

הטכניון - מכון טכנולוגי לישראל
הפקולטה להנדסת חשמל

אותות אקראיים 044202

משה זכאי ז"ל, משה סידי, אדם שורץ, עפר זיתוני, נרי מרחב

מהדורת תשע"ו 2015/6

© 2003, 2006, 2007, 2008, 2009, 2016 אדם שורץ

השימוש בחומר זה מותר לצרכים אישיים של לימוד ומחקר. בקשות לשימוש למטרות הוראה או למטרות אחרות יש להפנות למחזיק זכויות היוצרים adam "at" ee.technion.ac.il

תודות

מהדורת 1989:
ברצוננו להודות לגברת אנט ברג ז"ל על עזרתה בהדפסת ובהכנת השרטוטים, לגברת יפה לוי על עזרתה בהדפסה ולגברת חנה ביסמוט על עזרתה בהכנת השרטוטים.
כמו כן ברצוננו להודות למר דורון שקד על עזרתו בעדכון החוברת.

מהדורת 2003:
תודות לגברת לזלי פרייס על הדפסת החוברת ולגברת חנה ביסמוט על הכנת השרטוטים.

מהדורות 2003 ו- 2006: תודה לרמי אתר ונרי מרחב, על תרומתם לתוכן חוברת זו.

מהדורת 2015/6: תודה לעידו זצ"בסקי.

גרסה זו הינה גרסה א'.

מהדורת 2015/6 של חוברת זו מוקדשת בזאת לזכרו של פרופ' מחקר משה זכאי ז"ל, אשר כתב את מהדורתה הראשונה.

תוכן עניינים

| | | |
|-----|---|-------|
| 3 | מבוא וחזרה על הסתברות | 1 |
| 3 | מבוא | 1.1 |
| 7 | מהלך ההרצאות | 1.2 |
| 7 | חזרה על הסתברות | 1.3 |
| 14 | וקטורים אקראיים | 1.4 |
| 20 | הפילוג הגאوسي | 1.5 |
| 26 | הסתברות מותנית | 1.6 |
| 29 | שערוך | 2 |
| 29 | מבוא | 2.1 |
| 30 | שערוך אופטימלי | 2.2 |
| 35 | שערוך לינארי אופטימלי | 2.3 |
| 42 | עקרון ההשלכה (עקרון האורתוגונליות) | 2.4 |
| 45 | תהליכים אקראיים בזמן בדיד | 3 |
| 45 | חוק ההסתברות של תהליך אקראי | 3.1 |
| 48 | תוחלת ומומנטים של תהליך אקראי | 3.2 |
| 53 | סטציונריות וארגודיות | 3.3 |
| 59 | שרשראות מרקוב | 4 |
| 59 | דוגמאות והגדרות | 4.1 |
| 62 | שרשראות מרקוב הומוגניות | 4.2 |
| 65 | חישוב הפילוג והסתברויות המעבר | 4.3 |
| 68 | מצבים נשנים וחולפים | 4.4 |
| 73 | קישוריות בין מצבים ופרוק מרחב המצב | 4.5 |
| 77 | סטציונריות של שרשראות מרקוביות | 4.6 |
| 82 | תהליכים אקראיים בזמן רציף | 5 |
| 82 | מבוא, הגדרות ודוגמאות | 5.1 |
| 87 | סטציונריות | 5.2 |
| 99 | תהליך אקראי גאوسي | 5.3 |
| 101 | מעבר תהליכים אקראיים דרך מערכות לינאריות | 5.4 |
| 103 | מעבר תהליכים אקראיים סטציונריים במובן הרחב דרך מערכות קבועות בזמן | 5.4.1 |
| 115 | הזת ספקטרום | 5.5 |

| | | |
|-----|--------------------------------------|-----|
| 116 | סינון לינארי אופטימלי | 5.6 |
| 127 | כמה מילים על ארגודיות | 5.7 |
| 132 | רעשים במערכות ובמעגלים חשמליים | 6 |
| 132 | רעש לבן | 6.1 |
| 136 | רוחב סרט אפקטיבי לרעש | 6.2 |
| 137 | רעש תרמי (רעש הנגד, רעש Nyquist) | 6.3 |
| 145 | רעש הדיודה (Shot noise) | 6.4 |
| 149 | רעש טרמי - פתוח מתוך מודל רעש הדיודה | 6.5 |
| 150 | אפיון רעש מגבר | 6.6 |
| 153 | מסננת מתואמת Matched Filter | 6.7 |
| 156 | חזרה נוספת על "מבוא להסתברות" | 8 |
| 156 | הסתברות | 8.1 |
| 157 | משתנה אקראי ופילוג | 8.2 |
| 158 | ווקטור אקראי | 8.3 |
| 159 | אי תלות סטטיסטית | 8.4 |
| 159 | תוחלת | 8.5 |
| 162 | מומנטים | 8.6 |
| 165 | פונקציה אפיינית | 8.7 |
| 166 | הסתברות ותוחלת מותנים | 8.8 |

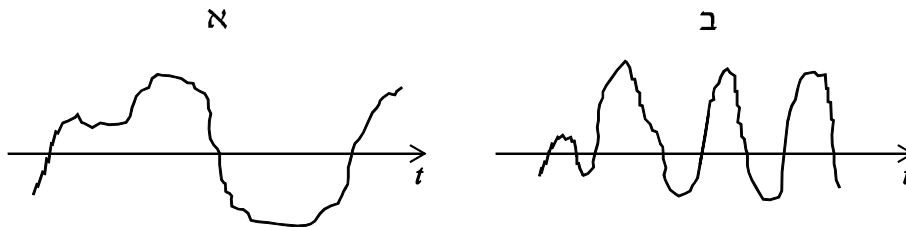
1 מבוא וחזרה על הסתברות

מטרת הקורס היא לחשוף את הסטודנטים לעקרונות ומושגי יסוד הנוגעים להתנהגות של תהליכים אקראיים: אלו אותות אשר צורתם נקבעת בין השאר על ידי גורם הסתברותי. הדגש הוא על נושאים כגון תהליכים מרקוביים, מעבר רעש במערכות לינאריות, רעש במעגלים חשמליים. נושאים אלו הם מרכזיים במגוון תחומים וישומים בהנדסת חשמל, כולל תקשורת, רשתות מחשבים, עיבוד אותות ועבוד תמונות, בקרה, מיכשור והתקנים אלקטרוניים. ואכן, מספר לא מבוטל של קורסים בשטחים שונים בפקולטה נשענים על חומר הקורס ב"אותות אקראיים", והקורס מהווה עבורם דרישת קדם.

1.1 מבוא

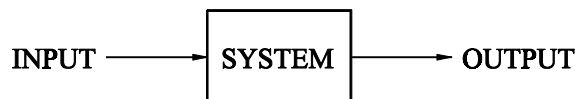
בתורת ההסתברות למדנו איך לאפיין אפיון הסתברותי ו"לתמצת" גדלים אקראיים שקראנו להם משתנים אקראיים או וקטורים אקראיים (חוק ההסתברות, מומנטים ועוד). המאפיין משתנים אלה הוא קיום פרמטר של "מזל" ω . דוגמה: תוצאת זריקת קוביה או מספר קוביות. בקורס זה נעסוק באפיון ותמצות תהליכים אקראיים. בתהליך אקראי (שמות נוספים: אות אקראי, פונקציה אקראית, תהליך סטוכסטי, סידרה אקראית) בנוסף לפרמטר ה"מזל", קיים גם פרמטר זמן t (דיסקרטי או רציף). דוגמאות: רעש מגבר, שיחות טלפון, גלי היס, מספר מכוניות העוברות בצומת, רעידות מכונית וכו'. אפשר לחשוב על תהליך אקראי כעל אוסף של משתנים אקראיים - או ווקטור של מ"א, כאשר כל רכיב קשור לזמן מסויים (האינדקס הוא זמן). כמובן שבהסתכלות כזו הווקטור הוא בדרך כלל בעל מימד אין סופי. נסמן תהליך אקראי באחת מהצורות הבאות: $X(t, \omega)$, $X_\omega(t)$ כאשר לעיתים נשמיט את הסימון ω ונרשום $X(t)$: זאת נעשה רק כאשר הדבר לא יגרום לבלבול.

עבור צורת גל דטרמיניסטית אנו יכולים לדבר על רוחב סרט ולהגיד (לפחות במקרים מסויימים) איזו משתי צורות גל "מהירה" יותר ע"י השוואת התמרת פוריה שלהן. במקרה ההסתברותי, נעיין בשתי צורות גל "טיפוסיות" המופיעות באיור 1.1. כאשר א'



איור 1.1: צורות גל

צורת גל טיפוסית של מכונה עם בולם זעזועים מסוג א', ו-ב' צורת גל טיפוסית עבור אותה מכונה עם בולם זעזועים מסוג ב'. הזעזועים אקראיים ואינם חוזרים על עצמם (תהליכים אקראיים). בהרגשה, תהליך ב' מהיר יותר ובעתיד נראה איך לתת לזה מובן. נראה בעיה זו כחלק מבעיה כללית ועקרונית יותר כדלקמן:



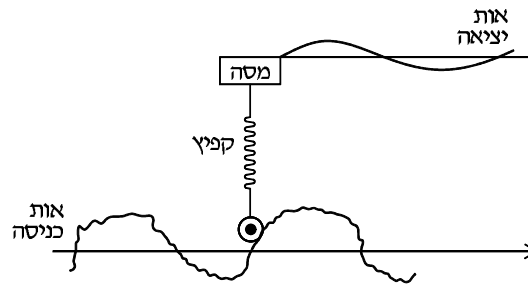
איור 1.2: מערכת

כאשר עסקנו בצורות גל דטרמיניסטיות עסקנו במודל המתואר בציור 1.2 ולמדנו (לפחות במקרה של מערכת לינארית, ובמיוחד במקרה של מערכת לינארית שאינה משתנה בזמן), איך לאפיין את המערכת ואיך לחשב את היציאה עבור כניסה שרירותית. במיוחד, עבור מערכת לינארית שאינה משתנה בזמן למדנו לאפיין את המערכת בשתי צורות: אפיון במרחב

הזמן (תגובת המערכת להלם - פונקצית דירק), ואפיון במרחב התדר (ע"י תגובה לערור הרמוני $e^{i\omega t}$). בעזרת אפיונים אלה ובעזרת עקרון הסופרפוזיציה, קבלנו את התגובה של המערכת לכל אות כניסה בשתי הצורות - במרחב התדר ובמרחב הזמן. בקורס זה נעסוק באפיון ההסתברותי של היציאה כאשר נתון האפיון ההסתברותי של הכניסה ואפיון המערכת. אנו נעסוק תמיד במערכת דטרמיניסטית, רק הכניסה, ולכן גם היציאה, יהיו תהליכים אקראיים. כלומר, בעוד שבקורסים כגון "מעגלים חשמליים" ו-"אותות ומערכות" עסקנו במכלול התכונות של צורת גל מסויימת, ובתגובת מערכת לינארית קבועה בזמן לאותה צורת גל, כאן אנו עוסקים במכלול ("אנסמבל") של צורות גל (צורה אחת לכל ערך של ω), אותו נתאר בעזרת תכונותיו הסטטיסטיות, וננתח מה קורה לאופי הסטטיסטי אחרי מעבר במערכת כזו. בהמשך כמובן נגדיר את מושגים אינטואיטיביים אלו בצורה מדוייקת.

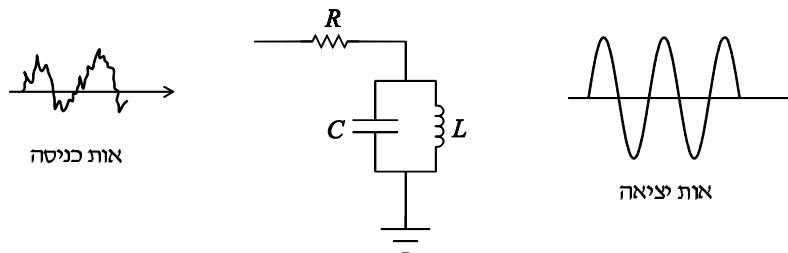
דוגמאות:

א. בולם זעזועים



איור 1.3: בולם זעזועים

ב. מעגל RLC



איור 1.4: מעגל RLC

הערה: בשני המקרים ציירנו את היציאה "פחות עצבנית" מהכניסה (אם כי לא בהכרח חלשה יותר, במגבר היא תהיה חזקה יותר). בהמשך ההרצאות נראה מדוע.

נחזור ונדגיש שאנו עוסקים במשפחות של פונקציות מבלי שיהיה לנו עניין בדגם מיוחד. הבעיה שלנו תהיה איך לאפיין אותות (משפחות) כאלה ואיך מערכות לינאריות מגיבות למשפחות כאלה.

דוגמאות נוספות

- מנקודת מבטו של מחשב המשמש כשרת תקשורת, משתמשים חדשים מתחברים בזמנים אקראיים. כיצד נתאר את מספר המשתמשים שהגיעו למחשב עד רגע t נתון? מצד אחד זהו תהליך, שכן יש תלות בזמן—אם נבדוק את מספר המשתמשים

זמן מאוחר יותר, יתכן שהמספר ישתנה. מצד שני, בכל רגע נתון מספר המשתמשים שהגיעו עד רגע זה הוא משתנה אקראי (הגדרה 8.9), משום שמשך הזמן בין הגעה להגעה הוא אקראי. "ליצור" שכזה אנו קוראים תהליך אקראי (הגדרה 3.2). הגישה ההנדסית לבעיות מעשיות בהנדסת חשמל (כמו הדוגמאות לעיל) היא לבנות מודל מתמטי עבור התנהגות האותות, או המערכות, אותם אנו רוצים לנתח או לתכנן. במקרים בהם יש אי וודאות מסויימת לגבי האותות או המערכות, אפשר לייצגה דרך מודל מסוג של תהליך אקראי.

- רעש אנלוגי במקלט רדיו (אות אנלוגי). מקור האקראיות: תהליכים פיסיקליים שונים היוצרים רעש, כגון תנועות אקראיות של אלקטרונים. ניתן לבנות מודלים לרעשים מסוגים שונים. מודל מרכזי בתקשורת הוא של משדר המפיק אות העובר בתווך (אוויר, למשל) בו נוספים לו רעשים, ולאחר מכן נקלט במקלט. משדר ומקלט חכמים מתוכננים כדי להקטין ככל האפשר את השפעת הרעש. לצורך כך יש להבין כיצד מתנהג הרעש כאשר הוא עובר את המסננים שבמקלט.
 - אות דיבור. מקור האקראיות: מודל מתמטי של אות עבורו לא קיים אפיון דטרמיניסטי. כאן מתעניינים באפשרות לזהות את המידע שבאות הדיבור (מה נאמר?), לשפר את איכות האות הנקלט (הקטנת רעשים כמו במודל התקשורת), ולדחוס את האות על מנת שניתן יהיה לנצל את ערוץ השידור הפיזיקלי בצורה מיטבית.
 - מכונית בתנועה. מקור האקראיות: מודל של כביש בתנאים מציאותיים (למרות שהכביש דטרמיניסטי!). אפיון סטטיסטי של מהמורות כביש יכל להוביל לתכנון מסנן (משכך) מתאים.
 - מערכת עקיבה. מקור האקראיות: תנועת הגוף (מטוס, אניה וכו') אשר אינה ידועה מראש, או מודל של שגיאות מדידה שונות. על מערכת העקיבה להתגבר על רעשי המדידה, ולא לאבד נעילה בעיניים. בתכנון מערכת כזו יש למצוא איזון נכון בין הקטנת רעשים מצד אחד, ליכולת עקובה אחרי מטרות דינמיות מצד שני.
- אנו רואים שתהליך אקראי מאופיין ע"י תלות בזמן (ולכן הוא "תהליך"), ואקראיות.

מודלים

כדי לראות כיצד נראה מודל מתמטי של תהליך אקראי, נזכר במושגים משתנה אקראי (הגדרה 8.9) פונקציה ואות. נתאר לנו אין-סוף מקלטים, שכולם הופעלו בזמן t_0 . כולם מאותו סוג, וכולם מכוונים לאותו תדר. נניח שאין שידור בתדר זה, ולכן במוצא המקלט יהיה רק הרעש שנוצר בו עצמו. אפשר, למשל, לחשוב על מרחב המדגם (הגדרה 8.1) Ω כעל אוסף כל המקלטים, כך שכל $\omega \in \Omega$ יהיה מקלט מסויים. נסמן ב- $X(t, \omega)$ את היציאה ממקלט מספר ω ברגע t . כך למשל, אם נקבע זמן מסויים, למשל $t = 2$, אזי בתלות ב"פרמטר המזל" ω , גודל זה $X(2, \omega)$ הוא משתנה אקראי. מצד שני, אם נקבע משדר מסויים, כלומר נקבע את $\omega = \omega_0$, אזי לפי המודל שלנו הפונקציה $X(t, \omega_0)$, היא פונקציה של המשתנה t בלבד: לפונקציה זו של משתנה הזמן (כאשר פרמטר המזל קבוע) קוראים פונקציית מדגם (הגדרה 3.2).

דוגמה 1.1 לפעמים אפשר לתת ביטוי מפורש לתהליך אקראי.

- יהיו Y ו- Z משתנים אקראיים. נגדיר $X(t, \omega) = Y(\omega) \cdot t + Z(\omega)$. זהו תהליך אקראי, ופונקציות המדגם (ω) קבוע, ומתבוננים בפונקציה של המשתנה t הן קווים ישרים.
- יהיו A ו- ϕ מ"א. נגדיר $X(t, \omega) = A(\omega) \sin(2\pi f t + \phi(\omega))$. צורות הגל (פונקציות המדגם) במקרה זה הן תנודות הרמוניות בעלות אמפליטודה (אקראית) A ופאזה (אקראית) ϕ . תהליך אקראי כזה הוא מודל של "גל נושא" - האות הסינוסי עליו מרכיבים (על ידי אפנון) מידע משודר. במקרים רבים אכן התדר ידוע אך הפאזה אינה ידועה מראש, ונוח לבנות לה מודל כמשתנה אקראי.

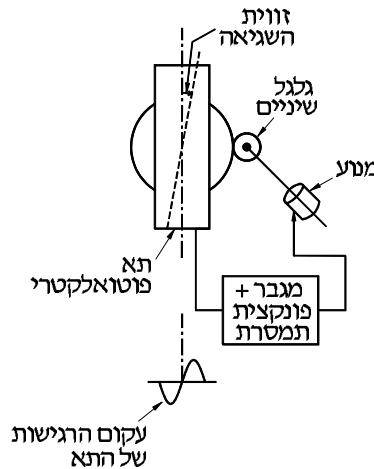
• יהיה N שלם חיובי קבוע או אקראי, ויהיו X_n משתנים אקראיים. נגדיר

$$X(t, \omega) = \sum_{n=1}^N X_n(\omega) \sin(n\nu_0 t + \phi_n(\omega))$$

כאשר ν_0 , התדר היסודי, הינו פרמטר. במקרה זה פונקציות המדגם הן סכום משוקלל של תנודות הרמוניות. אם $N = N(\omega)$ הוא אקראי, אזי מספר האברים בסכום תלוי ב- ω . תהליך זה יכול להיות קירוב לאות דיבור, אשר לעיתים מתנהג בצורה קרובה למחזורית.

חשוב לציין שלא כל תהליך אקראי ניתן לייצוג פשוט כמו בדוגמאות לעיל. בפרט, לא כל ת"א תלוי במספר סופי של מ"א.

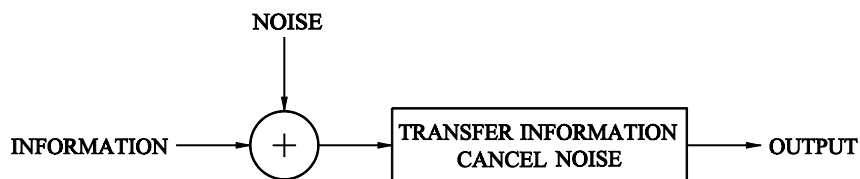
דוגמא נוספת: מערכת לעקיבה אחרי כוכב (או רובוט העוקב אחרי משהו) בנויה כמצויר:



איור 1.5: מערכת עקיבה

אור הכוכב נופל על התא הפוטואלקטרי. באם אור הכוכב אינו נופל על מרכז התא (שגיאה אפס), נוצר אות שגיאה חיובי או שלילי בהתאם לכיוון השגיאה. אות השגיאה מוגבר ומפעיל את המנוע בכיוון לאפוס השגיאה. מערכת שבאופן עקרוני דומה למערכת כזו מופיעה במערכות עקיבה של מכ"ם בטווח ובזווית וכן במעגל הנקרא "חוג נעול פאזה" המשמש למטרות שונות וחשובות ביותר בתקשורת. נחזור למערכת העקיבה המצוירת ונניח שהיא נמצאת על פלטפורמה מתנדנדת ולכן יש שגיאות עקיבה אחרי הכוכב בשל תנודות הפלטפורמה. כמו כן יש רעש שנוצר בתא הפוטואלקטרי שאף הוא גורם שגיאות עקיבה. המטרה היא עקיבה נאמנה אחרי הכוכב. נניח שיש לנו בקרת מהירות, כלומר, מהירות המנוע פרופורציונלית לזווית השגיאה בין ציר הטלסקופ והכוכב (בכיוון אפוס השגיאה). נניח שהמערכת יציבה: האם רצוי הגבר גדול או קטן? יש כאן ניגוד בין הדרישה לעקיבה נאמנה (אפוס מהיר של שגיאות העקיבה בגלל תנודות הפלטפורמה), הדורשת הגבר גבוה, לבין הקטנת שגיאות העקיבה בשל רעש התא הפוטואלקטרי, הדורשת הגבר נמוך. איך נאפיין את האות? במקרה שאנו עוסקים בו האות הוא תנועת הכוכב, או תנועת הפלטפורמה, או תנועת הגוף שעליו נעול הרובוט. זה נראה די טבעי ומובן מאליו שרעש טעון אפיון הסתברותי. מה שאולי פחות מובן מאליו הוא שגם האות הרצוי (המידע) הוא בעצם גודל אקראי.

בעבר הרחוק יותר השתמשו במודל מאד נאיבי לאות ורעש בתקשורת. במודל זה, סינוס מתאר את האות הרצוי (המידע) וסינוס אחר (או תהליך אקראי) מתאר את הרעש. מודל זה אינו מספיק וכיום הן בתקשורת והן בבקרה רואים הן את האות הרצוי והן את הרעש, כל אחד כתהליך אקראי. בתקשורת יהיה, לכן, המודל הטבעי:



איור 1.6: מודל תקשורת

1.2 מהלך ההרצאות

חלק א': מבוא וחזרה על הסתברות.

חלק ב': בחלק זה נעסוק בבעיות כגון הבעיה הבאה: נתונות המדידות Y_1 ו- Y_2 ,

$$Y_1 = X + N_1 \quad ; \quad Y_2 = X + N_2$$

X מ"א שרוצים לדעת את ערכו, N_1, N_2 רעשים; אין לנו גישה ישירה ל- X ואנו יודעים את Y_1, Y_2 בלבד. המטרה היא לתת "ניחוש חכם" ל- X על סמך ידיעת Y_1, Y_2 (וחוקי ההסתברות של (X, N_2, N_1)). בעיות מסוג זה נקראות שערוד. אפשר היה לשער כי הניחוש המיטבי הוא הממוצע בין שתי המדידות. בהמשך נראה כי זהו לא תמיד הניחוש הכי מוצלח שאפשר. פתרון בעיה זו מהווה בסיס לשאלה קשה יותר – בעיית הסינון – בה אנו משערכים את ערך התהליך בזמן מסוים, על סמך ערכי כל המדידות. שערוד כזה מהווה בסיס לפתרון בעיית התקשורת ובעיית העקיבה אשר הזכרנו.

חלק ג': בחלק השלישי (והעיקרי) של הקורס נעסוק בתהליכים אקראיים, אפיונם ומעברם דרך מערכות לינאריות. נתחיל בתהליכים בזמן בדיד: פילוגים, מומנטים ותכונות שונות שלהם. נרחיב בנושא שרשראות מרקוב – תהליכים בעלי ערכים דסקרטיים ו"זכרון" של צעד אחד. לבסוף נעסוק בתהליכים בזמן רציף, כולל השפעת מערכות לינאריות על תהליכים כאלו.

חלק ד': בחלק זה נעסוק באפיון הרעש הפיזיקלי התרמי הבסיסי – רעש הנגד, רעש הדיודה, אפיון רעש מגברים ותכונות הרעש של שרשרת מגברים.

1.3 חזרה על הסתברות

מרחב הסתברות

* מרחב המדגם $\Omega = \{\omega\}$: אוסף הכולל את כל התוצאות האפשריות של ניסוי. דוגמה: במקרה של קוביה מרחב מדגם אפשרי הוא מרחב עם 6 אלמנטים. במקרה של רולטה: אוסף הנקודות על היקף מעגל היחידה. במקרה של רעש היציאה ממגבר בקטע הזמן $[0, 1]$, מרחב מדגם יכול להיות אוסף כל הפונקציות הרציפות בקטע הזמן $[0, 1]$. מרחב המדגם חייב להיות עשיר מספיק כדי לתאר כל תוצאה אפשרית, אך בדרך כלל זהו מודל אשר אנו יכולים לבחור, ולעיתים נוח לבחור אותו כך שיהיה גדול מהמינימום הדרוש.

* מרחב המאורעות $F = \{A\}$: אוסף של תת קבוצות של Ω . מרחב המאורעות מתאר את המידע שניתן לקבל אודות תוצאות הניסויים יהיה $\omega_1 \in \Omega$, כלומר, ω_1 מדגם מסוים. בדוגמאות לעיל, עבור המקרה של קוביה אנו יכולים ליחס ל- ω_1

הסתברות שונה מאפס. בשתי הדוגמאות האחרות ההסתברות להופעת ω_1 מסוים היא אפס ולכן את ההסתברות עלינו ליחס למאורעות ולא לדגמים. לדוגמה, נוכל לשאול עבור הרעש מהמגבר מה ההסתברות ש $\int_0^1 n^2(t)dt < 3$. על מנת שאפשר יהיה לבנות תורה מבוססת של הסתברות יש לדרוש שאוסף תת הקבוצות של Ω , $F = \{A\}$ יקיים את התנאים הבאים:

1. $\Omega \in F$, כלומר תת הקבוצה Ω של Ω היא מאורע.
 2. אם A_1, A_2 מאורעות (ז.א. $A_i \in F$) אזי גם $A_1 \cup A_2$ מאורע ($A_1 \cup A_2 \in F$).
 3. אם A_i שייך ל- F אזי גם A_i^c שייך ל- F .
 4. אחוד ניתן להמנות של מאורעות אף הוא מאורע, כלומר, אם $A_i \in F$, $i = 1, 2, \dots$ גם $\cup A_i \in F$.
- מדרישות אלו נובע כי F סגור גם תחת חיתוך.

הערה: האוסף F אינו חייב להכיל את כל תת הקבוצות של Ω . אפשר לחשוב על מאורעות כעל גדלים הנתנים למדידה. כך, אם למשל אנו מודדים אנרגיה אזי ניתן לבחור את המאורעות כך שלא יבחינו בין שני אותות בעלי סימן הפוך. למשל, בדוגמה של הקוביה $\Omega = \{1, 2, 3, 4, 5, 6\}$, והאוסף $F = \{\Omega, \emptyset, \{1, 2, 3\}, \{4, 5, 6\}\}$ הוא מרחב מאורעות לגיטימי (בדוק!).

הערה: מאורעות A, B ($A \in F, B \in F$) המקיימים $A \cap B = \emptyset$ נקראים מאורעות זרים. חשוב לא להתבלבל בין המושג של מאורעות זרים למושג של מאורעות בלתי תלויים; שים לב כי מאורעות זרים אשר לשניהם הסתברות חיובית הם תמיד תלויים (סטטיסטית: ראה הגדרה בהמשך) שכן אם קרה A אזי ידוע בוודאות כי B לא קרה.

פונקציית הסתברות $P(A)$: עבור כל מאורע A , כאשר $A \in F$ מקיים: $0 \leq P(A) \leq 1$; $P(A^c) + P(A) = 1$ אם נתון A_1, A_2, \dots עבור כל $i \neq j$ אזי $A_i \cap A_j = \emptyset$; $P(\cup A_i) = \sum_i P(A_i)$; כמו כן $P(\Omega) = 1$.

* השלטה $\{\Omega, F, P\}$ נקראת מרחב הסתברות.

כאמור, פונקציית ההסתברות מוגדרת רק עבור מאורעות (כלומר אלמנטים של F), ולא בהכרח מוגדרת עבור קבוצות אחרות. בפרט אין הכרח שההסתברות של $\{\omega\}$ —דגם בודד—תהיה מוגדרת.

משתנה אקראי (מ"א) הוא פונקציה $X(\omega)$, $\omega \in \Omega$ על מרחב המדגם כך שעבור כל a ממשי $\{\omega : X(\omega) \leq a\}$ הוא מאורע. כלומר אם נמדוד תוצאות ניסוי על ידי ערכי המ"א, נקבל תמיד מאורעות.

דוגמאות של מ"א על רעש המגבר (נניח שרעש היציאה מהמגבר, $n(t)$ הוא צורת גל רציפה)

$$1. n(0.25) = n(t)|_{t=0.25} \text{ (הרעש ברגע } t = 0.25 \text{)}$$

$$2. X = \int_0^1 n^2(t)dt \text{ (האנרגיה של אות הרעש בתחום הזמן } [0, 1] \text{)}$$

$$3. Y = \max_{t \in [0, 1]} n(t) \text{ (הערך המכסימלי של אות הרעש בתחום הזמן } [0, 1] \text{)}$$

הטענות בדוגמאות הן אינטואיטיביות אך הן טעונות הוכחה (ההוכחה מתבססת על רציפות הדגמים של הרעש).

פונקציית פילוג

מכיוון שלפי הגדרת מ"א $\{\omega : X(\omega) \leq a\}$ הוא מאורע, ניתן להגדיר עבורו הסתברות. להסתברות זו כפונקציה של a קוראים פונקציית הפילוג של X . פונקציית הפילוג של המשתנה האקראי $X(\omega)$ מגדירה את חוק ההסתברות של $X(\omega)$:

$$F_X(a) = \text{Prob}\{X(\omega) \leq a\}$$

מוגדרת תמיד כפונקציה של a (הוא רק אינדקס), היא מונוטונית (לא יורדת), וניתן להוכיח שהיא רציפה מימין. בגבולות מקבלת פונקציית הפילוג את הערכים הבאים: $F_X(-\infty) = 0, F_X(+\infty) = 1$. עבור מ"א בדיד $F_X(a)$ עולה בקפיצות (ציייר את $F_X(a)$ עבור קובית משחק). אם קיים $f_X(\alpha)$ כך ש- $F_X(\alpha) = \int_{-\infty}^{\alpha} f_X(\theta)d\theta$ לכל α אזי $f_X(\alpha)$ נקראת פונקציית הצפיפות או הפילוג הסגולי (פ"ס) של X . בפרט אם $F_X(a)$ גזיר, אזי $f_X(a)$ היא הנגזרת של $F_X(a)$. מהגדרת פונקציית הפילוג נובע מיד שעבור $a_1 < a_2$, מתקיים $\text{Prob}\{a_1 < X \leq a_2\} = F_X(a_2) - F_X(a_1)$, (כי המאורע $\{X \leq a_2\}$ הוא איחוד של המאורעות הזרים $\{a_1 < X \leq a_2\}$ ו- $\{X \leq a_1\}$). מכאן גם נובעת המונוטוניות של $F_X(a)$.

תוחלת

עבור מ"א בדיד שערכיו האפשריים הם $\{\alpha_i\}$ התוחלת מוגדרת כ-

$$m_x = \bar{X} = E[X] = \sum_i \alpha_i \text{Prob}\{X = \alpha_i\},$$

ובתנאי שהביטוי מוגדר היטב, כלומר בתנאי ש- $\sum_i |\alpha_i| \text{Prob}\{X = \alpha_i\} < \infty$.

עבור מ"א עם פ"ס:

$$m_X = \bar{X} = E[X] = \int_{-\infty}^{\infty} \alpha f_X(\alpha) d\alpha$$

וזאת בהנחה ש- $\int_{-\infty}^{\infty} |\alpha| f_X(\alpha) d\alpha < \infty$. באופן כללי עבור כל n , תהיה $\alpha_i^{(n)}$; $i = 1, 2, \dots, m$; (כאשר m תלוי ב- n) חלוקה סופית של הקטע $(-n, n)$; $\alpha_i^{(n)} < \alpha_{i+1}^{(n)}$; נניח שמתקיים $\max_i |\alpha_{i+1}^{(n)} - \alpha_i^{(n)}| \rightarrow 0$ כאשר $n \rightarrow \infty$. (ז.א. ככל ש- n גדל החלוקה נעשית עדינה יותר).

נגדיר:

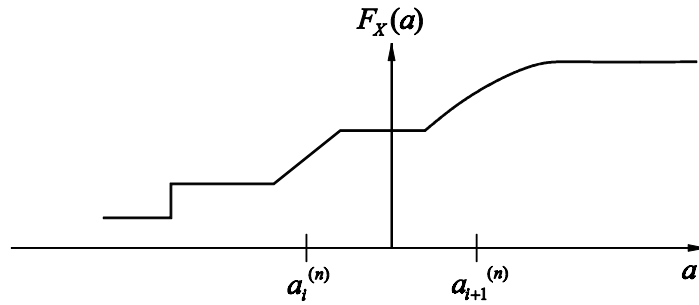
$$E[X] = \int_{-\infty}^{\infty} \alpha dF_X(\alpha) = \lim_{n \rightarrow \infty} \sum_i \alpha_i^{(n)} [F_X(\alpha_{i+1}^{(n)}) - F_X(\alpha_i^{(n)})]$$

בהנחה שהגבול קיים ואינו תלוי בסדרת החלוקות. בדוק שעבור המקרים הקודמים (בדיד, פ"ס) ההגדרה האחרונה אכן נותנת את התוצאה הנכונה.

$$E[Y] = E[g(X)] = \int_{-\infty}^{\infty} \alpha dF_Y(\alpha)$$

טענה חשובה ללא הוכחה:

$E[Y] = \int_{-\infty}^{\infty} g(\alpha) dF_X(\alpha)$ דהיינו, במקרה זה, על מנת לחשב את התוחלת של Y - אין צורך לחשב קודם את פונקציית



איור 1.7:

הפלוג של Y ומתוך פילוג זה לחשב את EY . אפשר לחשב את EY ישירות מתוך ידיעת פונקציית הפילוג של X (הדרך השניה היא כמעט תמיד יותר קצרה).

הערה: לא לכל מ"א יש תוחלת. לדוגמא, עבור מ"א עם פ"ס

$$f_X(\alpha) = \begin{cases} 0, & \alpha \geq 0 \\ \frac{2}{\pi(1+\alpha^2)}, & \alpha < 0 \end{cases}.$$

$$\int_{-\infty}^{\infty} \alpha f_X(\alpha) d\alpha = -\infty$$

מתקיים:

במקרה זה נאמר כי למ"א אין תוחלת—והכוונה היא שאין לו תוחלת סופית. יתכן מקרה פתולוגי אף יותר—יש מקרים שלא ניתן כלל להגדיר את האינטגרל (כאשר הגבול לעיל לא קיים). למשל, אם פונקציית הצפיפות היא

$$f_X(\alpha) = \frac{1}{\pi(1+\alpha^2)}, \quad -\infty < \alpha < \infty \quad (1.1)$$

אזי התוחלת אינה מוגדרת, שכן הגבול של

$$\int_{a_n}^{b_n} \frac{\alpha}{\pi(1+\alpha)} d\alpha$$

תלוי בקצב היחסי שבו $a_n \rightarrow -\infty$ ו- $b_n \rightarrow \infty$.

תכונות פשוטות של התוחלת

א. יהיה A מאורע ונסמן ב- I_A מ"א המקבל ערכים 0 או 1 בלבד כדלקמן:

$$I_A = \begin{cases} 1 & \omega \in A \\ 0 & \omega \in A^c \end{cases}$$

I_A נקרא הפונקציה המציינת (או פונקציית האינדיקטור) של המאורע A ומהגדרת התוחלת (או ישירות מנוסחת התוחלת למ"א בדיד) נקבל

$$EI_A = \int \alpha dF(\alpha) = 0 \cdot \text{Prob}(A^c) + 1 \cdot \text{Prob}(A) = \text{Prob}(A)$$

ב. אם $X = \text{const}$, כלומר X דטרמיניסטי או במילים אחרות משתנה אקראי מנוון, אזי מתקיים:
 $E[\text{constant}] = \text{constant}$

ג. לינאריות: X, Y מ"א; a, b קבועים. אזי: $E[aX + bY] = aE[X] + bE[Y]$ (ולא חשוב אם X, Y תלויים אם לאו).

ד. מונוטוניות: אם $X(\omega) \geq Y(\omega)$ לכל ω אזי $EX \geq EY$.

מומנטים מסדר גבוה יותר

מומנט שני: $E[X^2]$

מומנט מסדר n : $E[X^n]$

מומנט מוחלט מסדר n : $E[|X|^n]$

מומנט מרכזי מסדר n : $E[(X - \bar{X})^n]$ (ובפרט $E[X - \bar{X}] = 0$)

שוונות, וריאנס: $E[(X - \bar{X})^2] = \text{Var}(X)$

סטית התקן: $\sigma_X = \sqrt{\text{Var}(X)}$

הגדרה: X מ"א מנורמל אם $E[X] = 0, E[X^2] = 1$ ולכן אם $Z = \frac{X - E[X]}{\sigma_X}$ אזי Z מ"א מנורמל.

עבור משתנה אקראי בודד, חוק ההסתברות $(F_X(a), a \in \mathbb{R})$, מהווה את האינפורמציה הסטטיסטית המלאה אודותיו. בעקבות זאת אפשר לדבר על EX, EX^2, \dots . בכוון ההפוך, אם נתונים EX, EX^2, EX^3, \dots ואם $E^{1/n}|X|^n$ אינו עולה מהר מדי, אזי אפשר להראות שהמומנטים מגדירים את חוק ההסתברות של X .

לצורך הרבה בעיות טכניות חוק ההסתברות המלא לא ידוע ולא הכרחי; מספיק לדעת את EX ואת EX^2 (שני המומנטים הראשונים, שבדרך כלל אינם מגדירים את חוק ההסתברות) היכולים לתת תמונה כללית ולאפשר פתרון הבעיות. למשל, X מ"א, Y מ"א המתאר מדידה של X . אזי $E(X - Y)^2 / EX^2$ נותן הערכה טובה על השגיאה.

