתי-תלוי. באופן יותר כללי, ניתן להראות כי חלק נכבד מתורת הרגרסיה ניתן להפעיל על מודלים אוטו-רגרסיביים, ובאופן אסימפטוטי התוצאות תהיינה התוצאות הנכונות.

נשים לב עוד כי את המכנה במשוואה (20) ניתן לקרב על ידי $\sum_{t=1}^{N}(y_t-\overline{y})^2=c_0$, ולכן

$$\widehat{\phi}_1 \approx \frac{c_1}{c_0}=r_1.$$

רווח סמך עבור ϕ_1 מתקבל מכיוון שבאופן אסימפטוטי סטיית התקן של $\widehat{\phi}_1$ היא $\sqrt{\frac{1-(\phi_1)^2}{N}}$. כאשר $\phi_1=0$ סטיית התקן היא $1/\sqrt{N}$, ולכן המבחן לבדיקת $\phi_1=0$ הוא האם $\widehat{\phi}_1=r_1$ נמצא בתחום $\pm 2/\sqrt{N}$.

כאשר $\phi_1=1$, כלומר, כאשר

$$y_t=\sum_{j=0}^{t-1}x_{t-j}+y_0,$$

ניתן להראות כי

$$N(\widehat{\phi}_1-1)=\frac{\frac{1}{N}\times\sigma_{t=2}^{N}y_{t-1}x_t}{\frac{1}{N^2}\sigma_{t=2}^{N}y_{t-1}^2}=O(1) \tag{21}$$

חסום כאשר אורך הסדרה שואף לאינסוף, והן המונה והן המכנה מתכנסים להתפלגות גבולית שאינה נורמלית. לכן כאשר אורך הסדרה שואף לאינסוף $\widehat{\phi}_1$ מתכנס לערכו האמיתי $\phi_1=1$, וקצב ההתכנסות הוא $\frac{1}{N}$, שהוא קצב מהיר יותר מהמקרה הסטציונרי.

עבור מודל אוטו-רגרסיבי מסדר שני, האומדים ל-μ, ϕ_1 ו-ϕ_2 הם:

$$\begin{aligned}\widehat{\mu} &\approx \overline{y},\\ \widehat{\phi}_1 &\approx \frac{r_1(1-r_2)}{1-r_1^2},\\ \widehat{\phi}_2 &\approx \frac{r_2-r_1^2}{1-r_1^2}.\end{aligned}$$

נזכיר עוד כי ניתנו שיטות לחישוב confidence regions עבור $(\widehat{\phi}_1,\widehat{\phi}_2)$, שלא נראה כאן.

5 מודל ARMA

הגדרה 5.1 נניח ש-$\{X_t\}$ הוא תהליך אקראי טהור עם תוחלת μ_X ושונות σ_X^2. תהליך $\{Y_t\}$ יקרא **תהליך ARMA מסדר** (p,q) אם קיימים קבועים $\theta_1,\ldots,\theta_q$ ו-$\phi_1,\ldots,\phi_p$ כך שלכל t מתקיים

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + X_t + \theta_1 X_{t-1} + \cdots + \theta_q X_{t-q}. \tag{22}$$

נסמן תהליך כזה על ידי $\mathrm{ARMA}(p,q)$.

על ידי שימוש באופרטור ההזזה לאחור נקבל הצגה שקולה לתהליך $\mathrm{ARMA}(p,q)$:

$$(1-\phi_1 B - \cdots - \phi_p B^p)Y_t = (1+\theta_1 B + \cdots + \theta_q B^q)X_t.$$

ואם נסמן $\Phi(B) = 1-\phi_1 B - \cdots - \phi_p B^p$ ו-$\Theta(B) = 1+\theta_1 B + \cdots + \theta_q B^q$ נקבל הצגה מקוצרת

$$\Phi(B)Y_t = \Theta(B)X_t,$$

או

$$Y_t = (\Phi(B))^{-1}\Theta(B)X_t.$$

נסמן $\Psi(B) = (\Phi(B))^{-1}\Theta(B)$ ונקבל

$$Y_t = \Psi(B)X_t.$$

בדרך זו קיבלנו הצגה קומפקטית של סדרת הפלט $\{Y_t\}$ כפונקציה של סדרת הקלט $\{X_t\}$. יתר על כן, מכיוון ש-$\Psi(B)$ הוא טור של חזקות של B, קיבלנו הצגה של התהליך כתהליך $\mathrm{MA}(\infty)$. אם נסמן $\Pi(B) = (\Theta(B))^{-1}\Phi(B) = (\Psi(B))^{-1}$ נקבל

$$X_t = \Pi(B)Y_t,$$

שהיא הצגה של סדרת הקלט כפונקציה של סדרת הפלט. בפרט, קיבלנו כאן הצגה של התהליך כתהליך $\mathrm{AR}(\infty)$.

ממשפט 3.1 מעמוד 22 אנו מסיקים את המסקנה הבאה.

מסקנה 5.2 יהא $\{Y_t\}$ תהליך ARMA המקיים את המשוואה $\Phi(B)Y_t = \Theta(B)X_t$.

- אם כל השורשים של הפולינום Φ נמצאים מחוץ לעיגול היחידה, התהליך הוא סטציונרי.
- אם כל השורשים של הפולינום Θ נמצאים מחוץ לעיגול היחידה, ניתן לשחזר את הסדרה $\{X_t\}$ מתוך הסדרה $\{Y_t\}$. כלומר, לפולינום $\Theta(B)\Phi(B)^{-1}$ יש רדיוס התכנסות לפחות 1. במקרה זה נאמר שהתהליך **הפיך**.

אנו נתעניין בתהליכי $\mathrm{ARMA}(p,q)$ המקיימים שכל השורשים של הפולינומים $\Phi(B)$ ו-$\Theta(B)$ נמצאים מחוץ לעיגול היחידה.

5.1 חישוב המקדמים של Ψ ושל Π

כפי שאמרנו, ניתן להציג תהליך ARMA כתהליך $\mathrm{MA}(\infty)$ וכתהליך $\mathrm{AR}(\infty)$ באופן הבא:

$$\begin{aligned} Y_t &= \Psi(B)X_t = \sum_{i=0}^{\infty} \psi_i X_{t-i}, \\ X_t &= \Pi(B)Y_t = \sum_{j=0}^{\infty} \pi_j Y_{t-j}. \end{aligned}$$

כזכור,

$$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p,$$

ו-

$$\Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

כעת נראה כי המקדמים (π_j) ו-(ψ_i) מקיימים משוואת הפרשים.
מכיוון ש-$\Phi(B)\Psi(B) = \Theta(B)$, ומכיוון ש-$\Phi$ הוא פולינום ממעלה p,

$$\theta_j = \phi_0\psi_j - \phi_1\psi_{j-1} - \cdots - \phi_p\psi_{j-p}, \quad \forall j \geq 0,$$

כאשר אנו מבינים כי $\psi_j = 0$ עבור $j < 0$. בפרט, הסדרה (ψ_i) מקיימת משוואת הפרשים מסדר p. עבור $j = 0$, מכיוון ש-$\theta_0 = \phi_0 = 1$ נקבל כי $\psi_0 = 1$.

נשים לב שמכיוון ש-Θ הוא פולינום ממעלה q, $\theta_j = 0$ עבור $j > q$.
באופן דומה, $\Phi(B) = \Theta(B)\Pi(B)$, ו-$\Theta$ הוא פולינום ממעלה q. לכן

$$1 = \phi_0 = \theta_0\pi_0,$$

ו-

$$-\phi_j = \theta_0\pi_j + \theta_1\pi_{j-1} + \cdots + \theta_p\pi_{j-p},$$

כאשר אנו מבינים כי $\pi_j = 0$ עבור $j < 0$.

מכיוון שהמקדמים של Ψ ו-Π הם פתרונות של משוואות הפרשים, אנו יודעים מהי הצורה הכללית שלהם. אם התהליך הוא סטציונרי והפיך, כל השורשים של Φ ו-Θ נמצאים מחוץ לעיגול היחידה, ואז המקדמים (ψ_i) ו-(π_j) שואפים לאפס. בפרט, בתהליך ARMA סטציונרי והפיך, האוטוקורלציה שואפת לאפס ככל שהפער גדל.

5.2 דוגמא: תהליך $\mathrm{ARMA}(1,1)$

נתבונן בתהליך $\mathrm{ARMA}(1,1)$ הבא:

$$Y_t = 0.5Y_{t-1} + X_t - 0.3X_{t-1}.$$

אזי $\Phi(B) = 1 - 0.5B$, ו-$\Theta(B) = 1 - 0.3B$. השורשים של Φ ו-Θ הם 2 ו-$3\frac{1}{3}$, והם מחוץ לעיגול היחידה. לכן תהליך זה הוא הפיך וסטציונרי. כמו כן

$$\begin{aligned}
\Psi(B) &= \frac{\Theta(B)}{\Phi(B)} = \frac{1-0.3B}{1-0.5B} \\
&= (1-0.3B)(1+0.5B+0.5^2B^2+\cdots) \\
&= 1+0.5B-0.3B+0.5^2B^2-0.3\cdot 0.5B^2+\cdots \\
&= 1+0.2B+0.2\cdot 0.5B^2+\cdots .
\end{aligned}$$

באופן כללי, המקדם ה-i של Ψ נתון על ידי

$$\psi_i = (0.5-0.3)\cdot 0.5^{i-1} = 0.2\cdot 0.5^{i-1}, \quad i \geq 1.$$

עבור תהליך $\mathrm{ARMA}(1,1)$ כללי,

$$Y_t - \phi_1 Y_{t-1} + X_t - \theta_1 X_{t-1}.$$

ולכן

$$(1-\phi_1 B)Y_t = (1+\theta_1 B)X_t.$$

כדי שהתהליך יהיה הפיך צריך לדרוש ש-$|\theta_1| < 1$. כדי שהתהליך יהיה סטציונרי צריך לדרוש ש-$|\phi_1| < 1$.

5.3 גדלים סטטיסטיים של תהליכי ARMA

כעת נחשב את התוחלת, השונות, ופונקצית המתאם המשותף של תהליך ARMA, ונזכיר גם את פונקצית המתאם המשותף החלקית. זיהוי הפרמטרים של תהליך ARMA אינו פשוט כמו זיהוי הפרמטרים של תהליכי AR ו-MA, ונדון בו בהמשך.

5.3.1 התוחלת והשונות

מכיוון ש-$Y_t = \Psi(B)X_t = \sum_{i=0}^{\infty}\psi_i X_{t-i}$, נקבל כי

$$\mathbf{E}[Y_t] = (\sum_{i=0}^{t}\psi_i)\mu_X,$$

ו-

$$\mathrm{Var}[Y_t] = (\sum_{i=0}^{t}\psi_i^2)\sigma_X^2.$$

נשים לב שנוסחאות אלו אינן שימושיות במיוחד, כל עוד לא יודעים מהו הסכום האינסופי...

ננסה לפתח נוסחא שימושית יותר עבור התוחלת. מכיוון ש-

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + X_t + \theta_1 X_{t-1} + \cdots + \theta_q X_{t-q},$$

על ידי לקיחת תוחלת משני האגפים נקבל

$$\begin{aligned}\mu &= \mathbf{E}[Y_t] = \sum_{j=1}^{p} \phi_j \mathbf{E}[Y_{t-j}] + \sum_{i=0}^{q} \theta_i \mu_X \\ &= \mu \times \sum_{j=1}^{p} \phi_j + \mu_X \sum_{i=0}^{q} \theta_i.\end{aligned}$$

לכן

$$\mu = \mu_X \frac{\sum_{i=0}^{q} \theta_i}{1 - \sum_{j=1}^{p} \phi_j}.$$

פיתוח נוסחא מדוייקת לשונות היא משימה קשה יותר, ונראה פתרון כלשהו בהמשך.

5.3.2 פונקצית המתאם המשותף

על ידי שימוש במשוואה (22) מעמוד 44 נקבל

$$\begin{aligned}Y_{t-k}Y_t &= \phi_1 Y_{t-k}Y_{t-1} + \cdots + \phi_p Y_{t-k}Y_{t-p} \\ &\quad + Y_{t-k}X_k + \theta_1 Y_{t-k}X_{t-1} + \cdots + \theta_q Y_{t-k}X_{t-q}. \qquad (23)\end{aligned}$$

נסמן $\gamma_k(Y,X) = \mathbf{E}[Y_{t-k}X_t]$, ונקבל

$$\gamma_k = \mathbf{E}[Y_{t-k}Y_t] = \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p} + \gamma_k(Y,X) + \theta_1\gamma_{k-1}(Y,X) + \cdots + \theta_q\gamma_{k-q}(Y,X). \qquad (24)$$

כעת נחשב את $\gamma_k(Y,X)$.

עבור $k = 0$ נקבל:

$$\gamma_0(Y,X) = \mathbf{E}[Y_tX_t] = \sigma_X^2,$$

וזאת כי $Y_{t-1}, \ldots, Y_{t-p}, X_{t-1}, \ldots, X_{t-q}$ אינם תלויים ב-X_t.

עבור $k > 0$, $Y_{t-k-1}, \ldots, Y_{t-k-p}$ אינם תלויים ב-X_t, ולכן נקבל כי $\gamma_k(Y,X) = 0$.

עבור $k < 0$,

$$\begin{aligned}\gamma_k(Y,X) &= \mathbf{E}[Y_{t-k}X_t] \\ &= \mathbf{E}[\phi_1 Y_{t-k-1}X_t + \cdots + \phi_p Y_{t-k-p}X_t \\ &\quad + X_{t-k}X_t + \phi_1 X_{t-k-1}X_t + \cdots + \phi_q X_{t-k-q}X_t].\end{aligned}$$

בפרט, עבור $k = -1$ נקבל

$$\gamma_{-1}(Y,X) = \mathbf{E}[Y_{t+1}X_t] = \phi_1\sigma_X^2 + \theta_1\sigma_X^2, \qquad (25)$$

עבור $k = -2$ נקבל

$$\gamma_{-2}(Y,X) = \mathbf{E}[Y_{t+2}X_t] = \phi_1\gamma_{-1}(Y,X) + \phi_2\sigma_X^2 + \theta_2\sigma_X^2,$$

ונוסחא דומה תתקבל לכל k שלילי.

מכיוון ש-$\gamma_k(Y,X)=0$ עבור $k>0$, נקבל מ-(23) כי עבור $k \geq q+1$ מתקיים

$$\gamma_k = \mathbf{E}[Y_{t-k}Y_t] = \phi_1\gamma_{k-1} + \cdots + \phi_p\gamma_{k-p},$$

ולכן מתקיימות המשוואות

$$\Phi(B)\gamma_k = 0, \qquad k \geq q+1.$$

על ידי חלוקה ב-γ_0, נקבל

$$\Phi(B)\rho_k = 0, \qquad k \geq q+1.$$

קיבלנו דרך רקורסיבית לחישוב פונקצית המתאם החלקי: עבור $k \leq q$ קיבלנו נוסחא מלאה ל-ρ_k, ועבור $k>q$ קיבלנו משוואת הפרשים. נזכור כי מטענות 3.5 ו-3.8 אנו יודעים כיצד נראה הפתרון הכללי של מערכת משוואות הפרשים זו.

נשים לב כי

$$\begin{aligned}
\mathrm{Var}(Y_t) &= \gamma_0 \\
&= \phi_1\gamma_1 + \cdots + \phi_p\gamma_p + \sigma_X^2 + \theta_1\gamma_{-1}(Y,X) + \cdots + \theta_q\gamma_{-q}(Y,X).
\end{aligned}$$

ניתן לפתור משוואה זו יחד עם p המשוואות עבור $\gamma_1,\ldots,\gamma_p$ כדי למצוא את γ_0.

5.3.3 פונקצית המתאם המשותף החלקית

כדי לחשב את פונקצית המתאם המשותף החלקית נשתמש בעובדה ש-

$$X_t = (\Theta(B))^{-1}\Phi(B)Y_t$$

וזהו טור אינסופי. מכאן ניתן להראות כי התנהגות החזקות הגבוהות נקבע בעיקר על ידי $(\Theta(B))^{-1}$, ולכן פונקצית המתאם המשותף החלקית, עבור k-ים גבוהים, מתנהגת כמו פונקצית המתאם המשותף החלקית של תהליך MA.

5.4 המשך דוגמא: תהליך $\mathrm{ARMA}(1,1)$

נחשב את פונקצית המתאם המשותף של תהליך $\mathrm{ARMA}(1,1)$. מהמשוואות הכלליות שפיתחנו בסעיף 5.3.2

$$\begin{aligned}
\gamma_0 &= \phi_1\gamma_1 + \sigma_X^2 + \theta_1\gamma_{-1}(Y,X), \\
\gamma_1 &= \phi_1\gamma_0 + \theta_1\sigma_X^2, \\
\gamma_k &= \phi_1\gamma_{k-1}, \qquad k \geq 2.
\end{aligned} \tag{26}$$

נציב את משוואה (25) במשוואה (26) ונקבל

$$\begin{aligned}
\gamma_0 &= \frac{1+\theta_1^2+2\theta_1\phi_1}{1-\phi_1^2}\sigma_X^2, \\
\gamma_1 &= \frac{(1+\phi_1\theta_1)(\phi_1+\theta_1)}{1-\phi_1^2}\sigma_X^2, \\
\gamma_k &= \phi_1\gamma_{k-1}, \qquad k \geq 2.
\end{aligned}$$

מכאן נקבל כי

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{(1+\phi_1\theta_1)(\phi_1+\theta_1)}{1+\theta_1^2+2\phi_1\theta_1},$$

ו-

$$\rho_k = \phi_1\rho_{k-1}, \qquad k \geq 2.$$

אנו רואים כי (ρ_k) יורדת אקספוננציאלית ל-0.

ומה קורה עם פונקצית המתאם המשותף החלקית? מתקבל כי $\phi_{11} = \rho_1$, ועבור $k > 1$ פונקצית המתאם המשותף החלקית מתנהגת כמו פונקצית המתאם המשותף החלקית של תהליך $\mathrm{MA}(1)$, והגורם השולט בה הוא גורם אקספוננציאלי מוחלש.

6 מודלים לא-סטציונריים

נתבונן בתהליך המקיים את המשוואה

$$\Phi(B)Y_t = \Theta(B)X_t. \tag{27}$$

אם כל השורשים של $\Phi(B)$ נמצאים מחוץ לעיגול היחידה, התהליך $\{Y_t\}$ הוא סטציונרי. סדרות רבות שאנו נתקלים בהן בחיי היומיום אינן סטציונריות. לכן המודלים שפיתחנו עד כה אינם שימושיים ללימוד סדרות כאלו. פתרון טבעי הוא להחליש את הנחת הסטציונריות.

דוגמא 6.1

לשם דוגמא נתבונן במודל

$$Y_t = \phi_1 Y_{t-1} + X_t,$$

וניקח לדוגמא את סדרת הרעש האקראי

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_10
0.1	-1.1	0.2	-2	-0.2	-0.8	0.8	0.1	0.1	-0.9

נניח כי $Y_0 = 0.7$, ונראה כיצד מתנהגת הסדרה $\{Y_t\}$ עבור ערכים שונים של ϕ_1:

ϕ_1	Y_0	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9	Y_{10}
$\frac{1}{2}$	$.7$	$.45$	$-.88$	$-.24$	-2.12	-1.26	-1.43	$.09$	$.14$	$.17$	$-.82$
2	$.7$	1.5	1.9	4	6	11.8	22.8	46.4	92.9	185.9	370.9
1	$.7$	$.8$	$-.3$	$-.1$	-2.1	-2.3	-3.1	-2.3	-2.2	-2.1	-3

אנו רואים כי התנהגות התהליך תלויה מאוד בערך של ϕ_1: כאשר ϕ_1 גדול, ההתנהגות היא אקספוננציאלית, בעוד כאשר ϕ_1 קטן, המצב אינו כך.

כעת ננתח את התלות ב-ϕ_1 בצורה מסודרת. קל לראות כי:

$$\begin{aligned} Y_1 &= \phi_1 Y_0 + X_1, \\ Y_2 &= \phi_1^2 Y_0 + \phi_1 X_1 + X_2, \\ &\dots \\ Y_k &= \phi_1^k Y_0 + \phi_1^{k-1} X_1 + \cdots + X_k, \quad k \geq 1. \end{aligned}$$

כאשר $|\phi_1| > 1$ האיבר המשפיע ביותר על הסכום הוא ϕ_1^k, שערכו המוחלט $|\phi_1|^k$ הולך וגדל (אם ϕ_1 שלילי, ϕ_1^k מתנדנד). התבנית הכללית שסדרה המיוצרת על ידי תהליך כזה תפגין היא גידול אקספוננציאלי, אך לרוב איננו נפגשים בגידול כזה בסדרות נצפות.

כאשר $|\phi_1| < 1$ הטור $\sum_{k=1}^{\infty} |\phi_1|^k$ מתכנס, ונקבל תהליך $\mathrm{AR}(1)$.

המקרה הנותר הוא כאשר $|\phi_1| = 1$. במקרה זה נטפל עכשיו.

נניח כי ל-$\Phi(B)$ יש שורש על שפת עיגול היחידה. לדוגמא, נניח כי 1 הוא שורש של $\Phi(B)$ מריבוי d, וכל שאר השורשים של $\Phi(B)$ נמצאים מחוץ לעיגול היחידה. במקרה כזה ניתן לכתוב את המודל בצורה הבאה:

$$\Phi(B)(1-B)^d Y_t = \Theta(B)X_t, \tag{28}$$

כאשר כל השורשים של $\Phi(B)$ הם מחוץ לעיגול היחידה.

נגדיר תהליך חדש (Z_t) באופן הבא:

$$Z_t = (1-B)^d Y_t.$$

עבור $d=1$,

$$Z_t = (1-B)Y_t = Y_t - Y_{t-1} = \nabla Y_t,$$

וזוהי סדרת ההפרשים של התהליך (Y_t). עבור $d=2$,

$$Z_t = (1-B)(1-B)Y_t = \nabla^2 Y_t,$$

וזוהי סדרת ההפרשים מסדר שני, וכנ"ל לכל $d \geq 3$.

נקבל כי

$$\Phi(B)Z_t = \Theta(B)X_t,$$

כלומר $\{Z_t\}$ הוא תהליך ARMA, וכל השורשים של הפולינום $\Phi(B)$ הם מחוץ לעיגול היחידה. בפרט התהליך $\{Z_t\}$ הוא סטציונרי.

מסקנה 6.2 אם (Y_t) הוא תהליך $\text{ARIMA}(p,d,q)$, אזי סדרת ההפרשים $Z_t = (1-B)^d Y_t$ היא תהליך $\text{ARMA}(p,q)$.

מכאן, שכדי לזהות תהליך ARIMA, נסתכל על סדרות ההפרשים $(1-B)^d Y_t$, עבור $d = 0,1,2,\ldots$, וננסה לזהות האם יש d עבורו סדרת ההפרשים $(1-B)^d Y_t$ היא תהליך ARMA.[1]

נשים לב מתוך הסדרה (Z_t) ניתן לשחזר את הסדרה (Y_t). עבור $d=1$ מתקיים $Z_t = \nabla Y_t = Y_t - Y_{t-1}$, ולכן $Y_t = Z_t + Y_{t-1}$, ומכאן נקבל

$$Y_t = \sum_{j=-\infty}^{t} Z_j.$$

באופן דומה, עבור $d=2$ נקבל $Z_t = \nabla^2 Y_t$, ולכן

$$Y_t = \sum_{i=-\infty}^{t} \sum_{j=-\infty}^{i} Z_j.$$

קיבלנו כי ניתן להציג את $\{Y_t\}$ כסכום (או אינטגרציה) d פעמים של תהליך סטציונרי $\{Z_t\}$. מכאן מקור השם תהליכים אוטו-רגרסיביים-אינטגרטיביים-מיצוע-נע, באנגלית AutoRegressive Integrated Moving Average processes, ובקיצור, תהליכי ARIMA.

בפרק זה נדון במשפחת התהליכים המקיימים את המשוואה

$$\Phi^*(B)Y_t = \Phi(B)(1-B)^d Y_t = \mu + \Theta(B)X_t. \tag{29}$$

בהצגה זו:

- האופרטור $\Phi^*(B)$ הוא אופרטור AR כללי.

[1] נראה בהמשך כיצד מזהים האם תהליך מסוים הוא תהליך ARMA.

- האופרטור $\Phi(B)$ הוא אופרטור AR. הוא מייצג תהליך סטציונרי (כלומר, כל שורשיו מחוץ לעיגול היחידה), והוא מסדר p.
- μ הוא קבוע.
- $\Theta(B)$ הוא אופרטור MA, הוא הפיך (כל שורשיו מחוץ לעיגול היחידה), והוא מסדר q.

משפחת התהליכים המקיימים את המשוואה (29) מסומנת ב-$\mathrm{ARIMA}(p,d,q)$. ההשפעה של הקבוע μ היא להוסיף מרכיב דטרמיניסטי למגמה של התהליך.

נתבונן בתהליך פשוט

$$(1-B)Y_t = \mu + X_t. \tag{30}$$

על ידי העברת אגפים נקבל

$$Y_t = Y_{t-1} + \mu + X_t.$$

זהו תהליך הילוך מקרי עם סחף (drift) μ. אם התוחלת μ חיובית, התהליך $\{Y_t\}$ נוטה לעלות, ואם התוחלת μ שלילית, התהליך $\{Y_t\}$ נוטה לרדת. אכן,

$$\begin{aligned} Y_t &= Y_{t-1} + \mu + X_t \\ &= (Y_{t-2} + \mu + X_{t-1}) + \mu + X_t \\ &= Y_{t-2} + 2\mu + X_t + X_{t-1} \\ &\quad \dots \\ &= Y_0 + t\mu + \sum_{i=0}^{t-1} X_{t-1}. \end{aligned}$$

קיבלנו כי לתהליך המוגדר במשוואה (30) יש מגמה קבועה עם שיפוע μ, ורעש לא-סטציונרי (כי שונות הרעש אינה קבועה).

באופן כללי, הכללת הקבוע μ תורמת לתהליך מגמה דטרמיניסטית מסדר גודל שהוא פולינום מדרגה d בזמן.
באופן מעשי, אם $\mathbf{E}[Y_t]$ אינה תלויה בזמן, נקבל כי $\mu = 0$.
אם $\mathbf{E}[Y_t]$ לינארית ב-t נקבל כי $d = 1$.
אם $\mathbf{E}[Y_t - Y_{t-1}]$ לינארית ב-t נקבל כי $d = 2$ (במקרה כזה, $\mathbf{E}[Y_t]$ היא פולינום ריבועי ב-t).

6.1 הצגות שונות של תהליך ARIMA

בסעיף זה נתבונן בשלוש הצגות שונות של תהליך ARIMA. כל הצגה מבליטה צד אחר של התהליך.
נציג את הערך הנוכחי Y_t של התהליך בעזרת:

1. ערכים קודמים של הסדרה $\{Y_j\}_{j<t}$, והערכים הנוכחי והקודמים של סדרת הרעש $\{X_j\}_{j\le t}$.
2. הערך הנוכחי וערכים קודמים של הרעש $\{X_{t-j}\}_{j\ge 0}$.
3. כממוצע משוקלל של ערכים קודמים של הסדרה $\{Y_{t-j}\}_{j\ge 0}$, והרעש הנוכחי X_t.

6.1.1 הצגה ראשונה

ראינו כבר כי תהליך ARIMA מקיים את המשוואה

$$\Phi^*(B)Y_t = \Phi(B)(1-B)^d Y_t = \Theta(B)X_t. \tag{31}$$

הפולינום $\Phi^*(B)$ ניתן להצגה כ-

$$\Phi^*(B) = \Phi(B)(1-B)^d = 1 - \phi_1^* B - \phi_2^* B^2 - \cdots - \phi_{p+d}^* B^{p+d}.$$

ולכן

$$Y_t = \sum_{i=1}^{p+d} \phi_i^* Y_{t-i} + X_t + \sum_{j=1}^{q} \theta_j X_{t-j}.$$

6.1.2 הצגה שניה

ננסה להציג את התהליך $\{Y_t\}$ כסכום של רעש אקראי:

$$Y_t = \sum_{i=0}^{\infty} \psi_i X_{t-i}. \tag{32}$$

אם נסמן $\Psi(B) = \sum_{i=1}^{\infty} \psi_i B^i$, נקבל כי ההצגה המבוקשת היא

$$Y_t = \Psi(B)X_t.$$

ממשוואה 31 רואים כי $\Psi(B) = (\Phi^*(B))^{-1}\Theta(B)$, ובאופן שקול $\Phi^*(B)\Psi(B) = \Theta(B)$. קיבלנו כי ניתן לחשב את המקדמים של Ψ מהמקדמים של Φ^* ושל Θ:

$$\left(1 - \phi_1^* B - \cdots - \phi_{p+d}^* B^{p+d}\right)\left(\psi_0 + \psi_1 B + \psi_2 B^2 + \cdots\right) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

בפרט, $\psi_0 = 1$. יתר על כן, עבור $j > \max\{p+d-1, q\}$ נקבל כי $\Phi^*(B)\psi_j = 0$, או באופן שקול $\Phi(B)(1-B)^d\psi_j = 0$. כלומר, הסדרה (ψ_j) היא פתרון של מערכת משוואות הפרשים, שאת פתרונה ראינו בטענות 3.5 ו-3.8. בפרט, הסדרה (ψ_j) היא סכום של אקספוננטים וגלי סינוס מוחלשים.

6.1.3 הצגה שלישית

כזכור, תהליך ARIMA נתון על ידי המשוואה

$$\Phi^*(B)Y_t = \Theta(B)X_t.$$

אם נסמן $\Pi(B) = (\Theta(B))^{-1}\Phi^*(B)$, נקבל

$$\Phi^*(B) = \Theta(B)\Pi(B). \tag{33}$$

אם $\Pi(B) = \pi_0 - \pi_1 B - \pi_2 B^2 - \cdots$, משוואה (33) נותנת

$$(1 - \phi_1^* B - \cdots - \phi_{p+d}^* B^{p+d}) = (1 + \theta_1 B + \cdots + \theta_q B^q)(\pi_0 - \pi_1 B - \pi_2 B^2 - \cdots). \tag{34}$$

ולכן $\pi_0 = 1$. יתר על כן, $\Pi(B)Y_t = X_t$, ולכן

$$Y_t = \sum_{j=1}^{\infty} \pi_j Y_{t-j} + X_t.$$

ממשוואה (34) מקבלים שעבור $j > \max\{p+d, q\}$

$$\Theta(B)\pi_j = 0.$$

מכאן אנו מקבלים כי גם הסדרה (π_j) היא פתרון של מערכת משוואות הפרשים, ולכן היא סכום של אקספוננטים וסינוסים מוחלשים.

הפולינומים $\Phi(B)$, $\Theta(B)$ ו-$\Psi(B)$ הם פולינומים ב-B. כאשר $d \geq 1$, אם נציב $B = 1$ נקבל כי

$$\Phi^*(1) = \Phi(1) \times (1-1)^d = 0.$$

מכיוון ש-$\Theta(1) \neq 0$ נקבל כי $\Pi(1) = 1$. אבל

$$\Pi(1) = \pi_0 - \sum_{i=1}^{\infty} \pi_i,$$

מכיוון ש-$\pi_0 = 1$ נקבל כי

$$\sum_{i=1}^{\infty} \pi_i = 1.$$

מכיוון שכל השורשים של Φ^* ו-Θ הם מחוץ לעיגול היחידה, מכיוון ש-$\Pi(B)Y_t = X_t$ נקבל כי

$$Y_t = X_t + \sum_{i=1}^{\infty} \pi_i Y_{t-i},$$

ולכן Y_t הוא סכום של הרעש הנוכחי וממוצע של ערכים קודמים של הסדרה.

מכיוון ש-(π_i) הם סכום של אקספוננטים וסינוסים מוחלשים, הסדרה π_i שואפת ל-0 (ובדרך כלל די מהר), ולכן השפעת ערכים רחוקים של הסדרה על Y_t היא קטנה.

6.1.4 דוגמא: תהליך $\mathrm{ARIMA}(1,1,1)$

מודל $\mathrm{ARIMA}(1,1,1)$ מקיים את המשוואה

$$(1-\phi B)(1-B)Y_t = (1+\theta B)X_t.$$

לכן

$$1 - (1+\phi)B + \phi B^2 = \Phi^*(B) = \Theta(B)\Pi(B) = (1+\theta B)(1 - \pi_1 B - \pi_2 B^2 - \cdots).$$

על ידי השוואת המקדמים של B נקבל: $-(1+\phi) = -\pi_1 + \theta$, ולכן

$$\pi_1 = \theta + 1 + \phi.$$

על ידי השוואת המקדמים של B^2 נקבל $\phi = -\pi_2 - \theta\pi_1$, ולכן

$$\pi_2 = -\phi - \theta\pi_1 = -\phi - \theta(\phi + 1 + \theta) = -(\phi+\theta)(1+\theta).$$

על ידי השוואת המקדמים של B^3 נקבל $0 = -\pi_3 - \theta\pi_2$, ולכן

$$\pi_3 = -\theta\pi_2 = +\theta(\theta+\phi)(1+\theta).$$

באותו אופן נקבל כי לכל $j > 3$,

$$\pi_j = -(-\theta)^{j-2}(\theta+\phi)(1+\theta).$$

כאשר $|\theta| < 1$ המקדמים (π_j) יורדים ל-0 בקצב גיאומטרי, ולכן רוב המידע על Y_t שניתן לקבל מערכים קודמים של הסדרה Y נמצא בערכים האחרונים של הסדרה.

6.1.5 דוגמא: תהליך $\mathrm{ARIMA}(0,1,1)$

תהליך $\mathrm{ARIMA}(0,1,1)$ נתון על ידי

$$\nabla Y_t = Y_t - Y_{t-1} = X_t + \theta X_{t-1}.$$

כאן $Y_t - Y_{t-1} = X_t + \theta X_{t-1}$, כלומר

$$(1-B)Y_t = (1+\theta B)X_t.$$

לכן

$$\psi(B) = (\Psi^*(B))^{-1}\Theta(B) = \frac{1+\theta B}{1-B} = (1+\theta B)(1+B+B^2+\cdots).$$

מכאן נקבל

$$\psi(B) = 1 + \sum_{j=1}^{\infty}(1+\theta)B^j.$$

אנו רואים כי הקורלציה בין Y_t ו-X_{t-j} אינה שואפת ל-0 כאשר j שואף לאינסוף, אלא היא קבועה (ושווה ל-$1+\theta$). כמו כן

$$\Pi(B) = (\Theta(B))^{-1}\Phi^*(B) == frac1 - B1 + \theta B = (1-B)(1-\theta B+\theta^2B^2-\theta^3B^3+\cdots).$$

לכן

$$\Pi(B) = 1-(1+\theta)B+\theta(1+\theta)B^2-\theta^2(1+\theta)B^3+\cdots.$$

7 חיזוי (forecasting)

בסעיף זה נתעניין בבעיה הבאה. נתונות לנו t תצפיות $y_1, \ldots, y_t$ מתהליך לא ידוע. נניח כי לאחר התצפית ה-t ברצוננו לחזות את ערך הסדרה y_{t+l}. הגודל l נקרא פער החיזוי. מכיוון שמעורבת כאן אי-וודאות, אנו נרצה לספק את התוחלת והשונות של החיזוי שלנו.

בסעיף זה נניח כי התהליך שלפנינו הוא תהליך $\mathrm{ARIMA}(p,d,q)$, כלומר, מקיים את המשוואה

$$\Phi^*(B)Y_t = \Phi(B)(1-B)^d Y_t = \Theta(B)X_t,$$

כאשר התהליך $\{X_t\}$ הוא תהליך רעש אקראי סטציונרי עם תוחלת 0 ושונות σ_X^2, והפולינום $\Phi(B)$ אינו מתחלק ב-$1-B$.

נזכור כי בתהליך ARIMA התצפית Y_{t+l} ניתנת להצגה בשלוש דרכים:

1. כסכום של תצפיות קודמות, של הרעש הנוכחי, ושל רעשים קודמים:

$$Y_{t+l} = \phi_1^* Y_{t+l-1} + \cdots + \phi_{p+d}^* Y_{t+l-p-d} + X_{t+l} - \theta_1 X_{t+l-1} - \cdots - \theta_q X_{t+l-1}. \tag{35}$$

2. כסכום משוקלל של הרעש הנוכחי והרעשים הקודמים:

$$Y_{t+l} = \sum_{j=0}^{\infty} \psi_j X_{t+l-j}. \tag{36}$$

3. כסכום של תצפיות קודמות והרעש הנוכחי:

$$Y_{t+l} = \sum_{j=1}^{\infty} \pi_j Y_{t+l-j} + X_{t+l}. \tag{37}$$

ננסה כעת לראות מה התחזית הטובה ביותר שניתן לתת שהיא לינארית בסדרה הנצפית. כלומר, נניח כי התחזית שלנו $\widehat{y}_t(l)$ היא פונקציה לינארית של כל התצפיות עד זמן t (כולל זמן t). מכיוון שבתהליך ARIMA התהליך $\{Y_t\}$ הוא פונקציה לינארית של תהליך הרעש $\{X_t\}$, נקבל כי גם התחזית שלנו היא פונקציה לינארית של תהליך הרעש:

$$\widehat{Y}_t(l) = c_l X_t + c_{l+1} X_{t-1} + \cdots ,$$

באשר $(c_j)_{j \ge l}$ הם קבועים כלשהם.

תחילה ננסה למצוא את הקבועים (c_j) עבורם תוחלת ריבועי השגיאה מתמזערת; כלומר, הגודל $\mathbf{E}\left[\left(Y_{t+l} - \widehat{Y}_t(l)\right)^2\right]$ מינימלי.

אם נשתמש במשוואה (36) נקבל כי

$$\begin{aligned}\mathbf{E}\left[\left(Y_{t+l}-\widehat{Y}_t(l)\right)^2\right] &= \mathbf{E}\left[\left(\sum_{j=0}^{\infty}\psi_j X_{t+l-j}-\sum_{j=0}^{\infty}c_{j+l}X_{t-j}\right)^2\right]\\ &= \mathbf{E}\left[\left(\sum_{j=0}^{l-1}\psi_j X_{t+l-j}+\sum_{j=0}^{\infty}(\psi_{j+l}-c_{j+l})X_{t-j}\right)^2\right]\\ &= (1+\psi_1^2+\cdots+\psi_{l-1}^2)\sigma_X^2+\sum_{j=0}^{\infty}(\psi_{l+j}-c_{l+j})^2\sigma_X^2,\end{aligned}$$

גודל זה מתמזער כאשר $c_{l+j}=\psi_{l+j}$ לכל j. במקרה זה נקבל

$$Y_{t+l}=(X_{t+l}+\psi_1 X_{t+l-1}+\cdots+\psi_{l-1}X_{t+1})+(\psi_l X_t+\psi_{l+1}X_{t-1}+\cdots). \tag{38}$$

נסמן

$$e_t(l):=X_{t+l}+\psi_1 X_{t+l-1}+\cdots+\psi_{l-1}X_{t+1},$$

ו-

$$\widehat{Y}_t(l):=\psi_l X_t+\psi_{l+1}X_{t-1}+\cdots.$$

אזי

$$Y_{t+l}=e_t(l)+\widehat{Y}_t(l).$$

בעוד שהגודל $e_t(l)$ אינו ידוע בזמן t, הרי שהגודל $\widehat{Y}_t(l)$ ידוע בזמן t. מכיוון ש-(X_t) סדרת רעש אקראי, התוחלת של $e_t(l)$ היא 0. לכן נאמר כי $\widehat{Y}_t(l)$ היא התחזית בזמן t בפער זמן l, ו-$e_t(l)$ היא טעות החיזוי.

ניתן להסיק מכאן מספר מסקנות.

נסמן ב-$\mathbf{E}[Y_{t+l}\mid Y_t,Y_{t-1},\ldots]$ את התוחלת המותנה של Y_{t+l} בהינתן המידע הקיים בזמן t. נזכור כי $\{X_t\}$ הם משתנים מקריים בלתי תלויים.

מכיוון ש-X_t בלתי תלוי ב-$\{X_j\}_{j<t}$, נקבל כי

$$\mathbf{E}[X_{t+j}\mid Y_t,Y_{t-1},\ldots]=0,\quad j>0.$$

אם ניקח תוחלת בשני האגפים של משוואה (38) נקבל

$$\mathbf{E}[Y_{t+l}]=\widehat{Y}_t(l)=\psi_l X_t+\psi_{l+1}X_{t-1}+\cdots. \tag{39}$$

כלומר, $\widehat{Y}_t(l)$ היא התחזית הממזערת את ממוצע ריבועי הטעות (mean square error) בזמן t, עבור פער זמן l, והיא שווה לתוחלת המותנה של Y_{t+l} בזמן t, בהינתן המידע עד זמן t ועד בכלל.

מסתבר כי התנאי $\mathbf{E}[X_{t+j}]=0$ הוא תנאי הכרחי כדי שהתוחלת המותנה של Y_{t+l} (שהיא תמיד ממזערת את ממוצע ריבועי ההפרשים) תהיה שווה לתחזית האופטימלית אליה ניתן להגיע בעזרת פונקציית חיזוי לינארית.

נזכור כי טעות החיזוי של $\widehat{Y}_t(l)$ בפער זמן l נתונה על ידי:

$$e_t(l)=X_{t+l}+\psi_1 X_{t+l-1}+\cdots+\psi_{l-1}X_{t+1}$$

מכיוון ש-$\mathbf{E}[e_t(l)] = 0$, התחזית היא לא-מוטה (unbiased). כמו כן, השונות של התחזית היא

$$V(l) = \mathrm{Var}(e_t(l)) = (1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2)\sigma_X^2. \tag{40}$$

משוואה זו תאפשר לנו לספק רווחי חיזוי, כלומר, קטע המכיל בהסתברות נתונה את התצפית העתידית.

ניתן להראות כי כל פונקציה לינארית $\sum_{l=1}^{L} w_l \widehat{Y}_t(l)$ של התחזיות ממזערת את ממוצע ריבועי ההפרשים של הפונקציה $\sum_{l=1}^{L} w_l \ Y_t(l)$.

לכן, אם לדוגמא אנו מנסים לחזות את המכירות של מוצר מסויים באמצעות תצפיות חודשיות מהעבר, ואם התחזיות $\widehat{Y}_t(1), \widehat{Y}_t(2), \widehat{Y}_t(3)$ עבור שלושת החודשים הבאים התקבלו מהתצפיות בדרך שתארנו למעלה, אזי $\widehat{Y}_t(1) + \widehat{Y}_t(2) + \widehat{Y}_t(3)$ היא התחזית עבור המכירות ברבעון הבא הממזערת את ממוצע ריבועי ההפרשים.

מהמשוואה

$$e_t(l) = X_{t+l} + \psi_1 X_{t+l-1} + \cdots + \psi_{l-1} X_{t+1}$$

וממשוואה (38) ניתן לראות כי

$$e_t(1) = X_{t+1} = Y_{t+1} - \widehat{Y}_t(1).$$

לכן הרעש $\{X_t\}$, שעד כה השתמשנו בו רק כרעש אקראי, הוא בעצם טעות החיזוי החד-שלבית. מכיוון שה-X_t הם בלתי תלויים, טעויות החיזוי החד-שלבית בין זמנים שונים הן בלתי-תלויות.

באופן כללי, טעויות החיזוי בפערי זמן גדולים מ-1 בזמנים שונים יהיו מתואמות (וזאת מכיוון שיש אותם גורמים X_t הנלקחים בחשבון בחישוב התחזית בזמנים שונים).

ההשלכה הפרקטית שיש להבחנה זו היא שייתכן וכל התחזיות שלנו תהיינה מוטות באותו כיוון.

הגדרה 7.1 לכל t קבוע, הפונקציה $l \mapsto \widehat{Y}_t(l)$ נקראת **פונקצית החיזוי** עבור זמן מוצא t.