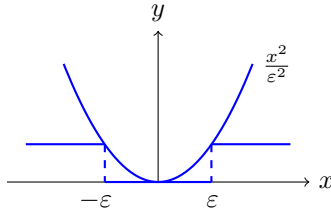
אי השוויון של צ'ביצ'ב

לפי תכונות א' וד' לעיל,

$$\begin{aligned} E[X^2] &\geq E[X^2 \mathbf{1}_{|X| \geq \varepsilon}] \\ &\geq \varepsilon^2 E[\mathbf{1}_{|X| \geq \varepsilon}] \\ &= \varepsilon^2 P(|X| \geq \varepsilon) \end{aligned}$$

ומכאן:

$$P(|X| \geq \varepsilon) = E[\mathbf{1}_{|X| \geq \varepsilon}] \leq \frac{EX^2}{\varepsilon^2}.$$



איור 1.8: המחשת אי-שוויון צ'בישב

אם נקח $X = Y - EY$ אזי $EX^2 = \text{Var } Y$ ומכאן:

$$P(|Y - EY| \geq \epsilon) \leq \frac{\text{Var } Y}{\epsilon^2}$$

אי שוויון זה שימושי כיוון שבמקרים רבים קל יותר לחשב מומנט שני של מ"א מאשר לחשב הסתברויות. כך ניתן לקבל הערכה (חסם עליון) על הסתברות של חריגה מהמוצע. לדוגמה, חישוב ההסתברות שמתוך 1,000,000 זריקות מטבע מספר ה"עץ" יהיה מעל 600,000 או מתחת 400,000 (כלומר סטיה של 100,000 מהמוצע) הוא קשה, אך קל לחשב חסם צ'ביצ'ב. זהו מקרה פרטי של משפט כללי יותר—ראה משפט 8.24.

הפונקציה האופינית של מ"א

הפונקציה האופינית של מ"א X , $\Phi_X(\nu)$, מוגדרת ע"י:

$$\Phi_X(\nu) = E[e^{j\nu X}] = E[\cos \nu X + j \sin \nu X]$$

(כאשר $j = \sqrt{-1}$). אם למשתנה האקראי X יש פילוג סגולי $f_X(\alpha)$ אזי

$$\Phi_X(\nu) = \int_{-\infty}^{\infty} e^{j\nu\alpha} f_X(\alpha) d\alpha$$

דהיינו $\phi_X(-\nu)$ היא התמרת פוריה של $f_X(\alpha)$. אפשר להראות שבכל מקרה, הפונקציה האופינית $\phi_X(\cdot)$ מגדירה חד משמעית את פונקציית הפילוג $F_X(\cdot)$.

יתכן ש- $E[X^n]$ לא קיים, אולם $\Phi_X(\nu)$ תמיד קיים כי $E[|\cos \nu X|]$ ו- $E[|\sin \nu X|]$ תמיד קיים (חסימות הסינוס והקוסינוס). בהנחות מתאימות (קיום מומנטים) נוכל לפרק את $e^{j\nu X}$ לטור חזקות ולהחליף את סדר הסיכום עם התוחלת ונקבל

$$\Phi_X(\nu) = \sum_{k=0}^{\infty} \frac{(j\nu)^k}{k!} E[X^k].$$

כלומר מתוך סדרת המומנטים אפשר לחשב את הפונקציה האופינית, וכאמור מתוכה ניתן לקבוע את הפילוג. הסקנו כי (בתנאים המתאימים) אוסף המומנטים קובע את הפילוג.

ושוב, בלי לבדוק תנאי קיום וכו', נגזור את הביטוי האחרון עבור $\Phi_X(\nu)$ לפי ν , ונקבל:

$$\phi_X(0) = 1; \quad \left. \frac{\partial \phi_X(\nu)}{\partial \nu} \right|_{\nu=0} = j EX; \quad \left(\frac{\partial^n \phi_X(\nu)}{\partial \nu^n} \right)_{\nu=0} = j^n EX^n$$

טענה: אם $Y = aX + b$ אזי $\phi_Y(\nu) = e^{j\nu b} \phi_X(a\nu)$

הוכחה:

$$\phi_Y(\nu) = E[e^{j\nu(aX+b)}] = e^{j\nu b} E[e^{i\nu a X}] = e^{j\nu b} \phi_X(\nu a)$$

שני משתנים אקראיים

אם Y, X מ"א

$$F_{X,Y}(x, y) = \mathbb{P}\{X \leq x, Y \leq y\}$$

$$f_{X,Y}(x, y) = \frac{\partial^2 F}{\partial x \partial y}$$

$$E[g(X, Y)] = \iint_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dx dy$$

ובמיוחד עבור $g(X, Y) = X^m Y^n$ נקבל את המומנטים המשולבים של X עם Y .

שני מ"א Y, X נקראים בלתי תלויים (ב"ת) סטטיסטית אם עבור כל x, y ממשיים מתקיים

$$\mathbb{P}\{X \leq x, Y \leq y\} = \mathbb{P}\{X \leq x\} \cdot \mathbb{P}\{Y \leq y\}$$

או

$$F_{X,Y}(x, y) = F_X(x) F_Y(y).$$

אם למשתנים יש צפיפות אזי שוויון זה שקול ל-

$$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y).$$

ללא הוכחה: Y, X ב"ת אמ"ם לכל שתי פונקציות חסומות g, h מתקיים $Eg(X)h(Y) = Eg(X)Eh(Y)$. מכאן נובע: עבור Y, X בלתי תלויים ו- $Z = X + Y$ מתקיים

$$\phi_Z(\nu) = \phi_X(\nu) \cdot \phi_Y(\nu)$$

הקווריאנס של זוג מ"א Y, X מוגדר ע"י:

$$\text{Cov}(X, Y) = E(X - \bar{X})(Y - \bar{Y})$$

אם $\text{Cov}(X, Y) = 0$ אומרים שהמ"א Y ו- X חסרי קורלציה או בלתי תלויים לינארית (בת"ל).

תרגיל: הוכח שאי תלות סטטיסטית גוררת אי תלות לינארית (בהנחה שהמומנטים קיימים), אבל ההיפך אינו בהכרח נכון.

דוגמה 1.2 יהי Θ מ"א המפולג אחיד בקטע $[0, 2\pi]$ ונגדיר $X = \sin \Theta$, $Y = \cos \Theta$. כמובן ש- X, Y תלויים סטטיסטית: אך קל לבדוק שהם חסרי קורלציה.

מקדם קורלציה

עבור שני משתנים אקראיים X_1 ו- X_2 נניח למען הפשטות $EX_1 = EX_2 = 0$. עבור איזה λ הביטוי $E(X_1 - \lambda X_2)^2$ מינימלי (או: כיצד ניתן לקרב את X_1 בעזרת פונקציה לינארית של X_2 כך שממוצע השגיאה הריבועית יהיה מינימלי)? ע"גזירה לפי λ נקבל ש- $\lambda^* = \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_2)}$ הוא האופטימלי. אם נציג את λ^* לביטוי עליו עשינו את המינימיזציה נקבל

$$E(X_1 - \lambda^* X_2)^2 = EX_1^2 - 2\lambda^* \text{Cov}(X_1, X_2) + (\lambda^*)^2 EX_2^2 = \text{Var } X_1 - \frac{(\text{Cov}(X_1, X_2))^2}{\text{Var}(X_2)} \geq 0$$

ומכאן קבלנו את אי שוויון קושי שזורץ

$$\text{Var}(X_1) \text{Var}(X_2) \geq (\text{Cov}(X_1, X_2))^2$$

למקדם

$$\rho = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1) \text{Var}(X_2)}}$$

קוראים מקדם הקורלציה ו- $|\rho| \leq 1$ לאור התוצאה האחרונה. אם X_1, X_2 בלתי תלויים לינארית, אזי $\rho = 0$. אם $\rho = 1$ אזי $X_2 = \alpha \cdot X_1$ עבור מקדם חיובי α כלשהו, ואם $\rho = -1$ אזי α הוא שלילי. במקרים אלו יש תלות לינארית במובן האלגברי. בהמשך נראה שביטוי זה מופיע בפתרון בעיות שנעסוק בהן.

1.4 וקטורים אקראיים

סימונים: $\underline{X}, \underline{a}$ יסמנו וקטורי עמודה. $\underline{B}, \underline{A}$ יסמנו מטריצות. יהיו $\underline{a}, \underline{b}$ וקטורים n מימדיים. אזי $\underline{b}^T \underline{a} = \sum a_i b_i$ הוא סקלר. לעומת זאת $\underline{a} \cdot \underline{b}^T$ היא מטריצה $n \times n$, שכן נזכר כי מכפלת מטריצות: $\underline{A}, \underline{B}$ מתאפשר להכפיל: $\underline{A} \cdot \underline{B}$ כאשר \underline{A} היא מטריצה $m \times n$ (m שורות n עמודות) ו- \underline{B} היא מטריצה $n \times k$. המטריצה המתקבלת היא $m \times k$.

יהי \underline{X} וקטור אקראי בעל n רכיבים

$$\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}; \quad \underline{X}^T = (X_1, X_2, \dots, X_n)$$

פונקציית הפילוג של הוקטור תסומן F_X —זוהי פונקציה של n משתנים, כמו גם הצפיפות (פ"ס) f_X כאשר הוא קיים. הגדרתם היא כמו עבור זוג מ"א לעיל. נגדיר תוחלת של וקטור על ידי

$$E\underline{X} \doteq \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix}.$$

כלומר תוחלת של וקטור מוגדרת להיות וקטור התוחלות. בצורה דומה תוחלת של מטריצה (אשר אבריה הם מ"א) מוגדרת להיות המטריצה אשר אבריה הם התוחלות של המ"א המתאימים.

* $i = 1, \dots, n, EX_i$ נקראים המומנטים מסדר ראשון ו- $E\underline{X}$ הוא וקטור הממוצע.

* $i = 1, 2, \dots, n, E|X_i|$ נקראים המומנטים המוחלטים מסדר ראשון.

* $E[\underline{X} \cdot \underline{X}^T] = \{E[X_i X_j]\}_{i,j=1}^n$ בצורה מטריצית שני. $1 \leq i \leq n, 1 \leq j \leq n, EX_i X_j$ היא מטריצת המומנטים מסדר שני.

* $E(X_i - EX_i)(X_j - EX_j)$ עבור $1 \leq i \leq n, 1 \leq j \leq n$ נקראים המומנטים המרכזיים מסדר שני, כלומר הקווריאנסים. בסימון מטריצני

$$\text{Cov}[\underline{X}, \underline{X}] \doteq E[(\underline{X} - E\underline{X})(\underline{X}^T - E\underline{X}^T)]$$

שים לב כי לפי ההגדרה $E[\underline{X}^T] = [E\underline{X}]^T$.

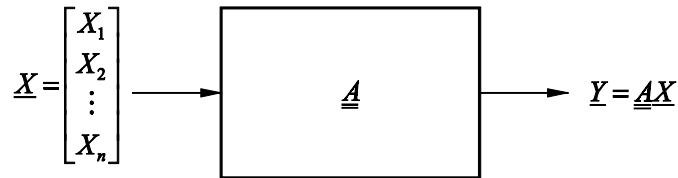
עבור וקטורים אקראיים, כאשר n גדול, חוק ההסתברות של וקטור \underline{X}

$$F_{X_1, X_2, \dots, X_n}(a_1, a_2, \dots, a_n) = \mathbb{P}\{X_1 \leq a_1, X_2 \leq a_2, \dots, X_n \leq a_n\}$$

(או בכתיב וקטורי $F_{\underline{X}}(\underline{a})$) יכול להיות מורכב ביותר ולא ידוע. אולם, כאמור, לבעיות רבות מספיק לדעת את המומנטים מסדר ראשון ושני. אחת הבעיות שנעסוק בהן בהמשך היא הבעיה הבאה: נתונים המומנטים מסדר ראשון ומסדר שני של \underline{X} והוקטור \underline{Y} מוגדר כדלקמן:

$$\underline{Y} = \underline{A}\underline{X} + \underline{b}$$

כאשר \underline{A} מטריצה קבועה, לא אקראית, ו- \underline{b} וקטור קבוע. מספר העמודות של \underline{A} צריך כמובן להיות שווה למספר השורות (אלמנטים) של \underline{X} . נניח לשם פשטות כי $b = 0$.



איור 1.9:

האם מתוך הנתון אפשר למצוא את המומנטים מסדר ראשון ומסדר שני של \underline{Y} ? שאלה זו היא כמובן הקדמה לשאלה מעניינת יותר—כאשר X הוא תהליך אקראי ו- A היא מערכת לינארית. השאלה כאן היא האם ניתן לחשב מתוך הנתונים את הממוצע והמומנט השני במוצא (Y) , תוך שימוש רק במידע על המומנט הראשון והשני בכניסה.

תוחלות

טענה: נתונים מטריצה \underline{A} וקטור \underline{b} בעלי מימדים מתאימים. אם $\underline{Y} = \underline{A}\underline{X} + \underline{b}$ אזי $E\underline{Y} = E[\underline{A}\underline{X} + \underline{b}] = \underline{A}E\underline{X} + \underline{b}$.

כלומר תוחלת היא פעולה לינארית.

הוכחה: התוצאה נובעת מהחשבון הישיר הבא:

$$E \left\{ \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & & & \\ \vdots & & & \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \right\} = E \left\{ \begin{pmatrix} a_{11}X_1 + a_{12}X_2 + \cdots + a_{1n}X_n + b_1 \\ \vdots \\ a_{m1}X_1 + a_{m2}X_2 + \cdots + a_{mn}X_n + b_m \end{pmatrix} \right\}$$

$$= \begin{pmatrix} a_{11}EX_1 + a_{12}EX_2 + \cdots + a_{1n}EX_n + b_1 \\ \vdots \\ a_{m1}EX_1 + a_{m2}EX_2 + \cdots + a_{mn}EX_n + b_m \end{pmatrix} = \underline{\underline{A}} \begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$$

הרחבה של הטענה האחרונה: תהיה $\underline{\underline{Y}}$ מטריצה אקראית (m שורות, n עמודות) עם אלמנטים אקראיים:

$$\underline{\underline{Y}} = \begin{pmatrix} Y_{11} & Y_{12} & \cdots & Y_{1n} \\ Y_{21} & & & \\ \vdots & & & \\ Y_{m1} & \cdots & Y_{mn} \end{pmatrix} = [Y_{ij}]$$

נזכר כי $E\underline{\underline{Y}} = [EY_{ij}]$. תהיה $\underline{\underline{A}}$ ו- $\underline{\underline{B}}$ מטריצות קבועות (לא אקראיות) בממדים מתאימים. אזי קל לראות ש- $E\underline{\underline{A}}\underline{\underline{Y}}\underline{\underline{B}} = \underline{\underline{A}}E\underline{\underline{Y}}\underline{\underline{B}}$ ולכן $E\underline{\underline{Y}}\underline{\underline{B}} = (E\underline{\underline{Y}}) \cdot \underline{\underline{B}}$ וכן $E\underline{\underline{A}}\underline{\underline{Y}} = \underline{\underline{A}}E\underline{\underline{Y}}$ כפל במטריצה דטרמיניסטית. הדבר כמובן אינו תמיד נכון עבור כפל במטריצה אקראית (מתי כן?).

מומנטים מסדר שני

נעבור כעת למומנטים מסדר שני וכאן נעסוק בוקטורים אקראיים בלבד. נעניין ב- $E[\underline{\underline{X}} \cdot \underline{\underline{X}}^T]$

$$E[\underline{\underline{X}} \cdot \underline{\underline{X}}^T] = E \left[\begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} (X_1, X_2, \dots, X_n) \right] = \begin{pmatrix} EX_1X_1 & EX_1X_2 & \cdots & EX_1X_n \\ EX_2X_1 & EX_2X_2 & \cdots & EX_2X_n \\ \vdots & \vdots & \ddots & \vdots \\ EX_nX_1 & \cdots & \cdots & EX_nX_n \end{pmatrix}$$

שים לב, מתקבלת מטריצה $n \times n$ סימטרית! מטריצה זו נקראת מטריצת המומנטים מסדר שני. בצורה דומה נגדיר את מטריצת הקווריאנס:

$$\{E(\underline{\underline{X}} - E\underline{\underline{X}})(\underline{\underline{X}} - E\underline{\underline{X}})^T\} = \{Cov(X_i, X_j)\}_{i,j} = \{E(X_i - \bar{X}_i)(X_j - \bar{X}_j)\}$$

הערה: עבור $n > 1$, $\det(\underline{X} \cdot \underline{X}^T)$ תמיד אפס, כיוון שזו מטריצה מדרגה 1. אבל זה לא בהכרח נכון ש- $\det(E\underline{X} \cdot \underline{X}^T) = 0$. משום שפעולת הדטרמיננטה אינה פעולה ליניארית! לדוגמא, אם $EX_1 = EX_2 = 0$, $EX_1X_2 = 0$, $EX_i^2 = 1$ עבור $i = 1, 2$, אזי מטריצת המומנטים היא מטריצת היחידה והדטרמיננטט שווה 1.

נניח שנתון ו"א \underline{X} וידוע $E\underline{X}$ וכן $E[\underline{X} \cdot \underline{X}^T]$. יהיה \underline{Y} מוגדר ע"י $\underline{Y} = \underline{A}\underline{X}$, אזי כבר הראינו בסעיף הקודם ש- $E\underline{Y} = \underline{A}E\underline{X}$. נשאלת השאלה מהי מטריצת המומנטים מסדר שני של \underline{Y} , כלומר,

$$E[\underline{Y} \cdot \underline{Y}^T] = ?$$

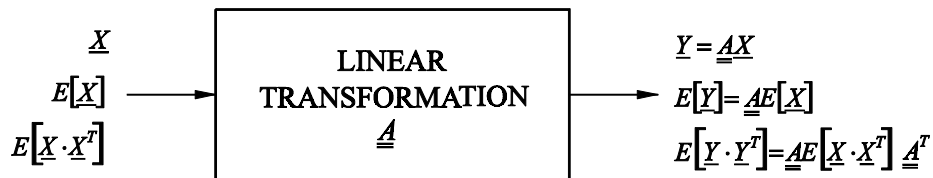
ניתן להשתכנע בנקל כי חשבון מפורש של כל איבר ואיבר במטריצה $\underline{Y} \cdot \underline{Y}^T$ יהיה מיגע¹. מושג המטריצות והכפל של מטריצות מאפשר לנו, לעומת זאת, להציג את התוצאה באופן פשוט מאד:

$$E[\underline{Y}\underline{Y}^T] = E[\underline{A}\underline{X}(\underline{A}\underline{X})^T] = E[\underline{A}\underline{X}\underline{X}^T\underline{A}^T] = \underline{A}E[\underline{X}\underline{X}^T]\underline{A}^T$$

ובצורה דומה:

$$\begin{aligned} (\text{Cov}(Y_i, Y_j))_{i,j} &= E[(\underline{Y} - E\underline{Y})(\underline{Y} - E\underline{Y})^T] = E[(\underline{A}\underline{X} - \underline{A}E\underline{X})(\underline{A}\underline{X} - \underline{A}E\underline{X})^T] \\ &= \underline{A}E[(\underline{X} - E\underline{X})(\underline{X} - E\underline{X})^T\underline{A}^T] = \underline{A}E[(\underline{X} - E\underline{X})(\underline{X} - E\underline{X})^T]\underline{A}^T \\ &= \underline{A}[\text{Cov}(X_i, X_j)]_{i,j}\underline{A}^T \end{aligned}$$

את התוצאות על המומנטים מסדר ראשון ושני כאשר מבצעים טרנספורמציה ליניארית על וקטור אקראי נסכם בציר 1.10.



איור 1.10:

תכונה בסיסית של מטריצת המומנטים מסדר שני

מטריצה ריבועית סימטרית C , $(n \times n)$, נקראת אי שלילית אם עבור כל וקטור n -מימדי \underline{a} , מתקיים:

$$\underline{a}^T \underline{C} \underline{a} \geq 0$$

(שים לב ש- $\underline{a}^T \underline{C} \underline{a}$ הוא סקלר) ואם, בנוסף, $\underline{a}^T \underline{C} \underline{a} = 0$ אך ורק כאשר כל רכיבי \underline{a} מתאפסים אזי \underline{C} נקראת חיובית או חיובית מוגדרת.

טענה: מטריצת המומנטים מסדר שני היא תמיד אי שלילית.

¹גרסה ב' מכילה דוגמא מפורטת יותר של כפל מטריצות.

הערה: מכיוון שמטריצת הקווריאנס היא מטריצת מומנטים מסדר שני, המשפט יראה שגם היא תמיד אי שלילית.

הוכחה: נקח $Y = \underline{a}^T \underline{X} = \underline{X}^T \underline{a}$. אזי Y הוא משתנה אקראי (סקלרי) ולכן:

$$0 \leq E[Y^2] = E[\underline{a}^T \underline{X} \underline{a}^T \underline{X}] = E[\underline{a}^T \underline{X} \underline{X}^T \underline{a}] = \underline{a}^T E[\underline{X} \underline{X}^T] \underline{a}$$

הערה: כפי שכבר הערנו, $\underline{X} \underline{X}^T$ הינה תמיד מטריצה סינגולרית אבל $E[\underline{X} \underline{X}^T]$ אינה בהכרח סינגולרית. על משמעות המקרה שבו $E[\underline{X} \underline{X}^T]$ סינגולרית נעמוד בהמשך. אפשר להראות שאם $E[\underline{X} \underline{X}^T]$ אינה סינגולרית אזי היא חיובית מוגדרת.

פונקציה אופיינית של וקטור אקראי

יהי \underline{X} וקטור אקראי ו- $\underline{\nu}$ וקטור דטרמיניסטי באותו מימד.

$$\underline{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} ; \quad \underline{\nu} = \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_n \end{pmatrix}$$

נגדיר

$$\phi_{\underline{X}}(\underline{\nu}) = E e^{j \sum_{k=1}^n X_k \nu_k} = E e^{j \underline{X}^T \underline{\nu}} = E e^{j \underline{\nu}^T \underline{X}} = E e^{j(\underline{X}, \underline{\nu})}$$

כאשר $(\underline{X}, \underline{\nu})$ מסמן את המכפלה הסקלרית של שני הוקטורים. זוהי מעין התמרת פוריה רב ממדית של פונקצית הפילוג (המשותפת). בפרט, אם לוקטור האקראי \underline{X} יש פונקצית צפיפות f_X אזי

$$\phi_{\underline{X}}(\underline{\nu}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} e^{j \sum_{k=1}^n \nu_k \alpha_k} f_X(\alpha_1, \dots, \alpha_n) d\alpha_1 \dots d\alpha_n .$$

ולכן, $\phi_{\underline{X}}(-\underline{\nu})$ היא בדיוק התמרת פוריה הרב-ממדית של פונקצית הצפיפות. חוק ההסתברות של \underline{X} מגדיר את $\phi_{\underline{X}}(\underline{\nu})$. בכיוון ההפוך, ללא הוכחה, ערכי הפונקציה $\phi_{\underline{X}}(\underline{\nu})$ עבור כל $\underline{\nu} \in \mathbb{R}^n$, מגדירים את חוק ההסתברות של הוקטור האקראי \underline{X} . זאת מאותו שיקול כמו במקרה החד ממדי—התמרת פוריה ההפוכה (רב ממדית) קובעת לחלוטין את פונקצית הפילוג. קל לראות שאם כל הרכיבים של \underline{X} בלתי תלויים אזי

$$\phi_{\underline{X}}(\underline{\nu}) = \prod_{i=1}^n \phi_{X_i}(\nu_i) \tag{1.2}$$

משום שאת הפונקציה האפיינית ניתן לרשום כמכפלה

$$\phi_{\underline{X}}(\underline{\nu}) = E \prod_{i=1}^n e^{j X_i \nu_i}$$

ותחת הנחת אי תלות, תוחלת המכפלה שווה למכפלת התוחלות. ולהיפך, אפשר להראות שאם הפונקציה האופיינית נתנת לייצוג כמכפלה כנ"ל, אזי רכיבי הוקטור \underline{X} הינם בלתי תלויים. זאת משום שהתמרת פוריה ההפוכה תראה שפונקצית הפילוג היא המכפלת של הפונקציות החד ממדיות. שם לב שבדרך כלל העובדה שתוחלת המכפלה שווה למכפלת התוחלות אינה גוררת אי תלות סטטיסטית! אך כאן נתון הרבה יותר—תוחלת המכפלה שווה למכפלת התוחלות עבור מספר גדול של ביטויים (אחד לכל ערך של $\underline{\nu}$) ובנוסף עבור פונקציות מיוחדות (אקספוננציאליות).

ראינו כיצד משנה טרנספורמציה לינארית את המומנט הראשון והשני. מסתבר שניתן לתת נוסחה סגורה גם עבור הפונקציה האפיינית.

שאלה: ידוע $\phi_{\underline{X}}(\underline{u})$, $\underline{u} \in \mathbb{R}^n$. נבצע את הטרנספורמציה המוכרת $\underline{Y} = \underline{A}\underline{X}$. מה נוכל לומר על $\phi_{\underline{Y}}(\underline{\nu})$? שים לב ש- X, Y יכולים להיות ממימד שונה!

תשובה:

$$\phi_{\underline{Y}}(\underline{\nu}) = E e^{j\underline{Y}^T \underline{\nu}} = E e^{j(\underline{A}\underline{X})^T \underline{\nu}} = E e^{j\underline{X}^T \underline{A}^T \underline{\nu}} = \phi_{\underline{X}}(\underline{A}^T \underline{\nu})$$

דוגמה 1.3 נראה טכניקה להעלמת משתנה: נתון הפלוג $F_{X_1, X_2, X_3}(a_1, a_2, a_3)$ של X_1, X_2, X_3 . מעוניינים רק ב- X_1, X_2 , אזי

$$F_{X_1, X_2}(a_1, a_2) = F_{X_1, X_2, X_3}(a_1, a_2, \infty)$$

נרשום $\underline{Y} = \underline{A}\underline{X}$ כאשר

$$\underline{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}; \quad n = 3$$

ועבור הפונקציה האופיינית נקבל:

$$\phi_{\underline{Y}}(\nu_1, \nu_2) = \phi_{\underline{X}} \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \nu_1 \\ \nu_2 \end{pmatrix} \right) = \phi_{\underline{X}}(\nu_1, \nu_2, 0)$$

מסקנה: באופן כללי, אם רוצים להעלים משתנה כגון X_i אזי לוקחים את $\phi_{\underline{X}}(\underline{\nu})$ ומציבים בו $\nu_i \equiv 0$.

דוגמה 1.4 סכום משוקלל של רכיבי ווקטור:

$$Y = \underline{a}^T \underline{X} = a_1 X_1 + \dots + a_n X_n = (a_1, \dots, a_n) \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

במקרה זה Y ו- ν הם חד מימדיים

$$\phi_Y(\nu) = \phi_{\underline{X}} \left(\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \nu \right) = \phi_{\underline{X}} \begin{pmatrix} a_1 \nu \\ \vdots \\ a_n \nu \end{pmatrix}$$

ובמיוחד, אם הרכיבים של \underline{X} ב"ת ו- $a_1 = a_2 = \dots = a_n = 1$ אזי

$$\phi_Y(\nu) = \phi_{X_1}(\nu) \phi_{X_2}(\nu) \dots \phi_{X_n}(\nu)$$

מסקנה: הפונקציה האופינית של סכום מ"א ב"ת היא מכפלת הפונקציות האופיניות.

נשים לב כי תוצאה זו שונה מ-(1.2): במקרה הקודם קיבלנו פונקציה אפינית של מספר משתנים. כאן כיוון שמדובר בסכום—ולכן במשתנה סקלרי—קיבלנו פונקציה אפינית של משתנה סקלרי.

הערה: יהיו Y, X מ"א. נפתח את האקספוננטים לטור טיילור ונקבל

$$E\left[e^{jX\nu_1+jY\nu_2}\right] = E\left\{\sum_{m=0}^{\infty} \frac{(j\nu_1)^m}{m!} X^m \cdot \sum_{k=0}^{\infty} \frac{(j\nu_2)^k}{k!} Y^k\right\}$$

ולכן אם נפתח את $\phi_{X,Y}(\nu_1, \nu_2)$ לטור חזקות ב- ν_1, ν_2 נוכל לקבל את המומנטים $E X^m Y^k$ (בתנאי שהם קיימים). באופן מעשי ניתן לקבל מומנטים כאלו על ידי גזירה של $\phi_{X,Y}(\nu_1, \nu_2)$ לפי שני המשתנים. מסיבה זו הפונקציה האפינית משמשת ככלי נוח לחישוב מומנטים גבוהים משותפים.

דוגמה 1.5 התמרות פוריה עוזרות להפוך בעיות המתוארות על ידי משוואות דיפרנציאליות, לבעיות אלגבריות. כך למשל חישוב תגובת התדר היא בעיה אלגברית. בצורה דומה ניתן להשתמש בפונקציה האפינית: למשל, נניח ש- X, N הם מ"א בת"ס (נחשוב על X כעל "אות רצוי" ועל N כ"רעש"), וידוע הפילוג של כל אחד מהם. נגדיר "מדידה רועשת" $Y = X + N$. חישוב הפילוג של Y כרוך בביצוע קונוולוציה—חישוב שאולי אינו פשוט אך ניתן ניתן עקרונית לביצוע. כעת נניח שידוע הפילוג של ה"רעש" N ושל ה"מדידה" Y , ואנו רוצים לחשב את הפילוג של ה"אות" X . לשם כך יש "להפוך" פעולת קונוולוציה: פעולה זו נקראת deconvolution. זאת לא ניתן, בדרך כלל, לעשות בצורה ישירה. אולם בגלל אי התלות הסטטיסטית נקבל

$$\phi_Y(\nu) = \phi_X(\nu) \cdot \phi_N(\nu) \quad \Rightarrow \quad \phi_X(\nu) = \frac{\phi_Y(\nu)}{\phi_N(\nu)}.$$

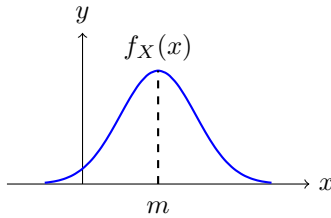
קיבלנו פתרון "מפורש".

תרגיל 1.6 חשב במפורש את הפילוג של X כאשר N מפולג אקספוננציאלית עם פרמטר μ ו- Y מפולג אקספוננציאלית עם פרמטר λ .

1.5 הפילוג הגאוסי

הפילוג הגאוסי חשוב במיוחד בהנדסת חשמל ממספר סיבות. ראשית, מודלים גאוסיים הם מקובלים עבור סוגי בעיות רבים, בפרט בתחומי התקשורת והבקרה, ונובעים פעמים רבות מתוך שיקולים פיסיקליים. שנית, קל יחסית לבצע חישובים עם פילוגים גאוסיים, ולכן משתמשים בו כקירוב נוח, וכן כאשר אנו מתעניינים רק במומנט ראשון ושני. הפילוג הגאוסי הוא גם המקרה הגרוע ביותר בהרבה בעיות (עוד נראה זאת), כך שאם אין אנו יודעים את הפילוג האמיתי במלואו, אבל כן יודעים את המומנט הראשון והשני, אזי ההנחה של פילוג גאוסי היא ברוח תכנון worst-case. לבסוף, משפט הגבול המרכזי מראה לנו כי הפילוג הגאוסי מהווה קירוב טוב לאוסף גדול של משתנים קטנים.

המקרה הסקלרי



איור 1.11: הצפיפות הגאוסית

יהא X מ"א גאوسی. כזכור, הצפיפות הגאוסית החד מימדית מוגדרת ע"י:

$$(1.3) \quad f_X(\alpha) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\alpha-m)^2}{2\sigma^2}}$$

$$E[X] = m, \quad \text{Var}(X) = \sigma^2$$

והפונקציה האופינית (ללא הוכחה) של מ"א גאوسی נתונה ע"י:

$$(1.4) \quad \phi_X(\nu) = e^{jm\nu - \frac{1}{2}\sigma^2\nu^2}$$

(מה בנוגע למקרה המנוון $\sigma^2 = 0$, כלומר $X \equiv m$.) עבור $Y = aX + b$

$$\phi_Y(u) = E e^{jaXu + jbu} = e^{jbu} e^{jamu - \frac{1}{2}a^2\sigma^2u^2} = e^{ju(b+am) - \frac{1}{2}a^2\sigma^2u^2}$$

ואם נסכים להרחיב את משפחת המשתנים האקראיים הגאוסים ע"י הכללת כל המ"א המנוונים ($\sigma^2 = 0$) לתוך המשפחה אזי נקבל את התוצאה שטרנספורמציה לינארית של מ"א גאוסי היא תמיד מ"א גאוסי (ללא צורך בהוספת דרישות מיוחדות). כפי שנראה בהמשך נוח וכדאי לעשות הכללה זו ולכן נגדיר את המ"א הגאוסי לפי (1.3) בצרוף המקרה המנוון, או ישירות לפי (1.4).

אנו נראה שלעיתים המעבר לפונקציה אפיינית מאפשר טיפול ישיר יותר בבעיה מאשר דרך הצפיפות. הדבר דומה לשימוש בהתמרת פוריה בנושאי ניתוח אותות.

המקרה הוקטורי

הגדרה: וקטור אקראי (ו"א) \underline{X} n -מימדי נקרא גאוסי אם עבור כל וקטור דטרמיניסטי n מימדי \underline{a} , המשתנה האקראי $\underline{a}^T \underline{X}$ הוא מ"א גאוסי. (אפשר גם לדרוש $|\underline{a}| = 1$; זה לא ישנה דבר. אז נוכל להגיד שו"א הוא גאוסי אם כל היטל שלו לכל כוון הוא מ"א גאוסי).

אם ו"א \underline{X} הוא גאוסי אזי כל רכיב שלו הוא היטל ולכן כל רכיב הוא מ"א גאוסי. מה עם ההפך? אפשר להביא דוגמה של וקטור אקראי שבמערכת קואורדינטות מסוימת כל הרכיבים הם גאוסיים אולם הוקטור האקראי איננו וקטור אקראי גאוסי! ולכן, אם נתון וקטור אקראי שכל רכיביו גאוסיים זה עדיין לא גורר שהוקטור הוא גאוסי.

דוגמה 1.7 יהי X מ"א גאוסי עם ממוצע אפס ו- S מ"א בינרי בת"ס ב- X המקיים $S = \pm 1$ בהסתברויות שוות. אזי $Y \doteq X \cdot S$ גאוסי, $X + Y$ אינו גאוסי, שכן הוא שווה 0 בהסתברות 1/2 ולכן (X, Y) אינו ווקטור גאוסי.

תכונות של וקטורים אקראיים גאומיים:

א. טענה: אם \underline{X} ו"א גאומי, אזי (1) גם $\underline{X} + \underline{b}$ ו"א גאומי; (2) גם $\underline{Y} = \underline{A}\underline{X}$ ו"א גאומי.

הוכחת (2):

$$(a, \underline{Y}) = (a, \underline{A}\underline{X}) = a^T \underline{A}\underline{X} = (\underline{A}^T a)^T \underline{X}$$

ולכן כל היטל של \underline{Y} בכיוון מסוים הוא, בסופו של דבר, היטל של \underline{X} (לא בהכרח באותו כיוון כמובן), ולכן \underline{Y} ו"א גאומי.

ב. טענה: \underline{X} ו"א גאומי n -מימדי אם ורק אם קיים וקטור n -מימדי \underline{m} ומטריצה $\underline{\Lambda}$ סימטרית ואי שלילית $n \times n$ כך שהפונקציה האופינית של \underline{X} נתונה ע"י:

$$\begin{aligned} \phi_{\underline{X}}(\underline{\nu}) &= e^{j(\underline{\nu}^T \underline{m}) - \frac{1}{2} \underline{\nu}^T \underline{\Lambda} \underline{\nu}} \\ &= e^{j \sum \nu_i m_i - \frac{1}{2} \sum_i \sum_k \nu_i \nu_k \lambda_{ik}} \end{aligned}$$

(***)

ואז מתקיים

$$E\underline{X} = \underline{m}$$

$$\underline{\Lambda} = E(\underline{X} - E\underline{X})(\underline{X} - E\underline{X})^T = [\lambda_{ij}]_{i,j} = [\text{Cov}(X_i, X_j)]_{i,j}$$

הוכחה: נניח ש- \underline{X} הוא וקטור גאומי. נסמן את וקטור הממוצעים ואת מטריצת הקווריאנס ב-

$$\underline{m} \doteq \mathbb{E}[\underline{X}]$$

$$\underline{\Lambda} \doteq \mathbb{E}[(\underline{X} - \underline{m})(\underline{X} - \underline{m})^T]$$

נסמן את אברי המטריצה ב- $\underline{\Lambda} \doteq \{\lambda_{ij}\}_{i,j}$. נבחר וקטור $\underline{\nu}$ ונגדיר מ"א חדש $Y = \underline{\nu}^T \underline{X}$. מהגדרת וקטור גאומי נובע כי Y מ"א גאומי. מלינאריות התוחלת נובע כי הממוצע שלו הוא

$$m_Y = \underline{\nu}^T \underline{m}$$

ומחישוב קודם של קווריאנסים נקבל שהווריאנס של Y הוא

$$\sigma_Y^2 = \underline{\nu}^T \underline{\Lambda} \underline{\nu} = \sum_{i,j=1}^n \nu_i \nu_j \lambda_{ij}$$

כוון ש- Y מ"א גאומי, הפונקציה האפיינית שלו בנקודה 1 היא

$$\phi_Y(1) = \mathbb{E}[e^{jY}] = e^{j \underline{\nu}^T \underline{m} - \frac{1}{2} \underline{\nu}^T \underline{\Lambda} \underline{\nu}}$$

אולם מהגדרת Y ,

$$\begin{aligned}\mathbb{E}[e^{jY}] &= \mathbb{E}\left[e^{j\sum_{i=1}^n \nu_i X_i}\right] \\ &= \mathbb{E}\left[e^{j\underline{\nu}^T \underline{X}}\right] \\ &= \phi_X(\underline{\nu})\end{aligned}$$

כיוון שהחישוב נכון לכל $\underline{\nu}$, הוכחנו כי לוקטור גאוסי יש פונקציה אפיינית כדרוש, כאשר הוקטור \underline{m} והמטריצה $\underline{\Lambda}$ המופיעים בפונקציה האפיינית הם בדיוק ווקטור הממוצעים ומטריצת הקווריאנס. להוכחת הכיוון השני, נניח שלפונקציה האפיינית יש את הצורה הנתונה. נתונים הוקטור \underline{m} והמטריצה $\underline{\Lambda}$. נבחר ווקטור \underline{a} ונגדיר שוב

$$Y = \underline{a}^T \underline{X}.$$

אם נוכיח כי לכל \underline{a} , Y הינו מ"א גאוסי, הוכחנו על-פי ההגדרה כי \underline{X} הוא גאוסי. מתכונות הפונקציה האפיינית,

$$\begin{aligned}\phi_Y(u) &= \phi_X(\underline{a} \cdot u) \\ &= \phi_X([a_1 \cdot u, a_2 \cdot u, \dots, a_n \cdot u]^T) \\ &= e^{j(\underline{a}^T \underline{m})u - \frac{1}{2}u^2(\underline{a}^T \underline{\Lambda} \underline{a})}\end{aligned}$$

כאשר השוויון האחרון נובע מההנחה על \underline{X} ובהצבת $\nu_i = a_i u$. מכאן נובע כי Y הוא מ"א גאוסי עם ממוצע $\underline{a}^T \underline{m}$ ושונות $\underline{a}^T \underline{\Lambda} \underline{a}$, וכיוון שהדבר נכון לכל \underline{a} , הרי מההגדרה \underline{X} הוא וקטור אקראי גאוסי. מהוכחת הכיוון הראשון אנו יודעים כי עבור ווקטור גאוסי, הוקטור \underline{m} והמטריצה $\underline{\Lambda}$ המופיעים בפונקציה האפיינית הם בדיוק ווקטור הממוצעים ומטריצת הקווריאנס.