7.1 שלוש הצגות לתחזית

התחזית ל-Y_{t+l} הממזערת את ממוצע ריבועי ההפרשים היא התוחלת המותנה של $\mathbf{E}[Y_{t+l}]$ בהינתן המידע בזמן t. ניתן כעת, בעזרת שלוש ההצגות של תהליך ARIMA, לתת שלושה ביטויים לתחזית זו. לשם פשטות הנוסחאות, נסמן

$$\langle Y_{t+l} \rangle = \mathbf{E}[Y_{t+l} \mid Y_t, Y_{t-1}, \ldots],$$

ו-

$$\langle X_{t+l} \rangle = \mathbf{E}[X_{t+l} \mid X_t, X_{t-1}, \ldots] = \begin{cases} 0 & l > 0, \\ X_{t+l} & l \le 0. \end{cases}$$

מכיוון שמתוך הסדרה $(Y_j)_{j \le t}$ ניתן לשחזר את הסדרה $(X_j)_{j \le t}$, נקבל כי

$$\mathbf{E}[X_{t+l} \mid Y_t, Y_{t-1}, \ldots] = \mathbf{E}[X_{t+l} \mid X_t, X_{t-1}, \ldots] = \langle X_{t+l} \rangle.$$

לכן על ידי שימוש במשוואה (35) מעמוד 56 נקבל

$$\begin{aligned} \langle Y_{t+l} \rangle &= \widehat{Y}_t(l) \\ &= \phi_1 \langle Y_{t+l-1} \rangle + \cdots + \phi_{p+d} \langle Y_{t+l-p-d} \rangle - \theta_1 \langle X_{t+l-1} \rangle - \cdots - \theta_q \langle X_{t+l-q} \rangle + \langle X_{t+l} \rangle. \end{aligned} \tag{41}$$

על ידי שימוש במשוואה (36) מעמוד 56 נקבל

$$\langle Y_{t+l}\rangle = \widehat{Y}_t(l) = \sum_{j=0}^{\infty}\psi_j\langle X_{t+l-j}\rangle. \tag{42}$$

על ידי שימוש במשוואה (37) מעמוד 56 נקבל

$$\langle Y_{t+l}\rangle = \widehat{Y}_t(l) = \sum_{j=1}^{\infty}\pi_j\langle Y_{t+l-j}\rangle + \langle X_{t+l}\rangle. \tag{43}$$

אם התהליך הפיך, כפי שראינו קודם הסדרה (π_j) מתכנסת ל-0. לכן, עבור רמת דיוק נתונה, התלות של Y_{t+l} בתצפיות Y_{t-j} עבור j גדול דיו ניתנת להזנחה. באופן מעשי, ה-π_j-לרוב שואפים לאפס די מהר.

כדי לחשב את התוחלת המותנה בביטויים שאך זה פיתחנו, נשים לב שלכל $j \geq 0$ מתקיים:

$$\begin{aligned}
\langle Y_{t-j}\rangle &= \mathbf{E}[Y_{t-j} \mid Y_t, Y_{t-1}, \ldots] = Y_{t-j}, \qquad j \geq 0,\\
\langle Y_{t+j}\rangle &= \mathbf{E}[Y_{t+j} \mid Y_t, Y_{t-1}, \ldots] = \widehat{Y}_t(j), \qquad j \geq 1,\\
\langle X_{t-j}\rangle &= \mathbf{E}[X_{t-j} \mid X_t, X_{t-1}, \ldots] = X_{t-j} = Y_{t-j} - \widehat{Y}_{t-j-1}(1), \qquad j \geq 0,\\
\langle X_{t+j}\rangle &= \mathbf{E}[X_{t+j} \mid X_t, X_{t-1}, \ldots] = 0, \qquad j \geq 1.
\end{aligned}$$

לכן, כדי לספק את התחזית $\widehat{Y}_t(l)$ בפער זמן l נשתמש בהצגת התהליך באחת משלושת הדרכים שהופיעו למעלה, ונשתמש בכללים הבאים:

1. כל Y_{t-j} עבור $j \geq 0$ שכבר נצפה במוצא זמן t נשאר כשהיה.
2. כל Y_{t+j} עבור $j \geq 1$ שעוד לא נצפה, מוחלף בתחזית $\widehat{Y}_t(j)$.
3. כל X_{t-j} עבור $j \geq 0$ שכבר ארע מוחלף בביטוי $Y_{t-j} - \widehat{Y}_{t-j-1}(1)$.
4. כל X_{t+j} עבור $j \geq 1$ שעוד לא ארע מוחלף ב-0.

לרוב הכי פשוט לעבוד עם משוואת ההפרשים (41). במקרה זה התחזית $\widehat{Y}_t(l)$, עבור $l \geq 1$, ניתנת לביטוי בצורה הבאה:

$$\widehat{Y}_t(l) = \sum_{j=1}^{p+d}\phi_j\widehat{Y}_t(l-j) - \sum_{j=l}^{q}\theta_j X_{t+l-j},$$

באשר $\widehat{Y}_t(-j) = \langle Y_{t-j}\rangle$ מציין את התצפית Y_{t-j} עבור $j \geq 0$, והגורמים של הממוצע-הנע עבור פער זמן $l > q$ מושמטים.

7.2 דוגמא לשימוש במשוואת ההפרשים (41)

בסעיף זה נראה מספר דוגמאות לשימוש במשוואת ההפרשים (41) לצורך חיזוי.

7.2.1 חיזוי עבור תהליך $\mathrm{ARIMA}(1,1,0)$

נתבונן בתהליך $\mathrm{ARIMA}(1,1,0)$ הנתון על ידי:

$$(1-0.8B)(1-B)Y_{t+1}=X_{t+1}.$$

כלומר

$$(1-1.8B+0.8B^2)Y_{t+1}=X_{t+1},$$

או באופן שקול

$$Y_{t+l}=1.8Y_{t+l-1}-0.8Y_{t+l-2}+X_{t+l}.$$

בפרט, $p=1$, $d=1$, $q=0$, $\phi_1=1.8$ ו-$\phi_2=-0.8$. ממשוואה (41), התחזית בזמן מוצא t נתונה על ידי:

$$\begin{aligned}\widehat{Y}_t(1) &= 1.8Y_t-0.8Y_{t-1},\\ \widehat{Y}_t(2) &= 1.8\widehat{Y}_t(1)-0.8Y_t,\\ \widehat{Y}_t(l) &= 1.8\widehat{Y}_t(l-1)-0.8\widehat{Y}_t(l-2),\quad l\geq 3\end{aligned} \tag{44}$$

קל לראות כי התחזיות ניתנות לחישוב באופן רקורסיבי.

מכיוון ש-$q=0$, אין גורמים של ממוצע-נע.

7.2.2 חיזוי עבור תהליך $\mathrm{ARIMA}(0,2,2)$

נתבונן בתהליך $\mathrm{ARIMA}(0,2,2)$ הנתון על ידי:

$$\nabla^2Y_t=(1-0.9B+0.5B^2)X_t,$$

כלומר

$$(1-2B+B^2)Y_t=(1-0.9B+0.5B^2)X_t.$$

אזי $p=0$, $d=2$, $q=2$, $\phi_1=2$, $\phi_2=-1$, $\theta_1=-0.9$, ו-$\theta_2=0.5$. נקבל כי

$$Y_{t+l}=2Y_{t+l-1}-Y_{t+l-2}+X_{t+l}-0.9X_{t+l-1}+0.5X_{t+l-2},$$

ולכן ממשוואה (41)

$$\begin{aligned}\widehat{Y}_t(1) &= 2Y_t-Y_{t-1}-0.9X_t+0.5X_{t-1},\\ \widehat{Y}_t(2) &= 2\widehat{Y}_t(1)-Y_t+0.5X_t,\\ \widehat{Y}_t(l) &= 2\widehat{Y}_t(l-1)-\widehat{Y}_t(l-2), l\geq 3.\end{aligned}$$

כדי לספק במקרה זה את התחזית, אנו צריכים לדעת את ערכי סדרת הרעש X_t ו-X_{t-1}. מכיוון שערכים אלו אינם ידועים, אנו אומדים אותם בדרך הבאה. נחשב

$$X_s=Y_s-\widehat{Y}_{s-1}(1)=Y_s-\left(\sum_{j=1}^{p+d}\phi_jY_{s-j}-\sum_{j=1}^{q}\theta_jX_{s-j}\right), s=p+d+1,\dots,t.$$

בנוסחא זו, אנו מציבים $Y_j = X_j = 0$ עבור $j < 1$, ומחשבים את X_s בצורה רקורסיבית.

ישנה שיטה מדוייקת יותר לחיזוי (ראה סעיף 7.5), בה אנו אומדים את הערכים של התהליכים $\{Y_t\}$ ו- $\{X_t\}$ שקרו לפני הסדרה הנצפית בשיטה הנקראת back-forcasting. בשיטה זו נקבל את התחזית הממזערת את ממוצע ריבועי הטעויות של $\mathbf{E}[Y_{t+l} \mid Y_t, \ldots, Y_1]$. אם מספר התצפיות הוא גדול, שתי השיטות נותנות בקירוב אותן התוצאות.

7.2.3 דיון

במקרה בו התהליך הוא $\mathrm{ARIMA}(p,d,0)$, כלומר $q = 0$, אין תלות של התהליך בערכים קודמים של הסדרות $\{Y_t\}$ או $\{X_t\}$. ממשוואה (41) רואים כי התחזית נתונה על ידי

$$\widehat{Y}_t(l) = \phi_1 \widehat{Y}_t(l-1) + \phi_2 \widehat{Y}_t(l-2) + \cdots + \phi_{p+d} \widehat{Y}_t(l-p-d), \quad l \geq 1,$$

כאשר

$$\widehat{Y}_t(1) = \phi_1 Y_t + \phi_2 Y_{t-1} + \cdots + \phi_{p+d} Y_{t+1-p-d}.$$

לכן תחזיות הנעשות בזמן מסויים תלויות אך ורק ב-$p+d$ התצפיות האחרונות.

באופן כללי, כשהתהליך הוא תהליך $\mathrm{ARIMA}(p,d,q)$ עם $q > 0$, התחזיות $\widehat{Y}_t(1), \widehat{Y}_t(2), \ldots, \widehat{Y}_t(q)$ תהיינה תלויות באופן ישיר בגורמי הרעש, אך התחזיות $\widehat{Y}_t(l)$ עבור $l > q$ לא תהיינה תלויות בגורמים אלו באופן ישיר, אלא רק באופן עקיף: $\widehat{Y}_t(l)$ תלוי ב-$\widehat{Y}_t(l-1)$, שתלוי ב-$Y_t(l-2)$, שתלוי באופן רקורסיבי ב-$\widehat{Y}_t(q)$, שתלוי בגורמי הרעש.

נסיים בהבחנה כי עבור תהליך ARIMA לא-סטציונרי, כלומר, כאשר $d \geq 1$, ניתן להציג את התחזית באופן נוסף. לדוגמא, כאשר $d = 1$ ניתן להציג $W_t = (1-B)Y_t$, או $Y_t = W_t + Y_{t-1}$. במקרה זה התהליך $\{W_t\}$ הוא תהליך $\mathrm{ARMA}(p,q)$ סטציונרי. לכן תחזיות עבור התהליך $\{Y_t\}$ ניתן להציג בצורה הבאה:

$$\widehat{Y}_t(l) = \widehat{W}_t(l) + \widehat{Y}_t(l-1) = \widehat{W}_t(l) + \widehat{W}_t(l-1) + \cdots + \widehat{W}_t(1) + Y_t,$$

כאשר התחזיות $\widehat{W}_t(l)$ מתקבלות כתחזיות עבור תהליך $\mathrm{ARMA}(p,q)$ בהסתמך על סדרת התצפיות $\{W_t\}$.

באופן דומה, עבור $d = 2$ נגדיר תהליך $\mathrm{ARMA}(p,q)$ על ידי

$$W_t = (1-B)^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2},$$

ונקבל

$$\widehat{Y}_t(l) = \widehat{W}_t(l) + 2\widehat{Y}_t(l-1) - \widehat{Y}_t(l-2) = \sum_{j=0}^{l-1} (j+1)\widehat{W}_t(l-j) + Y_t + l(Y_t - Y_{t-1}).$$

בצורה דומה ניתן לקבל נוסחאות לתחזית לכל $d \geq 1$.

7.3 חישוב ועדכון תחזיות

נניח כי נתונה לנו סדרה עתית $(y_1, y_2, \ldots, y_t)$, ואנחנו צריכים לחזות את ערך הסדרה בפערי זמן $l = 1, 2, \ldots, L$. אנו צריכים לחזות את ערך הסדרה בכל פער זמן l, ולספק סטית תקן של התחזית, כדי שאפשר יהיה לחשב רווחי חיזוי (probability limits). אנו נראה שיטה בה אפשר לעדכן את התחזיות בזמן $t+1$ מתוך התחזיות בזמן t.

7.3.1 חישוב הקבועים (ψ_j)

לצורך חישוב רקורסיבי זה נצטרך לחשב את הקבועים $\psi_1, \psi_2, \ldots, \psi_{L-1}$. ניתן לחשב קבועים אלו על ידי שימוש מתוך המשוואה $\Phi(B)\Psi(B) = \Theta(B)$, כפי שראינו בסעיף 6.1.2 בעמוד 53. אם נציג כל פולינום כטור חזקות נקבל

$$(1 - \phi_1 B - \cdots - \phi_{p+d} B^{p+d})(1 + \psi_1 B + \psi_2 B^2 + \cdots) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q).$$

אם אנו יודעים את הערכים של $\phi_1, \ldots, \phi_{p+d}, \theta_1, \ldots, \theta_q$, כלומר, אם אנו יודעים מהו התהליך ממנו יוצרה הסדרה, קל לחשב את המקדמים $\psi_1, \psi_2, \ldots$ בדרך הבאה:

$$\begin{aligned}
\psi_1 &= \phi_1 - \theta_1, \\
\psi_2 &= \phi_1 \psi_1 + \phi_2 - \theta_2, \\
&\cdots \\
\psi_j &= \phi_1 \psi_{j-1} + \cdots + \phi_{p+d} \psi_{j-p-d} - \theta_j.
\end{aligned}$$

בחישוב זה, $\psi_0 = 1$, $\psi_j = 0$ עבור $j < 0$, ו-$\theta_j = 0$ עבור $j > q$. לכל $j > \max\{p + d - 1, q\}$ מתקיימת משוואת ההפרשים

$$\psi_j = \phi_1 \psi_{j-1} + \phi_2 \psi_{j-2} + \cdots + \phi_{p+d} \psi_{j-p-d},$$

המאפשרת חישוב מהיר של ה-(ψ_j) באופן רקורסיבי.

לדוגמא, עבור התהליך $(1 - 1.8B + 0.8B^2)Y_t = X_t$ מקבלים

$$(1 - 1.8B + 0.8B^2)(1 + \psi_1 B + \psi_2 B^2 + \cdots) = 1,$$

ולכן $p = 2$, $d = 0$, $q = 0$, $\phi_1 = 1.8$ ו-$\phi_2 = -0.8$. נקבל אם כן

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= 1.8 \\
\psi_j &= 1.8\psi_{j-1} - 0.8\psi_{j-2}, \qquad j = 2, 3, 4, \ldots.
\end{aligned}$$

7.3.2 שימוש בקבועים (ψ_j) לעדכון התחזיות

ממשוואה (39) מעמוד 57 ניתן להציג את התחזיות $\widehat{Y}_{t+1}(l)$ ו-$\widehat{Y}_t(l+1)$ בדרך הבאה:

$$\begin{aligned}
\widehat{Y}_{t+1}(l) &= \psi_l X_{t+1} + \psi_{l+1} X_t + \psi_{l+2} X_{t-1} + \cdots, \\
\widehat{Y}_t(l+1) &= \psi_{l+1} X_t + \psi_{l+2} X_{t-1} + \cdots.
\end{aligned}$$

נחסר את שתי המשוואות ונקבל

$$\widehat{Y}_{t+1}(l) = \widehat{Y}_t(l+1) + \psi_l X_{t+1}. \qquad (45)$$

מכיוון ש-

$$X_{t+1} = Y_{t+1} - \widehat{Y}_t(1),$$

נקבל כי

$$\widehat{Y}_{t+1}(l) = \widehat{Y}_t(l+1) + \psi_l(Y_{t+1} - \widehat{Y}_t(1)), \tag{46}$$

כלומר, כאשר אנו מקבלים את התצפית Y_{t+1} אנו יכולים לעדכן את התחזית לזמן $t+l+1$ בקלות רבה.
אם כבר חישבנו את התחזיות $\widehat{Y}_t(1), \ldots, \widehat{Y}_t(L)$, וקיבלנו את Y_{t+1}, נוכל בעזרת משוואה (46) לחשב את $\widehat{Y}_{t+1}(1), \ldots, \widehat{Y}_{t+1}(L-1)$. חישוב $\widehat{Y}_{t+1}(L)$ אינו יכול להתבצע בשיטה זו, אך ניתן לקבלו בעזרת משוואת ההפרשים (41) מעמוד 58 מהתחזיות הקודמות לפער זמן זה.

7.3.3 חישוב רווחי חיזוי

משוואה (40) מעמוד 58 אומרת כי:

$$V(l) = \mathrm{Var}(e_t(l)) = (1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{l-1}^2)\sigma_X^2 = \left(\sum_{j=0}^{l-1} \psi_j^2\right) \sigma_X^2.$$

הנחנו כי הרעש האקראי הוא נורמלי, כלומר לכל t המשתנה המקרי X_t מתפלג $N(0, \sigma_X^2)$. אזי ההתפלגות המותנית של Y_{t+l} בהינתן $Y_1, \ldots, Y_t$ היא נורמלית עם תוחלת $\widehat{Y}_t(l)$ ושונות $\left(\sum_{j=1}^{l-1} \psi_j^2\right) \sigma_X^2$. בפרט, ניתן לספק רווחי חיזוי, קרי, קטעים שבהסתברות נתונה מכילים את התצפית.

7.3.4 חישוב תחזיות בעזרת ההצגה ההופכית

נזכור כי אם התהליך הפיך ניתן להציגו בדרך הבאה:

$$X_t = \Pi(B) Y_t = (1 - \pi_1 B - \pi_2 B^2 - \pi_3 B^3 - \cdots) Y_t.$$

אם יודעים את משוואת התהליך הנתון, ניתן לקבל את המקדמים (π_j) על ידי פתרון המשוואה

$$\Theta(B)(1 - \pi_1 B - \pi_2 B^2 - \cdots) = \Phi(B).$$

בסעיף זה נראה כיצד ניתן לנצל הצגה זו כדי לספק תחזיות. הדרך אותה נראה כעת אינה טובה יותר מהדרך אותה ראינו בסעיפים הקודמים.

נזכור כי משוואה (43) מעמוד 59 אומרת:

$$\langle Y_{t+l} \rangle = \widehat{Y}_t(l) = \sum_{j=1}^{\infty} \pi_j \langle Y_{t+l-j} \rangle + \langle X_{t+l} \rangle.$$

נסמן $\widehat{Y}_t(-j) = Y_{t-j}$ עבור $j \le 0$, ונקבל כי

$$\begin{aligned}
\widehat{Y}_t(l) &= \sum_{j=1}^{\infty} \pi_j \widehat{Y}_t(l-j) \\
&= \pi_1 \widehat{Y}_t(l-1) + \cdots + \pi_{l-1} \widehat{Y}_t(1) + \pi_l Y_t + \pi_{l+1} Y_{t-1} + \cdots .
\end{aligned}$$

אם נציב $l = 1$ נקבל

$$\widehat{Y}_t(1) = \pi_1 Y_t + \pi_2 Y_{t-1} + \pi_3 Y_{t-2} + \cdots .$$

גם את התחזיות בפערי זמן גדולים ניתן להביע כצרופים לינאריים של התצפיות $\{Y_t\}$. לדוגמא, עבור $l=2$ נקבל

$$\begin{aligned}
\widehat{Y}_t(2) &= \pi_1 \widehat{Y}_t(1) + \pi_2 Y_t + \pi_3 Y_{t-1} + \cdots \\
&= \pi_1 \sum_{j=1}^{\infty} \pi_j Y_{t-j+1} + \sum_{j=1}^{\infty} \pi_{j+1} Y_{t-j+1} \\
&= \sum_{j=1}^{\infty} \pi_j^{(2)} Y_{t-j+1},
\end{aligned}$$

באשר

$$\pi_j^{(2)} = \pi_1 \pi_j + \pi_{j+1}, \qquad j \geq 1.$$

בדרך זו ניתן להציג

$$\widehat{Y}_t(l) = \sum_{j=1}^{\infty} \pi_j^{(l)} Y_{t-j+1},$$

באשר המקדמים $(\pi_j^{(l)})$ ניתנים לחישוב רקורסיבי על ידי המשוואה

$$\pi_j^{(l)} = \pi_{j+l-1} + \sum_{h=1}^{l-1} \pi_h \pi_j^{(l-h)},$$

ו-$\pi_j^{(1)} = \pi_j$.

7.4 דוגמאות לחישוב התחזיות ולעדכונן

בסעיף זה נסתכל במספר דוגמאות של תהליכים, ונראה כיצד מחשבים את התחזיות וכיצד מעדכנים אותן. נזכור כי על ידי שימוש במשוואת ההפרשים הגענו לביטוי הבא עבור התחזית (ראה משוואה (41) מעמוד 58):

$$\begin{aligned}
\langle Y_{t+l} \rangle &= \widehat{Y}_t(l) \\
&= \phi_1 \langle Y_{t+l-1} \rangle + \cdots + \phi_{p+d} \langle Y_{t+l-p-d} \rangle - \theta_1 \langle X_{t+l-1} \rangle - \cdots - \theta_q \langle X_{t+l-q} \rangle + \langle X_{t+l} \rangle.
\end{aligned} \tag{47}$$

7.4.1 תהליך $\mathrm{ARIMA}(0,1,1)$

נתבונן בתהליך $\mathrm{ARIMA}(0,1,1)$ המקיים את המשוואה

$$\nabla Y_t = (1+\theta B) X_t.$$

במקרה זה $p=0$, $d=1$, $q=1$ ו-$\theta_1=\theta$. על ידי שימוש במשוואה (47) נקבל

$$\begin{aligned}
\widehat{Y}_t(1) &= Y_t + \theta X_t, \\
\widehat{Y}_t(l) &= \widehat{Y}_t(l-1), \quad l \geq 2.
\end{aligned} \tag{48}$$

בפרט, התחזיות לכל פערי הזמן זהות.

מכיוון ש-$Y_t = \widehat{Y}_{t-1}(1) + X_t$, ניתן לעדכן את התחזיות עם קבלת תצפית חדשה באופן הבא:

$$\widehat{Y}_t(l) = \widehat{Y}_{t-1}(l) + (1+\theta)X_t = \widehat{Y}_{t-1}(l) + (1+\theta)(Y_t - \widehat{Y}_{t-1}(1)).$$

לכן, אם התחזית $\widehat{Y}_{t-1}(1)$ שונה מהתצפית Y_t, מעדכנים את כל התחזיות בהפרש בין השתיים כפול הגורם $1-\theta$.
דרך נוספת לכתוב את משוואה (48) היא

$$\widehat{Y}_t(l) = (1-\theta)Y_t + \theta\widehat{Y}_{t-1}(l).$$

כלומר,התחזית החדשה עבור פער זמן l היא ממוצע משוקלל של התחזית הקודמת והתצפית החדשה. אם θ קרוב ל-0 המשקל של התצפית החדשה גדול והשפעת התצפיות הקודמות על התחזית קטנה, בעוד שאם θ קרוב ל-1 המשקל של התחזית הקודמת גדול, והתצפית החדשה משפיעה אך במעט על התחזית החדשה.

עבור מודל זה חישבנו בסעיף 6.1.5 בעמוד 55 כי

$$\psi_j = 1+\theta \quad \forall j \geq 1.$$

ממשוואה (40) מעמוד 58 השונות של התחזיות נתונה על ידי

$$V(l) = \sigma_X^2\left(1 + (l-1)(1+\theta)^2\right).$$

7.4.2 תהליך $\mathrm{ARIMA}(0,2,2)$

נתבונן בתהליך $\mathrm{ARIMA}(0,2,2)$ המקיים את המשוואה

$$(1-2B+B^2)Y_t = \nabla^2 Y_t = (1-\theta_1 B - \theta_2 B^2)X_t.$$

במודל זה $p=0$, $d=2$, $q=2$, $\phi_1=2$ ו-$\phi_2=-1$. על ידי שימוש במשוואה (41) מעמוד 58 נקבל

$$\begin{aligned}
\widehat{Y}_t(1) &= 2Y_t - Y_{t-1} - \theta_1 X_t - \theta_2 X_{t-1}, \\
\widehat{Y}_t(2) &= 2\widehat{Y}_t(1) - Y_t - \theta_2 X_t, \\
\widehat{Y}_t(l) &= 2\widehat{Y}_t(l-1) - \widehat{Y}_t(l-2), \quad l \geq 3.
\end{aligned} \tag{49}$$

כעת נחשב את הסדרה (ψ_j). ראשית, נשים לב כי מכיוון ש-

$$(1-2B+B^2)(1+2B+3B^2+4B^3+\cdots) = 1$$

נקבל כי

$$\begin{aligned}
\psi(B) &= \frac{\Theta(B)}{\Phi^*(B)} \\
&= (1-\theta_1 B - \theta_2 B^2)(1+2B+3B^2+4B^3+\cdots) \\
&= 1 + (2-\theta_1)B + (3-2\theta_1-\theta_2)B^2 + (4-3\theta_1-2\theta_2)B^3 + \cdots.
\end{aligned}$$

לכן

$$\psi_j = (j+1) - j\theta_1 - (j-1)\theta_2 = 1 + \theta_2 + j(1-\theta_1-\theta_2).$$

ממשוואה (45) מעמוד 62 עדכון התחזיות נעשה על ידי המשוואה

$$\begin{aligned}\widehat{Y}_{t+1}(l) &= \widehat{Y}_t(l+1) + 1 + \theta_2 + j(1-\theta_1-\theta_2)X_{t+1}\\ &= \widehat{Y}_t(l+1) + 1 + \theta_2 + j(1-\theta_1-\theta_2)(Y_{t+1} - \widehat{Y}_t(1)).\end{aligned}$$

ממשוואה (40), ומכיוון ש-$\psi_j = 1 + \theta_2 + j(1-\theta_1-\theta_2)$ מקבלים כי

$$V(l) = \sigma_X^2\left(1 + (l-1)(1+\theta_2)^2 + \frac{1}{6}(l-1)(2l-1)(1-\theta_1-\theta_2)^2 + l(l-1)(1+\theta_2)(1-\theta_1-\theta_2)\right).$$

7.4.3 תהליך $\mathrm{ARIMA}(p,d,0)$

נתבונן בתהליך $\mathrm{ARIMA}(p,d,0)$ המקיים את המשוואה

$$\Phi(B)(1-B)^d Y_t = X_t.$$

ממשוואה (41) בעמוד 58, לכל פער זמן קבוע $l \geq 1$ פונקצית התחזית עבור פער זמן l מקיימת את המשוואה

$$\Phi(B)\widehat{Y}_t(l) = 0.$$

נשים לב כי עבור t קבוע, הסדרה $(\widehat{Y}_t(l))_{l\geq 1}$ היא פתרון של מערכת משוואות הפרשים, ולכן היא סכום של גורמים אקספוננציאלים וגלי סינוס דועכים.

לדוגמא, אם התהליך נתון על ידי המשוואה

$$(1-B)Y_t = X_t,$$

כלומר

$$Y_t = Y_{t-1} + X_t,$$

אזי התחזית נתונה על ידי

$$\widehat{Y}_t(l) = Y_t,$$

כלומר, התחזית שווה לתצפית הנוכחית.

7.5 ההצגה ההפוכה של תהליך ARMA

תהליך ARMA מקיים את המשוואה

$$Y_t = X_t + \sum_{i=1}^{q} \theta_i X_{t-i} - \sum_{j=1}^{p} \phi_j Y_{t-j}.$$

על ידי העברת אגפים נקבל

$$\phi_{t-p} Y_{t-p} = X_t + \sum_{i=1}^{q} \theta_i X_{t-i} - \sum_{j=0}^{p-1} \phi_j Y_{t-j},$$

עם זיהוי $\phi_0 = 1$. אם נציב $t+p$ במקום t נקבל

$$\phi_{t-p} Y_t = X_{t+q} + \sum_{i=1}^{q} \theta_i X_{t+p-i} - \sum_{j=0}^{p-1} \phi_j Y_{t+p-j}.$$

התהליך $\{X_t\}$ הוא תהליך רעש אקראי, ולכן ניתן להזיזו בזמן; כלומר, ניתן להציב $W_t = X_{t+p}$ לכל t. נקבל, אם כן,

$$\phi_{t-p} Y_t = \sum_{i=0}^{q} \theta_{q-i} X_{t+i} - \sum_{j=1}^{p} \phi_{p-1-j} Y_{t+j},$$

ובאופן שקול

$$Y_t = \sum_{i=0}^{q} \frac{\theta_{q-i}}{\phi_{t-p}} X_{t+i} - \sum_{j=1}^{p} \frac{\phi_{p-1-j}}{\phi_{t-p}} Y_{t+j}.$$

כלומר, הצגנו את התצפית הנוכחית Y_t כצרוף לינארי של תצפיות עתידיות, גורם רעש נוכחי, וגורמי רעש עתידיים. במילים אחרות, הפכנו את כיוון הזמן. על ידי שימוש בטכניקת החיזוי שראינו, ניתן לחזות את התצפיות Y_0, Y_{-1}.

הסיבה לחישוב זה היא הסיבה הבאה. טכניקת החיזוי שפיתחנו דורשת לדעת את התצפיות $Y_0, \ldots, Y_{-p}$ בכדי לחזות את התצפית של Y_{t+l}.

לכן, בשלב ראשון נתבונן בסדרה ההפוכה, נניח כי $Y_{t+1} = Y_{t+2} = \ldots = 0$, ונספק תחזיות לתצפיות $Y_0, \ldots, Y_{-p}$. בעזרת תחזיות אלו נספק תחזיות לתצפיות $Y_{t+1} = Y_{t+2} = \ldots = 0$.

8 בניית מודל

בפרקים הקודמים הצגנו מספר מודלים המתארים סדרות עתיות, וראינו מספר תכונות שסדרות המיוצרות ממודלים אלו מקיימות. יתר על כן, בהינתן המודל המייצר את הסדרה העתית, ראינו כיצד ניתן לחזות ערכים עתידיים של הסדרה. אך כיצד נדע מהו המודל ממנו יוצרה הסדרה העתית?

בסעיף זה נדון במספר שיטות לזיהוי מודל שיסביר כיצד יוצרה הסדרה העתית. כאן נתעסק בסדרות שאין בהם גורם עונתי.

שיטות זיהוי מטבען אינן מדוייקות, והן אך ורק מרמזות על מודלים שכדאי להמשיך ולבדקם. מטרתנו כאן היא לקבל רעיון כללי לגבי הערכים של d, p ו-q של מודל ARIMA כללי שיסביר את הסדרה באופן האופטימלי, והערכים האופטימליים של המקדמים. המודל הראשוני שנקבל ישמש כנקודת פתיחה לשיטות פורמליות יעילות יותר.

התכונה הבסיסית של הסדרה שתשמש אותנו לבניית המודל היא פונקצית המתאם המשותף ופונקצית המתאם המשותף החלקית.

הרעיון הבסיסי בבניית מודל הוא למצוא את המודל הפשוט ביותר המסביר באופן סביר את הסדרה. לכן תחילה נבדוק את ההשערה שהמודל סטציונרי.

8.1 הנראות של תהליכי ARMA ו-ARIMA

יהא $\{Y_t\}$ תהליך $\mathrm{ARMA}(p,q)$ עם מקדמים ידועים. ונכתוב אותו בצורה הבאה:

$$X_t = Y_t - \phi_1 Y_{t-1} - \phi_p Y_{t-p} + \theta_1 X_{t-1} + \cdots + \theta_q X_{t-q}.$$

אם תוחלת התהליך אינה $\mu = 0$, נוריד את התוחלת מכל התצפיות.

אם הערכים של $Y_0, Y_{-1}, \ldots, Y_{-p+1}, X_0, X_{-1}, \ldots, X_{-q+1}$ היו נתונים, היינו יכולים לחשב את X_t, לכל $t \geq 1$.

לכן, לכל הצעה של מקדמים $\theta_1, \ldots, \theta_q, \phi_1, \ldots, \phi_p$, ובהינתן ערכים התחלתיים של $Y_0, Y_{-1}, \ldots, Y_{-p+1}, X_0, X_{-1}, \ldots, X_{-q+1}$, ניתן לחשב את ערכי סדרת הרעש $\{X_t\}$.

לשם פשטות הנוסחאות, נסמן

$$\begin{aligned}
\vec{\theta} &= (\theta_1, \ldots, \theta_q), \\
\vec{\phi} &= (\phi_1, \ldots, \phi_q), \\
\vec{X}_0 &= (X_0, X_{-1}, \ldots, X_{-p+1}), \\
\vec{Y}_0 &= (Y_0, Y_{-1}, \ldots, Y_{-q+1}).
\end{aligned}$$

מכיוון שמניחים כי סדרת הרעש האקראי מתפלגת נורמלית, וערכיה בלתי תלויים, פונקצית הצפיפות המשותפת של הסדרה $\{X_t\}$ פרופורציונית ל-

$$p(X_1, X_2, \ldots, X_t) \propto \frac{1}{\sigma_X^n} \exp\left(-\sum_{t=1}^{N} \frac{X_t^2}{2\sigma_X^2}\right).$$

מכאן נקבל כי פונקצית הלוג-נראות המותנית, בהינתן $\vec{\theta}, \vec{\phi}, \vec{X}_0, \vec{Y}_0, \sigma_X^2$, היא

$$l_*(\vec{\theta}, \vec{\phi}, \vec{X}_0, \vec{Y}_0, \sigma_X^2) = -n \ln(\sigma_X) - \frac{S_*(\vec{\theta}, \vec{\phi}, \vec{X}_0, \vec{Y}_0, \sigma_X^2)}{2\sigma_X^2},$$

באשר

$$S_*(\vec{\theta},\vec{\phi},\vec{X}_0,\vec{Y}_0,\sigma_X^2)=\sum_{t=1}^{N}X_t^2(\vec{\theta},\vec{\phi},\vec{X}_0,\vec{Y}_0).$$

נשים לב כי אם σ_X^2, השונות של תהליך הרעש, קבועה, אזי l_* היא פונקציה לינארית של S_*. הגודל S_* נקרא **conditional least squares estimates**.

כיצד נקבעים הערכים ההתחלתיים $\vec{X}_0,\vec{Y}_0$? דרך נאיבית היא לאתחל ערכים אלו ב-0. דרך טובה יותר היא לקבלם בעזרת השיטה שתוארה בסעיף 7.5.

אם הסדרה ארוכה, והשורשים של Θ רחוקים משפת עיגול היחידה, שתי השיטות תיתנה תוצאות דומות. אם, לעומת זאת, השורשים קרובים לעיגול היחידה, ההתנהגות המקומית של התהליך דומה להתנהגות של תהליך לא סטציונרי, וייתכנו הבדלים גדולים בין שני האיתחולים.

הנראות הלא-מותנית של תהליך ARMA היא מעט יותר מורכבת, ולא ניכנס כאן לנוסחא המדויקת שלה.

אם $\{Y_t\}$ הוא תהליך $\mathrm{ARIMA}(p,d,q)$ עם מקדמים ידועים, נגדיר את סדרת ההפרשים

$$Z_t=\nabla^d Y_t.$$

אזי $\{Z_t\}$ הוא תהליך $\mathrm{ARMA}(p,q)$, וניתן לחשב את הנראות שלו כפי שראינו קודם.

כאשר המקדמים $\vec{\theta}$ ו-$\vec{\phi}$ אינם ידועים, יש למצוא את הערכים עבורם לוג-הנראות מתמזערת. בספרות (ובתוכנות הסטטסיטיות) ניתן למצוא אלגוריתמים המחשבים אומדים למקדמים אלו.

8.2 זיהוי בעזרת פונקצית המתאם המשותף ופונקצית המתאם המשותף החלקית

בסעיף זה נראה כיצד לנצל את פונקצית המתאם המשותף ופונקצית המתאם המשותף החלקית כדי לבנות מודל לסדרה עתית נתונה.

8.2.1 זיהוי d

יהא $\{Y_t\}$ תהליך $\mathrm{ARIMA}(p,0,q)$ המקיים את המשוואה

$$\Phi(B)Y_t=\Theta(B)X_t.$$

פונקצית המתאם המשותף של התהליך מקיימת את משוואת ההפרשים

$$\Phi(B)\rho_k=0,\qquad \forall k>q.$$

יתר על כן, אם $\Phi(B)=\prod_{i=1}^{p}(1-G_iB)$, ואם כל השורשים של פולינום זה שונים, אזי הפתרונות של משוואת ההפרשים הם מהצורה

$$\rho_k=A_1G_1^K+A_2G_2^k+\cdots+A_pG_p^k,\quad \forall k>q-p.$$

אם בנוסף התהליך סטציונרי, השורשים $G_1,\ldots,G_p$ נמצאים בתוך עיגול היחידה. קל לראות כי ρ_k שואפים ל-0 כאשר k גדל. מהירות השאיפה תלויה במרחק בין השורשים $G_1,\ldots,G_p$ ושפת עיגול היחידה: ככל שהשורשים רחוקים יותר משפת עיגול היחידה, השאיפה ל-0 מהירה יותר.

לשם דוגמא, אם $p=1$ ו-$G_1=1-\delta$, באשר δ הוא מספר חיובי קטן, אזי כל עוד $k\delta$ אינו גדול, $\rho_k\approx A_1(1-k\delta)$, ולכן ρ_k דועך ל-0 אך בצורה לינארית ב-k.

לכן, אם אנו רואים כי פונקצית המתאם המשותף אינה דועכת ל-0 בקצב מהיר, ייתכן ולפחות אחד השורשים של הפולינום Φ קרוב לשפת עיגול היחידה. סיבה אפשרית נוספת לתופעה זו היא שהתהליך אינו סטציונרי.

מסקנה אופרטיבית: אם פונקצית המתאם המשותף אינה דועכת מהר ל-0, ייתכן והתהליך אינו סטציונרי, ויש להסתכל על סדרת ההפרשים.

כדי לזהות את d, נחשב לכל $d \geq 0$ את סדרת ההפרשים מסדר d:

$$W_t^{(d)} = \nabla^d Y_t.$$

נחשב את פונקצית המתאם המשותף של $W^{(d)}$. נחפש d עבורו פונקצית המתאם המשותף שואפת ל-0 מהר.

באופן מעשי, d הוא בדרך כלל 0, 1 או 2.

8.2.2 זיהוי p ו-q

לזיהוי p ו-q, ניזכר בהתנהגות פונקצית המתאם המשותף ופונקצית המתאם המשותף החלקית עבור תהליך מיצוע נע, עבור תהליך אוטו-רגרסיבי, ועבור תהליך אוטו-רגרסיבי-מיצוע-נע (ראה הטבלא המצורפת קודם לכן).

- תהליך אוטו-רגרסיבי $\mathrm{AR}(p)$: פונקצית המתאם המשותף לא מתאפסת, ופונקצית המתאם המשותף החלקית מתאפסת עבור פער גדול מ-p.
- תהליך מיצוע-נע $\mathrm{MA}(q)$: פונקצית המתאם המשותף מתאפסת עבור פער גדול מ-q, ופונקצית המתאם המשותף החלקית לא מתאפסת.
- תהליך אוטו-רגרסיבי-מיצוע-נע $\mathrm{ARMA}(p, q)$: פונקצית המתאם המשותף ופונקצית המתאם המשותף החלקית אינן מתאפסות.

התנהגות פונקצית המתאם המשותף של תהליך אוטו-רגרסיבי-מיצוע-נע $\mathrm{ARMA}(p, q)$ היא כסכום של אקספונ-נטים וגלי סינוס דועים עבור פער גדול מ-$q - p$.

ההתנהגות פונקצית המתאם המשותף החלקית של תהליך אוטו-רגרסיבי-מיצוע-נע $\mathrm{ARMA}(p, q)$ היא כסכום של פונקציה מעריכים ופונקצית סינוס דועכת עבור פער גדול מ-$p - q$.

באופן מעשי, סדרות עתיות רבות ניתנות להסבר על ידי תהליך $\mathrm{ARIMA}(1, d, 1)$, ולכן חשוב להכיר את פונקצית המתאם המשותף ופונקצית המתאם המשותף החלקית של תהליך זה (ראה איור מצורף).

8.3 זיהוי p, q ואמידת המקדמים באמצעות הנראות המקסימלית

בסעיף 8.1 ראינו כי ניתן לחשב, לכל ניחוש של p, d, q ולכל ניחוש של מקדמים $\vec{\theta}, \vec{\phi}$ את הנראות. כמו כן, קיימים בספרות אלגוריתמים שונים המחפשים, עבור הצעה ספציפית של p, d, q, את וקטורי המקדמים $\vec{\theta}, \vec{\phi}$ עבורם הנראות מתמקסמת.

נניח שנתונה לנו סדרה עתית, ואנו מניחים כי היא יוצרה על ידי תהליך ARMA. כיצד נמצא את הגדלים האופטימ-ליים עבור p, q לסדרה הנתונה. דרך אחת שהוצעה על ידי Akaike היא לחשב לכל p, q את הגודל הבא:

$$\mathrm{AIC}_{p,q} = \frac{-2 \ln(\text{maximum likelihood}) + 2(p + q + 1)}{N},$$

ולבחור את p,q עבורם גודל זה מתמזער. בגלל הגורם הראשון, ננסה למקסם את הנראות. הגורם השני מביא בחשבון את מספר המקדמים (דרגות החופש) במודל. הוא מוסיף "קנס"(penalty) לכל דרגת חופש שנוסיף למודל. ניתן להראות כי גודל זה שווה בקירוב ל-

$$\mathrm{AIC}_{p,q} \approx \ln(\widehat{\sigma}_X^2) + (p+q+1)\frac{2}{N} + \text{קבוע}.$$

אפשרות שניה שהועלתה בספרות על ידי Schwartz היא למזער את הגודל הבא:

$$\mathrm{BIC}_{p,q} = \ln(\widehat{\sigma}_X^2) + (p+q+1)\frac{\ln(N)}{N}.$$

בחישוב BIC הקנס על תוספת כל דרגת חופש גדול יותר.

חיסרון של תהליך זה הוא שיש לחשב את הנראות של מספר מודלים, וחישוב זה עלול להיות יקר בזמן. בספרות הציעו מספר שיטות לחישוב מהיר של הנראות המקסימלית עבור מספר ערכים של p ו-q.

8.4 הקשר בין פונקצית המתאם המשותף והאומדים שלה

האומדים של פונקצית המתאם המשותף הם בעלי שונות גבוהה, והם מתואמים (highly autocorrelated). לכן, גם אם פונקצית המתאם המשותף התאורטית דועכת ל-0, האומדים יכולים להיות שונים מ-0, וייתכן וכולם (או חלקם) יהיו גבוהים או נמוכים מהערכים התאורטיים.

לכן יתכן ולא נוכל לזהות בוודאות את המודל הנכון, וייתכן ונצטרך להמשיך לשלב הבא עם מספר מודלים אפשריים.

מכיוון שאיננו יודעים את הערכים התאורטיים של פונקצית המתאם המשותף, ומכיוון שהאומדים שאנו מחשבים שונים מהערכים התאורטיים, חשוב לקבל הערכה על ההפרש בין האומד לבין הערך התאורטי. באופן ספציפי, אנו צריכים לפתח דרך להעריך האם פונקצית המתאם המשותף מתאפסת עבור פער גדול דיו, או שמא היא לא מתאפסת.

עבור פערים גדולים, ותחת ההנחה שהתהליך הוא מיצוע-נע מסדר q, ניתן להשתמש בקירוב (4) מעמוד 17, כאשר האומדים של פונקצית המתאם המשותף מחליפים את האוטו-קורלציות התאורטיות:

$$\widehat{\sigma}^2[r_k] \approx \frac{1}{N}\left(1 + 2(r_1^2 + r_2^2 + \cdots + r_q^2)\right), \quad k > q.$$

כאשר התהליך הוא אוטו-רגרסיבי מסדר p, ניתן להראות כי סטיית התקן של האומד לפונקצית המתאם המשותף החלקית היא

$$\widehat{\sigma}^2(\widehat{\phi}_{kk}) \approx \frac{1}{N}, \quad k > p.$$

Anderson הוכיח כי בתנאים מסויימים ההתפלגות של האומד לפונקצית המתאם המשותף, שהערך התאורטי שלו הוא 0, היא נורמלית. תוצאה דומה נכונה עבור האומד של פונקצית המתאם המשותף החלקית. ניתן להשתמש בעובדות אלו כמדריך לא פורמלי לקביעה האם התהליך הוא אוטו-רגרסיבי או מיצוע-נע, ולקביעת סדר התהליך.