ג. נניח ש- $\underline{\Lambda}$ מטריצה אלכסונית, $\lambda_{ik} = \sigma_i^2 \delta_{ik}$ (כאשר $\delta_{ik} = 0$ עבור $i \neq k$ ו $\delta_{ii} = 1$ עבור כל i). אזי משמע שהמשתנים האקראיים X_1, X_2, \dots בלתי תלויים לינארית בזוגות, כלומר $\text{Cov}(X_i, X_k) = 0, i \neq k$. אם בנוסף \underline{X} ו"א גאוסי, אזי

$$\phi_{\underline{X}}(\underline{\nu}) = e^{j\sum \nu_k m_k - \frac{1}{2}\sum \lambda_{kk} \nu_k^2} = \prod_{k=1}^n e^{j\nu_k m_k - \frac{1}{2}\lambda_{kk} \nu_k^2}$$

ולכן, תמיד נכון ש- $\underline{\Lambda}$ אלכסונית אם ורק אם קיימת אי תלות לינארית בזוגות אולם במקרה הגאוסי מתקיימת גם אי תלות סטטיסטית, ולכן עבור וקטור אקראי גאוסי אי תלות לינארית גוררת אי תלות סטטיסטית.

הערה: אם \underline{X} הוא ו"א, שרכיביו הם בת"ס וכל אחד מהם הוא מ"א גאוסי, אזי \underline{X} הוא ו"א גאוסי. לעומת זאת המשתנים X, Y שבדוגמה 1.7 הם חסרי קורלציה, כל אחד מהם הוא גאוסי אך הוקטור אינו גאוסי! כלומר, וקטור גאוסי שרכיביו חסרי קורלציה—בהכרח הרכיבים בת"ס, אולם וקטור שרכיביו גאוסיים אינו בהכרח ווקטור גאוסי, אפילו אם רכיביו חסרי קורלציה.

ד. שאלה: מתי m הרכיבים הראשונים של ווקטור אקראי גאוסי n -מימדי בלתי תלויים ב- $(n - m)$ הרכיבים הנותרים?

תשובה (ללא הוכחה): כאשר למטריצת הקווריאנס הצורה:

$$\underline{\underline{\Lambda}} = \begin{pmatrix} m \times m & 0 \\ 0 & (n-m) \times (n-m) \end{pmatrix}$$

ה. שאלה: אם \underline{X} ו"א גאוסית, רכיבים ב"ת, $\underline{Y} = \underline{A}\underline{X}$ אזי בדרך כלל הרכיבים של \underline{Y} הם תלויים. מה בכיון ההפוך? האם ניתן להגיע לווקטור אשר רכיביו ב"ת" על ידי טרנספורמציה לינארית? לכך יש חשיבות כי אם הרכיבים ב"ת" אזי בבעיות רבות ניתן לטפל בכל רכיב בנפרד, ובכך מפשטים הן את הבעיה והן את סיבוכיות החישובים. התשובה נתונה ע"י: טענה: אם \underline{X} ו"א גאוסית n -מימדי (לשם פשטות נדון כאן במקרה של תוחלת אפס אבל זה כלל לא חשוב), אזי קיימת מטריצה לא סינגולרית $n \times n$, נקרא לה \underline{D} כך ש- $\underline{Y} = \underline{D}\underline{X}$ והרכיבים של \underline{Y} בלתי תלויים סטטיסטית! הערה: יתכן שחלק מהרכיבים של \underline{Y} יהיו קבועים (ולא מ"א) – כך שבאופן מעשי, החלק האקראי של \underline{Y} יהיה בעל מימד קטן ממש מהמימד של \underline{X} . למשל, $\underline{X} = \begin{pmatrix} X_1 \\ X_1 \end{pmatrix}$ כאשר X_1 מ"א גאוסית ו \underline{Y} מוגדר ע"י:

$$\underline{Y} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \sqrt{2} & \sqrt{2} \\ -\sqrt{2} & \sqrt{2} \end{pmatrix} \begin{pmatrix} X_1 \\ X_1 \end{pmatrix} = \begin{pmatrix} \sqrt{2}X_1 \\ 0 \end{pmatrix}$$

אזי הרכיבים של \underline{Y} בלתי תלויים. נסמן

$$\underline{\underline{D}} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} ; \det \underline{\underline{D}} = 1$$

שים לב ש- $\underline{\underline{D}}$ מקיימת

$$\underline{\underline{D}}\underline{\underline{D}}^T = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \underline{\underline{D}}^T \underline{\underline{D}}$$

ולכן $\underline{\underline{D}}^T = \underline{\underline{D}}^{-1}$.

לצורך הוכחת הטענה, נקבל ללא הוכחה את המשפט הבא: אם $\underline{\underline{\Lambda}}$ מטריצה סימטרית לא שלילית אזי קיימת מטריצת אוניטרית ($\underline{\underline{D}}^T = \underline{\underline{D}}^{-1}$) ולכן כמובן לא סינגולרית) כך שהמטריצה $\underline{\underline{D}}\underline{\underline{\Lambda}}\underline{\underline{D}}^T$ היא אלכסונית, כלומר

$$\underline{\underline{D}}\underline{\underline{\Lambda}}\underline{\underline{D}}^T = \begin{pmatrix} c_1 & \dots & 0 \\ \vdots & c_2 & \vdots \\ 0 & \dots & c_n \end{pmatrix} \quad (1.5)$$

מיד נחזור ונעיין ב-(1.5) אולם לפני זה נוכיח שלוקטור האקראי \underline{Y} המתקבל מתוך $\underline{Y} = \underline{D}\underline{X}$ רכיבים ב"ת לינארית (זכור $E\underline{X} = 0$):

$$E[\underline{Y}\underline{Y}^T] = E[\underline{D}\underline{X}\underline{X}^T \underline{D}^T] = \underline{\underline{D}}\underline{\underline{\Lambda}}\underline{\underline{D}}^T = \begin{pmatrix} c_1 & \dots & 0 \\ \vdots & c_2 & \vdots \\ 0 & \dots & c_n \end{pmatrix}$$

כעת נוסיף את הנחת הגאוסיות ולכן אי תלות לינארית גוררת אי תלות סטטיסטית, ולכן הרכיבים של \underline{Y} בלתי תלויים. (באופן כללי יותר:

$$E e^{j\underline{\nu}^T \underline{Y}} = E e^{j\underline{\nu}^T \underline{D} \underline{X}} = e^{j\underline{\nu}^T \underline{D}^{m-\frac{1}{2}} \underline{\nu}^T \underline{D} \underline{\Lambda} \underline{D}^T \underline{\nu}} = e^{j\underline{\nu}^T \underline{D}^{m-\frac{1}{2}} \sum_i \nu_i^2 c_i}$$

ולכן הרכיבים של \underline{Y} בלתי תלויים גם אם התוחלת של \underline{X} שונה מאפס.)

נחזור ל-(1.5): איך מוצאים את המטריצה \underline{D} ? נסתפק בהערה הבאה: נסמן $\underline{D}^T = (d_1, d_2, \dots, d_n)$ כאשר d_i וקטור n -מימדי. אזי נוכל לרשום:

$$\underline{\underline{\Lambda}} \underline{\underline{D}}^T = (\underline{\underline{\Lambda}} d_1, \underline{\underline{\Lambda}} d_2, \dots, \underline{\underline{\Lambda}} d_n)$$

$$\underline{\underline{D}}^T \begin{pmatrix} c_1 & \dots & 0 \\ \vdots & c_2 & \vdots \\ 0 & \dots & c_n \end{pmatrix} = (c_1 d_1, c_2 d_2, \dots, c_n d_n)$$

לכן נובע מתוך (1.5), היות ו- \underline{D} אוניטרית:

$$\underline{\underline{C}} = \underline{\underline{D}} \underline{\underline{\Lambda}} \underline{\underline{D}}^T$$

$$\underline{\underline{D}}^T \underline{\underline{C}} = \underline{\underline{\Lambda}} \underline{\underline{D}}^T$$

$$(c_1 d_1, c_2 d_2, \dots, c_n d_n) = (\underline{\underline{\Lambda}} d_1, \underline{\underline{\Lambda}} d_2, \dots, \underline{\underline{\Lambda}} d_n)$$

כאשר

$$\underline{\underline{C}} \triangleq \begin{pmatrix} c_1 & \dots & 0 \\ \vdots & c_2 & \vdots \\ 0 & \dots & c_n \end{pmatrix}$$

וע"י השוואת עמודות נקבל $\underline{\underline{\Lambda}} d_i = c_i d_i$. לכן, הוקטורים d_i פותרים את המשוואה $\underline{\underline{\Lambda}} d = cd$, במילים אחרות c_i הם הערכים העצמיים של $\underline{\underline{\Lambda}}$ ו- d_i הם הוקטורים העצמיים של $\underline{\underline{\Lambda}}$ (ולכן גם מתקיים $\underline{\underline{D}}^T = \underline{\underline{D}}^{-1}$). מן הראוי לציין שמטריצה \underline{D} זו אינה בהכרח המטריצה היחידה שמלכסנת את מטריצת הקווריאנס של ו"א נתון.

1. ללא הוכחה: אם $\underline{\underline{\Lambda}}$ לא סינגולרית אזי הפילוג הסגולי ה- n מימדי הגאוסית נתון ע"י:

$$f_{\underline{X}}(a) = \frac{1}{(2\pi)^{n/2} (\det \underline{\underline{\Lambda}})^{1/2}} e^{-\frac{1}{2} (a-m)^T \underline{\underline{\Lambda}}^{-1} (a-m)} = \frac{1}{(2\pi)^{n/2} (\det \underline{\underline{\Lambda}})^{1/2}} e^{-\frac{1}{2} \sum (a_i - m_i)(a_j - m_j) \theta_{ij}}$$

כאשר: $\underline{\underline{\Lambda}}^{-1} = [\theta_{ij}]$

שים לב כי כדי לרשום את הצפיפות בצורה מפורשת יש לחשב את ההופכי של המטריצה $\underline{\underline{\Lambda}}$. לצורך חישוב מפורש של הפונקציה האפיינית חישוב כזה אינו דרוש.

לנוסחה זו ניתן להגיע על ידי חישוב הצפיפות של ו"א גאוסית באותו מימד אך עם רכיבים בת"ס, וביצוע טנספורמציה לינארית על ידי מטריצה A המקיימת $\underline{\underline{\Lambda}} = AA^T$

סיכום

- א. תכונת הגאוסיות נשמרת תחת טרנספורמציות לינאריות.
- ב. הפילוג של ו"א גאوسی X נקבע חד משמעית ע"י $\underline{\mu}, \underline{\sigma}$ (שני המומנטים הראשונים).
- ג. אי תלות לינארית במקרה הגאואסי שקולה לאי תלות סטטיסטית.
- ד. ניתן לעבור למ"א ב"ת על ידי טרנספורמציה לינארית-יוניטרית.

1.6 הסתברות מותנית

Y, X מ"א; נניח X מ"א בדיד אזי הפילוג המותנה של Y כאשר נתון X היא

$$F_{Y|X}(y|x) = \mathbb{P}\{Y \leq y | X = x\} = \frac{\mathbb{P}\{Y \leq y, X = x\}}{\mathbb{P}\{X = x\}}$$

ואם X מ"א רציף

$$F_{Y|X}(y|x) = \lim_{\varepsilon \rightarrow 0} \frac{\mathbb{P}\{Y \leq y; x < X \leq x + \varepsilon\}}{\mathbb{P}\{x < X \leq x + \varepsilon\}}$$

פילוג סגולי מותנה

$$f_{Y|X}(y|x) = \frac{\partial F_{Y|X}(y|x)}{\partial y} = \frac{f_{Y,X}(y,x)}{f_X(x)}$$

ובמיוחד, כאשר Y, X בלתי תלויים (והצפיפויות קיימות) אזי

$$f_{Y|X}(y|x) = f_Y(y) \quad f_{X|Y}(x|y) = f_X(x).$$

אי-תלות מותנית

Y, X ו- Z מ"א; Y ו- X בלתי-תלויים סטטיסטית בהנתן Z אם

$$F_{X,Y|Z}(x,y|z) = F_{X|Z}(x|z)F_{Y|Z}(y|z)$$

ואם המ"א בעלי צפיפות אזי תנאי זה שקול לתנאי על הצפיפויות:

$$f_{X,Y|Z}(x,y|z) = f_{X|Z}(x|z)f_{Y|Z}(y|z)$$

תוחלת מותנית

$$E[Y|X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) \cdot dy$$

ובאופן כללי יותר

$$E[Y|X = x] = \int_{-\infty}^{\infty} y dF_{Y|X}(y|x)$$

שיב לב שהתוחלת של Y מותנה ב- $x = X$ היא פונקציה של x . אפשר להראות שגם במקרה זה קיים הקשר:

$$E[g(Y)|X = x] = \int_{-\infty}^{\infty} g(y)f_{Y|X}(y|x)dy$$

ובאופן כללי יותר

$$E[g(Y)|X = x] = \int_{-\infty}^{\infty} g(y)dF_{Y|X}(y|x)$$

ושבו זו פונקציה של ההתניה. כלומר אם אנו יודעים את הפילוג (או הצפיפות) המותנים של Y בהנתן $X = x$, אזי לצורך חישוב התוחלת המותנית של $Z = g(Y)$ בהנתן $X = x$ אין צורך לחשב את הפילוג (או הצפיפות) המותנים של Z בהנתן $X = x$.

כמו תוחלת רגילה, גם התוחלת המותנית היא ליניארית. אם $Y = aZ_1 + bZ_2$ אזי לא קשה להראות ש:

$$E[Y|X = x] = aE[Z_1|X = x] + bE[Z_2|X = x]$$

לכל ערך של x . כמו כן, כאשר X ו- Y בלתי תלויים, מתקיים: $E[Y|X = x] = E[Y]$.

מכאן שאם $Y = X + N$ כאשר X ו- N ב"ת אזי

$$E[Y|X = x] = E[X|X = x] + E[N|X = x] = x + E[N]$$

כאמור $E[Y|X = x]$ היא פונקציה של x , נסמן פונקציה זאת ב- $\Psi(x)$. אם כעת נציב בפונקציה זאת את המשתנה האקראי X עצמו, תהיה התוצאה, $\Psi(X)$, משתנה אקראי חדש. נהוג לסמן משתנה אקראי זה ב- $E[Y|X] = \Psi(X)$ כדהיינו: $E[Y|X] = \Psi(X)$ כאשר $\Psi(x) = E[Y|X = x]$.

שים לב: כאשר X ו- Y בלתי תלויים, $E[Y|X] = E[Y]$.

תכונות נוספות של תוחלת מותנית:

$$E[X|X = x] = x \quad (\alpha)$$

$$E[h(Y)g(X)|X = x] = g(x)E[h(Y)|X = x] \quad (\beta)$$

ובאופן כללי יותר: $E[h(X, Y)g(X)|X = x] = g(x)E[h(x, Y)|X = x]$

$$E[Y] = E[E[Y|X]] \quad (\gamma)$$

הערה: תכונות (ב) ו-(ג) נקראות תכונות החלקה. זהו כלי חשוב במיוחד, משום שבהרבה מקרים קל יותר לחשב את התוחלת בשני צעדים מאשר בצעד אחד. נעייך בניסוח הבא של תכונת החלקה:

$$E[h(X, Y)g(X)] = E[g(X)E[h(X, Y)|X]] \quad (1.6)$$

הוכחה:

$$\begin{aligned} E[g(X)E[h(X, Y)|X]] &= \int_{-\infty}^{\infty} g(x) \left[\int_{-\infty}^{\infty} h(x, y) f_{Y|X}(y|x) dy \right] f_X(x) dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) h(x, y) f_{Y|X}(y|x) f_X(x) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) h(x, y) f_{Y, X}(y, x) dy dx = E[g(X)h(X, Y)] \end{aligned}$$

תוצאות חשובות ללא הוכחה:

א. אם $\underline{Y} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ ו"א גאוסי אזי חוק ההסתברות של X_1 כאשר נתון X_2 הוא גם כן גאוסי. (ואז, בנוסחת הפילוג הסגולי המותנה של X_1 נתון X_2 או בנוסחת הפונקציה האופינית המותנית, תופיע התוחלת המותנית של X_1 נתון X_2 והפיזור המותנה. מתברר שבמקרה זה הפיזור המותנה איננו תלוי בצורה מפורשת בהתניה).

ב. אם (X, \underline{Y}) ו"א אקראי גאוסי ($n+1$ מימדי) $EX = 0, E\underline{Y} = 0$ אזי קיים וקטור n מימדי \underline{a} כך ש:

$$E[X|\underline{Y}] = \underline{a}^T \underline{Y}$$

נראה את ההוכחה בהמשך.

הערה: שכנעו את עצמכם שמתוצאה זאת נובע, שאם (X, \underline{Y}) הוא וקטור אקראי גאוסי, ללא הדרישה של תוחלת אפס, אזי קיים \underline{a} כך ש:

$$E[X|\underline{Y}] = E[X] + \underline{a}^T (\underline{Y} - E[\underline{Y}])$$

(במידה וקשה לכם להשתכנע, נסו תחילה לחשוב על המקרה בו \underline{Y} וקטור חד-מימדי).

דוגמה 1.8 נתונים זוג מ"א גאויסיים במשותף עם צפיפות

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma^2(1-\rho^2)}(x_1^2 - 2\rho x_1 x_2 + x_2^2)}.$$

בהתאם לסימונים שלנו

$$\underline{m} = 0, \quad \Lambda = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

מכאן ניתן לחשב את הצפיפות של x_1 ואת הצפיפות המותנית:

$$\begin{aligned} f_{X_1}(x_1) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x_1^2}{2\sigma^2}} \\ f_{X_2|X_1}(x_2|x_1) &= \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_1}(x_1)} = \frac{1}{\sigma\sqrt{2\pi(1-\rho^2)}} e^{-\frac{1}{2\sigma^2(1-\rho^2)}(x_2 - \rho x_1)^2} \end{aligned}$$

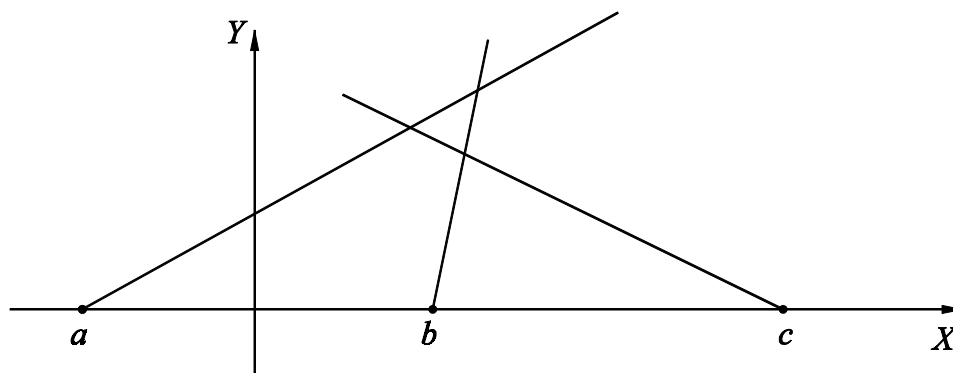
בהנתן x_1 המשתנה x_2 הוא גאוסי עם תוחלת ρx_1 והווריאנס הוא $\sigma^2(1-\rho^2)$: שים לב שהווריאנס כלל אינו תלוי בערך (המידדה) של x_1 ! המצב דומה גם במקרה הכללי.

2 שערור

2.1 מבוא

כיצד פועל מכשיר GPS (Global Positioning System)? המכשיר קולט שידורים מלווינים, ועל ידי השוואת זמן הקליטה (ביחס לזמן השידור) הוא אומד את המרחק מכל לוויין. נתון זה של מרחק אומר לנו כי אנו נמצאים על פני כדור, שרדיוסו הוא המרחק הנמדד. אם יש לנו שתי מדידות, אנו נמצאים בנקודה שהיא חיתוך של שני כדורים: אוסף זה של נקודות מתאר בדרך כלל מעגל. אם ידוע הגובה שלנו (למשל מעל פני הים), אזי קל לראות כי שתי מדידות מדוייקות של מרחק משני לווינים מספיקות כדי לפענח את המיקום—הגובה נתן חיתוך עם כדור נוסף שהוא כדור הארץ. ללא מידע על גובה, מספיקות שלוש מדידות. אולם באופן מעשי, המדידות אינן מושלמות. כיצד משפרים את הדיוק? על ידי הוספת מדידות—מרחק מלווינים נוספים. כיצד לשקלל את כל הנתונים ולקבל הערכה של המקום? זוהי בדיוק בעיית השערור.

דוגמה דומה אך פשוטה יותר היא של אתור קורן: משלוש הנקודות a , b ו- c נקבע כוון לקורן (שמיקומו אינו ידוע ורוצים לקבוע אותו).



איור 2.1:

אם היו לנו רק שתי מדידות למשל, אחת מ- a ואחת מ- c , היינו קובעים את המיקום כנקודת החיתוך (במקרה של מדידות עם רעש זו אינה הקביעה הטובה ביותר תמיד אבל כרגע נתעלם מזה), אולם כאשר שלושת המדידות יוצרות שלוש נקודות חיתוך נשאלת השאלה איך לקבוע את המיקום המשוער של הקורן (השערור), ובנוסף, מה יהיה סדר הגודל של שגיאת השערור.

"שערור"—הערכה מחדש—עוסק בהערכה או השערה לגבי גודל רצוי. מובן אחר: הכלאה של המילים "להעריך" ו-"לשער". בסעיף 1.1 ראינו דוגמאות בהן רצינו "לנקות" אות מרעשים (לסנון), להעריך מרחק למטרה על פי מדידות רועשות וכד'. בפרק זה נניח את היסודות לנושא זה.

ניסוח כללי של בעיית השערור:

א. נתון הוקטור האקראי \underline{Y} (וקטור המדידות), רוצים לשערך את הוקטור האקראי \underline{X} (אותו איננו יכולים למדוד בצורה ישירה), ו- $(\underline{Y}, \underline{X})$ הוא ו"א. המידע שבידינו הוא הפילוגים המשותפים של $(\underline{Y}, \underline{X})$. משערך כלשהו (טוב או רע) הוא

פונקציה של המדידה אשר נסמן ב- $\hat{X} = \phi(\underline{Y})$.

לשם פשוטות נעסוק במקרה ש- X הוא מ"א (ולא ו"א). היינו רוצים למצוא שיטה להעריך את X מתוך \underline{Y} , כלומר למצוא פונקציה ϕ של \underline{Y} כך ש- $\phi(\underline{Y})$ יהיה קרוב ל- X במובן כלשהו. בצורה פורמלית:

ב. הגדרת קריטריון טיב: בהנתן מדד לשגיאה $g(\cdot, \cdot)$, עבור כל משערך $\phi(\underline{Y})$, נגדיר $E[g(X, \phi(\underline{Y}))]$ השגיאה הממוצעת (לפי הקריטריון g). לדוגמה, עבור $g(x, \hat{x}) = (x - \hat{x})^2$ השגיאה הממוצעת היא $E[(X - \phi(\underline{Y}))^2]$. דוגמה אחרת (תוחלת הערך המוחלט של השגיאה): $g(x, \hat{x}) = |x - \hat{x}|$, והשגיאה הממוצעת היא $E[|X - \phi(\underline{Y})|]$. בתקשורת ספרתית, למשל, האות הרצוי X מקבל את הערכים 0, 1 וכדי לקבל סיכוי מרבי לשערך נכון בוחרים $g(x, \hat{x}) = 0$ אם $x = \hat{x}$ ו- $g(x, \hat{x}) = 1$ אם $x \neq \hat{x}$, כלומר, $E[g(X, \phi(\underline{Y}))]$ מקבל מובן של סיכוי טעות.

ג. הבעיה: לאחר שהחלטתם על קריטריון טיב, g , מצא $\phi(\cdot)$ (פונקציה של n משתנים) כך שהשגיאה הממוצעת תהיה מינימלית.

כדי שניתן יהיה להפעיל את הגישה הזו יש להקפיד על שתי הנקודות הבאות: (1) ידוע הפילוג המשותף $\mathbb{P}_{X,Y}$, או לחילופין ידועים פילוג הסתברות מלכתחילה \mathbb{P}_X (a-priori) וכן הפילוג המותנה $\mathbb{P}_{Y|X}$; (2) קובעים קריטריון שגיאה ואז שואפים למצוא משערך אופטימלי על פי קריטריון זה. לאחר מכן מחשבים את השגיאה הממוצעת של המשערך שנמצא. נשים לב שהשגיאה אינה ביחס לערך "אמיתי" של הגודל הרצוי—אותו איננו יודעים—אלא היא ממוצעת לפי פילוג הערכים האפשריים. בדוגמה של הקורן, השגיאה הממוצעת אינה תלויה במיקום הקורן אלא ממוצעת ע"פ כל מיקומי הקורן לפי הפילוג מלכתחילה.

לגבי (2) נשים לב כי בחירות קריטריון שונות יובילו למשערכים אופטימליים שונים. לגבי (1) קיימת נקודה עקרונית: בהרבה מקרים הגיוני שנוכל ליחס ל- X פילוג מלכתחילה (פילוג אפריורי) אולם אין הדבר כך בכל מקרה. למשל, התגלה כוכב לכת חדש ורוצים למדוד את המרחק אליו. איזה מובן ניתן ליחס לפילוג מלכתחילה של המרחק מאתנו לכוכב זה? מאידך, בבעיות אחרות (כגון בעיות בתקשורת) יש מובן לפילוג האפריורי. כאשר אנו מניחים שיש מובן לקיום הסתברות מלכתחילה הבעיה נקראת בעיה בייסיאנית (ע"ש חוק Bayes).

2.2 שערך אופטימלי

נחזור לבעיית השערך שהוצגה במבוא: X רצוי, \underline{Y} מצוי, קיים פילוג הסתברות משותף ל- \underline{Y} ו- X . המשערך ל- X כאשר נמדד \underline{Y} מסומן ב- $\hat{X} = \phi(\underline{Y})$. קובעים סיפרת טיב $g(\cdot, \cdot)$ ובעזרתה מגדירים את השגיאה הממוצעת $E[g(X, \phi(\underline{Y}))]$. עבור קריטריון השגיאה הריבועית הממוצעת $g(x, \hat{x}) = (x - \hat{x})^2$ הגודל ε המוגדר כ- $\varepsilon \triangleq X - \phi(\underline{Y})$ נקרא שגיאת השערך, ε^2 הוא השגיאה הריבועית ו- $E[\varepsilon^2]$ הוא השגיאה הממוצעת של המשערך. בכל מקרה מחפשים את $\phi_0(\cdot)$ שעושה מינימיזציה לשגיאה הממוצעת. מעתה והלאה נבחר בקריטריון השגיאה הריבועית הממוצעת ובכל מקום שלא נתיחס לקריטריון במפורש יהיה זה קריטריון השגיאה הריבועית הממוצעת.

טענה 2.1 המשערך $\phi_0(\cdot)$ שעבורו $E[(X - \phi_0(\underline{Y}))^2]$ הוא מינימלי (על פני כל המשערכים $\phi(\cdot)$) הוא:

$$\hat{X}_{\text{opt}} = \phi_0(\underline{Y}) = E[X|\underline{Y}].$$

המשמעות היא כמובן שמחשבים את $E[X|\underline{Y} = y]$ ומציבים בפונקציה שהתקבלה את ערך המדידה \underline{Y} .

להוכחת טענה זו ניגש מיד לאחר הצגת שלוש דוגמאות שבהן מחושב המשעריך האופטימלי במפורש.

דוגמא א' (סתם דוגמא טכנית):

$$f_{X,Y}(x, y) = x + y \quad ; \quad 0 \leq x, y \leq 1$$

$$f_Y(y) = \int_0^1 (x + y) dx = \frac{1}{2} + y$$

$$E(X|Y = y) = \int_0^1 x \frac{x + y}{\frac{1}{2} + y} dx = \frac{\frac{1}{3} + \frac{y}{2}}{\frac{1}{2} + y}$$

דוגמא ב': X, Y משתנים אקראיים גאוסים במשותף. נחשב את $E[X|Y]$.

נסמן:

$$X_c = X - E[X] \quad ; \quad Y_c = Y - E[Y]$$

ונבנה משתנה אקראי כדלקמן:

$$Z \doteq X_c - \frac{\text{Cov}(Y, X)}{\text{Var}(Y)} Y_c$$

מתוך הגדרת Z ברור ש- $E[Z] = 0$ ו- $E[Z Y_c] = 0$. מהגדרת גאוסיות במשותף קל לראות ש- Z ו- Y_c גאוסיים במשותף, ומחוסר הקורלציה נובע שהם בלתי תלויים. מכאן שגם Y, Z בלתי תלויים ו- $E[Z|Y] = E[Z] = 0$, ולכן:

$$(2.1) \quad 0 = E \left[X_c - \frac{\text{Cov}(Y, X)}{\text{Var}(Y)} Y_c \mid Y \right]$$

$$(2.2) \quad = E[X|Y] - E[X] - \frac{\text{Cov}(Y, X)}{\text{Var}(Y)} (Y - E[Y])$$

מכאן:

$$E[X|Y] = E[X] + \frac{\text{Cov}(Y, X)}{\text{Var}(Y)} (Y - E[Y])$$

ולכן המשעריך האופטימלי של X מתוך Y נתון על ידי:

$$(2.3) \quad \phi_0(y) = E[X] + \frac{\text{Cov}(Y, X)}{\text{Var}(Y)} (y - E[Y])$$

$$(2.4) \quad \hat{X}_{opt} = E[X] + \frac{\text{Cov}(Y, X)}{\text{Var}(Y)} (Y - E[Y]).$$

שים לב לכך שהמשעריך האופטימלי במקרה זה, הוא פונקציה ליניארית של המדידות. (השווה עם השורות האחרונות של סעיף 1.6). לכן \hat{X}_{opt} הוא מ"א גאוס, עם ממוצע $E[X]$.

דוגמא ג': דוגמא פשוטה אולם חשובה שמהווה בסיס לנושא הסינון הלינארי-אופטימלי שנלמד בהמשך. חשוב לזכור אותה כולל התוצאות המסומנות ב-*:

$$(2.5) \quad Y = X + N$$

$$(2.6) \quad E[X] = E[N] = 0$$

$$(2.7) \quad E[X^2] = \sigma_X^2, \quad E[N^2] = \sigma_N^2$$

X, N גאוסיים ב"ת (X הוא הסיגנל, N הרעש, מודדים את Y כלומר את הסיגנל טבול ברעש ורוצים לשערך את הסיגנל הנקי X).

הנחנו ש- X, N גאוסיים ב"ת, ולכן הם גאוסיים במשותף (תוחלת אפס). על-פי דוגמא ב' לעיל, מתקיים:

$$\hat{X}_{opt} = E[X|Y] = E[X] + \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}(Y - E[Y]) = \frac{\text{Cov}(Y, X)}{\text{Var}(Y)}Y$$

$$\text{Cov}(X, Y) = E[X(X + N)] = E[X^2] + E[XN] = \sigma_X^2$$

$$\text{Var}(Y) = E[(X + N)^2] = E[X^2] + 2E[XN] + E[N^2] = \sigma_X^2 + \sigma_N^2$$

$$(2.8) \quad \boxed{\hat{X}_{opt} = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2}Y}$$

דרך אחרת: נשתמש במשפט שהופיע ללא הוכחה בסוף סעיף 1.6. לכן נוכל לרשום:

$$E[X|Y] = cY$$

כלומר, אנו כבר יודעים שהמשערך האופטימלי הינו לינארי ולכן נותר רק למצוא את הקבוע האופטימלי c . על מנת למצוא את c , נבצע מינימיזציה על השגיאה הרבועית הממוצעת:

$$E[(X - cY)^2] = E[(X - cX - cN)^2] = (1 - c)^2\sigma_X^2 + c^2\sigma_N^2 - 2c(1 - c)E[XN]$$

כיוון ש- $\hat{X} = c(X + N)$, הרי שהמינימיזציה מחפשת פשרה (על ידי התאמת c) בין האיבר הראשון, המייצג אי שינוי של X (על ידי בחירה $c = 1$) מחד לבין האיבר השני, שמייצג את סילוק הרעש (על ידי בחירה $c = 0$) מאידך. מינימיזציה נותנת (ע"י גזירה והשוואה ל-0),

$$c^* = \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2} \quad (*)$$

שהוא מספר בין 0 ל-1 התלוי ב"עוצמות" היחסיות של האות והרעש. במבט ראשון על (2.5) נראה שהשערך הטוב ביותר של X הוא Y אבל (*) נותן תוצאה אחרת ובמבט שני הגיונית יותר; כאשר עוצמת הרעש נמוכה (σ_N קטן), המשערך הטוב ביותר ל- X הוא באמת המדידה; אך כאשר עוצמת הרעש גבוהה (σ_N גדול), המדידה חסרת משמעות למעשה ולכן המשערך הטוב ביותר ל- X הוא הממוצע שלו (במקרה הנוכחי - אפס).

השגיאה הנתרת (במקרה של ווריאנס כללי של X) היא:

$$\begin{aligned}
 (2.9) \quad E[(X - cY)^2] &= E[(X - c(X + N))^2] \\
 (2.10) \quad &= \left(1 - \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2}\right)^2 \sigma_X^2 + \left(\frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2}\right)^2 \sigma_N^2 \\
 (*) \quad &= \frac{\sigma_X^2 \sigma_N^2}{\sigma_X^2 + \sigma_N^2}
 \end{aligned}$$

הכללה למקרה של מדידה וקטורית

$E(X|\underline{Y})$. נחשב את $\underline{\Lambda}_Y = \text{Cov}\{\underline{Y}\}$ ונניח כי $\underline{Y} = (Y_1, \dots, Y_n)^T$ ו" \underline{Y}, X) גאוסית, ונניח כי $\underline{Y}_c = \underline{Y} - E\underline{Y}$ היא מטריצה לא סינגולרית. נחשב את $E(X|\underline{Y})$.
נגדיר

$$\begin{aligned}
 \underline{Y}_c &= \underline{Y} - E\underline{Y} \\
 X_c &= X - EX \\
 Z &\doteq X_c - \underline{Y}_c^T \underline{\Lambda}_Y^{-1} E[X_c \underline{Y}_c].
 \end{aligned}$$

Z ו- \underline{Y}_c גאוסיים במשותף, $EZ = 0$ וכן $E\underline{Y}_c Z = 0$. לכן Z, \underline{Y}_c ב"ת.
מכאן $E(Z|\underline{Y}) = EZ = 0$.
כלומר:

$$\begin{aligned}
 E(X_c|\underline{Y}) &= E\left[\underline{Y}_c^T \underline{\Lambda}_Y^{-1} E[X_c \underline{Y}_c]|\underline{Y}\right] \\
 &= \underline{Y}_c^T \underline{\Lambda}_Y^{-1} EX_c \underline{Y}_c.
 \end{aligned}$$

בצורה מפורשת יותר

$$E[X|\underline{Y}] = EX + (\underline{Y} - E\underline{Y})^T \underline{\Lambda}_Y^{-1} \begin{pmatrix} \text{Cov}(Y_1, X) \\ \vdots \\ \text{Cov}(Y_n, X) \end{pmatrix} \quad (2.11)$$

תרגיל 2.2 הרחב את התוצאה האחרונה למקרה של משתנה ווקטורי. כלומר הראה כי אם $(\underline{X}, \underline{Y})$ הם וקטורים גאוסיים במשותף אזי את הווקטור $E[\underline{X} | \underline{Y}]$ ניתן לחשב על ידי

$$E[\underline{X} | \underline{Y}] = E[\underline{X}] + [(\text{Var } \underline{Y})^{-1} \text{Cov}(\underline{Y}, \underline{X})]^T [\underline{Y} - E\underline{Y}] \quad (2.12)$$

נחזור לטענה 2.1, ונוכיח אותה.

הוכחה: נדרש להראות שלמשערך כל שהוא ϕ מתקיים: $E(\phi(\underline{Y}) - X)^2 \geq E(\phi_0(\underline{Y}) - X)^2$. עבור משערך כל שהוא $\phi(\cdot)$ (כאשר $\phi_0(\underline{Y})$ היא התוחלת המותנית של X כאשר נתון \underline{Y}), מתקיים:

$$\begin{aligned} E[(X - \phi(\underline{Y}))^2] &= E[(X - \phi_0(\underline{Y}) + \phi_0(\underline{Y}) - \phi(\underline{Y}))^2] \\ &= E[(X - \phi_0(\underline{Y}))^2] + E[(\phi_0(\underline{Y}) - \phi(\underline{Y}))^2] + 2E[(X - \phi_0(\underline{Y}))(\phi_0(\underline{Y}) - \phi(\underline{Y}))] \end{aligned}$$

נעיין באיבר האחרון $E[(X - \phi_0(\underline{Y}))(\phi_0(\underline{Y}) - \phi(\underline{Y}))]$; נבצע קודם תוחלת מותנית (מותנה ב- \underline{Y}) ואח"כ תוחלת על \underline{Y} . לפי (ב) ו (ג) נקבל

$$E[(X - \phi_0(\underline{Y}))(\phi_0(\underline{Y}) - \phi(\underline{Y}))] = E[(\phi_0(\underline{Y}) - \phi(\underline{Y}))E[X - \phi_0(\underline{Y})|\underline{Y}]]$$

ומתקיים

$$E[X - \phi_0(\underline{Y})|\underline{Y}] = E[X|\underline{Y}] - \phi_0(\underline{Y}) = \phi_0(\underline{Y}) - \phi_0(\underline{Y}) = 0$$

כלומר האיבר השלישי מתאפס. בנוסף, $E[(\phi_0(\underline{Y}) - \phi(\underline{Y}))^2] \geq 0$ ולכן:

$$E[(X - \phi(\underline{Y}))^2] \geq E[(X - \phi_0(\underline{Y}))^2]$$

ומכאן ברור ש- $\phi_0(\underline{Y})$ הוא המשערך הטוב ביותר על פי קריטריון השגיאה הריבועית הממוצעת (אולי יש טוב כמוהו אולם אין טוב ממנו).

בכך השלמנו את הוכחת הטענה.

טענה 2.3 המשערך האופטימלי הוא יחיד, במובן הבא. נסמן ב- ϕ_0 את משערך התוחלת המותנית ויהי ϕ_1 משערך אחר אשר

$$E[(\phi_0(Y) - \phi_1(Y))^2] = 0 \text{ אזי } \phi_0 = \phi_1.$$

הוכחה: נשים לב כי מהוכחת טענה 2.1 נובע כי לכל משערך, ובפרט ל- ϕ_1 , מתקיים:

$$E[(X - \phi_1(\underline{Y}))^2] = E[(X - \phi_0(\underline{Y}))^2] + E[(\phi_0(\underline{Y}) - \phi_1(\underline{Y}))^2]$$

ומכיוון שלפי ההנחה לשני המשערכים שגיאה ריבועית ממוצעת זהה, נקבל כי בהכרח $E[(\phi_0(\underline{Y}) - \phi_1(\underline{Y}))^2] = 0$, כנדרש. נשים לב כי היחידות היא באשר לתוצאה של השערך, ולא באשר למבנה של המשערך. כדי להבין נקודה זו נתבונן במקרה פשוט ביותר של שערך על סמך שתי מדידות: $X = Y_1 = Y_2$. כמובן ששני המשערכים $\phi_0(\underline{Y}) = Y_1$, $\phi_1(\underline{Y}) = Y_2$ הם אופטימליים—בשני המקרים תוצאת השערך היא X למרות שהנוסחה עבור המשערך שונה.

נעבור כעת לדוגמה נוספת לחישוב מפורש של המשערך האופטימלי. דוגמא זו עוסקת בשערך של תוצאת זריקת קוביה "הוגנת" $i = 1, \dots, 6$; $p_i = 1/6$ לפי קריטריון השגיאה הריבועית המינימלית.

א. ללא מדידות המשערך האופטימלי הוא הממוצע, כלומר: $E[X] = 3.5$, והשגיאה הנתרת

$$E[(X - EX)^2] = E[X^2] - (E[X])^2 = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) - \left(3\frac{1}{2}\right)^2 = \frac{91}{6} - \left(3\frac{1}{2}\right)^2 = 2.72$$

ב. אנו מקבלים אינפורמציה: מישהו מספר לנו אם התוצאה X היא אחד מהשלשה 4, 5, 6 או ש- X אחד מהשלשה 1, 2, 3. נגדיר מ"א Y כדלקמן:

$$Y = 0 \text{ if } X = 4 \text{ or } 5 \text{ or } 6$$

$$Y = 1 \text{ if } X = 1 \text{ or } 2 \text{ or } 3$$

מתקיים:

$$\mathbb{P}\{X = 1|Y = 0\} = \mathbb{P}\{X = 2|Y = 0\} = \mathbb{P}\{X = 3|Y = 0\} = 0$$

$$\mathbb{P}\{X = 4|Y = 0\} = \mathbb{P}\{X = 5|Y = 0\} = \mathbb{P}\{X = 6|Y = 0\} = 1/3$$

לכן: $E[X|Y = 0] = (4 + 5 + 6)1/3 = 5$ כמו כן

$$\mathbb{P}\{X = 1|Y = 1\} = \mathbb{P}\{X = 2|Y = 1\} = \mathbb{P}\{X = 3|Y = 1\} = 1/3$$

$$\mathbb{P}\{X = 4|Y = 1\} = \mathbb{P}\{X = 5|Y = 1\} = \mathbb{P}\{X = 6|Y = 1\} = 0$$

לכן $E[X|Y = 1] = 2$. לסיכום אם $Y = 1$ אזי $\hat{X}_{\text{opt}} = 2$ ואם $Y = 0$ אזי $\hat{X}_{\text{opt}} = 5$. השגיאה הנותרת תהיה:

$$E[\varepsilon^2] = E[(X - E[X|Y])^2] = E[X^2] - E[(E[X|Y])^2] = \frac{91}{6} - \frac{1}{2}(5^2 + 2^2) = \frac{2}{3}$$

שגיאה קטנה משמעותית מהשגיאה אשר קיבלנו כאשר השערך היה ללא מידע.

אם נשנה את הערכים של הקוביה, למשל הערכים יהיו 0, 2, 3, 4, 5, 7 אזי חישוב דומה יתן $E[X|Y = 0] = 16/3$ וכן $E[X|Y = 1] = 5/3$, כלומר קיבלנו ערכים שאינם ערכים של הקוביה. בפרט $\mathbb{P}(\hat{X}_{\text{opt}} = X) = 0$. הסיבה לכך היא שקריטריון השגיאה הריבועית דואג לשגיאה ריבועית ממוצעת קטנה, אך אינו מותאם להשגת שיוויון.

2.3 שערך לינארי אופטימלי

במקרים רבים חישוב התוחלת המותנית קשה ביותר. בכל מקרה, חישוב זה דורש ידע של חוק ההסתברות המשותף של \underline{Y} ו- X . כאשר חוק זה אינו ידוע או כאשר חישוב התוחלת המותנית אינו אפשרי, לא ניתן למצוא את המשערך האופטימלי של X בהינתן \underline{Y} . במקרים כאלה מסתפקים בדרך כלל במשערך טוב פחות, אך ניתן לחישוב. להלן נתרכז במשפחת משערכים שהם פונקציות לינאריות של המדידות. בסעיף זה איננו מניחים ש- X, \underline{Y} הוא ווקטור גאוס, ואף לא נצטרך להניח שידוע הפילוג המשותף שלהם. אנו נסתפק בידיעת המומנטים מסדר ראשון והמומנטים המשותפים מסדר שני.

המקרה הסקלרי

נתחיל מן המקרה הפשוט שבו Y הוא מ"א ולא ו"א, כלומר השערך מתבצע על סמך מדידה אחת. במקרה זה למשערך שהוא פונקציה לינארית של Y יש באופן כללי את הצורה הבאה:

$$\hat{X}^l = aY + b$$

כאשר a ו- b הם קבועים כלשהם.

את בעיית מציאת המשעריך הלינארי האופטימלי נגדיר כדלקמן:

בעיה: מצא משעריך מהצורה $\hat{X}^l = aY + b$ כך ששגיאת השערוך הריבועית הממוצעת תהיה מינימלית. במילים אחרות, מצא קבועים a ו- b כך שהביטוי

$$E[\varepsilon^2] = E[(X - \hat{X}^l)^2] = E[(X - aY - b)^2]$$

יהיה מינימלי. שים לב שזו בעיה הרבה יותר פשוטה ממצאת המשעריך האופטימלי כי כאן יש למצוא שני קבועים בלבד ולא פונקציה שלמה $\hat{X} = \phi(Y)$.

פתרון: נרשום פעם נוספת את הביטוי לשגיאת השערוך הריבועית הממוצעת:

$$E[\varepsilon^2] = E[(X - aY - b)^2] = E[X^2] - 2aE[XY] - 2bE[X] + a^2E[Y^2] + 2abE[Y] + b^2$$

אם נגזור את הביטוי לעיל לפי b ונשווה את הנגזרת ל-0 נקבל ש- b האופטימלי, b^* , צריך לקיים:

$$b^* = E[X] - aE[Y]$$

בהצבת b^* לתוך הביטוי לשגיאת השערוך הריבועית הממוצעת נקבל:

$$E[\varepsilon^2] = E[(X_c - aY_c)^2] = E[X_c^2] - 2aE[X_cY_c] + a^2E[Y_c^2]$$

כאשר $Y_c = Y - E[Y]$ ו- $X_c = X - E[X]$. שוב, ע"י גזירת הביטוי לעיל לפי a והשוואה ל-0 נקבל ש- a האופטימלי, a^* , צריך לקיים:

$$a^* = \frac{E[X_cY_c]}{E[Y_c^2]} = \frac{\text{Cov}(Y, X)}{\text{Var}(Y)}$$

ולכן המשעריך הלינארי האופטימלי הוא:

$$\hat{X}_{\text{opt}}^l = E[X] + \frac{\text{Cov}(Y, X)}{\text{Var}(Y)} [Y - E[Y]]$$

(מה הקשר בין תוצאה זאת לדוגמה ב' בסעיף 2.2? ראה הערה מס' 4 בהמשך). השגיאה הריבועית הממוצעת המתקבלת במקרה זה היא:

$$E[\varepsilon_{\min}^2] = \text{Var}(X) - \frac{[\text{Cov}(Y, X)]^2}{\text{Var}(Y)} = \text{Var}(X)(1 - \rho^2)$$

כאשר ρ מקדם הקורילציה; $|\rho| \leq 1$.

הערות:

1. שים לב שלצורך חישוב המשעריך הלינארי האופטימלי יש צורך לדעת רק מומנטים מסדר ראשון ושני.
2. משעריך אופטימלי לעולם אינו גרוע יותר ממשעריך לינארי אופטימלי; משעריך לינארי אופטימלי לעולם אינו גרוע יותר ממשעריך ללא מדידה כלל.

3. כאשר $\text{Cov}(Y, X) = 0$ המשערך הלינארי האופטימלי עם מדידה Y הוא $E[X]$, כלומר, אותו מספר כאילו ולא נמדד דבר. מכאן שכאשר X ו- Y חסרי קורלציה, Y אינו עוזר בשערך לינארי של X .

4. במקרים מסוימים המשערך הלינארי האופטימלי הוא המשערך האופטימלי. לדוגמה, כאשר X ו- Y בלתי תלויים. דוגמה אחרת היא המקרה הגאוסי (וזאת ומאחר וכבר צוין, ללא הוכחה שבמקרה הגאוסי $E[X|Y] = c_0 Y$ עבור קבוע c_0 מסוים). דוגמאות נוספות יובאו בתרגילים.

המקרה הוקטורי

נעבור למקרה בו השערך מתבצע על סמך וקטור מדידות, \underline{Y} . לשם פשוטות נניח ש- $E[X] = 0$ וכן $E[\underline{Y}] = 0$. במקרה זה למשערך, שהוא פונקציה לינארית של \underline{Y} , יש באופן כללי את הצורה הבאה:

$$\hat{X}^l = \underline{a}^T \underline{Y} = \sum_{i=1}^n a_i Y_i$$

כאשר $\underline{a}^T = (a_1, a_2, \dots, a_n)$ הם קבועים כלשהם. בעית מציאת המשערך הלינארי האופטימלי במקרה זה היא:

בעיה: מצא משערך מהצורה $\hat{X}^l = \underline{a}^T \underline{Y}$ כך ששגיאת השערך הריבועית הממוצעת תהיה מינימלית. במילים אחרות, מצא קבועים (a_1, a_2, \dots, a_n) כך שהביטוי

$$E[\varepsilon^2] = E[(X - \hat{X}^l)^2] = E[(X - \underline{a}^T \underline{Y})^2]$$

יהיה מינימלי (על פני כל הוקטורים הקבועים \underline{a}).

הערה: בדוק שההנחה של ממוצע אפס (עבור המדידה ועבור המשתנה הלא-ידוע) גורמת לכך שאין צורך ב"איבר חופשי" (קבוע דטרמיניסטי) נוסף.

פתרון: נרשום פעם נוספת את הביטוי לשגיאת השערך הריבועית הממוצעת:

$$E[\varepsilon^2] = E\left[\left(X - \sum_{j=1}^n a_j Y_j\right)^2\right]. \quad (2.13)$$

להלן ניתן הוכחה ישירה, אך מעט מיגעת. בהמשך נציג הוכחה חלופית, בשיטה שהיא חשובה לכשעצמה. אם נגזור את הביטוי לעיל לפי a_i ונשווה ל-0 את הנגזרת נקבל שערכי a_i האופטימליים, a_i^* צריכים לקיים:

$$0 = \frac{\partial E[\varepsilon^2]}{\partial a_i^*} = -2E\left[Y_i \left(X - \sum_{j=1}^n a_j^* Y_j\right)\right] = -2E[X Y_i] + 2 \sum_{j=1}^n a_j^* E[Y_i Y_j] \quad 1 \leq i \leq n$$

הערה: נסמן ב- $*$ את המקדמים האופטימליים, ומטעמי נוחות נכתוב a_i^* כאשר נתייחס לאיבר i -י מתוך וקטור מקדמים אופטימלי \underline{a} , ונכתוב \underline{a}_* כאשר נתייחס לווקטור המקדמים האופטימלי.

אנו רואים כי למשערך הלינארי האופטימלי יש את התכונה כי שגיאת השערך

$$\varepsilon = X - \sum_{j=1}^n a_j^* Y_j \quad (2.14)$$

חסרת קורלציה (בת"ל) עם כל אחת מהמדידות Y_i . לנקודה זו נחזור בהמשך.

תרגיל 2.4 בדוק כי a_i^* המאפס את הנגזרות החלקיות של שגיאת השערוך הריבועית, מביא אכן למינימום (ולא למקסימום).

בכתיב וקטורי:

$$\begin{pmatrix} E[Y_1^2] & E[Y_1Y_2] & \dots & E[Y_1Y_n] \\ E[Y_1Y_2] & E[Y_2^2] & \dots & E[Y_2Y_n] \\ \vdots & \vdots & \ddots & \vdots \\ E[Y_1Y_n] & E[Y_1Y_2] & \dots & E[Y_n^2] \end{pmatrix} \begin{pmatrix} a_1^* \\ \vdots \\ a_n^* \end{pmatrix} = \begin{pmatrix} E[XY_1] \\ E[XY_2] \\ \vdots \\ E[XY_n] \end{pmatrix}$$

ובצורה מקוצרת:

$$(2.15) \quad E[\underline{Y}\underline{Y}^T] \cdot \underline{a}_* = E[\underline{Y}X]$$

נסמן $\underline{\underline{\Lambda}}_Y = E[\underline{Y}\underline{Y}^T]$ אזי

$$(2.16) \quad \underline{\underline{\Lambda}}_Y \underline{a}_* = \underline{b} = \begin{pmatrix} \text{Cov}(X, Y_1) \\ \vdots \\ \text{Cov}(X, Y_n) \end{pmatrix}$$

כאשר באגף השמאלי מופיע וקטור n מימדי ובאגף הימני מכפלה של מטריצה $n \times n$ בוקטור n מימדי. כיוון שמדובר במשתנים עם תוחלת אפס ניתן לרשום זאת גם בצורה $\text{Cov}(\underline{Y}, X) = \underline{\underline{\Lambda}}_Y \cdot \underline{a}_*$. בהנחה שמטריצת הקווריאנס $\underline{\underline{\Lambda}}_Y$ אינה סינגולרית, יש פתרון (והוא יחיד) עבור \underline{a}_* (במקרה הסינגולרי נטפל אח"כ). פתרון זה נתון ע"י

$$(2.17) \quad \underline{a}_* = \underline{\underline{\Lambda}}_Y^{-1} E[\underline{Y}X] = \underline{\underline{\Lambda}}_Y^{-1} \underline{b}$$

והמשערוך הלינארי האופטימלי הוא

$$(2.18) \quad \hat{X}_{\text{opt}}^l = \underline{a}_*^T \underline{Y} = \sum_{j=1}^n a_j^* Y_j = \underline{b}^T \underline{\underline{\Lambda}}_Y^{-1} \underline{Y}$$

מה השגיאה הריבועית הממוצעת המינימלית, $E[(X - \underline{a}_*^T \cdot \underline{Y})^2] = ?$. דרך אחת לחשב את השגיאה הנותרת היא לרשום את (2.15) בצורת $\underline{b} = \underline{\underline{\Lambda}}_Y \cdot \underline{a}_*$ כאשר $\underline{b} = E[X\underline{Y}]$, ואז להציב לתוך (2.13). דרך אחרת, אשר נחזור אליה בצורה כללית בהמשך, היא הבאה: מהגדרת השגיאה,

$$(2.19) \quad E[X^2] = E \left[\sum_{i=1}^n a_j^* Y_j + \varepsilon \right]^2$$

$$(2.20) \quad = E \left[\sum_{i=1}^n a_j^* Y_j \right]^2 + E[\varepsilon^2] + 2 \left[\sum_{i=1}^n a_j^* E[Y_j \varepsilon] \right].$$

אולם ראינו כי כאשר בוחרים את המקדמים האופטימליים a_j^* (ורק במקרה זה!) נקבל $E[Y_j \varepsilon] = 0$. קיבלנו לכן שהשגיאה הריבועית המינימלית עבור שיערוך לינארי היא

$$(2.21) \quad (E[\varepsilon^2])_{\min} = E[X^2] - 2\underline{b}^T \underline{a}_* + \underline{a}_*^T \underline{\Lambda}_Y \underline{a}_* = E[X^2] - \underline{a}_*^T \underline{\Lambda}_Y \underline{a}_*$$

$$(2.22) \quad = E[X^2] - \underline{b}^T \underline{\Lambda}_Y^{-1} \underline{\Lambda}_Y \underline{\Lambda}_Y^{-1} \underline{b}$$

$$(2.23) \quad = E[X^2] - \underline{b}^T \underline{\Lambda}_Y^{-1} \underline{b}.$$

(2.17), (2.18) ו-(2.21) נותנים את התשובה המלאה לבעיית השיערוך הלינארי במקרה הוקטורי כאשר המומנטים מסדר ראשון מתאפסים (לפחות כאשר $\underline{\Lambda}_Y$ הפיכה).

נשים לב כי פיתוח זה גם מראה כי במקרה שמטריצת הקווריאנס הפיכה, יש רק משערוך לינארי אופטימלי אחד (בדומה ליחידות של המשערוך האופטימלי).

מה זה מזכיר? זה בדיוק הפתרון שקיבלנו למשערוך אופטימלי במקרה הגאומטרי. האם זה מקרי? מדוע זה קורה? בהנתן (X, \underline{Y}) עם צפיפות משותפת $f_{X, \underline{Y}}(x, \underline{y})$, יהי (X_G, \underline{Y}_G) וא"ג עם

$$\underline{\Lambda}_{\underline{Y}_G} = \underline{\Lambda}_{\underline{Y}}, \quad \text{Cov}(X_G, \underline{Y}_G) = \text{Cov}(X, \underline{Y})$$

(נניח שכל התוחלות הן אפס לשם פשטות), אזי

$$\hat{X}_{opt}^{lin}(\underline{y}) \stackrel{(1)}{=} (\hat{X}_G)_{opt}^{lin}(\underline{y}) \stackrel{(2)}{=} (\hat{X}_G)_{opt}(\underline{y}) \stackrel{(3)}{=} \text{Cov}(X_G, \underline{Y}_G)^T \underline{\Lambda}_{\underline{Y}_G}^{-1} \underline{y} \stackrel{(4)}{=} \text{Cov}(X, \underline{Y})^T \underline{\Lambda}_{\underline{Y}}^{-1} \underline{y}$$

כאשר (1) נובע מכך שהמשערוך תלוי רק במומנטים עד סדר שני, והם זהים במקרה זה, (2) נובע מכך שהמשערוך האופטימלי הינו לינארי, את (3) הוכחנו בעבר, והמעבר האחרון, (4), נובע מכך שלמשתנים יש אותם מומנטים מסדר שני, כפי שהגדרנו.

כלומר, ישנה דואליות (או שקילות מלאה) בין שני המקרים הבאים:

1. מקרה גאומטרי ומשערוך אופטימלי.

2. מקרה כללי ומשערוך אופטימלי לינארי.

תרגיל: שכנע את עצמך כי בין כל הוקטורים האקראיים (X, \underline{Y}) עם סטטיסטיקה מסדר שני נתונה, הוא"ג הוא הגרוע ביותר לשערוך אופטימלי, כלומר - השגיאה הריבועית הממוצעת שלו היא הגבוהה ביותר.

נענין כעת במקרה ש- $E[X] \neq 0, E[\underline{Y}] \neq 0$. במקרה זה המשערוך הלינארי האופטימלי הוא:

$$\hat{X}_{opt}^l = E[X] + \sum_{i=1}^n a_i^* (Y_i - E[Y_i]) = E[X] + \underline{b} \underline{\Lambda}_Y^{-1} (\underline{Y} - E[\underline{Y}])$$

כאשר ה- a_i^* נתונים ע"י פתרון משוואה (2.17) עבור המשתנים הממורכזים (ממוצע אפס) והשגיאה הנותרת תהיה:

$$(E[\varepsilon^2])_{\min} = E[(X - E[X])^2] - E[(\underline{a}_*^T (\underline{Y} - E[\underline{Y}]))^2]$$

(הוכח תוצאות אלה).

ההערות שניתנו במקרה הסקלרי תופסות גם כאן. במקרה זה (מטריצת קווריאנס שאינה סינגולרית), כמו גם במקרה הסקלרי, ברור שהמשעריך הלינארי האופטימלי הוא יחיד—יש לנו משוואה מפורשת הקובעת חד משמעית את המקדמים, והם קובעים את מבנה המשעריך.

דוגמה 2.5 נתונים

$$E[X] = E[Y_1] = E[Y_2] = 0 \quad n = 2$$

$$E[Y_1 Y_2] = 0; \quad E[Y_1^2] = E[Y_2^2] = 1; \quad E[XY_1] = 5; \quad E[XY_2] = 7$$

אזי

$$\begin{pmatrix} 5 \\ 7 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix}$$

ומכאן

$$a_1^* = 5, \quad a_2^* = 7$$

$$\hat{X}_{\text{opt}}^l = 5Y_1 + 7Y_2$$

$$(E[\varepsilon^2])_{\min} = E[(X - \underline{Y}^T \underline{a}_*)^2] = E[X^2] - E[(5Y_1 + 7Y_2)^2] = E[X^2] - 25 - 49$$

מדוע לא ייתכן שהביטוי האחרון יהיה שלילי? (זכור: $E[X^2]E[Y^2] \geq (E[XY])^2$, מטריצת הקווריאנס אינה שלילית).

דוגמה 2.6 נחשוב כעת על "אות" X המשודר בשני ערוצים. בערוץ הראשון מתווסף רעש N_1 ואנו מודדים $Y_1 = X + N_1$, ובערוץ השני מתווסף רעש N_2 ואנו מודדים $Y_2 = X + N_2$. נניח שהאות המקורי והרעשים חסרי קורלציה, ולכולם ממוצע אפס ווריאנס 1. נרשום ונחשב

$$E[X] = E[N_1] = E[N_2] = 0$$

$$E[X^2] = E[N_1^2] = E[N_2^2] = 1$$

$$E[Y_1 Y_2] = E[(X + N_1)(X + N_2)] = E[X^2] = 1$$

$$E[Y_1^2] = E[X^2] + E[N_1^2] = 2$$

$$E[Y_2^2] = E[X^2] + E[N_2^2] = 2$$

$$E[XY_1] = E[X(X + N_1)] = E[X^2] = 1$$

$$E[XY_2] = E[X(X + N_2)] = E[X^2] = 1$$

אזי

$$\begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$$

ומכאן

$$a_1^* = \frac{1}{3}, \quad a_2^* = \frac{1}{3}$$

$$\hat{X}_{\text{opt}}^l = \frac{Y_1 + Y_2}{3}$$

$$(E[\varepsilon^2])_{\min} = E\left[\left(X - \underline{Y}^T \underline{a}_*\right)^2\right] = E[X^2] - E\left[\left(\frac{1}{3}Y_1 + \frac{1}{3}Y_2\right)^2\right] = 1 - \frac{1}{9}E[Y_1 + Y_2]^2 = 1 - \frac{1}{9}(2 + 2 + 2) = \frac{1}{3}.$$

תוצאה זו שונה מה"ניחוש הטבעי" $\hat{X} = (Y_1 + Y_2)/2$. השגיאה של "ניחוש" זה היא

$$E\left[X - \frac{Y_1 + Y_2}{2}\right]^2 = 1 \quad (2.24)$$

כלומר שגיאה גדולה פי 3.

דוגמה 2.7 נמשיך את הרעיון של שתי מדידות של אותו גודל, עם אותם ממוצעים וסטיות תקן של האות והרעש כמו בדוגמה הקודמת, אך כעת $Y_1 = X + N_1$, $Y_2 = X + N_1 + N_2 = Y_1 + N_2$. כעת נקבל $E[Y_1^2] = 2$, $E[Y_2^2] = 3$. $E[XY_1] = E[XY_2] = 1$ המשערך האופטימלי נקבע על ידי

$$\begin{pmatrix} a_1^* \\ a_2^* \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix} \quad (2.25)$$

כלומר $\hat{X} = Y_1/2$ אינו תלוי כלל ב- Y_2 ! זאת משום ש- $Y_2 = Y_1 + N_2$ הוא "גרסה רועשת" של Y_1 , אשר אינה יכולה להוסיף מידע אודות X מעבר לזה שהתקבל מ- Y_1 .

מטריצת קווריאנס סינגולרית

כמובטח, נעבור כעת לטפל במקרה שבו מטריצת הקווריאנס, $\underline{\underline{\Delta}}_Y$, היא סינגולרית. במקרה זה קיים וקטור קבוע n מימדי, \underline{h} שאינו ווקטור האפס, כך ש $\underline{\underline{\Delta}}_Y \underline{h} = 0$ כי השורות של $\underline{\underline{\Delta}}_Y$ תלויות. לכן:

$$E\left[(\underline{h}^T \underline{Y})^2\right] = \underline{h}^T \underline{\underline{\Delta}}_Y \underline{h} = 0$$

נעיין בוקטור \underline{h} ונניח (ללא הגבלת הכלליות) ש- $h_n \neq 0$ ולכן נוכל להכפיל בקבוע כך ש- $h_n = 1$ ואז

$$E\left[(Y_n + h_{n-1}Y_{n-1} + \dots + h_1Y_1)^2\right] = 0$$

ולכן

$$Y_n = -h_{n-1}Y_{n-1} - h_{n-2}Y_{n-2} - \dots - h_1Y_1$$

דהיינו, אם נתונים Y_1, \dots, Y_{n-1} אזי נוכל לנחש את Y_n ללא שגיאה. במלים אחרות, באם ידועים Y_1, \dots, Y_{n-1} אזי אין צורך בידיעת Y_n . כלומר יש כאן יתירות: כל פונקציה לינארית של Y_1, \dots, Y_n אפשר לייצג כפונקציה לינארית של Y_1, \dots, Y_{n-1} : המשתנה Y_n אינו מוסיף מידע לצורך השערוך הלינארי. לכן, במקום לשערך את X כשנתון Y_1, \dots, Y_n מספיק לשערך את X כשנתון Y_1, \dots, Y_{n-1} . אם $h_n = 0$ אבל $h_i \neq 0$ עבור i כלשהוא יהיה החשבון בדיוק אותו דבר. מוסר השכל: אם \underline{A}_Y סינגולרית אזי קיים i_0 (לפחות אחד) כך שאפשר לותר על Y_{i_0} מבלי לקלקל את השערוך. אם המטריצה החדשה $(n-1) \times (n-1)$ שמתקבלת ע"י מחיקת העמודה והשורה i_0 היא מטריצה לא סינגולרית - נמשיך בשערוך לפי התוצאות שבידינו עבור הבעיה המוקטנת. אם המטריצה החדשה עדיין סינגולרית, נמשיך בתהליך הצימצום עד שנגיע למטריצה לא סינגולרית.

שימו לב שמקרה זה אינו דומה לדוגמה 2.7 בה המשערך "מתעלם" מאחת מהמדידות כיוון שהיא אינה תורמת עקב היותה "מדידה אחרת בתוספת רעש". סינגולריות מבטאת חוסר יחידות של הפתרון. ואכן, במקרה של סינגולריות ניתן לרשום משערכים שונים, חלקם תלויים מפורשות ב- Y_n וחלקם לאו, כאשר למעשה הם זהים. למשל, אם $Y_3 = Y_1 + Y_2$ ואם מצאנו משערך אופטימלי $\hat{X} = Y_1 + Y_2 + Y_3$ אזי $\hat{X} = 2Y_3$ וזוהי צורה אחרת לרשום את אותו משערך אופטימלי.

כלומר, בשתי הדוגמאות, המשערך "מתעלם" מאחת המדידות, אבל מסיבות שונות לחלוטין. במקרה אחד, מדידה מסוימת לא משפרת בהנתן המדידות האחרות; במקרה השני, מדידה מסוימת היא צירוף לינארי של שאר המדידות, ולכן ניתן לבטא את הפתרון, בסופו של דבר, כצירוף לינארי של פחות מדידות.

2.4 עקרון ההשלכה (עקרון האורתוגונליות)

מצאנו שהמשערך הלינארי האופטימלי של X כשנמדד \underline{Y} הוא $\hat{X}_{\text{opt}}^l = \underline{a}_*^T \underline{Y}$ כאשר הקבועים \underline{a}_* מקיימים:

$$E[\underline{Y} \cdot X] = E[\underline{Y} \cdot \underline{Y}^T] \cdot \underline{a}_* \quad (2.26)$$

או בצורה מפורטת יותר

$$E[(X - \underline{a}_*^T \underline{Y}) \cdot Y_i] = 0 \quad \text{עבור כל } i \quad (2.27)$$

כלומר שגיאת השערוך $\varepsilon = X - \underline{a}_*^T \underline{Y}$ חסרת קורלציה עם כל (פונקציה לינארית של) המדידה.

חשוב להדגיש כי תכונה זו שקולה לתכונת האופטימליות, כלומר - זה "אם ורק אם". המטרה שלנו כעת היא לנסות לקבל תובנה לעובדה זו מנקודת מבט גאומטרית, על-ידי הסתכלות על בעיה גאומטרית הדומה מאד לבעיית השערוך.

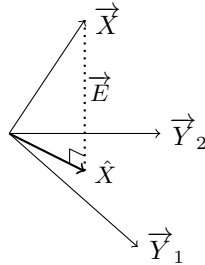
מדוע יש לצפות כי שגיאת השערוך של המשערך האופטימלי תהיה חסרת קורלציה עם המדידות? ומדוע תכונה דומה קיימת במקרה הלינארי? כדי לקבל תובנה גאומטרית, נסתכל על תופעה מקבילה: הטלה של וקטור (אוקלידי) על תת מרחב. כפי שהגדרנו בעיית שערוך לינארי של מ"א X על סמך ו"א \underline{Y} , בה אנו מחפשים וקטור מקדמים \underline{a} עבורו השגיאה הריבועית $E[(X - \underline{a}^T \underline{Y})^2]$ מינימלית, כך נגדיר בעית הטלה גאומטרית, בה מחפשים וקטור מקדמים כך שהשגיאה הריבועית $X - \underline{a}^T \underline{Y}$ היא מינימלית, כאן המדידות \underline{Y} הן אוסף של וקטורים: נפרט כעת את הבעיה.

יהיה \underline{X} וקטור דטרמיניסטי במרחב האוקלידי ה- m ממדי \mathbb{R}^m . יהיו $\underline{Y}_1, \dots, \underline{Y}_n$ וקטורים דטרמיניסטיים באותו מרחב m ממדי. נעיין בתת המרחב $M(\underline{Y}_1, \dots, \underline{Y}_n)$ הנפרש ע"י $\sum_{i=1}^n \theta_i \underline{Y}_i$ (כאשר המשתנים θ_i מקבלים את כל הערכים הממשיים האפשריים) ונניח שזה תת מרחב ממש של \mathbb{R}^m (ז.א. לא \mathbb{R}^m עצמו). נסמן ב- \hat{X} את ההשלכה של \underline{X} על M כלומר, הנקודה ב- M שהיא הקרובה ביותר ל- \underline{X} (לדוגמה, נחשוב על המקרה בו $n = m - 1$ ו- \underline{Y}_i הוא וקטור לאורך הקואורדינטה ה- i . M יהיה אם כן אוסף הנקודות עבורן הקואורדינטה האחרונה שווה אפס).

טענה: \hat{X} הוא ההיטל, של \underline{X} על M אם ורק אם עבור כל i מתקיים: $(\underline{X} - \hat{X}) \perp \underline{Y}_i$. כלומר אם ורק אם "שגיאת השערוך" ניצבת למדידות.

ניתן להשתכנע בכוננות הטענה על-ידי הפעלת משפט פיתגורס לחישוב השגיאה²:

$$\vec{E} \perp \vec{Y}_1, \vec{E} \perp \vec{Y}_2, \|\vec{X}\|^2 = \|\vec{\hat{X}}\|^2 + \|\vec{E}\|^2$$



איור 2.2: משפט פיתגורס: אלמלא היתה ניצבות, ניתן היה לפרק את השגיאה לרכיב הנפרש במרחב ולרכיב שניצב למרחב, ואז לרכיב הניצב (השגיאה החדשה) היתה נורמה קטנה יותר.

דרך נוספת להמחשת הטענה היא על-ידי כתיבת השגיאה מפורשות והבאתה למינימום:

$$\begin{aligned} \min_{\underline{a}} \|\underbrace{\underline{X} - \sum_j a_j Y_j}_{\varepsilon}\|^2 &= \min_{\underline{a}} (\underline{X} - \sum_j a_j Y_j, \underline{X} - \sum_j a_j Y_j) \\ 0 &= \frac{\partial \|\varepsilon\|^2}{\partial a_j} = -2(Y_j, \underline{X} - \sum_j a_j Y_j) = -2(Y_j, \varepsilon) \\ 0 &= (\underline{X} - \sum_j a_j Y_j, Y_j) \end{aligned}$$

²גרסה ב' מכילה הוכחה מעט שונה למשפט זה

זוהי בעיה זהה עם ההקבלות הבאות:

$$E[Y_i Y_j] \leftrightarrow (Y_i, Y_j)$$

$$E[\underline{X} Y_j] \leftrightarrow (\vec{X}, Y_j)$$

ניצבות \leftrightarrow חוסר קורלציה

מעניין לחזור ולעיין במשעריך האופטימלי (הלא-דווקא לינארי) בהקשר של עקרון האורתוגונליות, ולראות שגם עבורו ישנו עקרון דומה, אם כי הרבה יותר רחב. שגיאת השערוך של המשעריך האופטימלי, $\varepsilon = X - E[X|Y]$, היא חסרת קורלציה, לא רק עם כל אחת מהמדידות, אלא עם כל פונקציה של המדידות, $g(Y)$:

$$\begin{aligned} E[(X - E[X|Y])g(Y)] &= E[E[(X - E[X|Y])g(Y)|Y]] \\ &= E[(E[X|Y] - E[X|Y])g(Y)] \\ &= 0 \end{aligned}$$

3 תהליכים אקראיים בזמן בדיד

"וקטור אקראי" (הגדרה 8.14) הוא וקטור שרכיביו הם משתנים אקראיים: זוהי הרחבה של מושג ה"משתנה אקראי" למספר מימדים. המושג "תהליך אקראי" הוא הרחבה נוספת, כאשר מספר המשתנים יכול להיות אינסופי. בנוסף, אנו מתייחסים לסדרת המשתנים כאילו הופיעו בנקודות זמן עוקבות (דגימות בזמן).

דוגמה 3.1 נניח שמשדרים אות ספרתי: כלומר, בכל יחידת זמן k משודרת הסיפרה 0 או הסיפרה 1. נניח שלאות זה מתווסף רעש: כלומר בזמן k אנו קולטים את $X(k) + N(k)$, המורכב מהגודל המשודר $X(k)$ ומרעש $N(k)$. האות שאנו קולטים לכן מורכב ממשתנים אקראיים: משתנה אחד בכל יחידת זמן. לאות כזה קוראים תהליך אקראי.

כלומר, תהליך אקראי בזמן בדיד הוא סדרה אינסופית של מ"א. אפשר לחשוב עליו גם כעל פונקציה אקראית של משתנה בדיד (משתנה הזמן). הסתכלות כזו ניתנת להרחבה לתהליך בזמן רציף—אשר הוא פונקציה אקראית של משתנה זמן רציף.

הגדרה 3.2 תהליך אקראי בזמן בדיד הוא סדרה של משתנים אקראיים $\{X(n)\}$ (כלומר n מקבל ערכים בדידים). פונקציית מדגם של התהליך האקראי היא כל פונקציה של משתנה הזמן n עבור $\omega = \omega_0$ קבוע. כלומר, זוהי כל פונקציה מהצורה

$$X(n, \omega_0), \quad \omega_0 \text{ קבוע}$$

כאשר n מקבל את כל הערכים השלמים החיוביים, או את כל הערכים השלמים. תהליך בזמן בדיד נקרא גם "סדרה עתית" (מלשון עת). מקובל לסמן שלמים באותיות i, j, k, l, m, n , ולכן נשתמש לעיתים באותיות אלו לציון משתנה הזמן של תהליך אקראי בזמן בדיד. בדרך כלל נסמן ת"א בזמן בדיד על ידי $X(n, \omega)$ או $X_n(\omega)$, אולם לעיתים קרובות נשתמש בשני הסימונים המקוצרים (השקולים) $X(n)$ או X_n וכו'.

3.1 חוק ההסתברות של תהליך אקראי

עבור וקטור אקראי, ניתן להגדיר פונקציית פילוג (הגדרה 8.15). אולם בבואנו להרחיב הגדרה זו לתהליך אקראי, אנו נתקלים בקשיים: כוון שתהליך אקראי יכול להיות מוגדר עבור איך-סוף נקודות זמן, לא ברור כיצד נחשב הסתברויות של מאורעות, המוגדרים על ידי אין סוף משתנים. אחת הבעיות מתוארת בתרגיל הבא.

תרגיל 3.3 נתונה סדרת משתנים אקראיים $\{X_i(\omega), i = 1, 2, \dots\}$ המתארים זריקות מטבע בלתי תלויות (1 מתאר עץ, 0 מתאר פלי). הראה כי נדע לחשב את ה"פילוג האיך-סופי"

$$(3.1) \quad \mathbb{P}\{X_1 \leq a_1, X_2 \leq a_2, \dots\}$$

לכל סידרה איך-סופית a_1, a_2, \dots אם ורק אם מתקיים התנאי הבא. לכל N , לכל סידרת זמנים (אינדקסים) $n_1 < n_2 < \dots < n_N$ ולכל סידרה a_1, a_2, \dots, a_N באורך N אנו יודעים לחשב את הפילוג של הווקטור

$$F_{X(n_1), X(n_2), \dots, X(n_N)}(a_1, a_2, \dots, a_N) = \mathbb{P}\{X(n_1) \leq a_1, X(n_2) \leq a_2, \dots, X(n_N) \leq a_N\}$$

רמז: בדוק בניפרד את המקרה בו מספר הפעמים ש- $a_i < 1$ הוא איך-סופי (ואז ההסתברות ב-(3.1) היא 0), ואת המקרה בו $a_i \geq 1$ פרט למספר סופי של פעמים (ואז ניתן לרשום את (3.1) על ידי אוסף סופי של מ"א).

הקושי העיקרי במקרה זה הוא כי בבואנו לחשב הסתברות של מאורע המוגדר ע"י מספר אינסופי של תנאים, בדרך כלל הסתברות זו תהיה שווה 0. לכן נוח להעזר בהכללה הבאה.

הגדרה 3.4 חוק ההסתברות של תהליך אקראי בזמן בדיד הוא אוסף כל פונקציות הפילוג המשותפות, המוגדרות עבור כל סדרת זמנים סופית. כלומר, זהו האוסף של כל הפילוגים במימד סופי

$$F_{X(n_1), X(n_2), \dots, X(n_N)}(a_1, a_2, \dots, a_N) = \mathbb{P}\{X(n_1) \leq a_1, X(n_2) \leq a_2, \dots, X(n_N) \leq a_N\}$$

עבור כל N , כל סדרת זמנים (אינדקסים) $n_1 < n_2 < \dots < n_N$ וכל סדרה a_1, a_2, \dots, a_N . הגודל $F_{X(n)}$ שהוא הפילוג של המשתנה האקראי $X(n)$ נקרא הפילוג החד-מימדי, או הפילוג השולי של התהליך בזמן n .

כמובן שלא כל אוסף פילוגים סופיים מתאר פילוג של תהליך. לדוגמה, נתבונן בתהליך המוגדר עבור זמנים חיוביים $\{X(1), X(2), \dots\}$. האם יתכן שמתקיימים שני השוויונים

$$\mathbb{P}\{X(1) \leq a\} = 0.5$$

$$\mathbb{P}\{x(1) \leq a, X(2) \leq b\} = 1$$

בו זמנית? זה ודאי לא יתכן, שכן למאורע הראשון בוודאי הסתברות גדולה יותר מאשר לשני. לכן ברור כי חייב להיות תנאי עקביות כלשהו.

אוסף פילוגים מתאר פילוג של תהליך אם ורק אם הוא מקיים את דרישת העקביות (קונסיסטנטיות) הבאה.