8.5 בדיקת המודל

לאחר שמצאנו מודל המסביר בצורה הטובה ביותר את הסדרה העתית שברשותנו, עלינו לבדוק אם המודל אכן מתאים לסדרה. ישנן מספר שיטות לבצע בדיקה כזו. כאן נתרכז באחת השיטות.

נניח שהתאמנו לסדרה העתית מודל $\mathrm{ARIMA}(p,d,q)$ עם הנוסחא

$$\Phi(B)\Delta^d Y_t = \Theta(B)X_t,$$

ואמדנו את המקדמים $\vec{\phi}$ ו-$\vec{\theta}$.

נסמן לשם פשטות

$$Z_t = \Delta^d Y_t.$$

כעת נחשב רקורסיבית את סדרת הרעש באופן הבא:

$$\widehat{X}_t = (\Theta(B))^{-1}\Phi(B)Z_t.$$

כלומר, אנו מחשבים

$$\widehat{x}_t = z_t - \sum_{i=1}^{p}\widehat{\phi}_j z_{t-i} + \sum_{j=1}^{q}\widehat{\theta}_j \widehat{x}_{t-j}.$$

אנו מאתחלים את הווקטורים $\{X_t\}$ ו-$\{Y_t\}$ (עבור $t \le 0$) או באפסים, או שאנו אומדים אותם כמתואר בסעיף ??.

אם המודל שחישבנו מתאים, אזי הסדרה $\{\widehat{x}_t\}$ צריכה להיות סדרה שיוצרה על ידי תהליך רעש אקראי, ולכן מאפייניה צריכים להיות דומים למאפיינים של סדרה כזו.

ניתן להראות כי אם אכן מדובר במודל $\mathrm{ARIMA}(p,d,q)$ אזי

$$\widehat{x}_t = x_t + O(\frac{1}{\sqrt{N}}),$$

כלומר, האומד לרעש בזמן n שווה לרעש עד כדי סדר גודל של $\frac{1}{\sqrt{N}}$. מכאן ניתן להראות כי אם היינו יודעים במדויק את מקדמי הפולינומים Φ ו-Θ, האומדים למקדמי המתאם המשותף של סדרת הרעש, $r_k(X)$, היו בלתי מתואמים, מתפלגים נורמלית עם תוחלת אפס ושונות $\frac{1}{N}$.

במציאות איננו יודעים מקדמים אלו. מסתבר כי במקרה זה, הגודל $\frac{1}{N}$ מהווה חסם עליון לשונות של $r_k(X)$, וככל ש-k גדל, השונות מתקרבת ל-$\frac{1}{N}$.

כיצד ניתן לנצל זאת לבדיקת מודל?

מסתבר שאם המודל מתאים הגודל

$$Q = (N-d)\sum_{k=1}^{K} r_k^2(\widehat{x})$$

מתפלג בקרוב $\chi^2(K-p-q)$, בעוד שאם המודל אינו מתאים גודל זה גדל. לכן ניתן לבדוק האם המודל מתאים על ידי חישוב Q, ובדיקת סבירות גודל זה לפי התפלגות חי-בריבוע.

הועלו טענות כי התפלגות חי-בריבוע אינה קירוב טוב דיו להתפלגותו של הגודל Q. כדי להתגבר על בעיה זו, הציעו שימוש בגודל

$$\widetilde{Q} = (N-d)(N-d+2)\sum_{k=1}^{K}\frac{r_k^2(\widehat{x})}{N-d-k}.$$

ניתן להראות כי קירוב להתפלגות $\widetilde{Q}$ הוא $\chi^2(K-p-q)$.

8.6 ניצול השאריות לתיקון המודל

נניח כי אנו משערים כי הסדרה העתית שברשותנו נוצרה מהתהליך הסטוכסטי

$$\Phi(B)Y_t = \Theta(B)X_t,$$

בעוד שבמציאות הסדרה נוצרה באמצעות התהליך הסטוכסטי

$$\Phi_*(B)Y_t = \Theta_*(B)X_t.$$

הדרך לאמת את המודל היא לחשב את סדרת השאריות

$$Z_t = (\Theta(B))^{-1}\Phi(B)Y_t,$$

ולבדוק האם היא מקיימת תכונות של סדרת רעש לבן. מכיוון שמתקיים

$$Y_t = (\Phi_*(B))^{-1}\Theta_*(B)Y_t,$$

נקבל כי מתקיים

$$Z_t = (\Theta(B))^{-1}\Phi(B)(\Phi_*(B))^{-1}\Theta_*(B)X_t.$$

מכיוון שכפל פולינומים הוא קומוטטיביף

$$Z_t = (\Theta(B))^{-1}(\Phi_*(B))^{-1}\Phi(B)\Theta_*(B)X_t,$$

ולכן

$$\Phi_*(B)\Theta(B)Z_t = \Phi(B)\Theta_*(B)X_t.$$

כלומר, סדרת השאריות שנחשב אינה סדרת רעש לבן. אם כל השורשים של $\Phi, \Theta, \Phi_*, \Theta_*$ נמצאים מחוץ לעיגול היחידה, הסדרה נוצרה מתהליך ARMA, ואם חלק מהשורשים של Φ_* ו-Θ הם 1, זהו תהליך ARIMA.

על ידי חישוב פונקצית המתאם המשותף ופונקצית המתאם המשותף החלקי של סדרת השאריות נוכל, אם כן, לקבל חיזוק להשערה שהמודל שפיתחנו הוא המודל ממנו יוצרה הסדרה.

9 ניתוח בתחום התדר

בסעיף זה נתאר שיטה שונה לניתוח סדרות עתיות. שיטה זו, השואבת את מקורותיה בניתוח פורייה (Fourier) של פונקציות ממשיות, מנסה לאתר תדרים בסדרה העתית.

תחילה נתאר שלוש נקודות שיש לשים אליהן לב בעת ניתוח בתחום התדר.

מכיוון שאנו דוגמים את הסדרה במרווחים מסויימים, לא נוכל לאתר תדרים גבוהים מאוד או נמוכים מאוד. כדי לראות זאת נתבונן בדוגמא הבאה. נניח, לדוגמא, כי אנו מודדים את הטמפרטורה בתל אביב בצהרי כל יום. אם נתונות לנו מדידות של למעלה משנה, נוכל לזהו כי לסדרה יש מחזוריות של 365. כמובן שלא נוכל לגלות כי לטמפרטורה יש גם מחזוריות יומית של 24 שעות. נוכל לגלות מחזוריות זו רק אם נגדיל את קצב הדגימה שלנו: אם נדגום את הטמפרטורה פעמיים ביום, בצהרי היום ובחצות הלילה, נוכל לגלות גם את המחזוריות היומית.

אם נתונות לנו מדידות של הטמפרטורה היומית בתל אביב ממאה השנים האחרונות, אנו נוכל לזהות קו מגמה: הטמפרטורה עולה עם השנים. אך האם זהו קו מגמה, או אולי לטמפרטורה יש מחזוריות של כמה אלפי שנים, וההתחממות הנוכחית היא חלק ממחזוריות ארוכת שנים?

בעייה נוספת שעולה בניתות בתחום התדר היא בעיית הזיוף (alias). בעייה זו מתעוררת כאשר בסדרה יש מחזוריות מסויימת, אך ניתן לזהות אותה אך ורק כמחזוריות מסדר שונה. כדי להדגים בעייה זו, נתבונן בדוגמא של מדידת הטמפרטורה. אנו מודדים את הטמפרטורה בתל אביב כל 23 שעות. כלומר, ביום ראשון בצהרי היום, ביום שני ב-11 בבוקר, ביום שלישי ב-10 בבוקר, וכו'. בנוסף למחזוריות השנתית (מחזוריות של $365 \times \frac{24}{23}$ מדידות) נזהה גם מחזוריות כל 24 מדידות: מכיוון ש-23 ו-24 זרים, המדידה הראשונה והמדידה ה-25 מתבצעות באותה שעה של היממה. מכאן נסיק כי יש מחזוריות של 24 מדידות, שהן 23 יממות. מהנתונים לא נוכל לדעת כי בעצם יש מחזוריות של יממה, שהיא $1\frac{1}{23}$ מדידות.

אם, באופן היפותטי, בעולמנו היתה לטמפרטורה מחזוריות של 23 יממות, הרי שהתרומה של המחזור בן היממה היתה מצטרפת לתרומה של המחזור בן 23 היממות, ולא היינו יכולים להפריד ביניהן.

נדון ביתר פירוט בנקודות אלו בהמשך.

הנוסחאות שמקבלים כאשר אורך הסדרה זוגי או אי-זוגי שונות מעט. נשים לב שתמיד ניתן לזרוק את התצפית הראשונה בסדרה, ובכך לשנות את זוגיות אורך הסדרה. אם הסדרה ארוכה דיה איננו מאבדים מידע רב על ידי זריקת התצפית הראשונה.

כדי לפשט את הנוסחאות, נניח בפרק זה כי אורך הסדרה הנצפית הוא זוגי.

נתחיל בפיתוח מספר זהויות טריגונומטיריות. לאחר מכן נסקור את ניתוח פורייה עבור פונקציות המוגדרות על קטע או על הישר כולו. לבסוף נתמקד בניתוח פורייה עבור סדרות עתיות.

9.1 מעט זהויות טריגונומטריות

נזכור כי $i = \sqrt{-1}$, וכי:

$$e^{i\omega} = \cos\omega + i\sin\omega, \qquad \forall\omega \tag{50}$$

ולכן

$$\sin\omega = \frac{e^{i\omega} - e^{-i\omega}}{2i}, \qquad \forall\omega \tag{51}$$

ו-

$$\cos\omega = \frac{e^{i\omega}+e^{-i\omega}}{2}, \qquad \forall\omega. \tag{52}$$

מכאן נקבל כי לכל ω

$$\begin{aligned}
\sum_{n=1}^{N} e^{i\omega n} &= e^{i\omega}\times\frac{1-e^{i\omega N}}{1-e^{i\omega}}\\
&= e^{i\omega}\times\frac{e^{i\omega N}-1}{e^{i\omega}-1}\\
&= e^{i\omega}\times\frac{e^{i\omega N/2}(e^{i\omega N/2}-e^{-i\omega N/2})/2i}{e^{i\omega/2}(e^{i\omega/2}-e^{-i\omega/2})/2i}\\
&= e^{i\omega(N+1)/2}\times\frac{\sin(\omega N/2)}{\sin(\omega/2)}\\
&= \cos(\omega(N+1)/2)\times\frac{\sin(\omega N/2)}{\sin(\omega/2)}+i\sin(\omega(N+1)/2)\times\frac{\sin(\omega N/2)}{\sin(\omega/2)}.
\end{aligned} \tag{53}$$

ממשוואה (50) נקבל כי לכל ω

$$\sum_{n=1}^{N} e^{i\omega n} = \sum_{n=1}^{N}\cos(\omega n)+i\sum_{n=1}^{N}\sin(\omega n). \tag{54}$$

מהשוואת משוואות (53) ו-(54) אנו מסיקים כי לכל ω

$$\sum_{n=1}^{N}\cos(\omega n) = \cos(\omega(N+1)/2)\times\frac{\sin(\omega N/2)}{\sin(\omega/2)}, \tag{55}$$

$$\sum_{n=1}^{N}\sin(\omega n) = \sin(\omega(N+1)/2)\times\frac{\sin(\omega N/2)}{\sin(\omega/2)}. \tag{56}$$

מכיוון ש-$\lim_{x\to 0}\frac{\sin x}{x}=1$ נובע כי

$$\lim_{x\to 0}\frac{\sin x}{\sin(x/N)} = \lim_{x\to 0}\frac{x}{x/N} = N.$$

עבור כל מספר שלם k נסמן

$$\omega_k = \frac{2\pi k}{N}.$$

אזי

$$\frac{\sin(\omega_k N/2)}{\sin(\omega_k/2)} = \frac{\sin(\pi k)}{\sin(\pi k/N)} = \begin{cases} N & k=0,\\ 0 & 0<k<\frac{N}{2}.\end{cases}$$

מכיוון שעבור $\omega=0$ מתקיים $\cos\omega=1$ ו-$\sin\omega=0$, נקבל כי

$$\sum_{n=1}^{N}\cos(\omega_k n) = \begin{cases} N & k=0,\pm\frac{N}{2},\pm N,\ldots\\ 0 & k\neq 0,\pm\frac{N}{2},\pm N,\ldots,\end{cases} \tag{57}$$

$$\sum_{n=1}^{N}\sin(\frac{2\pi kn}{N}) = 0, \qquad k\in\mathbf{N}. \tag{58}$$

נזכור את הזהויות הטריגונומטריות הבאות: לכל שתי זוויות ω, λ,

$$\begin{aligned}
\cos\omega\cos\lambda &= \frac{1}{2}\left(\cos(\omega+\lambda)+\cos(\omega-\lambda)\right),\\
\sin\omega\sin\lambda &= \frac{1}{2}\left(\cos(\omega-\lambda)-\cos(\omega+\lambda)\right),\\
\sin\omega\cos\lambda &= \frac{1}{2}\left(\sin(\omega+\lambda)+\sin(\omega-\lambda)\right).
\end{aligned}$$

על ידי שימוש במשוואות (57) ו-(58) נקבל

$$\sum_{n=1}^{N}\cos(2\pi kn/N)\cos(2\pi jn/N) = \begin{cases} N & k=j=0 \text{ או } k=j=N/2,\\ N/2 & k=j\neq 0 \text{ וגם } k=j\neq N/2,\\ 0 & k\neq j.\end{cases} \tag{59}$$

$$\sum_{n=1}^{N}\sin(2\pi kn/N)\sin(2\pi jn/N) = \begin{cases} 0 & k=j=0 \text{ או } k=j=N/2,\\ N/2 & k=j\neq 0 \text{ וגם } k=j\neq N/2,\\ 0 & k\neq j.\end{cases} \tag{60}$$

$$\sum_{n=1}^{N}\sin(2\pi kn/N)\cos(2\pi jn/N) = 0, \quad \text{לכל } k,j\,. \tag{61}$$

9.2 ניתוח פורייה - זמן רציף

9.2.1 המשפט הבסיסי

ניתוח פורייה עוסק בקירוב פונקציה ממשית על ידי סכום של סינוסים וקוסינוסים. נניח כי $f:[-\pi,\pi]\to\mathbf{R}$ היא פונקציה ממשית. השאלה הבסיסית בניתוח פורייה היא אם קיימים קבועים $(a_0,a_1,a_2,\ldots,b_1,b_2,\ldots)$ כך שאם נגדיר לכל k פונקציה $f_k:[-\pi,\pi]\to\mathbf{R}$ על ידי:

$$f_k(t)=\frac{a_0}{2}+\sum_{r=1}^{k}(a_r\cos(rt)+b_r\sin(rt)) \tag{62}$$

יתקיים שסדרת הפונקציות $\{f_k\}$ שואפת נקודתית ל-f (החלוקה של a_0 ב-2 היא טכנית לחלוטין, והסיבה לעשותה תתברר בפסקה הבאה).

המשפט הבסיסי של ניתוח פורייה הוא המשפט הבא.

משפט 9.1 תהא $f:[-\pi,\pi]\to\mathbf{R}$ פונקציה רציפה המקיימת $f(-\pi)=f(\pi)$. נסמן

$$\begin{aligned}
a_0 &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)dt\\
a_r &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(rt)dt, \qquad r=1,2,\ldots\\
b_r &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin(rt)dt, \qquad r=1,2,\ldots
\end{aligned}$$

אזי סדרת הפונקציות המוגדרת ב-(62) מתכנסת נקודתית ל-f.

ניתן להראות שאם לפונקציה f יש נקודת אי-רציפות ב-t_0, אך היא רציפה מימין ל-t_0 ומשמאל ל-t_0, אזי[2]

$$\lim_{k\to\infty} f_k(t_0) = \frac{\lim_{t\to t_0+} f(t) + \lim_{t\to t_0-} f(t)}{2}.$$

כלומר, הגבול שווה לממוצע של הערך של f מימין ל-t ולערך של f משמאל ל-t_0.

כדי להשתמש בניתוח פורייה עבור סדרות עתיות, נתאים את הניתוח למקרה בו הפרמטר t הוא בדיד ולא רציף. לאחר הצגת מספר זהויות טריגונומטריות נראה כיצד ניתן לפרק את הסדרה העתית למחזורים בתדרים שונים. לאחר מכן נגדיר את הפריודוגרם, נחקור את הפריודוגרם של סדרות פשוטות, ולבסוף נראה את מגבלות השיטה.

9.2.2 פונקציה המוגדרת בקטע חסום

בסעיף זה נציג במפורש את הפירוק של פונקציה $f:[-\pi,\pi]\to\mathbf{R}$ לרכיביה המחזוריים. לאחר מכן נכליל את ההצגה למקרה ש-f מוגדרת על קטע $[-T,T]$ שאינו דווקא הקטע $[-\pi,\pi]$.

אנו נניח כי f היא רציפה ומקיימת $f(-\pi)=f(\pi)$. כפי שהערנו קודם, הניצתוח נכון גם אם f אינה רציפה, אך כל בכל נקודת אי-רציפות הגבולות מימין ומשמאל קיימים.

ממשפט 9.1 נקבל

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k\cos(kt) + b_k\sin(kt)).$$

נגדיר

$$c_k = \begin{cases} \frac{a_k - ib_k}{2} & k\geq 1, \\ \frac{a_0}{2} & k=0, \\ \frac{a_k+ib_k}{2} & k\leq -1. \end{cases}$$

נציב את משוואות (51) ו-(52) ונקבל

$$\begin{aligned} f(t) &= \frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k\frac{e^{ikt}+e^{-ikt}}{2} + b_k\frac{e^{ikt}-e^{-ikt}}{2i}) \\ &= \sum_{k=-\infty}^{+}\infty c_k e^{ikt}. \end{aligned}$$

נשים לב כי

$$c_k = \int_{-\pi}^{\pi} f(s)e^{-iks}ds.$$

מכאן נקבל את ההצגה הבאה: לכל $t\in[-\pi,\pi]$ מתקיים

$$f(t) = \sum_{k=-\infty}^{+}\infty\left(\int_{-\pi}^{\pi} f(s)e^{-iks}ds\right)e^{ikt}. \quad (63)$$

נניח כעת כי g מוגדרת על הקטע $[-T,T]$. נגדיר פונקציה $f:[-\pi,\pi]\to\mathbf{R}$ באופן הבא:

$$f(x) = g(\frac{xT}{\pi}).$$

[2] הסימון $\lim_{t\to t_0-}$ משמעו כי הפרמטר t שואף ל-t_0, אך מקבל רק ערכים הקטנים ממש מ-t_0. המשמעות של הסימון $\lim_{t\to t_0+}$ אנלוגית.

לכן
$$g(t) = f(\frac{t\pi}{T}).$$
יתר על כן, על ידי ההצבה $u = \frac{sT}{\pi}$ הנותנת $du = \frac{dsT}{\pi}$ נקבל
$$\begin{aligned} \int_{-\pi}^{\pi} f(s)e^{-iks}ds &= \int_{-T}^{T} f(\frac{u\pi}{T})e^{iku\pi/T}\frac{\pi}{T}du \\ &= \frac{\pi}{T}\int_{-T}^{T} g(u)e^{iku\pi/T}du. \end{aligned}$$
הפונקציה f מקיימת את התנאים לפיתוח (63), ולכן לכל $t \in [-\pi, \pi]$ מתקיים
$$f(t) = \sum_{k=-\infty}^{+}\infty \left(\int_{-\pi}^{\pi} f(s)e^{-iks}ds\right)e^{ikt}.$$
נציב כעת את g במקום f ונקבל כי לכל $t \in [-T, T]$ מתקיים
$$g(t) = \frac{\pi}{T}\sum_{k=-\infty}^{+}\infty \left(\int_{-T}^{T} g(u)e^{iku\pi/T}du\right)e^{ikt\pi/T}.$$

נשים לב כי בפירוק זה אנו מפרקים את g לרכיבים מחזוריים, עם מחזורים $(\frac{k\pi}{T})_{k=0}^{\infty}$ (המחזור $-\frac{k\pi}{T}$ זהה למחזור $\frac{k\pi}{T}$).

9.2.3 המקרה הכללי

בסעיף הקודם סקרנו את ניתוח פורייה כאשר הפונקציה מוגדרת על קטע חסום. ביישום שלנו, אנו דוגמים פונקציה שאינה מוגדרת על קטע חסום, אלא לכל זמן, הן בעבר והן בעתיד. אמנם אנו נקבל רק מספר סופי של תצפיות, אך הפונקציה עצמה מוגדרת לכל t. בסעיף זה נראה מהי ההכללה של השיטה למקרה זה.

תהא $g : \mathbf{R} \to \mathbf{R}$ פונקציה כלשהי. כדי שניתן יהיה לפרק את הפונקציה לרכיבים מחזוריים, יש לדרוש כי הפונקציה רציפה, ואינטגרבילית.[3] (יש תנאים חלשים יותר המספיקים לצרכינו, אך אנו מציגים כאן את התנאים הפשוטים ביותר שמספיק לדרוש).

לכל מספר ממשי חיובי T ניתן להסתכל על הפונקציה g כמוגדרת אך ורק על הקטע $[-T, T]$. עבור הפונקציה המצומצמת נקבל כי לכל $t \in [-T, T]$ מתקיים
$$g(t) = \frac{\pi}{T}\sum_{k=-\infty}^{+}\infty \left(\int_{-T}^{T} g(u)e^{iku\pi/T}du\right)e^{ikt\pi/T}.$$

כעת ניתן ל-T לשאוף לאינסוף. במקרה זה הגורם $\frac{1}{T}$ שואף לאפס. מכיוון שאנו מפרקים את g למחזורים $(\frac{k\pi}{T})_{k=0}^{\infty}$, הרי שכאשר T שואף לאינסוף אנו נקבל את כל המחזורים האפשריים.

אם נגדיר לכל $\omega \in \mathbf{R}$
$$C_\omega = \int_{-\infty}^{+\infty} g(t)e^{-2\pi i\omega t}dt,$$

[3] פונקציה $g : \mathbf{R} \to \mathbf{R}$ נקראת אינטגרבילית אם $\int_{-\infty}^{+\infty} |g(t)|dt < +\infty$.

נקבל כי לכל $t \in \mathbf{R}$

$$g(t) = \int_{-\infty}^{+\infty} C_\omega e^{2\pi i \omega t} d\omega. \tag{64}$$

שוויון זה מתקיים רק אם שני האינטגרים האחרונים סופיים (לכל t ולכל ω). תנאי מספיק לכך הוא שהפונקציה g היא אינטגרבילית.

הגודל C_ω הוא מספר מרוכב, וערכו מודד את העוצמה של הרכיב ω.

9.2.4 הקשר בין העוצמה לפונקצית המתאם המשותף

בסעיף זה נחשב את הערך המוחלט של העוצמה, c_ω. בחישוב זה, c^* הוא הצמוד של המספר המרוכב c.

$$\begin{aligned} |c_\omega|^2 &= c_\omega c_\omega^* \\ &= \left(\int_{-\infty}^{+\infty} g(t) e^{-2\pi i \omega t} dt \right) \left(\int_{-\infty}^{+\infty} g(s) e^{2\pi i \omega s} ds \right). \end{aligned}$$

נבצע כעת חילופי משתנה: $u = t - s$:

$$|c_\omega|^2 = \int_{u=-\infty}^{+\infty} \left(\int_{t=-\infty}^{+\infty} g(t) g(t-u) dt \right) e^{-2\pi i \omega u} du.$$

האינטגרל הפנימי $\int_{t=-\infty}^{+\infty} g(t)g(t-u)dt$ הוא, עד כדי ההפחתת התוחלת, מקדם המתאם המשותף בפער u של g. נוסחא זו היא המוטיבציה להגדרת הספקטרום של סדרה עתית שנראה בהמשך.

9.3 הצגת פורייה עבור סדרה עתית

בסעיף זה נקשר בין התאוריה של ניתוח פרוייה בזמן רציף, ובין ניתוח פורייה לסדרות עתיות.

כאשר אנו מנתחים סדרה עתית, אנו מניחים כי ברקע ישנו תהליך סטוכסטי $Y(t)$, ואנו צופים במספר סופי של דגימות מריאליזציה אחת של התהליך. ניתן להסתכל על המצב בדרך הבאה. תחילה אלוהים בוחר ריאליזציה אחת של התהליך, כלומר לכל t את הערך של התהליך בזמן t, $y(t)$. לאחר מכן, הסטטיסטיקאי בוחר מספר סופי של זמנים $(t_n)_{n=1}^N$ ודוגם את הפונקציה y בזמנים אלו. כלומר, הסטטיסטיקאי מקבל את התצפיות $y(t_1), \ldots, y(t_N)$.

מטרת הסטטיסטיקאי היא ללמוד תכונות של התהליך $Y(t)$. מכיוון שאלוהים בוחר מראש רק ריאליזציה אחת של התהליך, $y(t)$, הדבר הטוב ביותר שהסטטיסטיקאי יכול לבקש הוא ללמוד תכונות של הפונקציה $y(t)$, ולקוות שהפונקציה $y(t)$ מייצגת היטב את התהליך $Y(t)$, כלומר, שכל התכונות של התהליך ניתנות לזיהוי בעזרת הריאליזציה, ורק הן.

לו הסטטיסטיקאי היה צופה בכל הפונקציה $y(t)$, הוא היה כול לפרקה לרכיביה המחזוריים (ראה משוואה (וסץ זןעיוע5)). לצערו של הסטטיסטיקאי, הוא מקבל רק N תצפיות.

לרוב התצפיות נלקחות בהפרשי זמן קבועים, כלומר $\Delta t = t_{n+1} - t_n$ אינו תלוי ב-n. כפי שנראה בהמשך, דבר זה מגביל את יכולתו של הסטטיסטיקאי לפרק את הסדרה לרכיבים מחזוריים. למרות קושי זה, מסתבר כי ניתן לפרק את הסדרה העתית לרכיבים מחזוריים. בשארית הפרק נראה כיצד ניתן לעשות זאת.

המשפט הבא הוא הצגת פורייה של סדרה עתית.

משפט 9.2 תהא $y_1,\dots,y_N$ סדרה כלשהי של מספרים ממשיים באורך זוגי. נסמן

$$\begin{aligned}
a_0 &= \frac{1}{N}\sum_{n=1}^{N} y_n, \\
a_k &= \frac{2}{N}\sum_{n=1}^{N} y_n \cos(\frac{2\pi kn}{N}), \qquad k=1,2,\dots,\frac{N}{2}, \\
b_k &= \frac{2}{N}\sum_{n=1}^{N} y_n \sin(\frac{2\pi kn}{N}), \qquad k=1,2,\dots,\frac{N}{2}.
\end{aligned}$$

אזי

$$y_n = \sum_{k=0}^{N/2}\left(a_k\cos(\frac{2\pi kn}{N}) + b_k\sin(\frac{2\pi kn}{N})\right), \tag{65}$$

נשים לב לדמיון שבין ההצגה (65) וההצגה (62) מעמוד 76. השוני בין ההצגות הוא שבהצגה (65) הכפלנו את המקדמים ב-2, אך אנו מסכמים רק על $k=0,\dots,N/2$, במקום על $k=0,\dots,N$. הסיבה לכך היא הסיבה הבאה. מכיוון ש-$\cos(\frac{2\pi(N-k)n}{N}) = \cos(\frac{2\pi kn}{N})$, נובע כי $a_{N-k}=a_k$, ולכן

$$a_{N-k}\cos(\frac{2\pi(N-k)n}{N}) = a_k\cos(\frac{2\pi kn}{N}).$$

לכן, במשוואה (65) תרומת הקוסינוס לאיבר ה-$N-k$ בסכום שווה לתרומת הקוסינוס לאיבר ה-k, ולכן מספיק לסכם על $k=0,\dots,N/2$ אם דואגים להכפיל את a_k ב-2. באופן דומה, מכיוון ש-$\sin(\frac{2\pi(N-k)n}{N}) = -\sin(\frac{2\pi kn}{N})$, נובע כי $b_{N-k}=-b_k$, ולכן

$$b_{N-k}\sin(\frac{2\pi(N-k)n}{N}) = b_k\sin(\frac{2\pi kn}{N}).$$

לכן תרומת הסינוס לאיבר ה-$N-k$ שווה לתרומת הסינוס באיבר ה-k.

הוכחה. נסמן

$$\begin{aligned}
A &= \sum_{k=0}^{N/2}\left(a_k\cos(\frac{2\pi kn}{N}) + b_k\sin(\frac{2\pi kn}{N})\right) \\
&= \sum_{k=0}^{N/2}(\frac{1}{N}\sum_{t=1}^{N} y_t\cos(\frac{\omega_0}{N})\cos(\frac{\omega_k n}{N}) \\
&\quad +\frac{1}{N}\sum_{t=1}^{N} y_t\cos(\frac{\omega N/2}{N})\cos(\frac{\omega_k n}{N}) \\
&\quad +\frac{2}{N}\sum_{t=1}^{N} y_t\cos(\frac{\omega_k t}{N})\cos(\frac{\omega_k n}{N}) \\
&\quad +\frac{2}{N}\sum_{t=1}^{N} y_t\sin(\frac{\omega_k t}{N})\sin(\frac{\omega_k n}{N}).
\end{aligned}$$

נשנה את סדר הסכימה ונקבל כי

$$\begin{aligned}A &= \frac{1}{N}\sum_{t=1}^{N} y_t \left(\sum_{k=0}^{N/2}\cos(\frac{\omega_k n}{N})\left(\cos(\frac{\omega_0}{N}) + \cos(\frac{\omega_{N/2}}{N}) + \sum_{k=1}^{N/2-1} 2\cos(\frac{\omega_k t}{N})\right)\right. \\ &\quad \left. + \sum_{k=1}^{N/2} 2\sin(\frac{\omega_k n}{N})\sin(\frac{\omega_k t}{N})\right).\end{aligned}$$

ממשוואות (51) ו-(52) אנו מקבלים כי

$$\cos\omega\cos\lambda + \sin\omega\sin\lambda = \cos(\omega-\lambda).$$

לכן נקבל כי

$$A = \frac{1}{N}\sum_{t=1}^{N}\sum_{k=0}^{N-1}\cos(\frac{\omega_k(n-t)}{N}).$$

מכיוון שהסכום הפנימי שווה ל-N אם $t = n$ ול-0 אחרת, נקבל כי גודל זה שווה ל-y_n, כנדרש. ■

משוואה (65) מציגה את הסדרה העתית כסכום של מחזורים בתדירויות $(\omega_k)_{k=0,\ldots,N/2}$. עולות מכאן מספר שאלות.

1. מדוע דווקא תדירויות אלו, ולא תדירויות אחרות?
2. האם ניתן למצוא, בנוסף למחזורים אלו, גם מחזורים נוספים?

נתחיל מהשאלה השנייה. במשוואה (65) ישנם N משוואות לינאריות, N נתונים (אברי הסדרה העתית $y_1,\ldots,y_N$) ו-N נעלמים (המקדמים של הסינוסים והקוסינוסים). מערכת משוואות זו היא הפיכה, ולכן בהינתן N המקדמים של הסינוסים והקוסינוסים ניתן לחשב את הסדרה העתית שייצרה אותם.

בפרט, לאחר שחישבנו את המקדמים של הסינוסים והקוסינוסים, כלומר, לאחר שווידאנו את העוצמה של כל תדירות $(\omega_k)_{k=0,\ldots,N/2}$, נקבעת הסדרה העתית, ולכן העוצמה של כל צדירות אחרת שנחשב נקבעת אף היא בצורה חד-ערכית.

מכאן שאין לחשב את העוצמה של יותר מ-N תדרים, שכן כל תדר נוסף שנחשב לא יגלה לנו אינפורמציה נוספת על הסדרה, אלא יהיה תלוי בתדרים האחרים שחישבנו.

בנוגע לשאלה הראשונה, אין הכרח לבחור דווקא את התדרים $(\omega_k)_{k=0,\ldots,N/2}$. ניתן לפרק את הסדרה העתית לסכום של כל N תדרים שונים שנבחר. היתרון בתדרים $(\omega_k)_{k=0,\ldots,N/2}$ הוא שעבורם קל לחשב את המקדמים (בעזרת אלגוריתם הנקרא "טרנספורם פורייה המהיר" (fast Fourier transform).

9.4 הפריודוגרם

נשים לב כי

$$a_0 = \frac{1}{N}\sum_{n=1}^{N} y_n = \widehat{\mu}$$

הוא ממוצע ערכי הסדרה, ו-

$$a_{N/2} = \frac{1}{N}\sum_{n=1}^{N}(-1)^n y_n.$$

על ידי שימוש בזהויות (50)-(52) מעמוד 74 ניתן להציג את הסדרה $\{y_n\}$ בדרך שקולה.
עבור $k \neq 0$,

$$a_k \cos(\frac{2\pi kn}{N}) + b_k \sin(\frac{2\pi kn}{N}) = a_k \frac{e^{2\pi ikn/N} + e^{-2\pi ikn/N}}{2} + b_k \frac{e^{2\pi ikn/N} - e^{-2\pi ikn/N}}{2i}$$
$$= \frac{a_k - ib_k}{2} e^{2\pi ikn/N} + \frac{a_k + ib_k}{2} e^{-2\pi ikn/N}.$$

לכן, אם נסמן

$$c_0 = a_0, c_{N/2} = a_{N/2}$$

$$c_k = \frac{a_k - ib_k}{2}, \quad 1 \leq k \leq \frac{N}{2}.$$

$$c_{-k} = \frac{a_k + ib_k}{2}, \quad 1 \leq k \leq \frac{N}{2}.$$

נקבל

$$y_n = \sum_{k=-N/2+1}^{N/2} c_k e^{i\omega_k n}, \tag{66}$$

מכיון שהתדר ω_k מופיע פעמיים בנוסחא (66), באיברים $\pm k$, אנו מגדירים את העוצמה של תדר זה באופן הבא.

הגדרה 9.3 **העוצמה** של התדר $\omega_k = \frac{2\pi k}{N}$ נתונה על ידי

$$f_k = \begin{cases} |c_k|^2 & k = 0, \frac{N}{2}, \\ |c_{-k}|^2 + |c_k|^2 = 2|c_k|^2 & 0 < k < \frac{N}{2}. \end{cases}$$

לכן מקבלים

$$f_k = \begin{cases} a_k^2 & k = 0, \frac{N}{2} \\ \frac{1}{2}(a_k^2 + b_k^2) & 0 < k < \frac{N}{2}. \end{cases}$$

לרוב נרצה לחשב את העוצמה של כל התדרים $\{\omega_k\}$. דרך מהירה לחישוב זה היא באמצעות טרנספורם פורייה המהיר (שלא נלמד כאן).

הגרף של הפונקציה $k \mapsto f_k$ המתאימה לכל תדר את עוצמתו נקרא הפריודוגרם של הסדרה העתית. גרף זה מאפשר לנו לראות אם יש תדרים חזקים בסדרה.

הגדרה 9.4 **הפריודוגרם** של סדרה עתית $\{y_t\}$ הוא הפונקציה I המתאימה לזווית ω_k את הגודל הבא:

$$I(\omega_k) = N \times f_k = \begin{cases} Na_0^2 & k = 0, \\ \frac{N}{2}(a_k^2 + b_k^2), & k = 1, 2, \ldots, \frac{N-1}{2}, \\ Na_{N/2}^2 & k = \frac{N}{2}. \end{cases}$$

9.5 פריודוגרם של תהליכים פשוטים

בסעיף זה נחשב את הפריודוגרם של סדרות הנוצרות ממספר תהליכים פשוטים.

9.5.1 תהליך קבוע

נניח כי $\{Y_n\}$ הוא תהליך דטרמיניסטי קבוע, כלומר, $Y_n = \mu$ לכל n, באשר $\mu \in \mathbf{R}$ קבוע. תהא $(y_1, y_2, \ldots, y_N)$ סדרה עתית באורך זוגי המיוצרת על ידי התהליך $\{Y_n\}$.

נחשב כעת את מקדמי פורייה (a_k, b_k).

$$a_0 = \frac{1}{N}\sum_{n=1}^{N} \mu \cos 0 = \mu.$$

נשתמש במשוואה (57) מעמוד 75 כדי לקבל לכל $k = 1, 2, \ldots, N/2$

$$a_k = \frac{2}{N}\sum_{n=1}^{N} \mu \cos(\omega_k n) = \frac{2\mu}{N}\sum_{n=1}^{N} \cos(\omega_k n) = 0.$$

נשתמש במשוואה (58) מעמוד 75 כדי לקבל לכל $k = 1, 2, \ldots, N/2$

$$b_k = \frac{2}{N}\sum_{n=1}^{N} \mu \sin(\omega_k n) = \frac{2\mu}{N}\sum_{n=1}^{N/2} \sin(\omega_k n) = 0.$$

קיבלנו כי הפריודוגרם הוא

$$I(\omega_k) = \begin{cases} N\mu^2 & k = 0 \\ 0 & k \neq 0. \end{cases}$$

9.5.2 הפריודוגרם של תהליך רעש אקראי

נניח כי $(x_1, x_2, \ldots, x_N)$ היא סדרה באורך זוגי שיוצרה מתהליך רעש אקראי $\{X_n\}$ עם תוחלת 0 ושונות σ_X^2.

במקרה זה המקדמים (a_k, b_k) הם משתנים מקריים, התלויים בסדרה הספציפית שקיבלנו. מכיוון ש-

$$a_0 = \frac{1}{N}\sum_{n=1}^{N} x_n = \overline{x}$$

הוא הממוצע של הסדרה, הרי ש-$\mathbf{E}[a_0] = 0$ ו-$\mathrm{Var}(a_0) = \mathrm{Var}\left(\frac{1}{N}\sum_{n=1}^{N} X_n\right) = \frac{\sigma_X^2}{N^2}$. מאי-שוויון צ'ביצ'ב נוכל לחסום את ההסתברות ש-a_0 יהיה רחוק מ-0:

$$\mathbf{P}(|a_0| > c) \leq \frac{\sigma_X^2}{Nc^2}.$$

באופן דומה, מכיוון ש-$\mathbf{E}[X_n] = 0$, נקבל לכל k

$$\mathbf{E}[a_k] = \frac{2}{N}\sum_{n=1}^{N} \mathbf{E}[X_n]\cos(n\omega_k) = 0,$$

$$\mathbf{E}[b_k] = \frac{2}{N}\sum_{n=1}^{N}\mathbf{E}[X_n]\sin(n\omega_k) = 0.$$

יתר על כן

$$\begin{aligned}
\mathrm{Var}(a_k) &= \mathrm{Var}\left(\frac{2}{N}\sum_{n=1}^{N}X_n\cos(n\omega_k)\right) \\
&= \frac{4}{N^2}\sum_{n=1}^{N}\cos^2(n\omega_k)\mathrm{Var}(X_n) \\
&\le \frac{4}{N^2}\sum_{n=1}^{N}\mathrm{Var}(X_n) \\
&= \frac{4}{N}\sigma_X^2.
\end{aligned}$$

באופן דומה ניתן לקבל אותו החסם עבור $\mathrm{Var}(b_k)$. ניתן להראות (ראה תרגיל בנושא ניתוח פורייה) כי השונות שווה ל-$2\sigma_X^2/N$.

קיבלנו כי בהסתברות גבוהה עוצמת כל תדר של תהליך רעש אקראי תהיה קטנה.

9.5.3 הפריודוגרם של גל סינוס

כעת נניח כי נתונה לנו סדרה $(y_1, y_2, \ldots, y_N)$ באורך זוגי שיוצרה באמצעות גל סינוס:

$$y_n = \sin(n\lambda),$$

עבור זווית $\lambda \in (0, 2\pi)$ כלשהי. אנו מניחים כי $\lambda \neq 0$, מכיוון שאם $\lambda = 0$ מקבלים $y_n = 0$ לכל n. נחשב את הפריודוגרם של סדרה זו.

הערה 9.5 המקרה בו הסדרה מיוצרת באמצעות גל קוסינוס הוא אנלוגי לחלוטין למקרה בו הסדרה מיוצרת באמצעות גל סינוס, ולכן לא נדון בו.

תחילה נניח כי $\lambda = \omega_j = \frac{2\pi j}{N}$ עבור $j = 1, \ldots, N/2$ כל שהוא. משוואות (59)-(61) מעמוד 76 גוררות כי

$$\begin{aligned}
a_0 &= 0, \\
a_k &= 0 \qquad k = 1, \ldots, N/2, \\
b_k &= 0 \qquad k \neq j, \\
b_j &= 1,
\end{aligned}$$

כלומר, במקרה זה הפריודוגרם מזהה לחלוטין את התדר של גל הסינוס ממנה מיוצרת הסדרה.

כעת נראה מה קורה כאשר λ שונה מ-$\omega_1, \ldots, \omega_{N/2}$.

על ידי שימוש במשוואה (56) מעמוד 75,

$$a_0 = \frac{1}{N}\sum_{n=1}^{N}\sin\lambda n = \frac{1}{N}\sin(\lambda(N+1)/2)\frac{\sin(\lambda N/2)}{\sin(\lambda/2)} \le \frac{1}{N\sin(\lambda/2)}. \tag{67}$$

מכיוון ש-$\lambda \neq 0$, ככל שגדל אורך הסדרה a_0 קטן ל-0.

$$\begin{aligned} a_{N/2} &= \frac{1}{N}\sum_{n=1}^{N}\cos(\pi n)\sin(\lambda n) \\ &= \frac{1}{2N}\sum_{n=1}^{N}\left(\sin((\lambda+\pi)n)+\cos((\lambda-\pi)n)\right). \end{aligned}$$

מכיוון ש-λ שונה מ-$\pm\pi$, כמו במשוואה (67) שני הסכומים הללו שואפים ל-0 כאשר N גדל.

$$\begin{aligned} a_k &= \frac{2}{N}\sum_{n=1}^{N}\sin\lambda n\cos\omega_k n \\ &= \frac{2}{N}\sum_{n=1}^{N}\left(\sin((\lambda+\omega_k)n)+\sin((\lambda-\omega_k)n)\right). \end{aligned}$$

מכיוון ש-λ שונה מ-$\pm\omega_k$, כמו במשוואה (67), שני הסכומים הללו שואפים ל-0 כאשר N גדל.

$$\begin{aligned} b_k &= \frac{2}{N}\sum_{n=1}^{N}\sin\lambda n\sin\omega_k n \\ &= \frac{2}{N}\sum_{n=1}^{N}\left(\cos((\lambda-\omega_k)n)-\cos((\lambda+\omega_k)n)\right). \end{aligned}$$

מכיוון ש-λ שונה מ-$\pm\omega_k$, כמו במשוואה (67), שני הסכומים הללו שואפים ל-0 כאשר N גדל. נשים לב שאם λ קרוב ל-$\pm\omega_k$, מקדם פוריה b_k יהיה רחוק מאפס.

ראינו מה קורה לכל מקדם פורייה מסוים כאשר אורך הסדרה גדל. אבל מה קורה עבור אורך סדרה קבוע? הניתוח שעשינו כאן מראה שככל ש-λ רחוק יותר מ-ω_k, כך גם קטנים מקדמי פורייה a_k ו-b_k.

לכן, הפריודוגרם יקבל ערכים גבוהים ככל ש-ω_k "קרוב יותר" ל-λ, וערכים נמוכים "רחוק" מ-λ. המושגים קרוב ורחוק במקרה זה תלויים באורך הסדרה. ככל שהסדרה ארוכה יותר, צריך ש-λ תהיה קרובה יותר ל-ω_k כדי שמקדם פורייה המתאים יהיה גבוה.

9.5.4 סכום של סדרות

נניח כי הסדרה הנצפית $(z_1,\ldots,z_N)$ היא סכום של שתי סדרות $(x_1,\ldots,x_N)$ ו-$(y_1,\ldots,y_2)$. כלומר, לכל t מתקיים

$$z_t = x_t + y_t.$$

נסמן ב-(a_k^x, b_k^x) את מקדמי פורייה של הסדרה $\{x_t\}$, ב-(a_k^y, b_k^y) את מקדמי פורייה של הסדרה $\{y_t\}$, וב-(a_k^z, b_k^z) את מקדמי פורייה של הסדרה $\{z_t\}$, אזי

$$\begin{aligned} a_0^z &= \frac{1}{N}\sum_{t=1}^{N} z_t \\ &= \frac{1}{N}\sum_{t=1}^{N}(x_t + y_t) \\ &= a_0^x + a_0^y. \end{aligned}$$

באופן דומה לכל $k = 1, \dots, N/2 - 1$

$$\begin{aligned} a_k^z &= \frac{2}{N}\sum_{t=1}^{N} z_t \cos\omega_k t \\ &= \frac{2}{N}\sum_{t=1}^{N}(x_t + y_t)\cos\omega_k t \\ &= a_k^x + a_k^y. \end{aligned}$$

באותו אופן נקבל כי $a_{N/2}^z = a_{N/2}^x + a_{N/2}^y$, ו-$b_k^z = b_k^x + b_k^y$ לכל $k = 1, \dots, N/2$.