הגדרה 3.5 דרישת עיקביות: לכל $N \geq 2$ ו- $1 \leq k \leq N$, לכל סידרת זמנים (אינדקסים) $n_1 < n_2 < \dots < n_N$ ולכל סידרה a_1, a_2, \dots, a_N מתקיים התנאי

$$\begin{aligned} & \mathbb{P}\{X(n_1) \leq a_1, \dots, X(n_{k-1}) \leq a_{k-1}, X(n_{k+1}) \leq a_{k+1}, \dots, X(n_N) \leq a_N\} \\ &= \mathbb{P}\{X(n_1) \leq a_1, \dots, X(n_{k-1}) \leq a_{k-1}, X(n_k) < \infty, X(n_{k+1}) \leq a_{k+1}, \dots, X(n_N) \leq a_N\} \end{aligned}$$

או, בסימון של פונקציות פילוג,

$$\begin{aligned} & F_{X(n_1), \dots, X(n_{k-1}), X(n_{k+1}), \dots, X(n_N)}(a_1, \dots, a_{k-1}, a_{k+1}, \dots, a_N) \\ &= F_{X(n_1), \dots, X(n_N)}(a_1, \dots, a_{k-1}, \infty, a_{k+1}, \dots, a_N) \end{aligned}$$

כלומר הפילוג השולי של פונקצית הפילוג מממדים גבוהים מתאים לפונקצית הפילוג בממדים נמוכים יותר.

תרגיל 3.6 הראה כי בהגדרה 3.4, מספיק להשתמש בסדרות זמן מהצורה $N_1, N_1 + 1, \dots, N_1 + N$.

מספר דוגמאות של אותות אקראיים בזמן בדיד.

דוגמה 3.7 תהיה $\{X(n), n = 1, 2, \dots\}$ סדרה של משתנים בלתי תלויים סטטיסטית ושווי פילוג (בקצרה) i.i.d = independent, identically distributed אפשר להתייחס לסדרה כאל תהליך אקראי בזמן בדיד, כאשר הזמנים הם שלמים $n = 1, 2, \dots$

תהליך i.i.d מופיע למשל כסידרת הזכיות בהפעלות חוזרות של מכונת הימורים. בנוסף, במקרים רבים מודל הרעש במדידות הוא כזה. הבה נראה כי הפילוג של המ"א $X(1)$ קובע את חוק ההסתברות של תהליך i.i.d, ונחשב חוק ההסתברות זה. נסמן ב- F_X את חוק הפילוג של המ"א $X(1)$ (שהוא גם חוק הפילוג של כל אחד מהמשתנים האחרים).

יהיו $\{a_i, i = 1, 2, \dots\}$ מספרים ממשיים כלשהם. אזי בגלל אי התלות, אם $n_1 < n_2 < \dots < n_k$ נקבל

$$F_{X(n_1), X(n_2), \dots, X(n_k)}(a_1, a_2, \dots, a_k) = F_{X(n_1)}(a_1) \cdot F_{X(n_2)}(a_2) \cdots F_{X(n_k)}(a_k) = \prod_{i=1}^k F_X(a_i)$$

דוגמה 3.8 הילוך אקראי, או הילוך שיכור הוא תהליך של תוספות בת"ס ושווי פילוג. נגדיר תהליך אקראי $\{Y(n), n = 0, 1, \dots\}$ על ידי

$$Y(0) = 0, \quad Y(n) = Y(n-1) + X(n) \quad n = 1, 2, \dots$$

כאשר המשתנים $\{X(n), n = 1, 2, \dots\}$ הם בת"ס. התהליך $Y(n)$ יקרא תהליך עם תוספות בת"ס. אם לא יאמר אחרת, אנו נניח שהמשתנים $\{X(n), n = 1, 2, \dots\}$ הם שווי פילוג.

תהליך עם תוספות בת"ס ושווי פילוג נקרא גם הילוך אקראי (random walk), או הילוך שיכור, drunkard walk (מטעמים מובנים). אפשר (ומקובל) להגדיר תהליך כזה דרך התוספות:

$$(3.2) \quad Y(n) = \sum_{i=1}^n X(i)$$

דוגמא פשוטה לתהליך כזה הוא הרווח המצטבר בסידרת הימורים במכונת מזל.

עבור תהליך כזה, אם הוא מקבל למשל ערכים שלמים, ניתן לרשום ביטוי פשוט עבור פונקצית ההסתברות. בסימונים של דוגמה 3.7,

$$\begin{aligned} & \mathbb{P}\{Y(1) = a_1, \dots, Y(k) = a_k\} \\ &= \mathbb{P}\{Y(1) = a_1, \dots, Y(k-1) = a_{k-1}, Y(k-1) + X(k) = a_k\} \\ &= \mathbb{P}\{Y(1) = a_1, \dots, Y(k-1) = a_{k-1}, X(k) = a_k - a_{k-1}\} \\ &= \mathbb{P}\{Y(1) = a_1, \dots, Y(k-1) = a_{k-1}\} \cdot \mathbb{P}\{X(k) = a_k - a_{k-1}\} \end{aligned}$$

בגלל אי התלות. בצורה דומה נמשיך ונפתח

$$\begin{aligned} & \mathbb{P}\{Y(1) = a_1, \dots, Y(k) = a_k\} \\ &= \mathbb{P}\{X(1) = a_1, X(2) = a_2 - a_1, \dots, X(k) = a_k - a_{k-1}\} \\ &= \mathbb{P}\{X(1) = a_1\} \cdot \mathbb{P}\{X(2) = a_2 - a_1\} \cdots \mathbb{P}\{X(k-1) = a_{k-1} - a_{k-2}\} \cdot \mathbb{P}\{X(k) = a_k - a_{k-1}\} \\ &= \prod_{n=1}^k \mathbb{P}\{X(n) = a_n - a_{n-1}\} \end{aligned}$$

כאשר לצורך המעבר האחרון מגדירים $a_0 = 0$. אם $\{X(n)\}$ הם משתנים בעלי צפיפות אזי משיקולים דומים נקבל

$$f_{\underline{Y}}(\underline{a}) = f_{Y_0}(a_0) \cdot \prod_{i=1}^n f_{X_i}(a_i - a_{i-1}) \quad (3.3)$$

דוגמה 3.9 נתון ת"א $X(n) = A \cos(\omega_0 n + \phi)$ כאשר A, ϕ הם מ"א עם פילוג משותף ידוע. במקרה כזה הפילוג שלהם קובע את חוק ההסתברות של התהליך $\{X(n)\}$. כך למשל, בהנתן סדרה (סופית) של קבועים a_1, \dots, a_n ניתן לחשב את הפילוג $\mathbb{P}\{A \cos(\omega_0 k + \phi) \leq a_k, 1 \leq k \leq n\}$ בדרך הבאה. עבור כל k הדרישה $A \cos(\omega_0 k + \phi) \leq a_k$ קובעת אילוץ על הערכים של A, ϕ . ההסתברות שכל האילוצים מתקיימים היא בדיוק הפילוג המבוקש.

הגדרה 3.10 תהליך אקראי $\{X(n)\}$ יקרא תהליך גאוזי אם לכל שלם חיובי N ולכל N שלמים n_1, \dots, n_N הוקטור האקראי $(X(n_1), \dots, X(n_N))$ הוא וקטור אקראי גאוזי.

למשל: תהליך i.i.d. עבורו הפילוג של כל אחד מהמשתנים האקראיים הוא גאוזי, הוא ת"א גאוזי (מדוע?).

3.2 תוחלת ומומנטים של תהליך אקראי

הגדלים הבסיסיים ביותר המתארים תהליך אקראי הם פונקציית התוחלת, הווריאנס ופונקציית אוטוקורלציה. עבור משתנים אקראיים, התוחלת (הגדרה 8.17), המומנטים, השונות (הגדרה 8.22), והקווריאנס (הגדרה 8.23), הם מספרים. בעזרתם נוכל להגדיר תכונות מקבילות של תהליך אקראי, הכוללות מומנטים עד (כולל) סדר שני, ע"י התבוננות במשתנה האקראי $X(n_1)$ או בזוג המשתנים $(X(n_1), X(n_2))$. הגדלים יהיו כמובן תלויים בזמנים n_1, n_2 .

הגדרה 3.11 עבור תהליך אקראי $\{X(n), -\infty < n < \infty\}$ (כאשר n שלם) נגדיר את

$$\mu_X(n) = \mathbb{E}[X(n)] \quad 1. \text{ פונקציית התוחלת}$$

$$\mathbf{R}_X(n_1, n_2) = \mathbb{E}[X(n_1) \cdot X(n_2)] \quad 2. \text{ פונקציית האוטוקורלציה}$$

$$\begin{aligned} \mathbf{K}_X(n_1, n_2) &= \mathbb{E}[(X(n_1) - \mu_X(n_1)) \cdot (X(n_2) - \mu_X(n_2))] \\ &= \mathbf{R}_X(n_1, n_2) - \mu_X(n_1) \cdot \mu_X(n_2) \end{aligned}$$

פונקציות התוחלת והאוטוקורלציה מספקות הערכות גסות על התהליך: לא רק הממוצע (קירוב דטרמיניסטי) והפיזור (ווריאנס), אלא גם הערכה של התלות הסטטיסטית (הלינארית) של ערכי התהליך בנקודות זמן שונות.

הבה נחשב פונקציות אלו עבור תהליך i.i.d (דוגמה 3.7) והילוך שיכור (דוגמה 3.8).

דוגמה 3.12 עבור התהליך $X(n)$,

$$\begin{aligned} \mu_X(n) &= \mathbb{E}[X(n)] \\ &= \mathbb{E}[X(1)] \end{aligned}$$

כיוון שהפילוג אינו תלוי ב- n , ולכן $\mu_X(n) = \mu_X$ אינו תלוי בזמן n . פונקצית האוטוקורלציה היא

$$\begin{aligned} \mathbf{R}_X(n_1, n_2) &= \mathbb{E}[X(n_1)X(n_2)] \\ &= \begin{cases} \mathbb{E}[(X(1))^2] & \text{אם } n_1 = n_2 \\ \mu_X^2 & \text{אם } n_1 \neq n_2. \end{cases} \end{aligned}$$

בגלל אי התלות בזמן, ובגלל אי התלות הסטטיסטית בין $X(n_1)$ לבין $X(n_2)$ כאשר הזמנים הם שונים. כיוון ש-

$$\mathbb{E}[(X(1))^2] = \mathbb{E}[(X(1) - \mu_X + \mu_X)^2] = \sigma_x^2 + 2\mathbb{E}[\mu_X(X(1) - \mu_X)] + \mu_X^2 = \sigma_x^2 + \mu_X^2 \quad (3.4)$$

קיבלנו את הנוסחה הכללית

$$\mathbf{R}_X(n_1, n_2) = \mu_X^2 + \sigma_x^2 \delta(n_1 - n_2) . \quad (3.5)$$

(מדובר בפונקציה ה- δ של קרונקר—בזמן בדיד—ולא בפונקציה המוכללת!). מכאן נובע מייד כי

$$\mathbf{K}_X(n_1, n_2) = \sigma_x^2 \delta(n_1 - n_2) . \quad (3.6)$$

נשים לב כי $\mathbf{K}_X(n_1, n_2)$ ו- $\mathbf{R}_X(n_1, n_2)$ תלויים במשתני הזמן רק דרך ההפרש $n_1 - n_2$.

זו דוגמה חשובה מאד שעוד נעסוק בה ובאקוויולנט שלה בזמן רציף.

עבור הילוך אקראי $Y(n)$, נעזר בייצוג (3.2) ונקבל בעזרת הלינאריות של התוחלת (טענה 8.20),

$$\begin{aligned} \mu_Y(n) &= \mathbb{E}[Y(n)] \\ &= \sum_{i=1}^n \mathbb{E}[X(i)] \\ &= \sum_{i=1}^n \mu_X(i) \end{aligned}$$

כוון שההפרשים הם שווי פילוג, נקבל את הנוסחה הפשוטה

$$\mu_Y(n) = n\mu_X$$

לגבי המומנט השני, נניח ללא הגבלת הכלליות: $n_1 \leq n_2$

$$\begin{aligned} \mathbf{R}_Y(n_1, n_2) &= E[Y(n_1)Y(n_2)] = E\left\{\left[\sum_{i=1}^{n_1} X(i)\right]\left[\sum_{j=1}^{n_2} X(j)\right]\right\} \\ &= E\left\{\left[\sum_{i=1}^{n_1} X(i)\right]\left[\sum_{j=1}^{n_1} X(j) + \sum_{j=n_1+1}^{n_2} X(j)\right]\right\} \\ (3.7) \quad &= E\left[\sum_{i=1}^{n_1} X(i)X(j)\right] + \left(E\left[\sum_{i=1}^{n_1} X(i)\right]\right)\left(E\left[\sum_{j=n_1+1}^{n_2} X(j)\right]\right) \\ (3.8) \quad &= n_1(\mu_X^2 + \sigma_X^2) + (n_1^2 - n_1)\mu_X^2 + n_1\mu_X(n_2 - n_1)\mu_X \\ (3.9) \quad &= n_1n_2\mu_X^2 + n_1\sigma_X^2 \end{aligned}$$

כאשר המעבר ב-(3.7) נובע מאי-תלות והמעבר ב-(3.8) נובע מההפרדה בין $i = j$ ל- $i \neq j$. כמו כן נשים לב שהפיתוח כולל את המקרה בו $n_1 = n_2$ כי סכום על קבוצה ריקה הוא 0. לכן:

$$\mathbf{K}_Y(n_1, n_2) = \mathbf{R}_Y(n_1, n_2) - (n_1\mu_X)(n_2\mu_X) = n_1\sigma_X^2$$

באופן כללי (מאחר שהנחנו $n_1 \leq n_2$):

$$\mathbf{R}_Y(n_1, n_2) = n_1n_2\mu_X^2 + \min\{n_1, n_2\}\sigma_X^2$$

$$\mathbf{K}_Y(n_1, n_2) = \min\{n_1, n_2\}\sigma_X^2$$

נשים לב כי אם $\{Y(n)\}$ הוא ת"א גאוסית אזי מההגדרות נובע כי $\mu_Y(n) = \mu_Y(k, n)$ קובעים את הפילוג של התהליך.

דוגמה כללית יותר מתקבלת אם בוחנים אלגוריתמים רקורסיביים. מקובל לרשום אלגוריתמים כאילו בצורה הבאה.

דוגמה 3.13 יהיה $\{X(n), n = 0, 1, \dots\}$ תהליך i.i.d. תהי g פונקציה נתונה, ונקבע את $Y(0) = y_0$. נגדיר תהליך ע"י $\{Y(n), n = 0, 1, \dots\}$

$$Y(n+1) = g(Y(n), X(n)), \quad n = 0, 1, \dots, \quad Y(0) = y_0$$

זהו ניסוח כללי של אלגוריתם רקורסיבי "בנוכחות רעש". נוסחה זו היא כללית מדי מכדי שנוכל לנתח אותה. אולם ניתן לנתח מקרים פשוטים, למשל:

תרגיל 3.14 יהי $\{X(n), n = 1, 2, \dots\}$ אוסף של מ"א בת"ס ושווי פילוג,

$$\mathbb{P}\{X(1) = 1\} = \mathbb{P}\{X(1) = -1\} = \frac{1}{2}$$

יהיה $Y(0) = 0$ ונגדיר פונקציה $g(y, x) = x - \frac{y}{2}$. עבור דוגמאות 3.7, 3.8 ו-3.13, חשב את (או רשום נוסחאות עבור) פונקציות התוחלת μ_X ו- $\mu_Y(n)$, ופונקציות האוטוקורלציה $R_X(n_1, n_2)$ ו- $R_Y(n_1, n_2)$. שרטט (רצוי בעזרת MATLAB) את פונקציות התוחלת, מספר פונקציות מדגם של התהליך, ואת הממוצע של פונקציות המדגם (אם פונקציות המדגם הן $X(n, \omega_1), X(n, \omega_2), \dots, X(n, \omega_k)$ אזי הממוצע של פונקציות המדגם הוא

$$\frac{1}{k} \sum_{n=1}^k X(n, \omega_n)$$

שהיא פונקציה של משתנה הזמן n). שרטט את פונקציות האוטוקורלציה $R_X(1, n)$ ו- $R_Y(1, n)$ וכן את הפונקציות

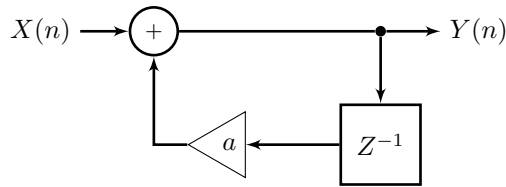
$$C_X(1, n; \omega) = X(1, \omega) \cdot X(n, \omega)$$

עבור מספר ערכים של ω ואת הממוצע של פונקציות אילו. הסק מסקנות.

דוגמה 3.15 נתבונן כעת במודל בו התהליך $\{Y(n)\}$ נוצר על ידי מעבר של רעש לבן במערכת לינארית מסדר ראשון, כך ש-

$$Y(n) = aY(n-1) + X(n), \quad n = 1, 2, \dots, \quad Y(0) = 0 \quad (3.10)$$

כאשר $|a| < 1$.



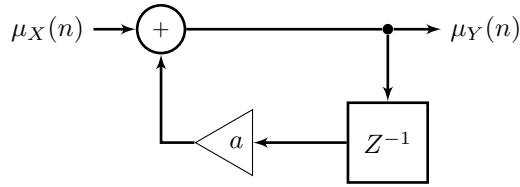
איור 3.1: המערכת היוצרת את התהליך $\{Y(n)\}$

נחשב את המומנטים של $\{Y(n)\}$. ניתן לרשום את $\{Y(n)\}$ כסכום משוקלל של $\{X(n)\}$ עם משקלים הנתונים על-ידי חזקות של a , אבל אפשר גם לחשב מומנטים על-ידי משוואת הפרש רקורסיביות:

$$(3.11) \quad \mu_Y(n) \doteq \mathbb{E} Y(n)$$

$$(3.12) \quad = \mathbb{E}[aY(n-1) + X(n)]$$

$$(3.13) \quad = a\mu_Y(n-1) + \mu_X(n) .$$

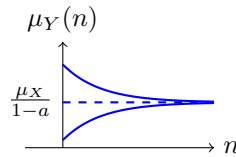


איור 3.2: המערכת היוצרת את סדרת התוחלות, $\{\mu_Y(n)\}$

קיבלנו משוואת הפרשים (דטרמיניסטית) מסדר ראשון בסדרה $\{\mu_Y(n)\}$ בתלות בסדרה $\{\mu_X(n)\}$. נשים לב שאפשר לחשוב על הסדרה $\{\mu_Y(n)\}$ כתגובה של אותה המערכת לכניסה שהיא סידרת הממוצעים $\{\mu_X(n)\}$ (עם תנאי ההתחלה המתאים - ממוצע תנאי ההתחלה!). בפרט, אם $\mu_X(n) = \mu_X$ אינו תלוי בזמן, אזי $\mu_Y(n)$ מתכנס לפתרון המשוואה

$$\mu_Y = a\mu_Y + \mu_X \quad (3.14)$$

כלומר ל- $\mu_Y = \mu_X / (1 - a)$ (ראו איור 3.3).



איור 3.3: התכנסות סדרת התוחלות עבור התהליך $\{Y(n)\}$ ותנאי התחלה שונים.

בצורה דומה ניתן לחשב את המומנט השני. לשם פשטות נניח $\mu_X(n) = 0$ לכל n . נקבל

$$\sigma_Y^2(n) = \mathbb{E} Y^2(n) \quad (3.15)$$

$$= \mathbb{E} [aY(n-1) + X(n)]^2 \quad (3.16)$$

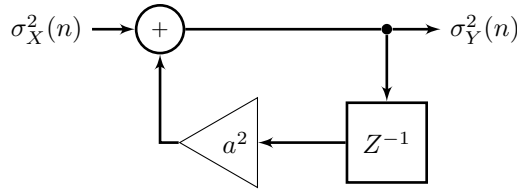
$$= a^2 \sigma_Y^2(n-1) + 2a \mathbb{E}[Y(n-1)X(n)] + \sigma_X^2(n) . \quad (3.17)$$

כיוון ש- $Y(n-1)$ הוא פונקציה של $\{X(k), k \leq n-1\}$ נובע מתכונות רעש לבן כי $X(n), Y(n-1)$ הם בת"ס. מכיוון שהממוצע של $X(n)$ הוא אפס, קיבלנו $\mathbb{E}[Y(n-1)X(n)] = 0$. לסיכום,

$$\sigma_Y^2(n) = a^2 \sigma_Y^2(n-1) + \sigma_X^2(n) . \quad (3.18)$$

כלומר, המומנט השני מתנהג כמו תגובה של מערכת ליניארית לכניסה שהיא המומנט השני של הכניסה האקראית, אולם כעת "המערכת" הליניארית היא עם פרמטר אחר- a^2 . בהמשך נראה כי זו התנהגות אפיינית. גם כאן, אם הווריאנס של $X(n)$ קבוע נקבל

$$\sigma_Y^2(n) \rightarrow \frac{\sigma_X^2}{1 - a^2} . \quad (3.19)$$



איור 3.4: המערכת היוצרת את סדרת השונויות, $\{\sigma_Y^2(n)\}$

תרגיל: מצא משוואת הפרש ל- $\mathbf{K}_Y(n, n - m)$ בתלות ב- n עבור m קבוע.

את המושגים בפרק זה העוסקים בתהליכים אקראיים ניתן להרחיב בצורה מיידית לתהליכים וקטוריים: למשל, ת"א וקטורי $\{\underline{X}(n)\}$ הוא אוסף של וקטורים אקראיים. חוק ההסתברות מוגדר דרך הוקטורים ואת שאר ההגדרות נרחיב בהתאם.

תרגיל 3.16 בידינו רכיב אלקטרוני לינארי אשר בתחום העבודה שלו מקיים את המשוואה $v = i + a$ כאשר הפרמטר a הוא קבוע אך ערכו אינו ידוע. נקבע את i על ערך קבוע (וידוע). את v אננו יכולים למדוד במדויק, שכן המדידה ה- n נותנת את $v(n) = v + X(n) = i + a + X(n)$, כלומר מדידה ועוד רעש. אנו יודעים כי הרעש $\{X(n), n = 1, 2, \dots\}$ מורכב ממשתנים בת"ס ושווי פילוג, עם ממוצע אפס. אנו מפעילים את האלגוריתם הבא לחיפוש הערך של a :

$$\begin{aligned} a(n+1) &= a(n) + \frac{1}{n} [v(n) - (i + a(n))] \\ &= a(n) + \frac{1}{n} [(i + a + X(n)) - (i + a(n))] \end{aligned}$$

שימו לב כי האלגוריתם ניתן לביצוע (הוא משתמש רק בגדלים ידועים או נמדדים).

פשט אלגוריתם זה על ידי ההצבה $b(n) = a(n) - a$. רשום נוסחה מפורשת עבור התהליך $b(n)$, ותאר את התנהגות האלגוריתם לאחר מספר איטרציות רב (n גדול).

3.3 סטציונריות וארגודיות

בקורסים קודמים, כמו "תורת המעגלים החשמליים" ו-"אותות ומערכות" למדנו שכאשר מזינים מערכת לינארית, קבועה בזמן ויציבה, באות DC או באות מחזורי, מקבלים תגובה המורכבת מתופעת מעבר דועכת פלוס תגובת מצב מתמיד, כאשר הראשונה שבהן קשורה בתנאי התחלה. מסתבר שישנה תופעה מקבילה עבור אותות כניסה אקראיים, הן בזמן רציף והן בזמן בדיד. כאשר לאות הכניסה ישנה "התנהגות סטטיסטית" אחידה או קבועה בזמן, באיזה-שהוא מובן הגיוני שמיד נגדיר במדויק, אזי אפשר לצפות שגם לאות המוצא (האקראי) תהיה התנהגות כזו, לאחר חלוף תופעות המעבר. רמזים לכך למעשה כבר התחלנו לראות בדוגמא 3.14. אותה התנהגות סטטיסטית אחידה בזמן, או קבועה בזמן, נקראת סטציונריות ולהלן הגדרתה:

הגדרה 3.17 ת"א בזמן בדיד $\{X(n), -\infty < n < \infty\}$ נקרא סטציונרי אם לכל n ולכל $m \geq 1$ שלמים, פונקציית הפילוג של הו"א

$$(X(n), X(n+1), \dots, X(n+m-1))$$

זהה לזו של

$$(X(1), X(2), \dots, X(m)).$$

במילים אחרות, פונקציית הפילוג של $(X(n), \dots, X(n+m-1))$ אינה תלויה ב- n - אינווריאנטית בזמן.

נאמר שהתהליך סטציונרי החל מזמן N אם התנאי מתקיים עבור כל $n \geq N$ וכל $m > 0$.

דוגמה 3.18 יהיה Y מ"א כלשהו, ויהיה Z מ"א המתאר הטלת קוביה, כלומר Z מקבל את הערכים $\{1, 2, 3, 4, 5, 6\}$ בהסתברות שווה.

1. נגדיר $X(n) = Y$ לכל n . זהו כמובן תהליך סטציונרי (אם כי לא מעניין).

2. יהיה B מ"א בינרי המקבל את הערכים $\{1, -1\}$ בהסתברות שווה. אזי התהליך

$$B(n) \doteq B \cdot (-1)^n \quad (3.20)$$

הוא ת"א סטציונרי. כדי לראות זאת נשים לב כי אם נזיז את התהליך ב- τ שהוא מספר זוגי, אזי הוא לא ישתנה כלל. לעומת זאת אם τ הוא אי זוגי, התהליך החדש נראה כאלו החלפנו את B ב- $(-B)$. אולם $(-B)$ מפולג בדיוק כמו B .

3. נגדיר כעת $X(n) = Y[1 + \sin(\pi\alpha n)]$. אם α שלם, אזי אנו במקרה הראשון והתהליך סטציונרי. אם α אינו שלם, אזי התהליך אינו סטציונרי: כדי לראות זאת מספיק להתבונן בפילוג החד-מימדי $F_{X(n)}$.

4. נגדיר $X(n) = \sin[\alpha\pi(n+Z-1)]$. אם $\alpha = 1/3$ אזי התהליך הוא סטציונרי. כדי לראות זאת, נבחר τ כלשהוא ונגדיר $\tilde{Z} = [Z + \tau] \bmod 6 + 1$ הוא שלם בין 0 ל-5 המקיים $m = [m] \bmod 6 + 6k$ עבור k שלם. אזי גם \tilde{Z} מתאר זריקת קוביה ולכן הפילוג של $\tilde{X}(n) = \sin[(\pi/3)(n + \tilde{Z} - 1)]$ זהה לפילוג של התהליך $X(n)$. מצב שני, מכיון ש- $\sin[(\pi/3)n]$ היא פונקציה מחזורית עם מחזור באורך של 6, מתקיים

$$\sin[(\pi/3)(n + Z - 1 + \tau)] = \sin[(\pi/3)(n + [Z + \tau - 1] \bmod 6)] \quad (3.21)$$

$$= \sin[(\pi/3)(n + \tilde{Z} - 1)] \quad (3.22)$$

ולכן $\tilde{X}(n) = \sin[(\pi/3)(n + Z - 1 + \tau)]$. כיוון שהפילוג של \tilde{X} זהה לפילוג של X , התהליך הוא סטציונרי. שים לב שהתוצאה תלויה בערך של α , ובאופן כללי (לערכים אחרים של α) התהליך אינו סטציונרי. התוצאה שקיבלנו היא מקרה פרטי של העקרון הבא: אות מחזורי אשר הפאזה שלו היא משתנה אקראי המפולג בצורה אחידה על מחזור שלם הוא תהליך סטציונרי. ההוכחה זהה לזאת לעיל.

5. התהליך $Y \sin(\alpha\pi n + Z)$ הוא סטציונרי עבור $\alpha = 1/3$, אך לא בהכרח עבור ערכים אחרים (לאילו ערכים נקבל סטציונריות?)

6. תהי $\{X(n)\}$ סדרת מ"א בת"ס ושוי פילוג (כולם מפולגים לפי F_X):

$$F_{X(n), X(n+1), \dots, X(n+m-1)}(x_1, \dots, x_m) = \prod_{i=1}^m F_X(x_i)$$

ללא תלות ב- n .

7. יהי $Y(n) = \sum_{i=1}^n X(i)$ כאשר $X(i)$ כמו בסעיף הקודם (הילוך שיכור). כאן ברור כי אין אפילו סטציונריות ל- $m = 1$ מכיון ש- $\mu_Y(n)$ ו- $\sigma_Y^2(n)$ תלויים ב- n .

סטציונריות היא באופן כללי תכונה שקשה לבדוק גם כאשר חוק ההסתברות המלא של התהליך ידוע במלואו, וזה גם די נדיר שהיא מתקיימת. קיימת צורה חלשה יותר של סטציונריות, שקל יותר לבדוק אותה ושמתיימת לעתים יותר קרובות, וגם ברוב האפליקציות היא מספיקה. זוהי סטציונריות מבחינת המומנטים מסדר ראשון ושני:

הגדרה 3.19 תהליך אקראי בזמן בדיד $\{X(n), -\infty < n < \infty\}$ נקרא סטציונרי במובן הרחב wide sense stationary אם

$$1. \mathbb{E} X(n) = \mu_X(n) = \mu_X: \text{ המומנט הראשון אינו תלוי בזמן:}$$

$$2. \mathbf{R}(n_1, n_2) \text{ ולכן גם } \mathbf{K}(n_1, n_2) \text{ תלויים רק בהפרש } n_2 - n_1.$$

סטציונריות כזו נקראת סטציונריות במובן הרחב, בנגוד לסטציונריות במובן "הצר" של חוק ההסתברות המלא. (השני גורר את הראשון אך לא להיפך). עוד נחזור למושגים הללו כשנדבר על ת"א בזמן רציף.

בהתייחס לדוגמה 3.18 כמובן שכל אות סטציונרי הוא גם סטציונרי במובן הרחב. ברור גם שהבדיקה קלה יותר—למשל חישובו מומנט ראשון ושני עבור תהליך i.i.d, והם מקיימים את התנאים. לעומת זאת, הילוך שיכור אינו סטציונרי במובן הרחב כיוון שהמומנט תלוי בזמן, וגם במקרה של ממוצע אפס המומנט השני תלוי בזמן.

ארגודיות

עד כה הנחנו שהפילוגים של המ"א או התהליכים ידועים לנו. כאשר אנו מנסים לתאר תופעה טיבעית (או טכנולוגית) ע"י מודל של משתנים אקראיים, אין לנו ידע שכזה. אפשר כמובן להניח הנחות, אולם כיצד נבדוק את ההנחות? אם נחשוב על המדידות שבידינו לבצע, הרי שנוכל למדוד רק פונקציית מדגם אחת של התהליך (ω נבחר ע"י הטבע, ואינו בידנו!).

אך יש משפחה של תהליכים עבורם ניתן להתבונן בפונקציית מדגם אחת (לאורך זמן ארוך מספיק) וללמוד מכך על הפילוגים של התהליך. זוהי משפחת התהליכים הארגודיים.

הגדרה 3.20 ארגודיות. תהליך אקראי סטציונרי $\{X(n), n = \dots, -2, -1, 0, 1, 2, \dots\}$ נקרא ארגודי אם מתקיים לכל k , ולכל פונקציה (חסומה) g של $k + 1$ משתנים:

$$(3.23) \quad \forall \varepsilon > 0, \quad \lim_{N \rightarrow \infty} \mathbb{P} \left\{ \left| \frac{1}{N} \sum_{n=1}^N g(X_n, \dots, X_{n+k}) - E[g(X_0, \dots, X_k)] \right| < \varepsilon \right\} = 1$$

אם בידינו תהליך ארגודי, נוכל למשל לבחור מספר a ולהגדיר פונקציה מציינת

$$\mathbf{1}_{(-\infty, a]}(x) = \begin{cases} 1 & \text{אם } x \leq a \\ 0 & \text{אחרת} \end{cases}$$

כיוון ש-

$$\mathbb{E}[\mathbf{1}_{(-\infty, a]}(X(0))] = 1 \cdot \mathbb{P}\{X(0) \leq a\} = F_{X(0)}(a)$$

קיבלנו דרך לקרב את הפילוג של המשתנים האקראיים, ע"י מדידה של התהליך לאורך זמן וחישוב הממוצע (האריתמטי, לא הסטטיסטי, שאינו ידוע!):

$$F_{X(0)}(a) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbf{1}_{(-\infty, a]}(X_n) = \lim_{N \rightarrow \infty} \frac{\{\text{מספר ה-} n \text{ ימים, } 1 \leq n \leq N \text{ עבורם } X_n \leq a\}}{N} \quad (3.24)$$

בצורה דומה ניתן לקרב פילוגים רב ממדיים של תהליך ארגודי. למשל

$$F_{X(0), X(k)}(a, b) = \lim_{N \rightarrow \infty} \frac{\{\text{מספר ה-} n \text{ ימים, } 1 \leq n \leq N \text{ עבורם } X(n) \leq a \text{ וכן } X(n+k) \leq b\}}{N} \quad (3.25)$$

המסקנה היא שעבור תהליך ארגודי, ניתן ללמוד את חוק הפילוג של התהליך מתוך התבוננות בפונקציית מדגם אחת!

דוגמה 3.21 לגבי הדוגמא ש- X_n i.i.d נסמן $Y_i = \underbrace{\mathbf{1}_{(-\infty, a]}(x_i)}_{g(x)}$, ואז

$$\mu_Y = E[Y] = E[\mathbf{1}_{(-\infty, a]}(X)] = \mathbb{P}\{X \leq a\} = F_X(a)$$

נראה כי $\frac{1}{N} \sum_{i=1}^N Y_i \rightarrow F_X(a)$ (תדירות יחסית אמפירית של $\{X_i \leq a\}$, לפחות במובן הבא:

$$\mathbb{P}\left\{\left|\frac{1}{N} \sum_{i=1}^N Y_i - F_X(a)\right| > \varepsilon\right\} \xrightarrow[N \rightarrow \infty]{\forall \varepsilon > 0} 0$$

הוכחה: לפי צ'בישב,

$$\begin{aligned} \mathbb{P}\left\{\left|\frac{1}{N} \sum_{i=1}^N Y_i - F_X(a)\right| > \varepsilon\right\} &\leq \frac{E\left[\left(\frac{1}{N} \sum_{i=1}^N Y_i - F_X(a)\right)^2\right]}{\varepsilon^2} \\ &= \frac{1}{\varepsilon^2} E\left[\left(\frac{1}{N} \sum_{i=1}^N (Y_i - F_X(a))\right)^2\right] \\ &= \frac{1}{\varepsilon^2} \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E[(Y_i - F_X(a))(Y_j - F_X(a))] \\ &= \frac{1}{N^2 \varepsilon^2} N \cdot E[(Y_1 - F_X(a))^2] \\ &= \frac{1}{N \varepsilon^2} \left[\mathbb{P}\{Y_1 = 0\} (0 - F_X(a))^2 + \mathbb{P}\{Y_1 = 1\} (1 - F_X(a))^2\right] \\ &= \frac{1}{N \varepsilon^2} \left[(1 - F_X(a)) F_X(a)^2 + F_X(a) (1 - F_X(a))^2\right] \\ &= \frac{1}{N \varepsilon^2} \cdot \underbrace{F_X(a)(1 - F_X(a))}_{\leq 1/4} \leq \frac{1}{4N \varepsilon^2} \rightarrow 0 \end{aligned}$$

לדוגמה, עבור מדגם של $N = 25 \times 10^6$, מקבלים שערך של $F_X(a)$ בדיוק של $\varepsilon = 10^{-3}$ ביותר מ-99% מהניסויים.

דוגמה 3.22 אם X מ"א, התהליך $\{X(n) = X, n = \dots, -1, 0, 1, \dots\}$ (אינו תלוי בזמן) אינו ארגודי, שכן הממוצע האריתמטי תמיד שווה ל- X .

דוגמה 3.23 התהליך

$$B(n) \doteq B \cdot (-1)^n \quad (3.26)$$

(משוואה (3.20)) הוא ארגודי, כיוון שלכל k ופונקציה g

$$(3.27) \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N g(B(n), \dots, B(n+k)) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N g(B(n), -B(n), \dots, (-1)^{k-1}B(n), (-1)^k B(n))$$

$$(3.28) \quad = \frac{1}{2} g(B(1), -B(1), \dots, (-1)^{k-1}B(1), (-1)^k B(1))$$

$$(3.29) \quad + \frac{1}{2} g(B(2), -B(2), \dots, (-1)^{k-1}B(2), (-1)^k B(2))$$

$$(3.30) \quad = \mathbb{E}[g(B(0), \dots, B(k))].$$

אולם שים לב כי ארגודיות (ואף סטציונריות) אינה נשמרת כאשר מחלקים את התהליך לבלוקים. כדי להבין זאת נגדיר תהליך חדש (ווקטורי)

$$A(n) \doteq (B(2n), B(2n+1)). \quad (3.31)$$

אזי $A(n) = A(n+1)$ ולכן זהו תהליך קבוע, והוא אינו ארגודי כפי שראינו בדוגמה הקודמת.

המודלים המקובלים בהנדסה הם בדרך כלל של אותות על פרק זמן ארוך, אולם עם נקודת התחלה ונקודת סיום. עבור תהליך כזה, אנו מניחים שפרק הזמן הוא ארוך מספיק כדי שתכונות כמו (3.23) יתקיימו בקירוב עבור T סופי. אנו גם מניחים ש"תופעות מעבר", אם ישנן, כבר דעכו.

דוגמה 3.24 יהי $\{X(n), n = 0, 1, \dots\}$ רעש לבן. נגדיר משתנה אקראי נוסף $Y(0)$ ונדרוש שיהיה בת"ס ב- n $\{X(n), n = 0, 1, \dots\}$ ע"י

$$Y(n+1) = \frac{1}{2}Y(n) + X(n), \quad n = 0, 1, \dots$$

תהליך כזה נקרא תהליך AR (תהליך Auto Regressive), כפי שראינו בעבר. נגדיר משתנה אקראי Z על ידי

$$Z = \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k X(k)$$

התהליך $Y(n)$ בדרך כלל אינו סטציונרי (אלא אם נתחיל אותו בתנאי התחלה מתאימים: אילו?). אולם עבור n גדול, הפילוג של $Y(n)$ קרוב לפילוג של Z . מסיבה זו, התהליך $Y(n)$ מקיים את הגירסה הבאה של תכונת הארגודיות (3.23):

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{n=0}^T g[Y(n)] = \mathbb{E}[g(Z)]$$

בקירוב הנדסי, אם נתבונן בתהליך החל מזמן N_0 גדול מספיק, התהליך יהיה בקירוב סטציונרי וארגודי.

עוד נחזור לנושאי הסטציונריות והארגודיות כאשר נגיע לתהליכים אקראיים בזמן רציף, ואז גם נדבר על תכונות נוספות של ת"א סטציונריים שלא נגענו בהם עדיין, והן תהיינה תקפות הן לזמן רציף והן לזמן בדיד.

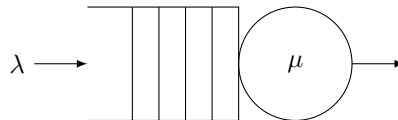
4 שרשראות מרקוב

שרשראות מרקוב מהוות משפחה של תהליכים אקראיים חשובים ומרכזיים, בשל היותם מודלים מתמטיים לתופעות בטבע ובהנדסה: רשתות תקשורת מחשבים, מערכות שירות, אלגוריתמים לדחיסה והתנהגות השקעות ובורסה הם דוגמה לתחומים בהם מודלים מרקוביים הם כלי מרכזי. מלבד היותם מודל עשיר, לשרשראות מרקוב תאוריה וכלים מתמטיים מפותחים ונוחים לשימוש.

להילוך שיכור (דוגמה 3.8) מבנה פשוט למדי. בפרט, אם ידוע שברגע n השיכור נמצא בנקודה $Y_n = y$, קל לחשב את הפילוג של מקומו ברגע $n + 1$: זאת משום שפילוג הצעד ידוע לנו. לצורך חישוב זה, אין זה משנה איזה מסלול עבר השיכור עד לזמן n . תכונה זו נקראת מרקוביות.