כלומר, קיבלנו כי מקדמי פורייה של הסדרה z, סדרת הסכום, הם הסכום של מקדמי פורייה של הסדרה x ושל הסדרה y.

9.5.5 מסקנות

מהחישובים שעשינו עד כה ניתן להסיק כיצד יראה הפריודוגרם של תהליכים מורכבים.
לדוגמא, אם הסדרה $(y_1, \dots, y_N)$ מיוצרת על ידי התהליך

$$Y_t = \mu + A\cos(\omega t) + B\sin(\lambda t) + X_t,$$

באשר μ, A, B הם קבועים, ו-$\{X_t\}$ הוא תהליך רעש טהור עם תוחלת 0 ושונות σ_X^2, אנו יודעים כיצד יראה הפריודוגרם:

- הגורם הקבוע μ יתרום μ ל-a_0, ולא יתרום דבר לשאר מקדמי פורייה.
- התרומה של הגורם $A\cos(\omega t)$ תהיה בעיקר למקדמים a_k המקיימים ש-ω_k קרוב ל-ω. ככל ש-ω_k מתרחק מ-ω התרומה תקטן. התרומה לגורמים b_k תהיה קטנה.
- התרומה של הגורם $B\sin(\omega t)$ תהיה בעיקר למקדמים b_k המקיימים ש-ω_k קרוב ל-λ. ככל ש-ω_k מתרחק מ-λ התרומה תקטן. התרומה לגורמים a_k תהיה קטנה.
- הרעש האקראי $\{X_t\}$ לא יתרום רבות לאף מקדם פורייה ספציפי.

מכאן אנו רואים כי אם התהליך שלנו הוא סכום של מספר גלי סינוס, קוסינוס ורעש אקראי, הפריודוגרם יגלה לנו זאת.

9.6 מגבלות הטכניקה

9.6.1 הרמוניות

אם יש רכיב מחזורי עם מחזור t, אזי יש מחזור של $2t$, $3t$ וכו'. המחזורים kt עבור $k \geq 2$ נקראים הרנמוניות של המחזור הבסיסי t. הדבר יתבטא בפריודוגרם על ידי קיום של גלים. מכאן, שכאשר קיימים בפריודוגרם מספר שיאים המופיעים במרחקים שווים האחד מהשני, יש להתייחס אך ורק לשיא הראשון, שאמור להיות חזק יותר מהאחרים.

9.6.2 מחזורים בתדרים השונים מ-$(\omega_k)_{k=0}^{N/2}$

ניתוח פורייה מאפשר לנו לזהות מחזורים בתדרים $(\omega_k)_{k=0}^{N/2}$. אם לסדרה יש רכיב מחזורי בתדר λ, באשר λ קרוב לאחד ה-ω_k (כאשר λ קרוב ל-ω_k אם ורק אם $N \times |\lambda - \omega_k|$ קטן) המקדם של ω_k בפיתוח פורייה יהיה גדול, ולכן נוכל לזהות מחזור זה.

אך אם הסדרה העתית מכילה רכיב מחזורי בתדר λ, ו-λ אינו קרוב לאף אחד מה-ω_k המקדמים בפיתוח פורייה לא יגלו לנו את קיום רכיב זה.

כדי להתגבר על בעייה זו ניתן לשנות את קצב הדגימה. קרי, אם הסדרה המקורית נוצרה על ידי דגימה של הטמפרטורה בתל אביב כל 53 דקות, מכיוון ש-53 זר ל-60, כדי לזהות מחזוריות זו יש לבצע $60 \times 24 = 1,440$ מדידות. אם נשנה את קצב הדגימה (לדגימה אחת כל 60 דקות, או 30 דקות) נוכל להסתפק בפחות דגימות כדי לזהות את המחזוריות של 24 שעות.

מסקנה אופרטיבית: כדי שניתוח פורייה יהיה משמעותי, יש להתאים את קצב הדגימה חמחזורים שמצפים למצוא בסדרה.

9.6.3 מחזורים קצרים: זיוף תדרים (aliasing)

נניח כי הסדרה העתית נוצרה על ידי דגימה תהליך סטוכסטי במרווחי זמן של Δt דקות.

ברור כי בכך מאבדים אינפורמציה על חלק מהמחזורים. לדוגמא, אם לסדרה יש רכיב מחזורי עם מחזור Δt, לא נוכל לראותו כלל. אם, לעומת זאת, לסדרה יש רכיב מחזורי עם מחזור $\frac{2}{3}\Delta t$, רכיב זה משלים שלושה מחזורים כל 2 תצפיות, ולכן ניתוח פורייה יצביע על קיום רכיב מחזורי עם מחזור $2\Delta t$.

יתר על כן, אם לסדרה יש שני רכיבים מחזוריים, האחד עם מחזור $2\Delta t$ והשני עם מחזור $\frac{2}{3}\Delta t$, ניתוח פורייה יצביע על קיום רכיב מחזורי עם מחזור $2\Delta t$, ועם עוצמה גדולה יותר מהעוצמה האמיתית של מחזור זה.

באותו אופן, גם רכיבים מחזוריים עם מחזור של $\frac{2}{5}\Delta t$, $\frac{2}{7}\Delta t$, וכו' יצורפו על ידי ניתוח פורייה לרכיב המחזורי עם מחזור $2\Delta t$.

באופן כללי, מכיוון שאנו מחשבים את מקדמי פורייה עבור הזוויות $(\omega_k)_{k=0}^{N/2}$, ניתוח פורייה מזהה רכיבים מחזורים עם מחזור $\Delta t, \ldots, \frac{N}{2}\Delta t$. לכל מחזור מזוהה $m\Delta t$ (עם $1 \leq m \leq \frac{N}{2}$), ולכל l הזר ל-m, העוצמה של המחזור $\frac{m}{k}\Delta t$ מצורפת על ידי ניתוח פורייה לעוצמת המחזור $m\Delta t$.

דרך אחת שנוסתה להתגבר על בעיית הזיוף היא בחירת זמני דגימה מקריים. כלומר, אם נסמן ב-t_n את הזמן בו נלקחת הדגימה n, אזי נבחר את זמני הדגימה כך ש-$(t_n - t_{n-1})$ יהיו בלתי תלויים ושווי התפלגות. אם ההתפלגות של $(t_n - t_{n-1})$ מקיימת מספר תכונות, אזי בעיית הזיוף לא עולה.

9.6.4 מחזורים ארוכים: אי-יכולת זיהוי

מכיוון שהסדרה העתית שבידינו היא סדרה באורך סופי N, אם לסדרה יש רכיב מחזורי עם מחזור ארוך מ-N לא נוכל בשום אופן לזהות מחזור זה.

אם יש מחזור באורך $\frac{N}{2} < m < N$, ואם $N-m$ גדול דיו כדי לאפשר הצלחת מבחנים סטטיסטיים, ניתן לזהות קיום רכיבים מחזוריים עם מחזור m, אך לא באמצעות ניתוח פורייה.

9.6.5 קו מגמה (trend)

אם הסדרה העתית מכילה קו מגמה, הדבר ישפיע על ניתוח פורייה בשני אופנים. ראשית, המקדם a_0 יהיה גבוה. שנית, כל שאר המקדמים יתגמדו.

מסקנה אופרטיבית: כדי שניתוח פורייה יהיה משמעותי, יש לנקות מבסדרה העתית את קו המגמה.

9.6.6 עונתיות (seasonallity)

אם יש בסדרה העתית רכיב עונתי חזק, המקדם המתאים בפיתוח פורייה יהיה חזק, אך כל שאר המקדמים יתגמדו.

מסקנה אופרטיבית: כדי שניתוח פורייה יוכל לזהות מחזורים מעבר לרכיב העונתי, יש לנקות את הרכיב העונתי מהסדרה.

9.7 הפריודוגרם ופונקצית המתאם המשותף

בחישוב הבא, c^* הוא הצמוד של המספר המרוכב c.

נזכור כי לכל $l = 1, 2, \ldots, \frac{N}{2} - 1$ מתקיים

$$\sum_{n=1}^{N} \cos(\omega_l n) = 0.$$

מכאן נובע כי

$$\sum_{n=1}^{N} \overline{y} \cos(\omega_l n) = 0,$$

ובאופן דומה

$$\sum_{n=1}^{N} \overline{y} \sin(\omega_l n) = 0.$$

משתי משוואות אלה נקבל כי

$$\sum_{n=1}^{N} \overline{y} e^{i\omega_l n} = 0.$$

מכיוון ש- $\widehat{\gamma}_k = \frac{1}{N}\sum_{t=1}^{N-k}(y_t-\overline{y})(y_{t+k}-\overline{y})$ (ראה סעיף 2.4 בעמוד 16) נקבל שלכל $l = 1, 2, \ldots, \frac{N}{2}-1$,

$$\begin{aligned}
|c_l|^2 &= c_l c_l^* \\
&= \frac{1}{N^2}\left(\sum_{t=1}^{N} y_t e^{-i\omega_l t}\right)\left(\sum_{u=1}^{N} y_u e^{i\omega_l u}\right) \\
&= \frac{1}{N^2}\left(\sum_{t=1}^{N} (y_t-\overline{y}) e^{-i\omega_l t}\right)\left(\sum_{u=1}^{N} (y_u-\overline{y}) e^{i\omega_l u}\right) \\
&= \frac{1}{N^2}\sum_{k=0}^{N}\sum_{t=1}^{N-k}(y_t-\overline{y})(y_{t+k}-\overline{y})(e^{-i\omega_l k}+e^{i\omega_l k}) \qquad (68) \\
&= \frac{1}{N}\widehat{\gamma}_0 + \frac{2}{N}\sum_{k=1}^{N}\widehat{\gamma}_k\cos(\omega_l k).
\end{aligned}$$

נשים לב עוד כי

$$\begin{aligned}
|c_l|^2 &= \frac{a_l - ib_l}{2}\times\frac{a_l+ib_l}{2} \\
&= \frac{a_l^2+b_l^2}{4} \\
&= \frac{I(\omega_l)}{2N}.
\end{aligned}$$

לכן ניתן להסיק כי

$$I(\omega_l) = 2\widehat{\gamma}_0 + 4\sum_{k=1}^{N}\widehat{\gamma}_k\cos(\omega_l k) = 2\sum_{k=-N}^{N}\widehat{\gamma}_k\cos(\omega_l k).$$

המוטיבציה להגדרה זו באה מהפיתוח (68).

הגדרה 9.6 הספקטרום של תהליך סטציונרי $\{Y_t\}$ עם פונקציית מתאם משותף (γ_k) היא הפונקציה $f : [-\pi,\pi] \to \mathbf{R}$ הנתונה על ידי

$$f(\omega) = \frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k e^{-i\omega k}.$$

את נוסחת הספקטרום נכתוב באופן מקוצר בצורה הבאה:

$$f(\omega) = \frac{1}{2\pi}\gamma e^{-i\omega}.$$

נשים לב כי עד כדי קבוע, הספקטרום $f(\omega_l)$ שווה ל-c_l, עוצמת התדר ω_l. לכן, ניתן היה לחשוב כי כאשר אורך הסדרה גדל, מקדמי ניתוח פורייה מקרבים את הספקטרום. למרות ש-

$$\lim_{N\to\infty} I(\omega) = f(\omega),$$

הרי שהשונות $\mathrm{Var}(I(\omega))$ אינה שואפת לאפס כאשר N שואף לאינסוף, ולכן $I(\omega)$ אינו אומד עקבי של $f(\omega)$ (consistent estimator).

טענה 9.7 $I(\omega)$ אינו אומד עקבי של $f(\omega)$.

הוכחה. נוכיח הטענה במקרה הפשוט ביותר, בו $\{Y_t\}$ היא סדרה אקראית, בה Y_t מתפלג נורמלית עם תוחלת 0 ושונות 1. מכיוון ש-a_k ו-b_k הם צרופים לינאריים של $\{Y_t\}$, נובע כי a_k ו-b_k מתפלגים נורמלית. ניתן להראות כי התוחלת של משתנים מקריים אלו היא 0 ושונותם $\frac{2}{N}$ (אם $n \neq \frac{N}{2}$). יתר על כן, מכיוון ש-$\{Y_t\}$ בלתי תלויים,

$$\begin{aligned}
\mathrm{Cov}(a_k, b_k) &= \frac{4}{N^2}\mathrm{Cov}\left((\sum_{n=1}^{N} Y_n \cos(\omega_l n))(\sum_{n=1}^{N} Y_n \sin(\omega_l n)) \right) \\
&= \frac{4}{N^2}\sum_{n=1}^{N}(\cos(\omega_l n)\sin(\omega_l n)) \\
&= \frac{2}{N^2}\left(\sum_{n=1}^{N}\sin(2\omega_l n)\right) = 0.
\end{aligned}$$

מכאן מקבלים כי (a_k, b_k) הם משתנים מקריים בלתי מתואמים עם תוחלת אפס. מכיוון שכל זוג משתנים מקריים בלתי מתואמים המתפלגים נורמלית הם בלתי תלויים, נובע כי (a_k, b_k) בלתי תלויים.

אם (Z_1, Z_2) הם משתנים מקריים נורמלים בלתי תלויים עם תוחלת 0 ושונות 1, אזי $Z_1 + Z_2$ מתפלג χ^2 עם שתי דרגות חופש. לכן המשתנה המקרי

$$\frac{N(a_k^2 + b_k^2)}{2} = I(\omega_l)2\pi$$

מתפלג χ^2 עם שתי דרגות חופש. השונות של משתנה מקרי המתפלג χ^2 עם m דרגות חופש היא $2m$, ולכן

$$\mathrm{Var}(I(\omega_l)2\pi) = 4,$$

או

$$\mathrm{Var}(I(\omega_l)) = \frac{1}{\pi^2}.$$

מכיוון שהשונות אינה שואפת ל-0 כאשר N גדל, האומד אינו עקבי. ■

9.8 אומדים לספקטרום

מכיוון שהפריודוגרם אינו אומד עקבי של הספקטרום, ניסו אנשים רבים לספק אומדים עקביים לספקטרום. למספר רב של אומדים יש הצורה הכללית הבאה:

$$\widehat{f}(\omega) = \frac{1}{\pi}\left(\lambda_0\widehat{\gamma}_0 + 2\sum_{k=1}^{M}\lambda_k\widehat{\gamma}_k\cos(\omega k)\right).$$

בתאור זה, (λ_k) הוא אוסף של משקלות הנקראים **חלון פיגור** (lag window), ו-M נקרא **נקודת קטימה** (truncation point). מתוך השוואה של אומד זה עם משוואה (68) אנו רואים כי כאן לא משתמשים במקדמי המתאם המשותף עבור פערים הגדולים מ-M, ומקדמי המתאם המשותף עם פער קטן מ-M מוכפלים בקבוע התלוי בפער.

כמובן שבחירה נכונה של המקדמים (λ_k) חשובה לשם קבלת אומד מוצלח. נציג כעת שתי בחירות אפשריות מקובלות.

9.8.1 חלון Tukey

חלון זה מוגדר על ידי המקדמים:

$$\lambda_k = \frac{1}{2}\left(1+\cos(\frac{\pi k}{M})\right), \qquad k=0,1,\ldots,M.$$

חלון זה ידוע גם בשמות Tukey-Hanning ו-Blackman-Tukey.

9.8.2 חלון Parzen

חלון זה מוגדר על ידי המקדמים:

$$\lambda_k = \begin{cases} 1-6\left(\frac{k}{M}\right)^2+6\left(\frac{k}{M}\right)^3 & 0\le k\le \frac{M}{2}, \\ 2\left(1-\frac{k}{M}\right)^2 & \frac{M}{2}\le k\le M. \end{cases}$$

האומדים המחושבים על ידי שני חלונות אלו דומים למדי. אין דרך מוסכמת להחליט על נקודת הקטימה M. ככל ש-M קטן יותר, השונות של $\widehat{f}(\omega)$ תהיה קטנה יותר, אך ההטיה שלו תהיה גדולה יותר. אם M יהיה קטן מדי, תכונות חשובות של $\widehat{f}(\omega)$ תוחלקנה, ולא ניתן יהיה לזהות אותן. כלל אצבע שימושי הוא לבחור את M להיות בערך $2\sqrt{N}$. חלון Parzen דורש M גבוה מעט יותר מאשר חלון Tukey.

Jenkins and Watts הציעו את דרך הניתוח הבאה:

1. להתחיל עם ערך נמוך עבור M, שיזהה היכן נמצא השיא של $f(\omega)$, אך יגרום לכך ש-$\widehat{f}(\omega)$ תוחלק.

2. להמשיך עם ערך גבוה עבור M, שיגרום לכך שיהיו ל-$f(\omega)$ שיאים רבים, חלקם מדומים (spurious).

3. לסיים עם ערך ביניים עבור M.

למרות שניתן לאמוד את $f(\omega)$ עבור כל ω, לרוב האמידה נעשית עבור ערכים $\omega=\frac{\pi k}{Q}$, באשר Q גדול דיו כדי שכל תכונותיה של $f(\omega)$ יזוהו. לרוב בוחרים $Q=M$.

9.8.3 החלקת הפריודוגרם

בסעיפים הקודמים ראינו כיצד לאמוד את הספקטרום על ידי שימוש בחלון מתאים. דרך אחרת, שהוצעה על ידי Daniell, היא לאמוד את $f(\omega)$ על ידי מיצוע מספר ערכים של הפריודוגרם:

$$\widehat{f}(\omega) = \frac{1}{2m+1}\sum_k I(\omega_k),$$

באשר $\omega_k=\frac{2\pi k}{N}$, ו-$k$ עובר על m שלמים עוקבים כך ש-ω_k סימטריים סביב ω. (כדי לאמוד את $f(\omega)$ עבור $\omega=0,\frac{N}{2}$, מסתכלים על הפריודוגרם באופן מחזורי).

מכיוון שערכים שכנים של הפריודוגרם הם בלתי תלויים באופן אסימפטוטי, נקבל כי השונות של $\widehat{f}(\omega)$ היא מסדר גודל של $\frac{1}{m}$. לעומת זאת, ניתן להראות כי

$$\mathbf{E}[\widehat{f}(\omega)] \approx \frac{1}{m}\sum_k f(\omega_k),$$

וגודל זה שווה ל-$f(\omega)$ רק אם הספקטרום לינארי. אם הספקטרום אינו לינארי, אך חלק מספיק, הטייה זו אינה משמעותית.

אין גם דרך מוסכמת לבחירת m: ככל ש-m גדול יותר השונות של $\widehat{f}(\omega)$ קטנה, אך ההטייה תגדל. אם m גדול מדי, תכונות מעניינות של $f(\omega)$, כמו, למשל, השיאים שלה, עלולים להיעלם. המלצה כללית היא לנסות עבור m מספר ערכים בסביבת $\frac{N}{40}$.

9.9 רווחי סמך עבור הספקרום

כאמור קודם, האומד הכללי לספקטרום הוא

$$\widehat{f}(\omega) = \frac{1}{\pi}\left(\lambda_0\widehat{\gamma}_0 + 2\sum_{k=1}^{M}\lambda_k\widehat{\gamma}_k\cos(\omega k)\right).$$

נסמן

$$\nu = \frac{2N}{\sum_{k=-M}^{M}\lambda_k^2}.$$

Jenkins and Watts הוכיח כי התפלגותו של המשתנה המקרי $\nu\frac{\widehat{f}(\omega)}{f(\omega)}$ היא באופן אסימפטוטי χ^2_ν: χ^2 עם ν דרגות חופש. לכן רווח סמך עבור הספקטוגרם הוא:

$$\mathbf{P}\left(\chi^2_{\nu,1-\alpha/2} < \nu\frac{\widehat{f}(\omega)}{f(\omega)} < \chi^2_{\nu,\alpha/2}\right) = 1-\alpha.$$

במילים אחרות, בהסתברות $1-\alpha$ הערך האמיתי של $f(\omega)$ נמצא בושח

$$\left[\frac{\nu\widehat{f}(\omega)}{\chi^2_{\nu,\alpha/2}}, \frac{\nu\widehat{f}(\omega)}{\chi^2_{\nu,1-\alpha/2}}\right].$$

רווח סמך זה הוא אסימפטוטי, כלומר, כאשר אורך הסדרה גדל רווח הסמך שואף לקטע זה. Neave הראה כי רווח סמך זה מדויק דיו גם עבור סדרות קצרות.

9.10 הספקטרום של תהליכים שכיחים

בסעיף זה נחשב בצורה תאורטית את עוצמת התדר של תהליכים שכיחים. נתחיל בהגדרת פונקצית יוצרת המתאם המשותף, שהיא שימושית מאוד לחישוב זה.

9.10.1 הפונקציה יוצרת המתאם המשותף (autocovariance generating function)

יהא $\{Y_t\}$ תהליך סטציונרי עם פונקצית מתאם משותף (γ_k). כלומר, $\gamma_0 = \mathrm{Var}(Y_t)$ היא השונות של התהליך, ולכל $k \neq 0$ מתקיים $\gamma_k = \mathbf{E}[Y_tY_{t+k}]$ היא השונות המשותפת בפער k.

הגדרה 9.8 הפונקציה יוצרת המתאם המשותף מוגדרת על ידי:

$$\gamma(B) = \sum_{k=-\infty}^{\infty}\gamma_k B^k.$$

שימושיות פונקציה זו נובעת מהעובדה ש-

$$f(w) = \frac{1}{2\pi}\gamma(e^{-i\omega}).$$

נחשב כעת את פונקצית יוצרת המתאם המשותף של תהליך ARMA. תהליך ARMA מקיים את המשוואה

$$Y_t = \sum_{j=0}^{\infty} \psi_j X_{t-j},$$

באשר $\psi_0 = 1$. יתר על כן, עבור תהליך זה

$$\gamma_k = \sigma_X^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}.$$

לכן

$$\begin{aligned} \gamma(B) &= \sigma_X^2 \sum_{k=-\infty}^{\infty} \sum_{j=0}^{\infty} \psi_j \psi_{j+k} B^k \\ &= \sigma_X^2 \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \psi_j \psi_i B^{j-i} \\ &= \sigma_X^2 \sum_{i=0}^{\infty} \psi_i B^i \sum_{j=0}^{\infty} \psi_j B^{-j} \qquad (69) \\ &= \sigma_X^2 \Psi(B)\Psi(B^{-1}). \end{aligned}$$

כאן הנחנו כי $\psi_j = 0$ עבור $j < 0$.