4.1 דוגמאות והגדרות

אחד המודלים הבסיסיים של רשתות מחשבים הוא תור עם שרת בודד, ונתייחס לגרסה שלו בזמן בדיד.



איור 4.1: תור עם שרת בודד

דוגמה 4.1 תור עם שרת בודד. אל תור יחיד מגיעים משתמשים חדשים לפי תהליך i.i.d. בינארי: כלומר, בכל רגע מגיע צרכן נוסף באופן בלתי תלוי ובהסתברות קבועה λ . שרת יחיד משרת את כל הצרכנים, ובכל יחידת זמן אחד מהם מסיים את עבודתו באופן בלתי תלוי, בהסתברות μ . נניח שברגע 0 היו $X(0)$ צרכנים, ונסמן ב- $X(n)$ את מספר המשתמשים אשר עדיין במערכת, ברגע n .

התור עם שרת בודד קרוב מאוד להילוך שיכור: כל עוד התור אינו ריק, פילוג הצעד הבא הוא קבוע: צעד של $+1$ אם מגיע משתמש ולא מסתיים שרות, כלומר בהסתברות $\lambda(1 - \mu)$, וצעד של -1 אם לא מגיע משתמש והסתיים שרות, כלומר בהסתברות $\mu(1 - \lambda)$. בהסתברות $\mu\lambda + (1 - \mu)(1 - \lambda)$ התור נשאר ללא שינוי: אם בגלל שלא קרה דבר (בהסתברות $(1 - \lambda)(1 - \mu)$), או משום שהיו גם הגעה וגם עזיבה (בהסתברות $\lambda \cdot \mu$). כאשר התור ריק אין כמובן עזיבות, ולכן פילוג הצעד הבא שונה ($+1$ בהסתברות λ , ו- 0 בהסתברות $1 - \lambda$). (הנחנו כאן שכאשר התור ריק, לקוח שמגיע לא יכול לקבל מייד שירות, אלא ממתין לנקודת הזמן הבאה.) לכן תהליך זה אינו הילוך שיכור. למרות זאת, התהליך שומר על תכונת המרקוביות: בהנתן גודל תור של $X(n)$ ברגע n , פילוג גודל התור ברגע $(n + 1)$ אינו תלוי באורך התור בעבר הרחוק יותר. בניגוד להילוך שיכור, החישוב תלוי במצב הנוכחי: החישוב כשהתור ריק שונה מהחישוב כאשר אינו ריק. ניתן לייצג את אורך התור גם בצורה הבאה: נסמן את אורך התור ברגע n ב- $Y(n)$ ויהיו $D(n)$ ו- $A(n)$ ת"א בינאריים $\{0, 1\}$ i.i.d. ובת"ס זה בזה המייצגים הגעות, וסיומי שירות, בהתאמה:

$$\mathbb{P}\{A_n = 1\} = \lambda, \quad \mathbb{P}\{D_n = 1\} = \mu \quad (4.1)$$

אזי אורך התור מקיים את המשוואה

$$Y(n+1) = Y(n) + A(n) - D(n)I[Y(n) > 0]. \quad (4.2)$$

מתאור זה ברור כי ההסטוריה אינה חשובה לצורך קביעת הפילוג בצעד הבא.

קיימים תהליכים רבים, הן בטבע והן בהנדסה, בעלי התכונה המרקובית, קרי שהמצב הנוכחי קובע את פילוג המצב הבא באופן חד משמעי. זה קורה בכל מבנה רקורסיבי מתאים.

בפרק זה נעסוק בשרשרות מרקוביות, כלומר בתהליכים מרקוביים המקבלים ערכים בדידים. הדבר ייקל מאד על ההבנה.

הגדרה 4.2 שרשרת מרקוב. יהי $\{X(n), n = 1, 2, \dots\}$ תהליך אקראי בזמן בדיד המקבל ערכים בדידים. התהליך יקרא שרשרת מרקוב אם לכל $n > k$ ו- $\{x_{n+1}, x_n, \dots, x_{n-k}\}$

$$\begin{aligned} \mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n, X(n-1) = x_{n-1}, \dots, X(n-k) = x_{n-k}\} \\ = \mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n\} \end{aligned}$$

כלומר, בהנתן הערך בהווה, העתיד אינו תלוי בעבר. נשים לב כי הפילוג בעתיד תלוי בעבר: חוסר התלות אשר בהגדרה היא אך ורק בהנתן המצב בהווה. חשוב לשים לב כי העתיד אינו תלוי בעבר כאשר ההווה נתון, אבל בהעדר התניה בהווה, העבר והעתיד בחלט יכולים להיות תלויים.

הביטוי $\mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n\}$ נקרא הסתברות המעבר של השרשרת. הגודל $X(n)$ נקרא המצב של התהליך ברגע n . אוסף המצבים האפשריים של השרשרת נקרא מרחב המצב, ונסמן אותו ב- S . ההגדרה תקפה כמובן גם לתהליך בזמן בדיד שאינו שלם, או לזמנים שלמים ולא דוקא חיוביים. לשם נוחיות בסימונים נניח בדרך כלל כי ערכי התהליך הם שלמים.

שים לב שגם עבור שרשרת מרקובית,

$$(4.3) \quad \mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n, X(n+2) = x_{n+2}\} \neq \mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n\}$$

תרגיל 4.3 יהי $\{Z(n)\}$ תהליך i.i.d המקבל ערכים ± 1 כאשר $\mathbb{P}\{Z(1) = 1\} = p$. יהי $\{Y(n)\}$ הילוך שיכור (דוגמה 3.8) עם הפרשים $\{Z(n)\}$. הראה כי שני התהליכים הם מרקוביים. חשב את הסתברויות המעבר והראה כי אינן תלויות בזמן (מדוע?). הראה כי עבור תהליך i.i.d מתקיים שוויון ב- (4.3), אולם הילוך שיכור אינו מקיים שוויון כזה.

למעשה, הפילוג בעתיד של תהליך מרקובי תלוי רק במידע האחרון שיש לנו עליו: נראה כי

$$\mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n, X(0) = x_0\} = \mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n\} \quad (4.4)$$

כלומר התלות היא במידע האחרון בלבד. ואכן

$$(4.5) \quad \mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n, X(0) = x_0\} = \frac{\mathbb{P}\{X(n+1) = x_{n+1}, X(n) = x_n, X(0) = x_0\}}{\mathbb{P}\{X(n) = x_n, X(0) = x_0\}}$$

$$(4.6) \quad \mathbb{P}\{X(n+1) = x_{n+1}, X(n) = x_n, X(0) = x_0\}$$

$$(4.7) \quad = \sum_{x_{n-1}} \cdots \sum_{x_1} \mathbb{P}\{X(n+1) = x_{n+1}, X(n) = x_n, X(n-1) = x_{n-1}, \dots, X(1) = x_1, X(0) = x_0\}$$

$$(4.8) \quad = \sum_{x_{n-1}} \cdots \sum_{x_1} \mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n, \dots, X(0) = x_0\} \mathbb{P}\{X(n) = x_n, \dots, X(0) = x_0\}$$

$$(4.9) \quad = \sum_{x_{n-1}} \cdots \sum_{x_1} \mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n\} \mathbb{P}\{X(n) = x_n, \dots, X(0) = x_0\}$$

$$(4.10) \quad = \mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n\} \mathbb{P}\{X(n) = x_n, X(0) = x_0\}$$

כאשר השורה לפני האחרונה מתקבלת בגלל תכונת המרקוביות. נציב ב-(4.5) ונקבל

$$\mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n, X(0) = x_0\} = \mathbb{P}\{X(n+1) = x_{n+1} \mid X(n) = x_n\} \quad (4.11)$$

באותה דרך של שילוב משפט בייס והתכונה המרקובית, ניתן להוכיח גם את התוצאות הבאות:

תרגיל 4.4 עבור שרשרת מרקוב הראה כי

$$(4.12) \quad \mathbb{P}\{X(n+m) = x_{n+m}, \dots, X(n+1) = x_{n+1} \mid X(n) = x_n, \dots, X(0) = x_0\} \\ = \mathbb{P}\{X(n+m) = x_{n+m}, \dots, X(n+1) = x_{n+1} \mid X(n) = x_n\} .$$

טענה 4.5 ללא הוכחה: עבור שרשרת מרקוב, לכל סדרת זמנים (שלמים) $n_1 < n_2 < \dots < n_k < n_{k-1}$ מתקיים

$$\mathbb{P}\{X(n_k) = x_k \mid X(n_{k-1}) = x_{k-1}, \dots, X(n_1) = x_1\} = \mathbb{P}\{X(n_k) = x_k \mid X(n_{k-1}) = x_{k-1}\} . \quad (4.13)$$

מסקנה: בהנתן שרשרת מרקוב $\{X_n\}$ וסדרת זמנים דטרמיניסטית $n_1 < n_2 < \dots$ גם X_{n_k} הינה שרשרת מרקוב. כלומר: תהליך מרקובי שעבר "דילול" נותר מרקובי.

בשרשרת מרקוב, בהנתן ההווה, העתיד והעבר הם בת"ס. הגדרה זו סימטרית ביחס לעבר ולהווה—ואכן התכונה המרקובית מתקיימת גם אם הופכים את ציר הזמן. נוכיח זאת³ ונראה שקיימת סימטריה עקרונית בין עבר ועתיד בשרשראות מרקוב. נסמן את המאורעות הבאים

$$A_n = \{X_0 = x_0, X_1 = x_1, \dots, X_{n-1} = x_{n-1}\}; \quad B_n = \{X_n = x_n\}$$

³גרסה ב' מכילה הוכחה מעט שונה למשפט זה

$$C_n = \{X_{n+1} = x_{n+1}, X_{n+2} = x_{n+2}, \dots, X_{n+k} = x_{n+k}\}$$

המסמנים עבר, הווה ועתיד בהתאמה. התכונה המרקובית:

$$\mathbb{P}\{C_n | A_n, B_n\} = \mathbb{P}\{C_n | A_n\}$$

נשים לב להצגה הבאה:

$$\mathbb{P}\{A_n, B_n, C_n\} \stackrel{(1)}{=} \mathbb{P}\{B_n\} \mathbb{P}\{A_n | B_n\} \mathbb{P}\{C_n | A_n, B_n\} \stackrel{(2)}{=} \mathbb{P}\{B_n\} \mathbb{P}\{A_n | B_n\} \mathbb{P}\{C_n | B_n\}$$

כאשר (1) נובע מחוק ביס ו-(2) נובע ממרקוביות. מכאן:

$$\mathbb{P}\{A_n, C_n | B_n\} = \frac{\mathbb{P}\{A_n, C_n, B_n\}}{\mathbb{P}\{B_n\}} = \mathbb{P}\{A_n | B_n\} \mathbb{P}\{C_n | B_n\}$$

נשים לב למסקנה:

$$\boxed{\mathbb{P}\{A_n, C_n | B_n\} = \mathbb{P}\{A_n | B_n\} \mathbb{P}\{C_n | B_n\}}$$

כלומר: בהנתן ההווה, B_n , העבר והעתיד בת"ס זה בזה. העבר והעתיד משחקים כאן תפקידים "שווי זכויות", ותכונת המרקוביות נשמרת גם עם היפוך ציר הזמן.

תרגיל 4.6 יהי $\{Z(n)\}$ תהליך i.i.d המקבל ערכים $\{-1, -2, \dots, -K\}$ בהסתברות $\mathbb{P}\{Z(n) = -j\} = p_j$. תהי $g = g(k, l)$ פונקציה המקבלת ערכים $\{1, 2, \dots, K\}$. הראה כי האלגוריתם הרקורסיבי

$$Y(n+1) = g(Y(n), Z(n))$$

(דוגמה 3.13) מייצר תהליך מרקובי $\{Y(n), n \geq 0\}$. חשב את הסתברויות המעבר.

המבנה שקיבלנו בתרגיל האחרון הוא כללי למדי, ובעזרתו ניתן לבנות מודלים לתופעות ותהליכים רבים כולל רשתות תקשורת ומחשבים, אותות דיבור, התנהגות הבורסה, מערכות שרות, תהליכי ייצור ועוד. מסתבר (אם כי לא נראה זאת כאן) שהמודל של אלגוריתם רקורסיבי עם תהליך i.i.d שקול למודל של תהליך מרקובי (שרשרת מרקובית אך עם ערכים לאו דוקא שלמים), במובן שכל אלגוריתם מתאר תהליך מרקובי, וכל תהליך מרקובי אפשר לתאר על ידי אלגוריתם רקורסיבי (לעיתים בעל תלות מפורשת בזמן), כאשר הרעש הוא לבן.

הסיבה לחשיבות של שרשרות מרקוב היא כפולה. מצד אחד כפי שראינו, מודל זה עשיר מספיק כדי לכסות מגוון רחב של תהליכים ומערכות. מצד שני כפי שנראה לשרשרת מרקוב יש מבנה מוגדר מספיק כדי לאפשר חישובים מפורשים, בצורה פשוטה למדי (לפחות מבחינה נומרית).

4.2 שרשראות מרקוב הומוגניות

נתרכז מעתה במקרה בו מספר המצבים סופי וכן הסתברויות המעבר אינן תלויות בזמן. להגבלה האחרונה מספר סיבות: כפי שנראה מייד, החישובים נעשים פשוטים בהרבה, ומצד שני רוב המודלים המעניינים אכן מקיימים תכונה זו. בנוסף, ניתן לייצג שרשרת מרקובית כזו על ידי רשת-או דיאגרמה. ייצוג זה תורם רבות להבנת התהליך.

ניראה כי חישובים של פילוגים במקרה זה אפשר לבצע על ידי פעולות אלגבריות פשוטות.

הגדרה 4.7 שרשרת מרקוב נקראת שרשרת הומוגנית אם לכל n

$$\mathbb{P}\{X(n+1) = j \mid X(n) = i\} = \mathbb{P}\{X(n) = j \mid X(n-1) = i\} \quad \forall i, j \in S = \{1, 2, \dots, N\}$$

כלומר אם הסתברויות המעבר אינן תלויות בזמן.

במקרה זה נשתמש באחד מהסימונים המקובלים הבאים:

$$p_{ij} \triangleq \mathbb{P}\{X(n+1) = j \mid X(n) = i\}$$

ונכנה אותן "הסתברות המעבר מ- i ל- j ". מטריצת הסתברויות המעבר P מוגדרת כ-

$$P = \{p_{ij}\}_{i,j=1}^N = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1N} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2N} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{N1} & p_{N2} & p_{N3} & \cdots & p_{NN} \end{bmatrix}$$

כאשר מספר המצבים כאן הוא N ובחרנו לקרוא להם $\{1, 2, \dots, N\}$. כלומר $S = \{1, 2, \dots, N\}$. זוהי מטריצה בה כל שורה מייצגת את המצב הנוכחי, והעמודה את המצב אליו עוברים. המינוח "שרשרת הומוגנית" (homogeneous Markov chain) אינו לגמרי סטנדרטי. יש הקוראים לשרשרת כזו "שרשרת סטציונרית" ("stationary Markov chain"). מינוח זה מטעה, שכן כפי שנראה בקרוב, שרשרת הומוגנית אינה בהכרח תהליך אקראי סטציונרי! מסיבה זו אנו נקפיד להשתמש במינוח "שרשרת הומוגנית".

כפי שנראה בפרק 4.3, חוק ההסתברות של שרשרת מרקובית נקבע באופן חד משמעי על ידי הפילוג ההתחלתי ואוסף הסתברויות המעבר, או בצורה שקולה על ידי וקטור הפילוג ההתחלתי ומטריצת הסתברויות המעבר. ייצוג נוסף הוא דרך דיאגרמת המעברים.

דוגמה 4.8 בהמשך לדוגמה 4.1, נניח שבתור יש מספר סופי $N-1$ של מקומות. במקרה זה, אם מגיע משתמש לתור כאשר התור מלא, הוא נחסם ולכן אינו מצטרף לתור. את אורך התור נוכל לייצג בעזרת המשוואה

$$Y(n+1) = Y(n) + A(n)I[Y(n) < N-1] - D(n)I[Y(n) > 0] . \quad (4.14)$$

לכן, הסתברויות המעבר הן:

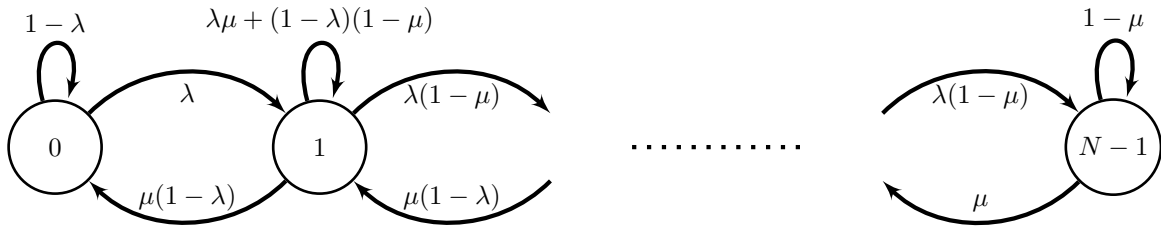
$$p_{ij} = 0 \quad \text{אם } |i - j| > 1$$

$$p_{i(i+1)} = \begin{cases} \lambda & \text{אם } i = 0 \\ \lambda(1 - \mu) & \text{אם } 0 < i < N - 1 \\ 0 & \text{אם } i = N - 1 \end{cases}$$

$$p_{ii} = \begin{cases} 1 - \lambda & \text{אם } i = 0 \\ 1 - \mu & \text{אם } i = N - 1 \\ (1 - \lambda)(1 - \mu) + \lambda \cdot \mu & \text{אחרת} \end{cases}$$

$$p_{i(i-1)} = \begin{cases} 0 & \text{אם } i = 0 \\ \mu(1 - \lambda) & \text{אם } 0 < i < N - 1 \\ \mu & \text{אם } i = N - 1 \end{cases}$$

(בהסתברויות אילו טמונה הנחה לגבי הסדר שבין הגעות ועזיבות. מהי?) ניתן לייצג הסתברויות אילו בצורה של דיאגרמת מעברים.



איור 4.2: שרשרת מרקוב: תור סופי

בדוגמה זו, מצבי השרשרת הם $\{0, 1, \dots, N - 1\}$. כל מצב מיוצג על ידי עיגול, שבתוכו שם המצב. כל חץ מתאר מעבר אפשרי בין מצבים, ולידו מספר, המייצג את הסתברות המעבר.

בצורה דומה ניתן לצייר דיאגרמת מעברים לכל שרשרת מרקוב סופית והומוגנית. בדיאגרמה כזו מספר הצמתים הוא כמספר המצבים של השרשרת. מספר הקשתות הוא כמספר המעברים שהסתברותם חיובית ממש.

טענה 4.9 הסתברויות המעבר של שרשרת מרקוב הומוגנית מקיימות את התנאים הבאים:

$$0 \leq p_{ij} \leq 1$$

$$\sum_{j \in S} p_{ij} = 1 \quad \forall i \in S.$$

כלומר סכום איברי כל שורה במטריצת המעברים הוא 1.

התנאי הראשון נובע מהגדרת הסתברות מותנית. התנאי השני טוען כי בהסתברות 1, לאחר צעד אחד נמצא את עצמנו במצב כלשהו. בסימון מטריצי אפשר לכתוב את התנאי השני כ-

$$P \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

כלומר, ווקטור העמודה $(1, 1, \dots, 1)^T$ הוא תמיד ווקטור עצמי (ימני) של מטריצת הסתברויות המעבר P , עם ערך עצמי 1.

4.3 חישוב הפילוג והסתברויות המעבר

מהו הסיכוי ששרשרת מרקובית תמצא במצב j לאחר n צעדים? ומהו הסיכוי ששרשרת מרקובית תעבור ממצב i למצב j ב- n צעדים? הבה נראה כי זאת ניתן לחשב בקלות. לשם הנוחיות, נגדיר סימון לפונקציית ההסתברות ולהסתברות המעבר במספר צעדים.

הגדרה 4.10 תהי $\{X(n), n = 0, 1, \dots\}$ שרשרת מרקוב, ונסמן את מצבי השרשרת ב- $S = \{1, 2, \dots, N-1\}$. נסמן $\nu_k(n) \doteq \mathbb{P}\{X(n) = k\}$ כעת נגדיר וקטור שורה

$$\begin{aligned} \underline{\nu}(n) &= (\nu_1(n), \nu_2(n), \dots, \nu_K(n)) \\ &= (\mathbb{P}\{X(n) = 1\}, \mathbb{P}\{X(n) = 2\}, \dots, \mathbb{P}\{X(n) = K\}) \end{aligned}$$

נסמן את הסתברות המעבר ממצב i למצב j ב- n צעדים ב-

$$p_{ij}^{(n)} = \mathbb{P}\{X(n) = j \mid X(0) = i\}$$

טענה 4.11 חישוב הפילוג והסתברויות המעבר של שרשרת הומוגנית וסופית.

ההסתברות למעבר ב- n צעדים מקיימת את המשוואה הרקורסיבית

$$p_{ij}^{(n)} = \sum_k p_{ik}^{(n-1)} p_{kj}, \quad p_{ij}^{(1)} = p_{ij} \xrightarrow{\text{כתיב מטריצי}} \begin{cases} P^{(n)} = P^{n-1} \cdot P \\ P^{(1)} = P \end{cases} \Rightarrow P^{(n)} = P^n$$

מסקנה 1: המטריצה של הסתברויות המעבר ב- n צעדים היא $\{p_{ij}^{(n)}\}_{ij} = P^n$, כלומר החזקה ה- n של מטריצת המעברים. קיבלנו לכן את נוסחת צ'פמן-קולמוגורוב (Chapman-Kolmogorov):

$$\begin{aligned} p_{ij}^{(n+m)} &= \sum_k p_{ik}^{(n)} \cdot p_{kj}^{(m)} \\ P^{(n+m)} &= P^n \cdot P^m \end{aligned}$$

כלומר, ההסתברות לעבור ממצב i בזמן 0 למצב j בזמן $n+m$ הוא סכום ההסתברויות של כל המסלולים העוברים דרך כל מצבי הביניים, k , בזמן n .

מסקנה 2: את פונקציית ההסתברות של התהליך ברגע n ניתן לחשב על ידי

$$\begin{aligned}\nu_j(n) &= \sum_k \mathbb{P}\{X(0) = k\} \cdot \mathbb{P}\{X(n) = j \mid X(0) = k\} \\ &= \sum_k \nu_k(0) \cdot p_{kj}^{(n)} \\ \underline{\nu}(n) &= \underline{\nu}(0) \cdot P^n\end{aligned}$$

כאשר הביטוי האחרון הוא בייצוג ווקטורי.

הוכחת הטענה⁴: נרשום מנוסחת ההסתברות השלמה

$$p_{ij}^{(n)} = \mathbb{P}\{X(n) = j \mid X(0) = i\} = \sum_k \mathbb{P}\{X(n) = j, X(n-1) = k \mid X(0) = i\}$$

כפי שראינו, מתכונת המרקוביות נקבל:

$$\begin{aligned}&= \sum_k \mathbb{P}\{X(n) = j \mid X(n-1) = k\} \cdot \mathbb{P}\{X(n-1) = k \mid X(0) = i\} \\ &= \sum_k p_{ik}^{(n-1)} p_{kj}\end{aligned}$$

ובכך סיימנו להוכיח את טענה 1. כיוון שלפי ההגדרה

$$P = \{p_{ij}\}_{ij}$$

נובע מסעיף 1 כי מטריצת המעבר בשני צעדים היא P^2 , ובאינדוקציה נקבל שהטענה נכונה לכל n . כיוון שכך, משוואת צ'פמן-קולמוגורוב נובעת מיידית מתכונות (כפל) של מטריצות. כדי להוכיח את סעיף 3 נשתמש בהגדרת הסתברות מותנית ונרשום

$$\begin{aligned}\nu_j(n) &= \mathbb{P}\{X(n) = j\} \\ &= \sum_k \mathbb{P}\{X(n) = j \mid X(0) = k\} \cdot \mathbb{P}\{X(0) = k\} \\ &= \sum_k \nu_k(0) \cdot p_{kj}^{(n)}\end{aligned}$$

או בסימון ווקטורי

$$\underline{\nu}(n) = \underline{\nu}(0) P^n$$

ובכך סיימנו את ההוכחה.

שים לב כי אם $\underline{\nu}(0)$ הוא ו"ע שמאלי של P אזי $\underline{\nu}(0) P = \underline{\nu}(0)$, וכנ"ל גם $\underline{\nu}(2)$ וכ'ו. כלומר, $\underline{\nu}(n)$ לא תלוי ב- n , ואז השרשרת המרקובית ההומוגנית היא סטציונרית.

⁴גרסה ב' מכילה הוכחה מעט שונה למשפט זה

אם כך, החישוב של פונקציית ההסתברות מסתכם בכפל מטריצות. כדי לדעת מהו הפילוג של התהליך, עלינו להיות מסוגלים לחשב את כל הפילוגים הרב-מימדיים. האם גם זאת ניתן לעשות בשיטות אלגבריות (כפל מטריצות)? נבדוק לדוגמה:

$$\begin{aligned}
 & \mathbb{P}\{X(5) = i, X(100) = j\} \\
 &= \sum_k \mathbb{P}\{X(100) = j, X(5) = i, X(0) = k\} \\
 &= \sum_k \mathbb{P}\{X(100) = j \mid X(5) = i, X(0) = k\} \cdot \mathbb{P}\{X(5) = i, X(0) = k\} \\
 &= \sum_k \mathbb{P}\{X(100) = j \mid X(5) = i\} \cdot \mathbb{P}\{X(5) = i \mid X(0) = k\} \cdot \mathbb{P}\{X(0) = k\} \\
 &= p_{ij}^{(95)} \cdot \sum_k p_{ki}^{(5)} \nu_k(0) = (P^{95})_{ij} \sum_k (P^5)_{ki} \nu_k(0) = (P^{95})_{ij} (\underline{\nu}(0) \cdot P^5)_i
 \end{aligned}$$

בצורה דומה ניתן לחשב כל פילוג רב ממדי. מכאן נובעת המסקנה החשובה הבאה:

טענה 4.12 חוק ההסתברות של שרשרת מרקובית הומוגנית $\{X(n), n \geq 0\}$ נקבע באופן חד משמעי על ידי ההסתברות בזמן 0, ומטריצת הסתברויות המעבר.

הוכחה: נבחר סדרת זמנים $\{n_1 < n_2 < \dots < n_k\}$ וסדרת מצבים $\{j_1, j_2, \dots, j_k\}$ כלשהם. נחשב את ההסתברות המשותפת:

$$\begin{aligned}
 & \mathbb{P}\{X(n_k) = j_k, X(n_{k-1}) = j_{k-1}, \dots, X(n_1) = j_1\} \\
 &= \mathbb{P}\{X(n_1) = j_1\} \cdot \left(\prod_{l=2}^k \mathbb{P}\{X(n_l) = j_l \mid X(n_1) = j_1, \dots, X(n_{l-1}) = j_{l-1}\} \right) \\
 &= \mathbb{P}\{X(n_1) = j_1\} \cdot \left(\prod_{l=2}^k \mathbb{P}\{X(n_l) = j_l \mid X(n_{l-1}) = j_{l-1}\} \right)
 \end{aligned}$$

בגלל המרקוביות. נמשיך ונקבל:

$$= (\underline{\nu}(0) P^{n_1})_{j_1} \cdot \left(\prod_{l=2}^k P_{j_{l-1} j_l}^{(n_l - n_{l-1})} \right)$$

כלומר, אפשר לחשב את הפילוג הרב-מימדי אם יודעים את הפילוג החד-מימדי (אותו ראינו שניתן לחשב מתוך הפילוג התחלתי והסתברויות המעבר), ואת הסתברויות המעבר.

משפט 4.13 יהי $\underline{\nu}$ פילוג שהוא וקטור עצמי שמאלי של מטריצת המעברים P עם ערך עצמי 1 כלומר $\underline{\nu} = \underline{\nu} P$. אזי השרשרת עם פילוג התחלתי $\underline{\nu}(0) = \underline{\nu}$ היא תהליך אקראי סטציונרי החל מזמן 0 (ראה הגדרה 3.17). כאשר ההנחה היא, כמובן, שכל איברי $\underline{\nu}$ אי שיליים ושהוא מנורמל להיות וקטור הסתברות.

הוכחה: ממשוואת צ'פמן קולמוגורוב נובע כי $\underline{p}(n) = \underline{p}(n-1)$. סטציונריות נובעת לכן מיידית מהוכחת משפט 4.12. במקרה זה $\underline{p}^n = \underline{p}$, ולכן:

$$\mathbb{P}\{X(n_k) = j_k, X(n_{k-1}) = j_{k-1}, \dots, X(n_1) = j_1\} = (\underline{p})_{j_1} \cdot \left(\prod_{l=2}^k P_{j_{l-1} j_l}^{(n_l - n_{l-1})} \right)$$

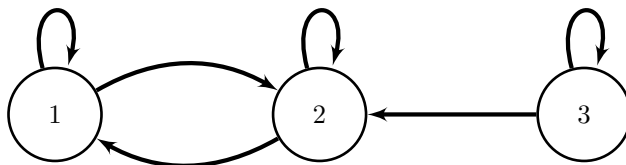
כאשר אגף שמאל תלוי רק בהפרש הזמנים. עוד נחזור לזה בהמשך עם הרחבה.

הערה: את החישובים שעשינו ניתן להרחיב למקרה של שרשרות שאינן סופיות. הקושי היחיד הוא בסימון המטריצי. אולם נחשוב על "מטריצות אינסופיות", אשר מקיימות את כללי החיבור והכפל הרגילים בין מטריצות. אזי כל הפיתוחים שעשינו תקפים, ולכן המסקנות נכונות גם עבור שרשרות לא סופיות.

4.4 מצבים נשנים וחולפים

על מנת להבין את הדינמיקה וההתנהגות של שרשראות מרקוב לטווח ארוך יש לעמוד על כמה מושגי יסוד ותכונות בסיסיות שלהן. ניתן לסווג את המצבים של שרשרת מרקוב לפי הקריטריון הבא: האם נחזור למצב נתון שוב ושוב, או שמא נבקר בו רק מספר סופי של פעמים?

דוגמה 4.14 בדוגמה שלפנינו לא רשומות הסתברויות המעבר, אולם כל קשת מתארת הסתברות חיובית.

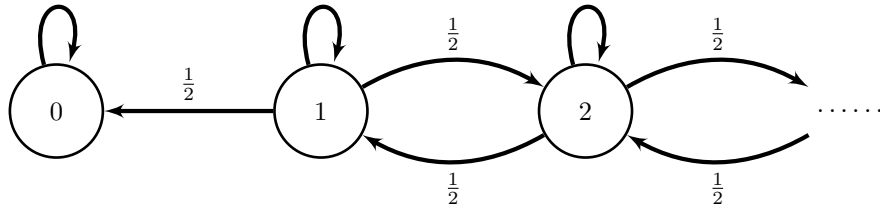


איור 4.3: מצבים חולפים ונשנים

אם נתחיל במצב 1 או 2, לעולם לא נגיע למצב 3. אם נתחיל במצב 3, אזי בכל צעד יש סיכוי לעבור למצב 2, וההסתברות להשאר במצב 3 צעדים n היא $p_{33}^n \rightarrow 0$, ולכן במוקדם או במאוחר נעזוב את מצב 3 לבלי שוב.

דוגמה 4.15 סופו של מהמר, או ה-Gambler's ruin. נניח שבכיסינו $X(0)$ שקלים, ואנו מהמרים בכל רגע על שקל אחד. בהסתברות $1/2$ נרוויח שקל, ובהסתברות זהה נפסיד שקל. יהי $\{X(n), n \geq 0\}$ סכום הכסף שבכיסינו לפני ההימור מספר n . זוהי שרשרת מרקובית, דומה לתור עם שרת יחיד או הילוך שיכור (דוגמה 4.1), אך עם הבדל קטן: כאשר יגמר הכסף שבכיסינו, יגמר הבילוי, ולא נוכל עוד להמר. כלומר, מצב "0" הוא מצב מיוחד: כאשר הגענו אליו, שם נשאר.

באשר לשאר המצבים, המצב אינו לגמרי ברור. האם יתכן שהמשחק ימשך לנצח? אם לא, פירוש הדבר כי מספר הפעמים שנהיה בכל מצב שאינו 0 הוא סופי (אקראי, כמובן).



איור 4.4: סופו של מהמר

נגדיר את ρ_{ji} להיות ההסתברות לעבור ממצב j בזמן אפס, למצב i בזמן כלשהו:

$$\rho_{ji} = \mathbb{P} \{ \exists n > 0 : X(n) = i, \text{ כלשהו } n > 0 \text{ עבור } | X(0) = j \}$$

$$= \sum_{n=1}^{\infty} \mathbb{P} \{ X(n) = i, X(n-1) \neq i, \dots, X(2) \neq i, X(1) \neq i | X(0) = j \}$$

המאורע עבורו אנו מחשבים הסתברות הוא שהביקור הראשון במצב i קרה בזמן n . השוויון מתקיים כיוון שהמאורעות הם זרים עבור ערכים שונים של פרמטר הזמן.

הגדרה 4.16 מצב i נקרא מצב נשנה (Recurrent) אם, כאשר השרשרת מתחילה במצב i , מובטח לנו שנשוב אליו לפחות פעם אחת בעתיד, כלומר אם $\rho_{ii} = 1$. מצב שאינו נשנה נקרא מצב חולף (Transient).

דוגמה 4.17 אם נתחיל במצב 1, לא נוכל לעבור למצב 2, ולהיפך. לפי ההגדרה, שני המצבים נשנים.



איור 4.5: מצבים נשנים

מצב 3 באיור 4.3 הוא מצב חולף, שכן יש הסתברות חיובית שנעזוב מצב זה בצעד ראשון - ובמקרה זה לא נחזור אליו לעולם.

הגדרה 4.18 תהליך מרקובי נקרא תהליך חולף אם כל מצביו חולפים.

תהליך המקיים את המשוואה

$$X(n+1) = X(n) + 1$$

הוא חולף, שכן אם נתחיל במצב כלשהו $X(0) = i$, לא נחזור אליו לעולם. כלומר, כל המצבים חולפים. ברור שתופעה כזו לא תיתכן אם מספר המצבים הוא סופי, שכן בכל רגע התהליך נמצא במצב כלשהו.

חזרה לדוגמא 4.14: מצב מס' 3 בדוגמא זו הוא חולף אם $p_{33} < 1$.

$$\rho_{33} = \underbrace{\mathbb{P}\{X(1) = 3 \mid X(0) = 3\}}_{p_{33} < 1} + \underbrace{\sum_{n \geq 2} \mathbb{P}\{X(n) = 3, X(n-1) \neq 3, \dots, X(1) \neq 3 \mid X(3) = 3\}}_0$$

לעומת זאת, מצב מס' 2 (ובאופן דומה גם מצב מס' 1) הינו נשנה:

$$\begin{aligned} \rho_{22} &= \mathbb{P}\{X(n) = 2, X(n-1) = X(n-2) = \dots = X(1) = 1 \mid X(0) = 2\} \\ &= p_{22} + \sum_{n \geq 2} \underbrace{(1 - p_{22})}_{p_{21}} p_{11}^{n-2} \underbrace{(1 - p_{11})}_{p_{12}} = p_{22} + (1 - p_{22}) \cdot \frac{1 - p_{11}}{1 - p_{11}} = 1 \end{aligned}$$

נשים לב כי לא ערכן המדויק של הסתברויות המעבר הוא הקובע האם המצב חולף או נשנה, אלא רק האם הן חיוביות או שוות לאפס.

לכאורה, לא הגבלנו את מספר הביקורים במצב חולף אם התחלנו במצב אחר. אולם בגלל המרקוביות, הדברים קשורים. קריטריון חשוב נוסף⁵ המבדיל בין מצבים נשנים לחולפים הינו התנהגות מספר הביקורים $N_i \triangleq \sum_{n=1}^{\infty} I\{X(n) = i\}$ במצב i :

טענה 4.19 מצב i הוא חולף אם ורק אם לכל j ,

$$\mathbb{E}[N_i \mid X(0) = j] < \infty \quad (4.15)$$

הוכחה: לצורך חישוב התוחלת נשתמש בקשר $\mathbb{E} N = \sum_{k=1}^{\infty} \mathbb{P}\{N \geq k\}$ (המתקיים למשתנה אקראי חיובי שערכיו שלמים - ראה טענה 8.21). באופן דומה, כמובן, $\mathbb{E}[N_i \mid X(0) = j] = \sum_{k=1}^{\infty} \mathbb{P}\{N_i \geq k \mid X(0) = j\}$. עבור $k = 1$, כלומר - לפחות לפחות ביקור אחד במצב i , נקבל

$$(4.16) \quad \mathbb{P}\{N_i \geq 1 \mid X(0) = j\} = \rho_{ji}$$

⁵ זהו כלי חשוב שישרת אותנו אח"כ בתיאור ההתנהגות של שרשראות.

על פי ההגדרה. עבור $k = 2$, כלומר - ביקור ב- i , נאמר בזמן m , ואחריו ביקור נוסף ב- i , נקבל:

$$\begin{aligned}
 & \mathbb{P}\{N_i \geq 2 \mid X(0) = j\} \\
 &= \sum_{m=1}^{\infty} \sum_{n=m+1}^{\infty} \mathbb{P}\{X(n) = i, X(n-1) \neq i, \dots, X(m+1) \neq i, X(m) = i, X(m-1) \neq i, \dots, X(1) \neq i \mid X(0) = j\} \\
 &= \sum_{m=1}^{\infty} \sum_{n=m+1}^{\infty} \mathbb{P}\{X(n) = i, X(n-1) \neq i, \dots, X(m+1) \neq i \mid X(m) = i, X(m-1) \neq i, \dots, X(1) \neq i, X(0) = j\} \\
 &\quad \times \mathbb{P}\{X(m) = i, X(m-1) \neq i, \dots, X(1) \neq i \mid X(0) = j\} \\
 &= \sum_{m=1}^{\infty} \sum_{n=m+1}^{\infty} \mathbb{P}\{X(n) = i, X(n-1) \neq i, \dots, X(m+1) \neq i \mid X(m) = i\} \\
 &\quad \times \mathbb{P}\{X(m) = i, X(m-1) \neq i, \dots, X(1) \neq i \mid X(0) = j\} \\
 &= \sum_{n=1}^{\infty} \underbrace{\mathbb{P}\{X(n) = i, X(n-1) \neq i, \dots, X(1) \neq i \mid X(0) = i\}}_{\text{הזזה של הזמנים ב-} m \text{ אחורה}} \\
 &\quad \times \sum_{m=1}^{\infty} \mathbb{P}\{X(m) = i, X(m-1) \neq i, \dots, X(1) \neq i \mid X(0) = j\} \\
 &= \rho_{ii} \rho_{ji}.
 \end{aligned}$$

כאשר במעבר הלפני אחרון השתמשנו בהומוגניות, ובאחרון בהגדרת ρ . באופן דומה, עבור k כללי, מקבלים:

$$\mathbb{P}\{N_i \geq k \mid X(0) = j\} = \rho_{ji} \cdot \rho_{ii}^{k-1}$$

האינטואיציה ברורה: צריך פעם אחת להגיע מ- j ל- i , מכאן נקבל את הגורם ρ_{ji} , ואח"כ "לעבור" מ- i לעצמו עוד $k-1$ פעמים. לכן:

$$(4.17) \quad \mathbb{E}[N_i \mid X(0) = j] = \sum_{k=1}^{\infty} \mathbb{P}\{N_i \geq k \mid X(0) = j\}$$

$$(4.18) \quad = \sum_{k=1}^{\infty} \rho_{ji} \cdot \rho_{ii}^{k-1} = \frac{\rho_{ji}}{1 - \rho_{ii}}$$

כעת, אם $\rho_{ii} < 1$ (מצב חולף), אזי ברור כי $\mathbb{E}[N_i \mid X(0) = j] < \infty$ לכל i . אם $\rho_{ii} = 1$ (מצב נשנה), אזי

$$\mathbb{E}[N_i \mid X(0) = j] = \infty$$

לפחות עבור $i = j$ (עבור j -ים אחרים יתכן $\rho_{ji} = 0$). ולהיפך, אם

$$\mathbb{E}[N_i \mid X(0) = i]$$

הוא סופי אזי $\rho_{ii} < 1$, ואם הוא אינסופי אזי $\rho_{ii} = 1$. בכך הוכחנו את הטענה.