9.10.2 הספקטרום של תהליך ARMA

יהא $\{Y_t\}$ תהליך ARMA המקיים את המשוואה

$$\Phi(B)Y_t = \Theta(B)X_t.$$

אם מגדירים $\Psi(B) = \frac{\Theta(B)}{\Phi(B)}$ מקבלים כי

$$Y_t = \Psi(B)X_t = \sum_{j=0}^{\infty} \psi_j X_{t-j}.$$

כפי שראינו במשוואה (69), הפונקציה יוצרת המתאם המשותף נתונה על ידי:

$$\gamma(B) = \sigma_X^2 \Psi(B)\Psi(B^{-1}) = \sigma_X^2 \frac{\Theta(B)\Theta(B^{-1})}{\Phi(B)\Phi(B^{-1})}.$$

אם התהליך סטציונרי, השורשים של $\Phi(B)=0$ נמצאים מחוץ לעיגול היחידה, ולכן $\sum_{k=1}^{\infty}|\gamma_k|<+\infty$. במקרה זה הספקטרום נתון על ידי

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi}\gamma(e^{-i\omega}) \\ &= \frac{\sigma_X^2}{2\pi}\times\frac{\Theta(e^{-i\omega})\Theta(e^{i\omega})}{\Phi(e^{-i\omega})\Phi(e^{i\omega})} \\ &= \frac{\sigma_X^2}{2\pi}\times\left|\frac{\Theta(e^{-i\omega})}{\Phi(e^{-i\omega})}\right|^2. \end{aligned} \tag{70}$$

אם התהליך הפיך השורשים של $\Theta(B)=0$ נמצאים מחוץ לעיגול היחידה, ולכן המונה אינו מתאפס. לכן

$$\frac{1}{f(\omega)}=\frac{2\pi}{\sigma_X^2}\times\left|\frac{\Phi(e^{-i\omega})}{\Theta(e^{-i\omega})}\right|^2$$

וזהו הספקטרום של תהליך $\mathrm{ARMA}(q,p)$.

9.10.3 הספקטרום של תהליך רעש אקראי

פונקצית המתאם המשותף של תהליך רעש אקראי $\{X_t\}$ עם תוחלת μ ושונות σ_X^2 נתונה על ידי

$$\gamma_k=\begin{cases}\sigma_X^2 & k=0\\ 0 & k\neq 0.\end{cases}$$

הפונקציה יוצרת המתאם המשותף נתונה על ידי

$$\gamma(B)=\sigma_X^2.$$

לכן הספקטרום נתון על ידי

$$f(\omega)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}\gamma_k e^{-i\omega k}=\frac{\sigma_X^2}{2\pi}.$$

זוהי פונקציה קבועה, שאינה תלויה ב-ω.

9.10.4 הספקטרום של תהליך $\mathrm{AR}(1)$

יהא $\{Y_t\}$ תהליך $\mathrm{AR}(1)$ המקיים את המשוואה

$$(1-\phi B)Y_t=X_t,$$

באשר $\{X_t\}$ תהליך רעש אקראי. הפונקציה יוצרת המתאם המשותף נתונה על ידי

$$\gamma(B)=\sigma_X^2\frac{1}{(1-\phi B)(1-\phi B^{-1})}.$$

מכיוון ש-$\Phi(B)=1-\phi B$, ממשוואה (70) נקבל כי הספקטרום נתון על ידי

$$\begin{aligned} f(\omega) &= \frac{\sigma_X^2}{2\pi}\times\frac{1}{(1-\phi e^{-i\omega})(1-\phi e^{i\omega})} \\ &= \frac{\sigma_X^2}{2\pi}\times\frac{1}{1+\phi^2-2\phi\cos\omega}. \end{aligned}$$

הגרף של הספקטרום תלוי בסימן של ϕ (האם ϕ חיובי או שלילי).

9.10.5 הספקטרום של תהליך $\mathrm{MA}(1)$

יהא $\{Y_t\}$ תהליך $\mathrm{MA}(1)$ המקיים את המשוואה

$$Y_t = (1+\theta B)X_t,$$

באשר $\{X_t\}$ תהליך רעש אקראי. הפונקציה יוצרת המתאם המשותף נתונה על ידי

$$\gamma(B) = \sigma_X^2(1+\theta B)(1+\theta B^{-1}).$$

לכן הספקטרום נתון על ידי

$$\begin{aligned} f(\omega) &= \frac{\sigma_X^2}{2\pi}(1+\theta e^{-i\omega})(1+\theta e^{i\omega}) \\ &= \frac{\sigma_X^2}{2\pi}(1+\theta^2+2\theta\cos\omega). \end{aligned}$$

גם במקרה זה הגרף של הספקטרום תלוי בסימן של ϕ.

נזכור כי עבור תהליך $\mathrm{MA}(1)$

$$\rho_k = \begin{cases} 1 & k=0 \\ \frac{\theta}{1+\theta^2} & k=\pm 1 \\ 0 & \text{אחרת} \end{cases}$$

לכן, אם θ שלילי, איברים סמוכים הם אנטי-קורלטיביים, ובספקטרום נראה ערכים גבוהים לתדירויות גבוהות. אם θ חיובי איברים סמוכים בסדרה הם מתואמים, ולכן הסדרה תיטה להיות חלקה: בספקטרום תדירויות נמוכות תקבלנה ערכים גבוהים.

9.10.6 הספקטרום של סכום של שני תהליכים

נניח כי התהליך $\{Z_t\}$ הוא סכום של שני תהליכים סטציונרים בלתי-תלויים $\{X_t\}$ ו-$\{Y_t\}$:

$$Z_t = X_t + Y_t.$$

נסמן ב-$\gamma^X(B), \gamma^Y(B), \gamma^Z(B)$ את הפונקציות יוצרות המתאם המשותף של שלושת התהליכים. בגלל ש-$\{X_t\}$ ו-$\{Y_t\}$ בלתי-תלויים

$$\begin{aligned} \gamma^Z(B) &= \sum_{k=-\infty}^{\infty} \gamma_k^Z(B)B^k \\ &= \sum_{k=-\infty}^{\infty} \mathrm{Cov}(Z_t, Z_{t+k})B^k \\ &= \sum_{k=-\infty}^{\infty} \mathrm{Cov}(X_t+Y_t, X_{t+k}+Y_{t+k})B^k \\ &= \sum_{k=-\infty}^{\infty} \left(\mathrm{Cov}(X_t, X_{t+k}) + \mathrm{Cov}(Y_t, Y_{t+k})\right) B^k \\ &= \sum_{k=-\infty}^{\infty} \gamma_k^X(B)B^k + \sum_{k=-\infty}^{\infty} \gamma_k^Y(B)B^k \\ &= \gamma^X(B) + \gamma^Y(B). \end{aligned}$$

מכאן נקבל כי
$$f^Z(\omega) = f^X(\omega) + f^Y(\omega).$$
כלומר, הספקטרום של סכום של שני תהליכים בלתי-תלויים הוא סכום הספקטרומים.

9.10.7 הספקטרום של תהליכים עם גורם עונתי

יהא $\{Y_t\}$ תהליך הנתון על ידי
$$(1-\phi B^{12})Y_t = X_t,$$
באשר $\{X_t\}$ תהליך רעש אקראי.

הפונקציה יוצרת המתאם המשותף נתונה על ידי
$$\gamma(B) = \sigma_X^2 \frac{1}{(1-\phi B^{12})(1-\phi B^{-12})},$$
ואם התהליך סטציונרי הספקטרום קיים ונתון על ידי
$$\begin{aligned} f(\omega) &= \frac{\sigma_X^2}{2\pi} \times \frac{1}{(1-\phi e^{-12i\omega})(1-\phi B^{12i\omega})} \\ &= \frac{\sigma_X^2}{2\pi} \times \frac{1}{1+\phi^2-2\phi\cos(12\omega)}. \end{aligned}$$
נשים לב כי $\cos(12\omega) = -1$ עבור $\omega = \frac{\pi(2k-1)}{12}$ כאשר $\omega = 1,2,3,4,5,6$, ו-$\cos(12\omega) = 0$ עבור $\omega = \frac{2\pi k}{12}$ כאשר $\omega = 1,2,3,4,5,6$. לכן, אם $\phi > 0$, לספקטרום יהיה שיא ב-$\omega = 0$, ובנוסף לכך שיאים במכפלות של תדר הבסיס $\frac{2\pi k}{12}$. הספקטרום יקבל מינימום בתדרים $\frac{\pi(2k-1)}{12}$. ניתוח הספקטרום כאשר $\phi < 0$ יכול להעשות באופן דומה.

יהא $\{Y_t\}$ תהליך הנתון על ידי
$$(1-\phi B^{12})(1-\mu B)Y_t = X_t,$$
באשר $\{X_t\}$ תהליך רעש אקראי.

בתהליך זה יש גם גורם עונתי, וגם מתאם בין תצפיות סמוכות. קל לראות כי הספקטרום נתון על ידי
$$f(\omega) = \frac{\sigma_X^2}{2\pi} \times \frac{1}{(1+\phi^2-2\phi\cos(12\omega))(1+\mu^2-2\mu\cos(12\omega))}.$$
הגרף של הספקטרום כאשר $\phi, \mu > 0$ מופיע באיור.

9.11 הערה אודות תהליך עונתי

נניח כי $\{Y_t\}$ הוא תהליך דטרמיניסטי הנתון על ידי
$$Y_t = \sin(\lambda t + \theta),$$
באשר $\lambda, \theta \in [-\pi, \pi]$. נשים לב תהליך זה אינו סטציונרי, ולכן פונקצית המתאם המשותף אינה מוגדרת. באופן דומה, הספקטרום של כל תהליך $\{Y_t\}$ דטרמיניסטי אינו קיים. בסעיף הבא נראה כי באופן מעשי כן ניתן לזהות תנודות עונתיות.

9.11.1 דיון

כאן חישבנו את הספקטרום של תהליכים שונים. בבעיות מעשיות איננו יודעים מראש מה התהליך שהסדרה הנצפית יוצרה ממנו: הספקטרום הוא כלי לזיהוי התהליך. במקרה כזה אנו נחשב את העוצמה של כל תדר, או את הפריודוגרם, וגדלים אלו ישמשו לקרב את הספקטרום. בעזרת הספקטרום ננסה לראות האם יש גורמים עונתיים.

9.12 תכונות סטציונריות של סכום של סינוסים

נתבונן בתהליך

$$Y_t = \sum_{i=1}^{M} A_i \sin(\omega_i t + C_i), \tag{71}$$

הנתון על ידי סכום של גלי סינוס, גל הסינוס ה-i הוא עם אמפליטודה $A_i > 0$, תדר $\omega_i \in [-\pi, \pi]$, ופאזה C_i המתפלגת אחיד בקטע $[-\pi, \pi]$. נניח כי כל ה-(ω_i) שונים האחד מהשני (אחרת, ניתן לחבר אותם ביחד). נחשב את התוחלת, השונות ופונקצית המתאם המשותף של תהליך זה.

התוחלת של Y_t נתונה על ידי:

$$\begin{aligned}
\mathbf{E}[Y_t] &= \sum_{i=1}^{M} A_i \mathbf{E}[\sin(\omega_i t + C_i)] \\
&= \sum_{i=1}^{M} \frac{A_i}{2\pi} \int_{-\pi}^{\pi} \sin(\omega_i t + x) dx = 0.
\end{aligned}$$

מכיוון שלכל זווית β

$$\int_{y=-\pi}^{\pi} \sin(\beta + y) dy = \int_{y=-\pi}^{\pi} \sin y dy = 0,$$

נובע שלכל שתי זוויות $\alpha \neq \beta$,

$$\int_{x=-\pi}^{\pi} \int_{y=-\pi}^{\pi} \sin(\alpha + x) \sin(\beta + y) dy dx = \int_{x=-\pi}^{\pi} \sin(\alpha + x) \left(\int_{y=-\pi}^{\pi} \sin(\beta + y) dy \right) dx = 0. \tag{72}$$

לכן לכל k שלם התוחלת של $Y_t Y_{t+k}$ נתונה על ידי:

$$\begin{aligned}
\mathbf{E}[Y_t Y_{t+k}] &= \sum_{i,j=1}^{M} A_i A_j \mathbf{E}[\sin(\omega_i t + C_i) \times \sin(\omega_j (t + k) + C_j)] \\
&= \sum_{i=1}^{M} A_i^2 \mathbf{E}[\sin(\omega_i t + C_i) \times \sin(\omega_i (t + k) + C_i)] \\
&= \sum_{i=1}^{M} A_i^2 \mathbf{E}\left[\frac{1}{2} \left(\cos(\omega_i k) - \cos(\omega_i (2t + k) + 2C_i) \right) \right] \\
&= \frac{1}{2} \sum_{i=1}^{M} A_i^2 \cos \omega_i k.
\end{aligned}$$

באופן דומה ניתן להוכיח כי ההתפלגות של $(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m})$ אינן תלויות ב-t לכל m קבוע ולכל $(t_1, \ldots, t_m)$ השונים זה מזה. בפרט, התהליך הנתון במשוואה (71) הוא סטציונרי.

9.13 פירוק של תהליכים סטציונריים

כפי שראינו בסעיף הקודם, התהליך

$$Y_t = \sum_{i=1}^{M} A_i \sin(\omega_i t + C_i) \qquad (73)$$

הוא סטציונרי. הבדל בסיסי בין תהליך זה לתהליך ARIMA הוא שבעוד כאן הרכיב המחזורי ניתן לחיזוי מראש בהינתן התצפיות, אין זה כך בתהליך ARIMA.

תהליך בו הרכיב המחזורי ניתן לחיזוי נקרא דטרמיניסטי, ותהליך בו הרכיב המחזורי אינו ניתן לחיזוי נקרא תהליך לא דטרמיניסטי. תהליכי ARIMA הם דוגמא לתהליכים לא-דטרמיניסטיים, והתהליך הנתון במשוואה (73) הוא דטרמיניסטי. ישנם תהליכים המכילים רכיבים דטרמיניסטים ולא-דטרמיניסטים, לדוגמא נתבונן בתהליך הבא:

$$Y_t = A\sin(0.1t + C) + \frac{\theta(B)}{\phi(B)}X_t,$$

בו C מתפלג אחיד בקטע $[-\pi,\pi]$, ו-$\frac{\theta(B)}{\phi(B)}X_t$ הוא תהליך ARMA הפיך וסטציונרי.

הסטטיסטיקאי Wold הוכיח את המשפט הבא, הנקרא **משפט הפירוק של Wold**:

משפט 9.9 כל תהליך סטציונרי $\{Y_t\}$ ניתן להצגה כסכום של שני תהליכים $Y_t = Y_t^{(d)} + Y_t^{(n)}$, באשר $\{Y_t^{(d)}\}$ הוא תהליך סטציונרי דטרמיניסטי, ו-$\{Y_t^{(n)}\}$ הוא תהליך סטציונרי לא-דטרמיניסטי.

10 תרגילים

10.1 תרגיל ראשון - מחקר כללי של סדרה עתית

התבונן בסדרה העתית $\{Y_t\}$ המופיעה בקובץ madad.xls. סדרה זו מתארת את מדד המחירים לצרכן בישראל בין השנים 1952 ו-2003 (בסיס 1951 = 100).

1. תאר באופן גרפי את הסדרות העתיות הבאות:
   - (i) Y_t,
   - (ii) $\log(Y_t)$,
   - (iii) $\nabla Y_t = Y_t - Y_{t-1}$,
   - (iv) $\nabla \log(Y_t) = \log(Y_t) - \log(Y_{t-1}) = \log(Y_t/Y_{t-1})$,
   - (v) Y_t/Y_{t-1}.
2. אילו מהסדרות נראות מתאימות ביותר להמשך ניתוח סטטיסטי? מדוע?
3. התאם קו מגמה לכל אחת מהסדרות (i), (iv) ו-(v), ובנה את סדרת השאריות. תאר אותן באופן גרפי. בחר את הסדרה שנראית מתאימה ביותר להנחת סטציונריות.
4. עבור הסדרה שבחרת בסעיף הקודם:

   (א) החלק על ידי ממוצע-נע מסדר $q = 2$ (ראה משוואה (1) בעמוד 8).

   (ב) החלק על ידי חציונים נעים מסדר $q = 2$ (ראה משוואה (2) בעמוד 8).

   האם ניתן לזהות מגמה נוספת?
   האם יש שנים (או תקופות) בעייתיות במיוחד?
   עד כמה נראית הנחת הסטציונריות סבירה עבור שאריות אלה?
5. צור את הקורלוגרם עבור סדרת השאריות.

10.2 תרגיל שני - מודלים סטציונריים

1. יהא $\{Y_t\}$ תהליך סטציונרי. נגדיר תהליך $\{Z_t\}$ באופן הבא:

$$Z_t = \sum_{i=0}^{n} a_i Y_{t-i}.$$

האם התהליך $\{Z_t\}$ הוא סטציונרי? הוכח או תן דוגמא נגדית.

2. יהא $\{X_t\}$ תהליך של רעש אקראי. מהי פונקצית המתאם המשותף של התהליך $\{Y_t\}$ המקיים את המשוואה

$$Y_t = X_t + 0.7X_{t-1} - 0.2X_{t-2}.$$

3. יהא $\{X_t\}$ תהליך של רעש אקראי. מהי פונקצית המתאם המשותף של התהליך $\{Y_t\}$ המקיים את המשוואה

$$Y_t = \sum_{k=0}^{q} X_{t-k}.$$

4. יהא $\{Y_t\}$ תהליך סטציונרי, המקיים $\rho_1 = 0.8$ ו-$\rho_2 = 0.4$. מה ניתן לאמר על ρ_3?

10.3 תרגיל שלישי - מודלים סטציונריים

1. התבונן בסדרה העתית המופיעה בקובץ seriesc.

(א) חשב את סדרות ההפרשים $\{\nabla Y_t\}$, $\{\nabla^2 Y_t\}$.

(ב) בדוק האם מודל אוטו-רגרסיבי או מודל מיצוע-נע יכולים להסביר כיצד אחת משלוש הסדרות יוצרה (הסדרה המקורית ושתי סדרות ההפרשים).

(ג) מה המודל הסטוכסטי שהכי סביר ייצר את הסדרה העתית שבקובץ?

10.4 תרגיל רביעי - תהליכי ARMA

1. נתבונן בתהליך $\{Y_t\}$ המקיים את המשוואה

$$Y_t = 1.2Y_{t-1} - Y_{t-2} + X_t - 0.5X_{t-1}.$$

(א) זהה איזה מודל ARMA מתאים לתהליך זה.

(ב) האם התהליך סטציונרי?

(ג) חשב את הפולינום Ψ.

(ד) חשב את הפולינום Π.

2. מצא עבור איזה c התהליך $Y_t = Y_{t-1} + cY_{t-2} + X_t$ סטציונרי, וחשב את פונקצית המתאם המשותף שלו.

3. הראה שהתהליך $Y_t = Y_{t-1} + cY_{t-2} - cY_{t-3} + X_t$ אינו סטציונרי לכל c.

4. את כל אחד מהמודלים הבאים:

(א) $Y_t = 0.3Y_{t-1} + X_t$,

(ב) $Y_t = X_t - 1.3X_{t-1} + 0.4X_{t-2}$,

(ג) $Y_t = 0.5Y_{t-1} + X_t - 1.3X_{t-1} + 0.4X_{t-2}$,

בטא באמצעות האופרטור B. קבע אם התהליך הפיך ו/או סטציונרי, והצג אותו כתהליך $\mathrm{MA}(\infty)$.

5. ברשותנו סדרה שיוצרה על ידי תהליך המקיים את המשוואה $Y_t = \phi_1 Y_{t-1} + X_t$. אנו מנסים להתאים לסדרה מודל $\mathrm{MA}(q)$. איזה סדר מודל יתן התאמה טובה? כיצד התשובה תלויה בערכו של ϕ_1?

6. נניח ש-$\{Y_t\}$ הוא תהליך $\mathrm{ARMA}(p_1, q_1)$, ו-$\{Z_t\}$ הוא תהליך $\mathrm{ARMA}(p_2, q_2)$. נגדיר תהליך $\{W_t\}$ על ידי

$$W_t = Y_t + Z_t.$$

האם $\{W_t\}$ הוא תהליך ARMA? אם כן, מאיזה סדר?

10.5 תרגיל חמישי - חיזוי

בתרגיל השלישי התבוננו בסדרה העתית המופיעה בקובץ seriesc, ופיתחנו עבורה שני מודלים אפשריים. עבור כל אחד משני המודלים הללו:

1. התג את המודל בשלוש ההצגות של תהליך ARIMA.
2. ספק תחזיות עבור האיברים הבאים בסדרה.
3. מה ניתן לאמר על שונות התחזיות?

10.6 תרגיל ששי - ניתוח פורייה

1. הראה כי פיתוח פורייה של הפונקציה x^2 בקטע $[-\pi, \pi]$ הוא

$$f(x) = \frac{\pi^2}{3} - 4\left(\frac{\cos(x)}{1^2} - \frac{\cos(2x)}{2^2} + \frac{\cos(3x)}{3^2} - \cdots\right).$$

2. הוכח את אי-שוויון פרסבל:

$$\frac{\sum_{t=1}^{N}(x_t - \overline{x})^2}{N} = \sum_{p=1}^{N/2-1} \frac{a_p^2 + b_p^2}{2} + a_{N/2}^2.$$

3. נניח כי $X_1, \ldots, X_n$ הם משתנים מקריים בלתי-תלויים שווי-התפלגות המתפלגים $N(\mu, \sigma^2)$. הוכח כי

$$a_p = \frac{2}{N}\sum_{t=1}^{N} X_t \cos(2\pi pt/N)$$

מתפלג $N(0, \frac{2\sigma^2}{N})$ עבור $p = 1, 2, \ldots, N/2 - 1$.