הערה: מכאן אנו רואים כי מספיק להסתכל על $\mathbb{E}[N_i | X(0) = i]$ (כלומר, $j = i$ במשוואה (4.15)) על מנת לקבוע אם מצב הוא חולף או נשנה.

כעת נשתמש בתוצאה זו כדי להראות את התוצאה (האינטואיטיבית) הבאה.

טענה 4.20 בשרשרת מרקובית עם מספר מצבים סופי ($N < \infty$), לפחות מצב אחד הוא נשנה.

הוכחה: נניח שהתחלנו ממצב כלשהו, j . נשים לב כי מספר הביקורים במצב כלשהו עד זמן n (לא כולל זמן 0) הוא בדיוק n , שכן בכל רגע התהליך נמצא במצב כלשהוא. לכן

$$\sum_i N_i = \infty$$

מתכונות התוחלת נובע כי

$$\infty = \mathbb{E} \left[\sum_i N_i | X(0) = j \right] = \sum_i \mathbb{E} [N_i | X(0) = j]$$

כיוון שמספר המצבים הוא סופי, לא יתכן שיתקיים בו זמנית

$$\mathbb{E} \left[\sum_i N_i | X(0) = j \right] = \infty, \quad \mathbb{E} [N_i | X(0) = j] < \infty \quad \text{לכל } i$$

ולכן לפחות מצב אחד אינו חולף, כלומר - קיים לפחות מצב אחד המקיים $\mathbb{E}[N_i | X(0) = j] = \infty$, ועל כן הוא נשנה. במלים פשוטות, מאחר שיש אינסוף זמנים $n = 1, 2, 3, \dots$ ומספר סופי, N , של מצבים, לא יתכן שמספר הביקורים בכלם הוא סופי. לעומת זאת, כאשר מספר המצבים אינסופי ($N = \infty$), בהחלט יתכן שכל המצבים חולפים, ואז השרשרת נקראת חולפת. למשל: $X(n+1) = X(n) + 1$. טענה זו נותנת גם שיטה נוספת לבדוק אם מצב הוא חולף, על ידי חישוב תוחלת מספר הביקורים בו.

$$\mathbb{E} [N_i | X(0) = i] = \sum_{n=1}^{\infty} p_{ii}^{(n)} = \sum_{n=1}^{\infty} [P^n]_{ii} \quad \text{טענה 4.21}$$

ולכן תנאי שקול נוסף להבחנה בין מצב חולף ומצב נשנה הוא האם $\sum_{n=1}^{\infty} p_{ii}^{(n)}$ הוא סופי או אינסופי.

הערה: מכאן ניתן לקבל נוסחה עבור ρ_{ii} : חלצו את ρ_{ii} מתוך:

$$\frac{\rho_{ii}}{1 - \rho_{ii}} = \sum_{n \geq 1} [P^n]_{ii}.$$

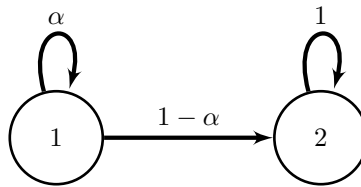
הוכחה: נתחיל במצב i ונזכר כי לפי הטענה הקודמת, מצב הוא נשנה אם ורק אם $\mathbb{E} [N_i | X(0) = i] = \infty$. נזכר בסימון

של פונקציה מציינת, ונחשב

$$\begin{aligned} \mathbb{E}[N_i | X(0) = i] &= \mathbb{E}\left[\sum_{n=1}^{\infty} I\{X(n) = i\} | X(0) = i\right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}[I\{X(n) = i\} | X(0) = i] \\ &= \sum_{n=1}^{\infty} \mathbb{P}\{X(n) = i | X(0) = i\} \\ &= \sum_{n=1}^{\infty} p_{ii}^{(n)} \end{aligned}$$

כדרוש.

דוגמה 4.22 נתבונן בשרשרת⁶:



$$P = \begin{bmatrix} \alpha & 1 - \alpha \\ 0 & 1 \end{bmatrix}$$

ברור כי:

$$P^n = \begin{bmatrix} \alpha^n & 1 - \alpha^n \\ 0 & 1 \end{bmatrix}$$

כעת,

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \sum_{n=1}^{\infty} [P^n]_{ii} = \begin{cases} \sum_{n=1}^{\infty} \alpha^n = \frac{\alpha}{1-\alpha} < \infty, & i = 1 \\ \sum_{n=1}^{\infty} 1 = \infty, & i = 2 \end{cases}$$

על כן, מצב מס' 1 הוא חולף ואילו מצב מס' 2 הוא נשנה.

הערה: באחד מתרגילי הבית אתם עשויים להתבקש לחשב את P^n עבור n כללי באמצעות משפט Cayley-Hamilton. החישוב של $\sum_{n=1}^{\infty} [P^n]_{ii}$ ממחיש בצורה ברורה מאד מדוע חשוב להיות מסוגלים לקבל ביטוי סגור עבור P^n לכל n .

4.5 קישוריות בין מצבים ופרוק מרחב המצב

ראינו שמה שקובע את הדינמיקה של השרשרת המרקובית לטווח ארוך זה מבנה המצבים החולפים והנשנים שלה, וזה נקבע על-פי החלוקה להסתברויות מעבר חיוביות או אפס.

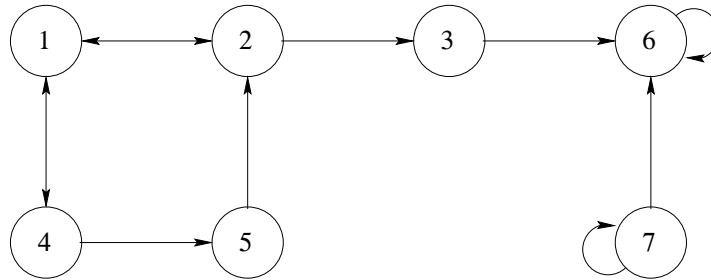
⁶גרסה ב' מכילה דוגמה שונה

הגדרה 4.23 אם $p_{ij}^{(n)} > 0$ עבור n כלשהוא או ש- $i = j$ (גם אם אין מעבר מ- i לעצמו) אזי נאמר ש- i מוביל ל- j , ונסמן $i \rightarrow j$. אם $i \rightarrow j$ וגם $j \rightarrow i$ נאמר ש- i ו- j מקושרים (communicating), ונסמן $i \leftrightarrow j$.

נשים לב כי לפי ההגדרה, כל מצב מקושר לעצמו גם אם הסתברות המעבר מהמצב לעצמו היא 0.

בסימון הקודם, קיים n כך ש- $p_{ij}^{(n)} > 0$ אם ורק אם $\rho_{ij} > 0$. מבחינת דיאגרמת המצבים, $i \rightarrow j$ אם (ורק אם) קיים מסלול בדיאגרמה המוביל מ- i ל- j . כתוצאה מכך, $i \leftrightarrow j$ אם ורק אם קיים בדיאגרמה מסלול סגור מ- i לעצמו, העובר דרך j .

דוגמה 4.24 בדוגמה זו (ראה איור 4.6), מצבים 1, 2 מקושרים, וגם מצבים 1, 4 מקושרים. פחות מידי לראות כי 1, 5 מקושרים: 1 מוביל ל-5 דרך $1 \rightarrow 4 \rightarrow 5$ וכו'. (אילו זוגות נוספים מקושרים?). מצב 2 מוביל למצב 3, ומצבים 3, 7 מובילים למצב 6 אולם מצבים 3, 6, 7 אינם מקושרים לשום מצב. ברור כי מצב 7 הוא מצב חולף, אך האם מצב 1 נשנה? בהמשך נראה כיצד להחליט על כך.



איור 4.6: מצבים קשורים

טענה 4.25 תכונות יחס הקשירות

1. $i \leftrightarrow i$ רפלקסיביות reflexive relation
2. $i \leftrightarrow j$ אם ורק אם $j \leftrightarrow i$ סימטריה symmetry
3. אם $i \leftrightarrow j$ וגם $j \leftrightarrow k$ אזי $i \leftrightarrow k$ טרנזיטיביות transitivity
4. אם $i \leftrightarrow j$ אזי שניהם נשנים או שניהם חולפים.

הוכחה: שלוש הטענות הראשונות נובעות מיידית מההגדרות (ראה דוגמה). נוכיח את 4. כיוון ש- $i \leftrightarrow j$ קיימים m, k כך ש-

$$p_{ij}^{(m)} > 0 \quad \text{וגם} \quad p_{ji}^{(k)} > 0$$

נניח ש- i נשנה ונבדוק את הקריטריון עבור j

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} \geq \sum_{n=1}^{\infty} p_{jj}^{(k+n+m)}$$

אי השוויון מתקיים שכן זהו אותו הסכום אלא שחסרים בו $k + m$ האיברים הראשונים. כעת, נעיין רק במסלולים המגיעים ל- i אחרי k צעדים, חוזרים ל- i אחרי עוד n צעדים, ושוב חוזרים ל- j אחרי m צעדים אחרונים:

$$\sum_{n=1}^{\infty} p_{jj}^{(k+n+m)} \geq \sum_{n=1}^{\infty} p_{ji}^{(k)} p_{ii}^{(n)} p_{ij}^{(m)} = p_{ji}^{(k)} \left(\sum_{n=1}^{\infty} p_{ii}^{(n)} \right) p_{ij}^{(m)}$$

אם i נשנה אזי ראינו כי

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$$

ולכן גם

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} = \infty$$

וקיבלנו כי j נשנה. מצד שני, אם j חולף אזי

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} < \infty$$

ואזי גם

$$\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$$

וגם i חולף. בכך הראינו את תכונה 4.

מסקנה מתכונות 3-1: יחס הקשירות הוא רפלקסיבי סימטרי וטרנזיטיבי ולכן הוא יחס שקילות. לכן הוא מחלק את מרחב המצב S ל"קבוצות שקילות". לפי תכונה 4, בכל קבוצה שקילות כל המצבים נשנים או שכולם חולפים.

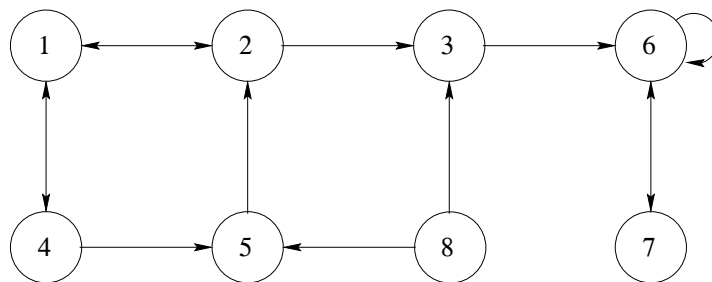
הגדרה 4.26 אוסף מצבים נקרא קבוצה סגורה אם ההסתברות לצאת מהקבוצה היא אפס. כלומר קבוצה היא סגורה אם מתקיים התנאי הבא. לכל מצב i בקבוצה ולכל מצב j שאינו בקבוצה, $p_{ij}^{(n)} = 0$ לכל n .

מהגדרה זו נובע כי בדיאגרמת המעברים, אין קשתות היוצאות מקבוצה סגורה. למשל בדוגמה של איור 4.6 מצבים 6, 7 מהווים קבוצה סגורה. בנוסף אוסף כל המצבים הוא (כמובן) קבוצה סגורה.

הגדרה 4.27 קבוצת הקשירות של מצב i היא קבוצת השקילות שלו ביחס לקשירות. כלומר זוהי קבוצת כל המצבים המקושרים ל- i . קבוצת קשירות A נקראת קבוצה סגורה אם כל מצב ב- A מוביל רק למצבים ב- A .

קבוצה נקראת נשנית אם כל מצביה נישנים.

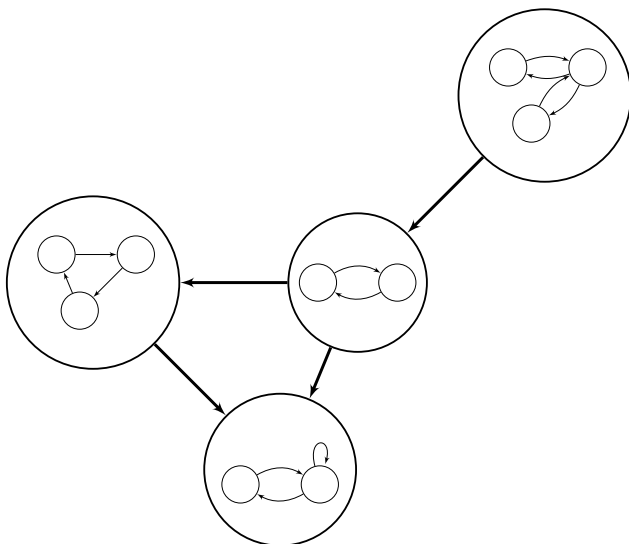
דוגמה 4.28 בדוגמה הבאה 8 מצבים. לכל מצב שנבחר נוכל להגדיר קבוצת קשירות. קבוצת הקשירות של מצב 7 מכילה מלבדו גם את מצב 6, ובגלל הסימטריה, קבוצת הקשירות של מצב 6 מכילה גם את מצב 7. לסיכום, איתרנו קבוצת קשירות: $\{6, 7\}$. קבוצה זו היא קבוצה סגורה, שכן המצבים $\{6, 7\}$ אינם מובילים לאף מצב מחוץ לקבוצה זו. לעומת זאת מצב 3 שייך לקבוצת הקשירות $\{3\}$, המכילה אותו בלבד: זאת כיוון שהוא אינו מקושר לאף מצב אחר. ניתוח זה תופס גם עבור מצב



איור 4.7: סיווג מצבים

8. לעומת זאת, קבוצת המצבים $\{1, 2, 4, 5\}$ היא אכן קבוצת קשירות, אך איננה קבוצה סגורה שכן מצב 2, השייך לקבוצה, מוביל למצב 3, שאינו שייך אליה.

באופן כללי, קבוצות קשירות הן מעין "סופר-מצבים" המחוברים ביניהם לכל היות במעברים חד-סטריים (אחרת, ניתן היה לאחדן), וללא לולאות. לקבוצות הסגורות יש רק "כניסות" ואין יציאות (איור 4.8).



איור 4.8: שרטוט סכמטי של שרשרת בעלת "סופר-מצבים".

מהו הקשר בין סגירות לנשנות? אם קבוצת קשירות היא סגורה, אזי היא נשנית, ואם היא אינה סגורה, אזי היא חולפת, כי אז יש מעבר (חד-סטרי) ממנה החוצה, ללא דרך חזרה.

טענה 4.29 (ללא הוכחה). כל שרשרת סופית נתנת לפירוק למספר סופי של קבוצות קשירות סגורות, וקבוצה נוספת (אולי ריקה) של מצבים חולפים השייכים לקבוצה שאינה סגורה. לפחות קבוצת קשירות אחת היא קבוצה סגורה. כל המצבים בכל קבוצת קשירות סגורה נשנים (וכאמור, כל המצבים שאינם בקבוצה סגורה חולפים).

בשרשרת עם מספר בן מניה של מצבים, מספר הקבוצות הסגורות לא בהכרח סופי, יתכן שאין אף קבוצת קשירות בעלת

יותר ממצב אחד, יתכן שאין אף קבוצה סגורה ויתכן שאין אף מצב נישנה, אפילו אם יש קבוצה סגורה.

הגדרה 4.30 שרשרת מרקובית נקראת שרשרת פריקה אם יש לה יותר מאשר קבוצה סגורה אחת. אחרת היא נקראת לא פריקה indecomposable. קבוצה נקראת irreducible אם כל המצבים מקושרים זה לזה, כלומר – כל המצבים יחד מהווים קבוצת קשירות סגורה אחת. מתקיים: $\text{irreducible} \rightarrow \text{indecomposable}$.

מסקנה: שרשרת מרקובית היא פריקה אם ורק אם יש בה לפחות שני מצבים נשנים שאינם מקושרים.

אם המצב ההתחלתי שייך לקבוצה סגורה, לעולם לא נצא מקבוצה זו. זה מסביר מדוע בכל קבוצה סגורה סופית יש מצב נשנה אחד לפחות, ואם הקבוצה היא קבוצת קשירות אזי כל מצביה נשנים. אולם יתכן שקיימת קבוצה סגורה נוספת, ואליה לא נגיע! זאת מכיוון שהמצב ההתחלתי הוא כזה. יתכן כמובן להתחיל במצב התחלתי מחוץ לקבוצה סגורה, ולהגיע אליה לאחר מספר צעדים אקראי. מצב התחלתי כזה הוא בהכרח חולף.

טענה 4.31 (ללא הוכחה). כל מצב מחוץ לקבוצה סגורה המוביל למצב בקבוצה, הוא מצב חולף.

אנו רואים, אם כך, שכדי לסווג מצבים לחולפים ונישנים, עלינו לנתח את דיאגרמת המעברים, אולם אין כל חשיבות לערכים המדוייקים של הסתברויות המעבר! תכונות אילו (מצב חולף או נישנה) נקבעות על ידי הקשתות בלבד (כאשר כמובן לא נצייר קשת כאשר הסתברות המעבר היא 0). כך למשל, בציר 4.7 רק מצבים 6, 7 הם נישנים, ושאר המצבים הם חולפים.

4.6 סטציונריות של שרשראות מרקוביות

הגדרה 4.32 פילוג \underline{p} נקרא פילוג סטציונרי או פילוג אינווריאנטי Stationary, invariant עבור שרשרת מרקובית אם $\underline{p}(0) = \underline{p}$ גורר $\underline{p}(n) = \underline{p}$ לכל $n > 0$. כלומר, אם נתחיל את השרשרת עם פילוג זה, הפילוג (החד-מימדי!) ישאר קבוע.

במבט ראשון, סוג הסטציונריות שקיבלנו הוא מוגבל-הוא נוגע רק לפילוג החד מימדי. אולם במקרה המרקובי ההומוגני, זה מספיק!

טענה 4.33 עבור שרשרת מרקוב הומוגנית,

1. הפילוג \underline{p} הוא פילוג סטציונרי אם ורק אם הוא מקיים $\underline{p} \cdot P = \underline{p}$, כלומר הוא ווקטור עצמי שמאלי של מטריצת המעברים, עם ערך עצמי 1.

2. אם $\underline{p}(T)$ הוא פילוג סטציונרי אזי השרשרת המרקובית היא תהליך סטציונרי החל מזמן T . אם בנוסף השרשרת מוגדרת לכל זמן $-\infty < n < \infty$ אזי השרשרת היא תהליך אקראי סטציונרי.

הוכחה: ראה משפט 4.13. נחזור על ההוכחה ביתר פירוט. $\underline{\nu} \cdot P = \underline{\nu}$ ו- $\underline{\nu}(0) = \underline{\nu}$ אזי

$$\begin{aligned}\underline{\nu}(n) &= \underline{\nu}(0) P^n \\ &= (\underline{\nu}(0) P) P^{n-1} \\ &= \underline{\nu}(0) P^{n-1} \\ &= (\underline{\nu}(0) P) P^{n-2} \\ &= \dots = \underline{\nu}(0) P = \underline{\nu}(0)\end{aligned}$$

ולכן לפי ההגדרה הפילוג הוא סטציונרי. מצד שני, אם הפילוג הוא סטציונרי אזי לפי ההגדרה

$$\begin{aligned}\underline{\nu} P &= \underline{\nu}(0) P \\ &= \underline{\nu}(1) = \underline{\nu}\end{aligned}$$

ובכך הוכחנו את טענה 1. כעת נזכר בהגדרת הסטציונריות 3.17 ונבדוק לפי ההגדרה. ראינו בהוכחת משפט 4.12 כי

$$\mathbb{P}\{X(n_j) = i_j, j = 1, \dots, k\} = \prod_{j=2}^k P_{i_{j-1} i_j}^{(n_j - n_{j-1})} \nu_{i_1}(n_1). \quad (4.19)$$

ביטוי זה אינו משתנה אם נזיז את כל הזמנים ב- τ : זאת כיוון ש- $n_j - n_{j-1} = (n_j + \tau) - (n_{j-1} + \tau)$, והראנו לעיל כי $\nu_{i_1}(n_1)$ כלל אינו תלוי בערך של n_1 כלומר

$$\mathbb{P}\{X(t_1 + \tau) = j_1\} = \nu_{j_1}(t_1 + \tau) = \nu_{j_1}(t_1) = \mathbb{P}\{X(t_1) = j_1\}$$

ולכן ההסתברות לעיל אינה תלויה ב- τ ובפרט ערכה שווה ב- T וב- T^{-1} . ולכן התהליך הוא סטציונרי החל מזמן T . הוכחת הטענה האחרונה מורכבת יותר ולא נכלול אותה כאן. היא נובעת מהעובדה כי על ידי הסתכלות על נקודות זמן בעבר הרחוק אפשר לראות כי בהכרח ההסתברות להיות במצב חולף היא אפס, ומהעובדה כי מנקודות זמן רחוקה בעבר ומכל פילוג, הפילוג החד ממדי מותכנס במהירות אקספוננציאלית לפילוג סטציונרי. להבנת התופעה פתור את התרגיל הבא.

נשים לב כי באופן כללי, אם השרשרת מוגדרת החל מזמן סופי $t_1 < T$ אזי לא ניתן להסיק סטציונריות גם לפני זמן T , כפי שממחישות הדוגמאות הבאות.

תרגיל 4.34 הראה כי השרשרות הבאות סטציונריות החל מזמן 0, אך אינן סטציונריות (כלומר אינן סטציונריות עבור $-\infty < n < \infty$).

1. שרשרת עם מצבים $\{1, 2\}$ עבורה כל (ארבע) הסתברויות המעבר שוות $1/2$ והפילוג ב- $n = -1$ הוא $\mathbb{P}\{X(-1) = 1\} = 1$. האם שרשרת זו יכולה להיות מוגדרת עבור $n < -1$?

2. שרשרת עם מצבים $\{1, 2\}$ עבורה הסתברויות המעבר הן $p_{12} = p_{22} = 1$ והפילוג ב- $n = -1$ הוא $\mathbb{P}\{X(-1) = 1\} = 1$. האם שרשרת זו יכולה להיות מוגדרת עבור $n < -1$? בהנתן $T > 0$ מצא שרשרת בעלת מספר מצבים סופי המוגדרת עבור $0 \leq n < \infty$ כך שהשרשרת סטציונרית החל מזמן T אך לא סטציונרית החל בזמנים לפני T .

הראה כי שרשרת הומוגנית סופית שאינה מחזורית המוגדרת עבור זמנים $-\infty < n < \infty$ היא בהכרח ת"א סטציונרי: זאת על ידי השלבים הבאים.

1. נסמן ב- T_r את קבוצת המצבים החולפים וב- N את מספר המצבים החולפים. הראה כי לכל מצב חולף i מתקיים

$$\mathbb{P}\{x(n) \in T_r \mid x(n-N) = i\} < 1. \quad (4.20)$$

הסק מכך כי עבור שרשרת המוגדרת על אוסף הזמנים האינסופי כאמור לעיל מתקיים, לכל n

$$\mathbb{P}\{x(n) \in T_r\} = 0. \quad (4.21)$$

2. כעת נתח כל קבוצה סגורה בנפרד. השתמש בעובדה כי הפילוג החד-ממדי של שרשרת סופית שכל מצביה מקושרים מתכנס (במהירות גאומטרית, ללא תלות בפילוג ההתחלתי) לפילוג הסטציונרי (שהוא יחיד במקרה זה). הסק כי הפילוג החד ממדי הוא בהכרח הפילוג הסטציונרי, בכל זמן. מכך הסק כי השרשרת היא ת"א סטציונרי.

מה קורה אם השרשרת מחזורית?

מתי קיים פילוג סטציונרי? ומתי הוא יחיד? ברור כי לשרשרת חולפת אין פילוג סטציונרי. להילוך שיכור סימטרי (כל צעד הוא ± 1 בהסתברות $1/2$ כל אחד) אין פילוג סטציונרי, כיוון שהפתרון החיובי היחיד (עד כדי קבוע) של המשוואה (האיך-סופית!)

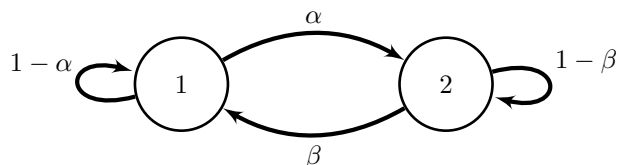
$$\underline{v} \cdot P = \underline{v}$$

הוא הווקטור $\underline{v} = (\dots, 1, 1, 1, \dots)$ אשר אינו פילוג. לעומת זאת, לתור שהגדרנו בתחילת הפרק (דוגמא 4.8) יש פילוג סטציונרי.

טענה 4.35 (ללא הוכחה). לכל שרשרת מרקוב סופית יש פילוג סטציונרי. השרשרת אינה פריקה אם ורק אם הפילוג הוא יחיד. הפילוג הסטציונרי מקבל ערכים חיוביים ממש רק על מצבים נשנים.

נשים לב כי שרשרת סופית לא פריקה, גם אם היא אינה irreducible עדיין יש לה פילוג סטציונרי יחיד. חוסר יחידות נגרם על ידי קיום שתי קבוצות סגורות (או יותר).

דוגמה 4.36 שרשרת כללית בעלת שני מצבים:



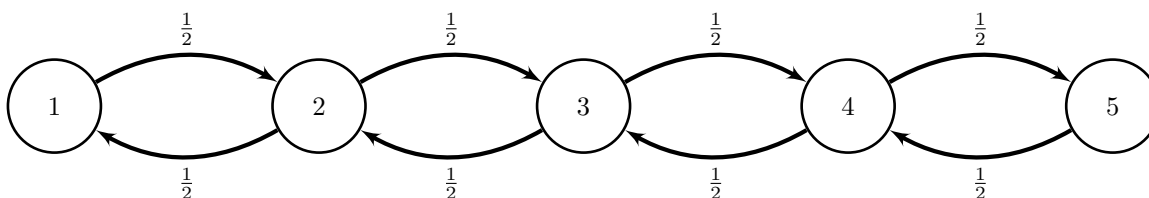
$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{pmatrix} \quad \underline{v} = \left(\frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta} \right)$$

הערה: קיום פילוג סטציונרי נובע מתוצאות באלגברה לינארית: למשוואה

$$P \underline{x} = \underline{x}$$

יש פיתרון (הווקטור $(\underline{x} = (\dots, 1, 1, 1, \dots))^T$), כלומר למטריצה P יש וקטור עצמי (ימני) עם ערך עצמי 1. מכך נובע שלמטריצה זו יש גם וקטור עצמי שמאלי עם ערך עצמי 1. משפט פרוך-פרובניוס מתורת המטריצות מבטיח כי, כיוון שאברי המטריצה P הם חיוביים, אזי לווקטור העצמי השמאלי (השייך לערך העצמי הגדול ביותר של P , שהוא 1) רכיבים חיוביים. כיוון שמדובר בווקטור סופי, ניתן לנרמל אותו כך שיהיה פילוג, ומטענה 4.33 זהו פילוג סטציונרי.

דוגמה 4.37 בדוגמה שלפנינו יש שתי קבוצות סגורות, ולכן הדוגמה אינה מקיימת את תכונת האי־פריקות. נניח שכל המעברים קורים בהסתברות $1/2$.



איור 4.9: פילוג סטציונרי

אזי קל לבדוק כי לכל $0 \leq \alpha \leq 1$, הפילוג

$$\underline{v} = (\alpha/2, \alpha/2, 0, (1 - \alpha)/2, (1 - \alpha)/2) \quad (4.22)$$

הוא פילוג סטציונרי. כלומר, הפילוג הסטציונרי אינו יחיד. באופן אינטואיטיבי, ניתן לחלק את ההסתברות הסטציונרית בין הקבוצות הסגורות כאוות נפשנו. החלוקה הפנימית של ההסתברות בתוך הקבוצה הסגורה נקבעת על ידי הסתברויות המעבר בתוך הקבוצה.

בדוגמה לעיל, אם נתחיל במצב 3 אזי בהסתברות $1/2$ נעבור בצעד הבא למצב 2 ואז לעולם לא נעבור לאחד מהמצבים 6, 7. הפילוג הסטציונרי הוא במקרה זה $1/4$ להיות בכל אחד מהמצבים 1, 2, 6, 7. אולם הסיכוי לעבור מ־1 ל־2 הוא 1, בעוד שהסיכוי לעבור מ־1 ל־6 הוא 0. עבור ω מסויים נמצא בסופו של דבר או רק במצבים 1, 2 או רק במצבים 6, 7.

כשם שחשוב להבהיר את השימוש במושג "סטציונריות" ביחס לשרשרת מרקובית, חשוב להדגיש גם את המשמעות המיוחדת של המושג "ארגודיות" בקשר זה. הנושא אינו חלק מחומר הקורס ולכן נסתפק בהגדרה והצגת התוצאה.

הגדרה 4.38 מצב i נקרא מחזורי אם המחלק המשותף הגדול ביותר (שנסמן ב־ d) של אוסף המספרים

$$\{n : p_{ii}^{(n)} > 0\} \quad (4.23)$$

הוא גדול ממש מ־1, ואז d נקרא המחזור של המצב. קבוצת קשירות נקראת מחזורית עם מחזור d אם כל מצביה מחזוריים עם מחזור d . שרשרת מרקובית לא־פריקה (indecomposable) שאינה מחזורית נקראת שרשרת ארגודית.

ניתן להראות כי אורך המחזור הוא תכונה של קבוצת הקשירות, כלומר לכל המצבים בקבוצת קשירות נתונה יש מחזור זהה (או שכולם אינם מחזוריים).

דוגמה 4.39 עבור שרשרת עם מצבים $\{1, 2, 3, 4, 5\}$, אם הסתברויות המעבר הן

$$P_{i(i+1)} = 1, \quad i = 1, 2, 3, 4, \quad P_{51} = 1$$

אזי לשרשרת מחזור $d = 5$. לעומת זאת אם הסתברויות המעבר הן

$$P_{12} = P_{23} = P_{34} = P_{41} = 1, \quad P_{55} = 1$$

אזי מצבים 1, 2, 3, 4 מחזוריים עם מחזור $d = 4$ ומצב 5 אינו מחזורי. לבסוף, אם

$$P_{12} = 1, \quad P_{21} = 1/2, \quad P_{23} = 1/2, \quad P_{34} = P_{42} = P_{55} = 1$$

אזי שום מצב אינו מחזורי.

שם לב כי שרשרת ארגודית אינה בהכרח תהליך ארגודי. אולם ניתן להראות כי אם הפילוג של שרשרת ארגודית שווה, ברגע כלשהוא, לפילוג הסטציונרי אזי השרשרת היא לא רק תהליך סטציונרי אלא גם תהליך ארגודי. יותר מכך, בשרשרת ארגודית, לכל פילוג התחלתי, פילוג המצב (כלומר הפילוג של $X(n)$) מתכנס, כאשר $n \rightarrow \infty$, לפילוג האינוריאנטי. קצב ההתכנסות הוא גאומטרי, כלומר לאחר t צעדים המרחק בין הפילוג לבין הפילוג האינוריאנטי הוא לכל היותר $C\beta^t$ כאשר C הוא קבוע, ו- β הוא (הערך המוחלט של) הערך העצמי השני בגדלו של מטריצת המעברים.

5 תהליכים אקראיים בזמן רציף

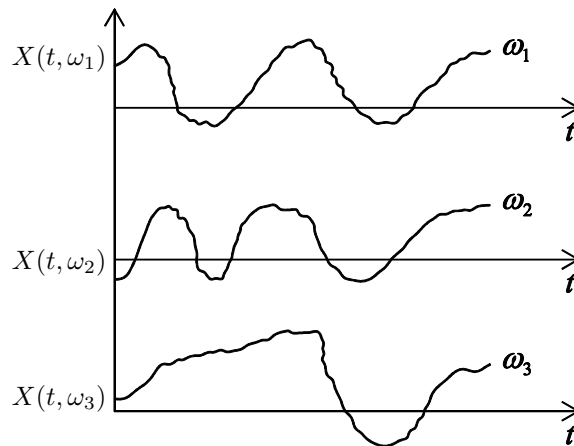
5.1 מבוא, הגדרות ודוגמאות

בפרק 3 עסקנו בתהליכים אקראיים בזמן בדיד, כלומר: סדרת משתנים אקראיים עם אינדקס זמן שהוא מספר שלם. אפשר להרחיב הגדרה זו ולעסוק באינדקס זמן שהוא קבוצת זמנים כלשהי. בפרק זה נעסוק באינדקס זמן רציף, כלומר: התהליך יוגדר עבור אינטרוול זמן $T_1 \leq t \leq T_2$, אשר יכול להיות סופי (כלומר T_1, T_2 הם מספרים), חצי אינסופי (כלומר $T_1 = -\infty$ או $T_2 = \infty$ אך לא שניהם) או אינסופי $T_1 = -\infty$ וכן $T_2 = \infty$.

הגדרה 5.1 תהליך אקראי בקטע $[T_1, T_2]$ הינו פונקציה אקראית של הזמן, $X(t, \omega)$, המוגדרת בתחום $T_1 \leq t \leq T_2$.

גם כאן נשתמש לפי הנוחיות בסימונים (השקולים) $X(t, \omega)$, $X_t(\omega)$ ולרוב נשמיט את פרמטר הזמן ונרשום את התהליך כ- $X(t)$ או X_t .

כפונקציה של פרמטר הזמן, $X(t, \omega)$ הוא מ"א, לדוגמא, $X(2, \omega)$ הוא מ"א. עבור ω קבוע, בקטע $T_1 \leq t \leq T_2$, $X(t, \omega)$ הוא פונקציה של t , הנקראת פונקצית מדגם (ראה ציור 5.1).



איור 5.1: שלוש פונקציות מדגם, בערכים שונים של ω

להלן מספר דוגמאות פשוטות של תהליכים אקראיים בזמן רציף, שהזכרנו עוד בפגישה הראשונה:

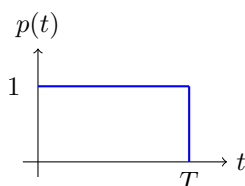
דוגמא 1: $X(t, \omega) = A(\omega)t + B(\omega)$ כאשר A, B מ"א. צורות גל טיפוסיות (פונקציות מדגם) במקרה זה הן קוים ישרים.

דוגמא 2: גל נושא עם פאזה אקראית (התלויה למשל בזמן ההפעלה של משדר) ואמפליטודה אקראית (התלויה למשל בהנחתה עקב גורמים פיזיקליים) ניתן לתאור כ- $X(t, \omega) = A(\omega) \sin(2\pi ft + \phi(\omega))$ כאשר A, ϕ מ"א. צורות הגל הטפוסיות במקרה זה הן תנודות הרמוניות בעלות אמפליטודה A ופאזה ϕ .

דוגמא 3: N קבוע (או אקראי) $X(t, \omega) = \sum_{n=1}^N X_n(\omega) \sin nt$ כאשר X_n הם מ"א. צורות גל טפוסיות במקרה זה הן סכום משוקלל של תנודות הרמוניות. זהו מודל מקובל לאות דיבור.

כמובן שלא כל תהליך אקראי ניתן לתאור פשוט כזה. בפרט, לא כל תהליך אקראי תלוי במספר סופי בלבד של פרמטרים אקראיים. לדוגמה, מודל מקובל לאות לקשורת ספרתי מוגדר בצורה הבאה. יהי $p(t)$ גל מרובע כלומר

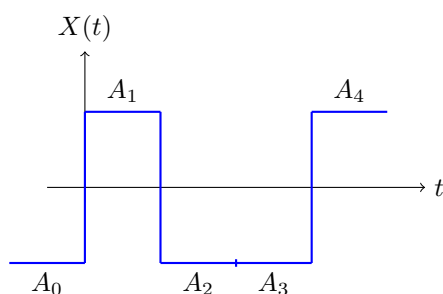
$$p(t) \doteq \begin{cases} 1 & 0 \leq t < T \\ 0 & \text{אחרת.} \end{cases} \quad (5.1)$$



איור 5.2: גל מרובע

אזי שידור של סידרה (אינסופית) של סיביות (משתנים אקראיים A_n המקבלים ערכים ± 1) המאפננים את הפולס המרובע נותנים את התהליך האקראי

$$X(t) = \sum_{n=-\infty}^{\infty} A_n p(t - nT). \quad (5.2)$$



איור 5.3: דוגמה לסדרת סיביות, A_n , המאופנת על ת"א $X(t)$.

חוק ההסתברות של תהליך אקראי:

יהיה $\{X_t, T_1 \leq t \leq T_2\}$ תהליך אקראי. נעיין באוסף חוקי ההסתברות כדלקמן: עבור כל n וכל $(T_1 \leq t_i \leq T_2)$ t_1, t_2, \dots, t_n

$$F_{X_{t_1}, \dots, X_{t_n}}(a_1, \dots, a_n) = \mathbb{P}\{X_{t_1} \leq a_1, \dots, X_{t_n} \leq a_n\}$$

אוסף זה של פונקציות פילוג נקרא חוק ההסתברות של התהליך האקראי X_t והוא חייב לקיים את חוק הקונסיסטנטיות:

$$F_{X_{t_1}, \dots, X_{t_{i-1}}, X_{t_{i+1}}, \dots, X_{t_n}}(a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_n) = F_{X_{t_1}, \dots, X_{t_n}}(a_1, \dots, a_{i-1}, \infty, a_{i+1}, \dots, a_n)$$

נביא עתה דוגמא חשובה של ת"א הנקרא תהליך פואסון. תהליך זה משמש מודל למספר רב של תופעות כגון קריאת מונה גייגר, מעבר מכוניות בנקודה בכביש, כניסת שיחות למרכזית טלפונים, מעבר מידע במערכות תקשורת מחשבים, מערכות תורים, פליטת אלקטרונים וכו'. על מנת להגדיר את תהליך פואסון נתאר לעצמנו ארוע המתרחש באקראי מדי פעם ונסמן ב N_t את המספר המצטבר של המאורעות המתרחשים בפרק הזמן $(0, t]$.

הגדרה 5.2 תהליך מנייה הוא תהליך המקבל ערכים שלמים, ואשר פונקציות המדגם שלו אינן יורדות. תהליך מנייה פשוט הוא תהליך אשר כל קפיצה (עליה) היא בגודל 1 בדיוק.

התהליך N_t הוא תהליך מנייה פשוט, דהיינו, הוא מונוטוני לא יורד ויכול לקבל ערכים שלמים $0, 1, 2, \dots$ בלבד. נשתמש בסימון המתמטי הבא: $o(\Delta t)$ הוא גודל זניח לגבי Δt כאשר $\Delta t \rightarrow 0$ (באופן כללי $f(\varepsilon)$ הוא $o(\varepsilon)$ אם $f(\varepsilon)/\varepsilon \rightarrow 0$ כאשר $\varepsilon \downarrow 0$). לגבי חוק ההסתברות של התהליך נניח את ההנחות הבאות:

הגדרה 5.3 תהליך פואסון הוא תהליך מנייה פשוט המקיים את התכונות הבאות:

(א) $\mathbb{P}\{(N_{t+\Delta t} - N_t) = 1\} = \lambda \cdot \Delta t + o(\Delta t)$ כלומר ההסתברות להתרחשות מאורע אחד בדיוק בקטע הזמן $(t, t + \Delta t]$ הוא $\lambda \cdot \Delta t + o(\Delta t)$ או בסימון שקול

$$\lim_{\Delta t \downarrow 0} \frac{\mathbb{P}\{N_{t+\Delta t} - N_t = 1\}}{\Delta t} = \lambda . \quad (5.3)$$

(ב) $\mathbb{P}\{(N_{t+\Delta t} - N_t) = 0\} = 1 - \lambda \cdot \Delta t + o(\Delta t)$ כלומר ההסתברות שלא התרחש אף מאורע בקטע הזמן $(t, t + \Delta t]$ הוא $1 - \lambda \cdot \Delta t + o(\Delta t)$

(ג) ארועים בקטעי זמן לא חופפים הם בלתי תלויים. (תהליך כזה נקרא תהליך בעל תוספות בלתי תלויות או הפרשים בלתי תלויים. תהליך של תוספות בלתי תלויות ראינו כבר בזמן בדיד: הילוך שיכור—דוגמה 3.8).

אנו נניח ש- $N_0 = 0$ אלא אם נאמר במפורש אחרת.

בשלב זה לא ברור שההגדרה סבירה, כלומר שהיא מגדירה באופן חד משמעי את הפילוג של התהליך. בהמשך נראה שאכן ההגדרה מספקת.

$$\mathbb{P}\{(N_{T+\Delta t} - N_t) \geq 2\} = \mathbb{P}\{\text{שני מאורעות או יותר}\} = o(\Delta t) \quad \text{מ- (א) ו- (ב) נובע ש:}$$

תזכורת: אם בניסוי בודד יש סיכוי הצלחה p וסיכוי כישלון q ($p + q = 1$) אזי ההסתברות ל- k הצלחות ב- n ניסויים בלתי תלויים נתונה ע"י הביטוי

$$\binom{n}{k} p^k q^{n-k} = \frac{n!}{(n-k)!k!} p^k q^{n-k} .$$

נחזור לתהליך פואסון. התהליך יתקבל בגבול של חוק ההסתברות הבינומי לעיל, כאשר נחלק את ציר הזמן לקטעים קטנים באורך Δ בצורה הבאה. את הקטע $(0, T]$ נחלק ל- $n = T/\Delta$ קטעי זמן. עבור קטעי הזמן הקצרים שנקבל, נפעיל את ההנחות (א) - (ג). לכן בכל קטע קטן נקבל (בקירוב) $p = \lambda\Delta$. נשים לב שעל מנת לקבל בדיוק $N_T = k$ ארועים בקטע $(0, T]$ צריכים $n - k$ קטעי זמן להיות ללא ארוע ("כשלוף"), ואילו ב k קטעי זמן צריך להיות ארוע ("הצלחה"); לכן, עבור k קבוע, כאשר $\Delta \rightarrow 0$ קטן $n = T/\Delta$ כיוון ש- $n \rightarrow \infty$ נקבל

$$(5.4) \quad \mathbb{P}\{N_T = k\} \cong \frac{n!}{(n-k)!k!} (\lambda\Delta)^k (1-\lambda\Delta)^{n-k} = \frac{n!}{(n-k)!k!} \left(\frac{\lambda T}{n}\right)^k \left(1 - \lambda\frac{T}{n}\right)^{n-k}$$

$$(5.5) \quad = \frac{n!}{(n-k)!n^k} \underbrace{\left(1 - \lambda\frac{T}{n}\right)^{-k}}_{\rightarrow 1} \frac{(\lambda T)^k}{k!} \underbrace{\left(1 - \lambda\frac{T}{n}\right)^n}_{\rightarrow e^{-\lambda T}}$$

נשים לב לגבולות של הביטויים שקיבלנו עבור $n \rightarrow \infty$:

$$\frac{n!}{(n-k)!n^k} = \frac{n(n-1)\cdots(n-k+1)(n-k)!}{nn\cdots n(n-k)!} = \underbrace{\left(\frac{n}{n}\right) \cdot \left(\frac{n-1}{n}\right) \cdots \left(\frac{n-k+1}{n}\right)}_{k \text{ גורמים, כ"א מהם שואף ל-1}} \rightarrow 1 \quad (5.6)$$

כאשר $n \rightarrow \infty$. כלומר

$$\frac{n!}{(n-k)!n^k} \rightarrow 1.$$

בנוסף

$$\left(1 - \lambda\frac{T}{n}\right)^{-k} \rightarrow 1; \quad \left(1 - \lambda\frac{T}{n}\right)^n \rightarrow e^{-\lambda T}$$

ולכן

$$\mathbb{P}\{N_T = k\} = \frac{(\lambda T)^k}{k!} e^{-\lambda T}.$$

זהו פילוג של מ"א פואסוני עם פרמטר λT , ולכן בפרט זהו פילוג. באותה צורה נקבל

$$\mathbb{P}\{N_{T_2} - N_{T_1} = k\} = \frac{[\lambda(T_2 - T_1)]^k}{k!} e^{-\lambda(T_2 - T_1)} = \mathbb{P}\{N_{T_2 - T_1} = k\}$$

כלומר, $N_{T_2} - N_{T_1}$ הוא בעל פילוג זהה לפילוג של $N_{T_2 - T_1}$. קל לראות שאכן מתקיימות הנחות (א)-(ב) לעיל. כאשר Δ קטן נקבל $\mathbb{P}\{N_{t+\Delta} - N_t = 1\} = \lambda\Delta e^{-\lambda\Delta} \approx 1 - \lambda\Delta$ נשים לב כי $e^{-\lambda\Delta} \approx 1 - \lambda\Delta$ (קירוב טיילור).

בשימוש בתכונות מ"א פואסוני ניתן לקבל כי:

$$\mathbb{E}[N_T] = \lambda T$$

$$\mathbb{E}[N_T^2] = (\lambda T)^2 + \lambda T$$

$$\text{Var}(N_T) = (\lambda T)^2 + \lambda T - (\lambda T)^2 = \lambda T$$

נוכח כעת את התכונה הראשונה לפי ההגדרה:

$$\mathbb{E}[N_T] = \sum_{k=0}^{\infty} k \mathbb{P}\{N_T = k\} = \sum_{k=0}^{\infty} k e^{-\lambda T} \frac{(\lambda T)^k}{k!}$$

נסמן $\tau = \lambda T$ ונציג את הטור המתקבל כנגזרת של טור חזקות:

$$\begin{aligned} \mathbb{E}[N_T] &= e^{-\tau} \tau \sum_{k=0}^{\infty} k \frac{\tau^{k-1}}{k!} = e^{-\tau} \tau \left(\sum_{k=0}^{\infty} \frac{\tau^k}{k!} \right)' \\ &= e^{-\tau} \tau (e^\tau)' = \tau = \lambda T. \end{aligned}$$

תכונה זו מסבירה מדוע λ נקרא פרמטר הקצב של התהליך. את חישוב המומנט השני, $\mathbb{E}[N_T^2]$, נשאיר כתרגיל.

תכונה מעניינת נוספת הינה הבאה, עבור $t_2 > t_1$:

$$\mathbb{E}[N_{t_2} N_{t_1}] = \mathbb{E}[N_{t_1}^2] + \mathbb{E}[N_{t_1}(N_{t_2} - N_{t_1})] = (\lambda t_1)^2 + \lambda t_1 + \lambda^2 t_1(t_2 - t_1) = \lambda t_1 + \lambda^2 t_1 t_2$$

ולכן

$$\text{Cov}\{N_{t_1}, N_{t_2}\} = \mathbb{E}[N_{t_1} N_{t_2}] - \underbrace{\mathbb{E} N_{t_1} \mathbb{E} N_{t_2}}_{\lambda^2 t_1 t_2} = \lambda t_1 = \lambda \min\{t_1, t_2\}.$$

השוו תוצאה זו לתוצאה 3.8 שקיבלנו עבור הילוך שיכור, גם הוא תהליך הפרשים בת"ס.

בגלל הנחה (ג), תהליך פואסון הוא תהליך של "תוספות בלתי תלויות", דהיינו: אם $t_1 < t_2 \leq t_3 < t_4$, אזי

$$\mathbb{P}\{N_{t_2} - N_{t_1} = k_1, N_{t_4} - N_{t_3} = k_2\} = \mathbb{P}\{N_{t_2} - N_{t_1} = k_1\} \mathbb{P}\{N_{t_4} - N_{t_3} = k_2\}$$

ולכן, עבור מסרק זמנים $t_1 < t_2 < t_3$ נקבל

$$\mathbb{P}\{N_{t_1} = k_1, N_{t_2} = k_2, N_{t_3} = k_3\} = \mathbb{P}\{N_{t_1} = k_1\} \mathbb{P}\{N_{t_2} - N_{t_1} = k_2 - k_1\} \mathbb{P}\{N_{t_3} - N_{t_2} = k_3 - k_2\}$$

ובאופן כללי עבור $0 < t_1 < \dots < t_n$ נגדיר $t_0 = 0, k_0 = 0$ ונקבל

$$\mathbb{P}\{N_{t_1} = k_1, \dots, N_{t_n} = k_n\} = \mathbb{P}\{N_{t_1} = k_1\} \mathbb{P}\{N_{t_2} - N_{t_1} = k_2 - k_1\} \dots \mathbb{P}\{N_{t_n} - N_{t_{n-1}} = k_n - k_{n-1}\}$$

$$\begin{aligned} &= \prod_{i=1}^n e^{-\lambda(t_i - t_{i-1})} \frac{[\lambda(t_i - t_{i-1})]^{k_i - k_{i-1}}}{(k_i - k_{i-1})!} \\ &= e^{-\lambda t_n} \prod_{i=1}^n \frac{[\lambda(t_i - t_{i-1})]^{k_i - k_{i-1}}}{(k_i - k_{i-1})!}. \end{aligned}$$

בצורה כזו אנו יכולים לקבל את פילוג ההסתברות המשותף של N_{t_1}, \dots, N_{t_n} לכל n ולכל t_1, \dots, t_n ולכן חוק ההסתברות של תהליך פואסון ידוע. בגלל הפרשים הבלתי תלויים חוק ההסתברות הוא פשוט. בדרך כלל אין הדבר כך. בהמשך נראה מקרה אחר שבו נוכל לתאר באופן פשוט את חוק ההסתברות של התהליך (תהליך גאוס).⁷

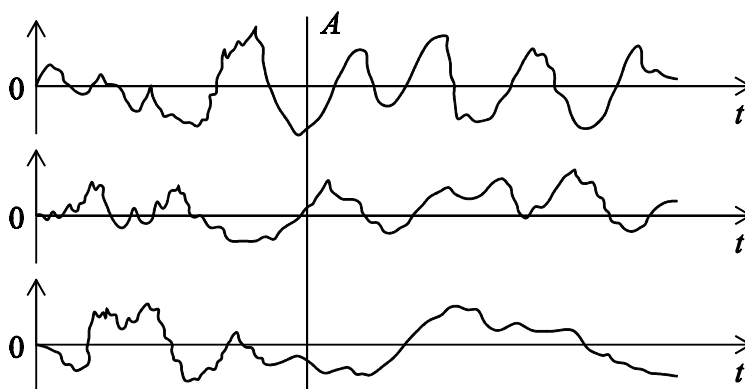
תרגיל 5.4 הוכיחו כי אם N_t ו- M_t הם שני תהליכי פואסון בת"ס עם קצבים λ ו- ν בהתאמה, אזי $N_t + M_t$ הוא תהליך

פואסון עם קצב $\lambda + \nu$. רמז: בדקו את ההנחות.⁷

⁷גרסה ב' מכילה כאן דוגמה נוספת.

5.2 סטציונריות

כבר עסקנו בנושא הסטציונריות בהקשר של ת"א בזמן בדיד. כעת אנו חוזרים לזה בהקשר של זמן רציף ועם יותר עומק ורוחב. נעיין ביציאה מאוסף ("אינסופי") של מגברים החל מרגע הפעלת המגבר. נניח שאין סיגנל וכל היציאה היא רעש שבא מהמגבר עצמו. באופן טיפוסי נקבל את הרישום הבא:



איור 5.4:

החלק מ- $t = 0$ ועד לסביבת הזמן $t = A$ מתאר את תהליך התחממות המגבר והוא מתאר לכן תופעת מעבר אקראית. לאחר מכן מגיע המגבר למצב הניתן לתאור כ"מצב מתמיד". נתאר ב- $\{X(t, \omega), t \geq 0\}$ את רעש היציאה של המקלט. נעיין ב- $\mathbb{P}\{X(t, \omega) \leq a\}$ עבור a קבוע. חוק ההסתברות $F_{X_t}(a)$ בדרך כלל תלוי ב- t , ואכן בדוגמא זו עבור $t < A$ יהיה החוק תלוי ב- t . לעומת זאת עבור $t \gg A$, כלומר, בתחום הזמן של "מצב מתמיד" סביר להניח ש- $F_{X_t}(a)$ יהיה למעשה בלתי תלוי ב- t . שים לב: אין פרוש הדבר ש- $X(t, \omega)$ בלתי תלוי ב- t עבור $t \gg A$ (ראה הצויר); רק חוק ההסתברות אינו תלוי ב- t . על מנת שנוכל לטפל בנוחיות בתחום הזמן המתאר את המצב המתמיד נגדיר את המושג של ת"א סטציונרי. כפי שהגדרנו עבור תהליכים בזמן בדיד (פרק 3),

הגדרה 5.5 התהליך $\{X(t, \omega), \infty < t < \infty\}$ נקרא ת"א סטציונרי אם עבור כל n , כל t_1, \dots, t_n וכל τ , חוקי ההסתברות של הוקטור האקראי:

$$\left(X_{t_1}, X_{t_2}, \dots, X_{t_n} \right)^T$$

ושל הוקטור האקראי:

$$\left(X_{t_1+\tau}, X_{t_2+\tau}, \dots, X_{t_n+\tau} \right)^T$$

זהים זה לזה, דהיינו:

$$F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(a_1, a_2, \dots, a_n) = F_{X_{t_1+\tau}, X_{t_2+\tau}, \dots, X_{t_n+\tau}}(a_1, a_2, \dots, a_n)$$

בצורה מקוצרת נאמר שת"א הוא סטציונרי אם חוק ההסתברות שלו אינו רינטי להזזת ציר הזמן (חוק ההסתברות תלוי בהפרשי זמנים בלבד). תהליך פואסון הוא דוגמא לת"א לא סטציונרי; אפשר לטעון שכל תהליך פיסיקלי איננו סטציונרי מאחר שהוא התחיל בזמן כלשהוא ואכן, המושג של ת"א סטציונרי הוא אידיאליזציה.

הערה 5.6 אם $X(t)$ ת"א סטציונרי אזי לכל $\Delta > 0$ התהליך בזמן בדיד $X(n\Delta)$ אשר מוגדר עבור הזמנים $n\Delta$ הוא סטציונרי, משום שהוא דגימה של תהליך סטציונרי, ותכונת הסטציונריות מיידית שכן היא נובעת מהסטציונריות של X כאשר אנו בודקים במסרקי זמנים המוגבלים להיות כפולות של Δ .

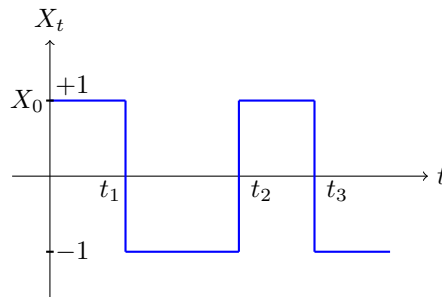
דוגמאות⁸

1. האם תהליך פואסון הינו סטציונרי? התשובה היא כמובן שלילית, ולו מהטעם הפשוט ש- $\mathbb{P}\{N_t = k\} = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$ בוודאי תלוי ב- t . אולם, ניתן להצביע על שני תהליכים לפחות הקשורים בתהליך פואסון שהם כן סטציונריים, ועל כך - בשתי הדוגמאות הבאות.

2. תהליך טלגרף אקראי מוגדר ע"י תהליך בינארי של קפיצות בזמנים אקראיים, שהם זמני האירועים הנמנים ע"י תהליך פואסון. בקצרה:

$$X_t = X_0 (-1)^{N_t}$$

כאשר X_0 מ"א בינארי המקבל את הערכים ± 1 בהסתברות $\frac{1}{2}$, ואילו $\{N_t\}$ תהליך פואסון, בת"ס ב- X_0 . פונקציית מדגם טיפוסית: לדוגמא, זהו מודל מקובל של חילופין בין עירור ורלקסציה של אלקטרונים באטום (הסופגים אנרגיית פוטון



איור 5.5: פונקציית מדגם של תהליך טלגרף אקראי.

ומאבדים אותה חליפות) או פלקטואציות תרמיות של כל מיני דפקטים במל"מ. זה גם מודל מקובל במערכות תקשורת. אם נתבונן במסרק זמנים כלשהו, $t_1 < t_2 < t_3$, אזי לדוגמא:

$$\begin{aligned} & \mathbb{P}\{X_{t_1} = 1, X_{t_2} = -1, X_{t_3} = -1\} \\ &= \underbrace{\mathbb{P}\{X_{t_1} = 1\}}_{1/2} \cdot \underbrace{\mathbb{P}\{X_{t_2} = -1 | X_{t_1} = 1\}}_{\substack{\mathbb{P}\{N_{t_2} - N_{t_1} \text{ is odd}\} \\ = \mathbb{P}\{N_{t_2 - t_1} \text{ is odd}\}}} \cdot \underbrace{\mathbb{P}\{X_{t_3} = -1 | X_{t_1} = 1, X_{t_2} = -1\}}_{\mathbb{P}\{N_{t_3} - N_{t_2} \text{ is even}\}} \end{aligned}$$

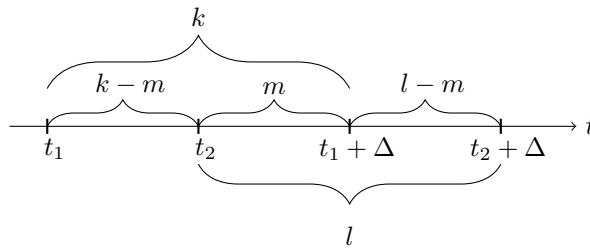
⁸גרסה ב' מכילה כאן דוגמאות שונות, חלקן מפורטות פחות.

וזה תלוי רק בהפרשי הזמנים $t_3 - t_2, t_2 - t_1$ וכו'.

3. $Y_t = N_t - N_{t-\Delta}$ (פרמטר $\Delta > 0$). מודד אירועים בחלון נע בגודל Δ . מה אפשר לומר, למשל, על $\mathbb{P}\{Y_{t_1} = k, Y_{t_2} = l\}$? ברור כי אם $|t_1 - t_2| \geq \Delta$, אזי הקטעים $(t_1, t_1 + \Delta)$ ו- $(t_2, t_2 + \Delta)$ הם בלתי חופפים, ואז לפי הנחה (ג) בהגדרת תהליך פואסון, Y_{t_1} ו- Y_{t_2} בת"ס:

$$\begin{aligned}\mathbb{P}\{Y_{t_1} = k, Y_{t_2} = l\} &= \mathbb{P}\{Y_{t_1} = k\} \mathbb{P}\{Y_{t_2} = l\} \\ &= e^{-\lambda\Delta} \frac{(\lambda\Delta)^k}{k!} \cdot e^{-\lambda\Delta} \frac{(\lambda\Delta)^l}{l!}\end{aligned}$$

ללא תלות כלל ב- t_1 וב- t_2 . אם $|t_1 - t_2| < \Delta$, אזי המצב הוא כדלקמן (נניח $t_1 < t_2$):



איור 5.6: זמנים בתהליך Y_t . ערכי התהליך רשומים על הקשתות.

$$\begin{aligned}\mathbb{P}\{Y_{t_1} = k, Y_{t_2} = l\} &= \sum_{m=0}^{\min(k,l)} \mathbb{P}\{N_{t_2} - N_{t_1} = k - m\} \cdot \mathbb{P}\{N_{t_1+\Delta} - N_{t_2} = m\} \cdot \\ &\quad \cdot \mathbb{P}\{N_{t_2+\Delta} - N_{t_1+\Delta} = l - m\} \\ &= \sum_{m=0}^{\min(k,l)} e^{-\lambda(t_2-t_1)} \frac{[\lambda(t_2-t_1)]^{k-m}}{(k-m)!} \cdot e^{-\lambda(t_1+\Delta-t_2)} \frac{[\lambda(t_1+\Delta-t_2)]^m}{m!} \cdot \\ &\quad \cdot e^{-\lambda(t_2+\Delta-t_1-\Delta)} \frac{[\lambda(t_2-t_1)]^{l-m}}{(l-m)!}\end{aligned}$$

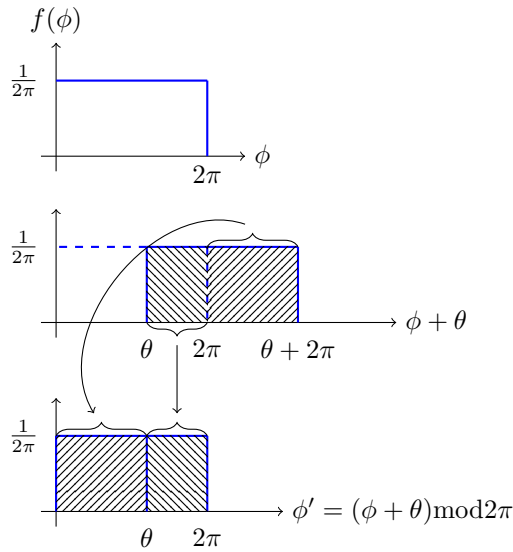
רואים שמה שקובע הוא רק הפרש הזמנים, $t_2 - t_1$, ולא הזמנים האבסולוטיים. כנ"ל כאשר מסתכלים על פילוג משותף של שלושה ערכי Y_t או יותר.

4. אות סינוסי עם פאזה אקראית: נוכיח כי

$$X(t, \omega) = A \sin(2\pi f_0 t + \phi)$$

כאשר $\phi \sim \mathcal{U}[0, 2\pi)$ הינו ת"א סטציונרי.

הוכחה מקוצרת: ראשית, נשים לב לעובדה שאם $\phi \sim \mathcal{U}[0, 2\pi)$ אזי לכל θ (דטרמיניסטי) גם $\phi' = (\phi + \theta) \bmod (2\pi)$ מפולג אחידות בקטע $[0, 2\pi)$. כתרגיל בית מומלץ, שכנעו עצמכם בנכונות טענה זו גם באופן פורמלי. ניתן להשתכנע בכך באמצעות איור 5.7. ברור כי $\cos(2\pi f_0 t + \phi)$ תלוי ב- ϕ רק דרך $\phi \bmod (2\pi)$ בגלל המחזוריות. כעת, נתון מאורע



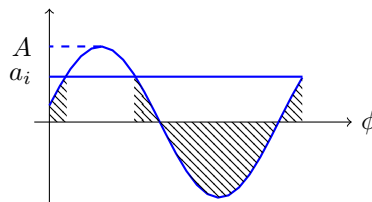
איור 5.7: הסטה דטרמיניסטית של המשתנה ϕ במודולו 2π מניבה משתנה עם אותו פילוג. טווח המשתנה החורג מהתחום $[0, 2\pi)$ מועבר על-ידי פעולת המודולו בחזרה לראשית.

$$\{X_{t_1} \leq a_1, \dots, X_{t_n} \leq a_n\}$$

כאשר כל האקראיות נובעת מ- ϕ . כלומר, המאורע הוא:

$$\{\phi : A \cos(2\pi f_0 t_1 + \phi) \leq a_1, \dots, A \cos(2\pi f_0 t_n + \phi) \leq a_n\} \cap [0, 2\pi)$$

כל אי שוויון כזה משרה אינטרוול, או איחוד אינטרוולים של ϕ בקטע $[0, 2\pi)$ ואנו מסתכלים על החתך של כולם עבור $i = 1, 2, \dots, n$. ההסתברות של המאורע, אם כן, היא מנת סכום האורכים של הקטעים הזרים ב- 2π , בגלל הפילוג האחד של ϕ . כעת, אם עוברים מסט זמנים (t_1, \dots, t_n) לסט זמנים $(t_1 + \tau, \dots, t_n + \tau)$, זה כמו להוסיף ל- ϕ הזזה של $\theta = 2\pi f_0 \tau$ ולקחת מודולו 2π , משום שהסינוס לא רגיש ללקיחת מודולו 2π (מחזוריות). כלומר - זה כמו להחליף את ϕ ב- $\phi' = (\phi + 2\pi f_0 \tau) \bmod (2\pi)$, שהוא גם-כן כעת מפולג אחידות בקטע $[0, 2\pi)$, ולכן הסתברות אותו מאורע לגבי ϕ' תהיה זהה.



איור 5.8: חישוב המאורע הנתון ע"י איחוד אינטרוולים זרים.

הוכחה מפורטת יותר: נגדיר את התהליך מעט אחרת, באופן שקול:

$$X(t, \omega) = A \cos(2\pi f_0 t + 2\pi\phi(\omega)), \quad \{-\infty < t < \infty\}$$

כאשר f_0, A קבועים (אינם אקראיים) ו- $\phi(\omega)$ מ"א אקראי המפולג באופן אחיד בתחום $[0, 1)$. נקבע n נקבע $a_1, \dots, a_n, t_1, \dots, t_n$ ונעיין במאורע B המוגדר כדלקמן:

$$B = \{\omega : X(t_1, \omega) \leq a_1, \dots, X(t_n, \omega) \leq a_n\}$$

נגדיר גם את Φ_B כדלקמן:

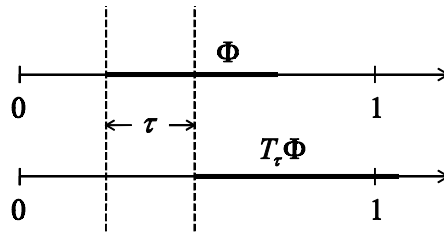
$$\Phi_B = \{\theta : \theta \in [0, 1)\} \cap \{A \cos(2\pi f_0 t_1 + 2\pi\theta) \leq a_1, \dots, A \cos(2\pi f_0 t_n + 2\pi\theta) \leq a_n\}$$

דהיינו: Φ_B הוא אוסף כל המספרים θ כך שאם $\phi(\omega) \in \Phi_B$ (כלומר, אם הגרלנו $\phi(\omega)$ כזה השייך ל- Φ_B) אזי ארע המאורע שקראנו לו B . באופן כללי $\mathbb{P}\{B\} = \mathbb{P}\{\phi \in \Phi_B\}$ ובמיוחד, אם Φ_B הוא קטע ב- $[0, 1)$ אזי $\mathbb{P}\{B\}$ הוא אורך הקטע, ואם Φ_B הוא אחד של קטעים לא חופפים אזי

$$\mathbb{P}\{B\} = \{\text{סכום ארכי הקטעים המהווים את } \Phi_B\}$$

תהיה Φ קבוצת נקודות כלשהיא ב- $[0, 1)$ נגדיר את ההזזה של Φ בכמות τ ע"י (ראה איור 5.9):

$$T_\tau \Phi = \{\phi : \phi - \tau \in \Phi\}$$

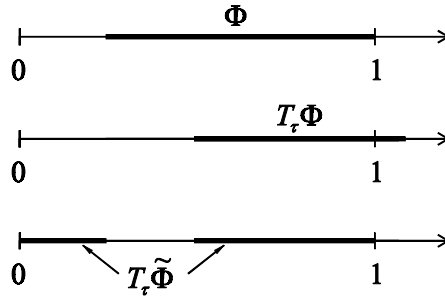


איור 5.9:

שים לב שאם גם Φ וגם $T_\tau \Phi$ ב- $[0, 1)$ אזי $\mathbb{P}\{\phi(\omega) \in \Phi\} = \mathbb{P}\{\phi(\omega) \in T_\tau \Phi\}$ וזאת מאחר ו- $\phi(\omega)$ מפולג אחיד בתחום $[0, 1)$. עבור מספר ממשי X נסמן ב- $[X]$ את החלק השלם של X (כגון אם $X = 2.3$, $[X] = 2$) וב- \tilde{X} נסמן $X - [X]$ (0.3 בדוגמא). לכן אם $X = -1.6$ אזי $[X] = -2$ ו- $\tilde{X} = 4$. שים לב ש- \tilde{X} תמיד בתחום $[0, 1)$. עבור קבוצת נקודות Φ ב- $(-\infty, \infty)$ נסמן ב- $\tilde{\Phi}$ את הקבוצה $\tilde{\Phi} = \{\tilde{\phi} : \phi \in \Phi\}$ (ראה איור 5.10)

כעת אם Φ קבוצת נקודות ב- $[0, 1)$ אזי

$$\mathbb{P}\{\phi(\omega) \in \Phi\} = \mathbb{P}\{\phi(\omega) \in (T_\tau \tilde{\Phi})\}$$



איור 5.10:

עבור כל τ ושוב, הדבר נכון כי מפולג אחיד. נחזור עכשיו ל-

$$\Phi_B = \left\{ \phi : A \cos(2\pi f_0 t_i + 2\pi\phi) \leq a_i, \quad i = 1, \dots, n \right\}$$

נעיין כעת ב-

$$\Phi_C = \left\{ \phi : A \cos(2\pi f_0(t_i + \tau) + 2\pi\phi) \leq a_i, \quad i = 1, \dots, n \right\}$$

אזי $\Phi_C = T_{-2\pi f_0 \tau} \Phi_B$ ולכן

$$\begin{aligned} \mathbb{P}\{X_{t_i+\tau} \leq a_i, \quad i = 1, \dots, n\} &= \mathbb{P}\{\phi \in \tilde{\Phi}_C\} = \mathbb{P}\{\phi \in T_{-2\pi f_0 \tau} \tilde{\Phi}_B\} \\ &= \mathbb{P}\{\phi \in \Phi_B\} = \mathbb{P}\{X_{t_i} \leq a_i, \quad i = 1, \dots, n\} \end{aligned}$$

ובזאת השלמנו את ההוכחה.

שתי ההוכחות אינן משתמשות למעשה בנתונים הייחודיים לדוגמה זו, ולכן הן מראות כי כל פונקציה (דטרמיניסטית) מחזורית עם מחזור T תהפוך לת"א סטציונרי אם ניתן לה פאזה המפולגת באופן אחיד על מחזור אחד שלם. לדוגמא:

5. התהליך הבא סטציונרי, עבור $\{A_n\}$ סדרה i.i.d של משתנים בינאריים $A_n \in \{\pm 1\}$, בת"ס ב- $U \sim \mathcal{U}[0, T]$:

$$X_t = \sum_{n=-\infty}^{\infty} A_n P(t - nT - U)$$

כאשר

$$P(t) = \begin{cases} 1, & 0 \leq t \leq T \\ 0, & \text{otherwise.} \end{cases}$$

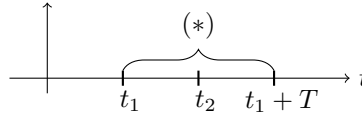
נקבל:

$$\begin{aligned} \mathbb{P}\{X_t = 1\} &= \frac{1}{2} \quad \forall t \\ \mathbb{P}\{X_{t_1} = 1, X_{t_2} = 1\} &= \begin{cases} \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}, & |t_1 - t_2| > T \\ \frac{1}{2} \left(1 - \frac{|t_1 - t_2|}{T}\right) + \frac{1}{4} \cdot \frac{|t_1 - t_2|}{T}, & |t_1 - t_2| \leq T \end{cases} \end{aligned}$$

כאשר $\frac{|t_1 - t_2|}{T}$ הינה ההסתברות שאחד מהזמנים $\{U + nT\}_{n=-\infty}^{\infty}$ נופל באינטרוול $[t_1, t_2]$ (כל עוד $|t_1 - t_2| \leq T$):

$$\mathbb{P}\{U \geq t_1 \bmod T \text{ AND } U \leq t_2 \bmod T\}$$

הכל תלוי בהפרש $(t_2 - t_1)$. אותו דבר לגבי שלשות של זמנים, וכו'.



איור 5.11: דוגמאת זמנים. בקטע (*) חייב להיות מעבר.

הגדרה 5.7 זוג תהליכים $\{X_t, Y_t\}$ נקראים סטציונרים במשותף (או, בצורה שקולה, הווקטור $\{(X_t, Y_t)\}$ נקרא סטציונרי) אם עבור כל n , כל $t_1, s_1, \dots, t_n, s_n$ וכל τ ,

$$F_{X_{t_1}, Y_{s_1}, \dots, X_{t_n}, Y_{s_n}}(a_1, b_1, \dots, a_n, b_n) = F_{X_{t_1+\tau}, Y_{s_1+\tau}, \dots, X_{t_n+\tau}, Y_{s_n+\tau}}(a_1, b_1, \dots, a_n, b_n)$$

בצורה זהה מגדירים סטציונריות במשותף של N ת"א, או סטציונריות של ווקטור של תהליכים.

נשים לב כי שני ת"א סטציונריים אינם בהכרח סטציונרים במשותף.

מומנטים

נעזוב לרגע את מושג הסטציונריות ונתרכז במומנטים מסדר ראשון ומסדר שני הקשורים לת"א. המקרים בהם ידוע מפורשות חוק ההסתברות של ת"א נדירים, אולם להרבה בעיות מספיק לדעת את המומנטים מסדר ראשון ושני (וכן כפי שנראה קיים המושג של תהליך אקראי גאוסי שעבורו ידיעת המומנטים מסדר ראשון ושני מגדירה את חוק ההסתברות).

הגדרות:

פונקצית התוחלת:

$$\mu_X(t) \triangleq \mathbb{E}[X(t)]$$

פונקצית האוטוקורלציה:

$$\mathbf{R}_X(t_1, t_2) \triangleq \mathbb{E}[X_{t_1} X_{t_2}]$$

פונקצית הקווריאנס:

$$\mathbf{K}_X(t_1, t_2) \triangleq \mathbb{E}[(X_{t_1} - \mu_X(t_1))(X_{t_2} - \mu_X(t_2))] = \mathbf{R}_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)$$

כפי שנראה בהמשך, לפונקציות הנ"ל יש חשיבות רבה לגבי תהליכים אקראיים. להלן נסכם מספר תכונות פשוטות וחשובות של פונקצית האוטוקורלציה.

$$1. \mathbf{R}_X(t_1, t_2) = \mathbf{R}_X(t_2, t_1)$$

$$2. \mathbf{R}_X(t_1, t_1) \geq 0$$

3. עבור כל n וכל t_1, \dots, t_n המטריצה:

$$\begin{pmatrix} \mathbf{R}_X(t_1, t_1) & \mathbf{R}_X(t_1, t_2) & \dots & \mathbf{R}_X(t_1, t_n) \\ \mathbf{R}_X(t_2, t_1) & & & \\ \vdots & & & \\ \mathbf{R}_X(t_n, t_1) & \dots & \dots & \mathbf{R}_X(t_n, t_n) \end{pmatrix}$$

מטריצה לא שלילית (מדוע? הוכח).

שלושת התכונות נכונות כמובן (כמקרה פרטי) גם עבור מטריצת הקווריאנס \mathbf{K} .

סטציונריות במובן הרחב

נחזור לסטציונריות; עבור ת"א סטציונרי ברור שמתקיים $\forall a \quad F_{X_t}(a) = F_{X_{t'}}(a)$ לכל t ו- t' לכן

$$\mathbb{E}[X_t^m] = \mathbb{E}[X_{t'}^m]$$

ולכן המומנטים מכל סדר שהוא הינם בלתי תלויים בזמן. בפרט, המומנט הראשון $\mu_X(t)$ אינו תלוי בזמן, כלומר

$$\mathbb{E}[X(t)] = \mu_X(t) = \mu_X(0) \equiv \mu_X$$

שים לב: אם המומנטים מכל סדר שהוא אינם תלויים בזמן אין הדבר גורר סטציונריות של התהליך, שכן המומנטים אינם מספקים מידע על הפילוג המשותף במספר נקודות זמן.

בנוסף $F_{X_{t_1}, X_{t_2}}(a_1, a_2) = F_{X_{t_1+\tau}, X_{t_2+\tau}}(a_1, a_2)$ ולכן לגבי המומנט השני חייב להתקיים:

$$\mathbb{E}[X_{t_1} X_{t_2}] = \mathbb{E}[X_{t_1+\tau} X_{t_2+\tau}]$$

לכן $\mathbf{R}_X(t_1, t_2) = \mathbf{R}_X(t_1 + \tau, t_2 + \tau)$ ואם נבחר $\tau = -t_1$ נקבל:

$$\mathbf{R}_X(t_1, t_2) = \mathbf{R}_X(0, t_2 - t_1) = \mathbf{R}_X(|t_2 - t_1|)$$

או

$$\mathbf{R}_X(t_1, t_1 + \tau) = R_X(\tau) = R_X(-\tau) = R_X(|\tau|)$$

ולכן גם:

$$\mathbf{K}_X(t_1, t_2) = K_X(|t_1 - t_2|); \quad \mathbf{K}_X(t_1, t_1 + \tau) = K_X(|\tau|)$$

דהיינו: עבור ת"א סטציונרי פונקציות האוטוקורלציה והקווריאנס הן פונקציות של משתנה אחד (והוא הפרש הזמנים) בלבד.

הגדרה: ת"א אקראי נקרא "סטציונרי במובן הרחב" אם $\mu_X(t)$ בלתי תלוי ב- t ו- $\mathbf{R}_X(t_1, t_2)$ תלוי ב- $t_1 - t_2$ בלבד דהיינו:

א. עבור כל t מתקיים $\mu_X(t) = \mu_X(0) = \mu_X$

ב. עבור כל t_1, t_2 מתקיים $\mathbf{R}_X(t_1, t_1 + \tau) = \mathbf{R}_X(0, \tau) = R_X(\tau)$

במילים אחרות: ת"א נקרא "סטציונרי במובן הרחב" אם מבחינת המומנטים מסדר ראשון ושני הוא "נראה כאילו היה סטציונרי". ברור שאם ת"א סטציונרי (ושני המומנטים הראשונים קיימים) אזי הוא גם סטציונרי במובן הרחב. ההיפך אינו נכון באופן כללי: אפשר לראות שישנם ת"א סטציונרים במובן הרחב שאינם סטציונריים (דוגמא 5.8).

בדרך כלל קל יותר להוכיח סטציונריות במובן הרחב מאשר סטציונריות. כדי להדגים זאת נבחר מ"א בלתי תלויים A, ϕ כך ש- ϕ מפולג אחיד בתחום $(0, 2\pi]$. נתבונן בתהליך $X_t = A(\omega) \cos(2\pi f_0 t + \phi(\omega))$. עבור $A(\omega)$ דטרמיניסטי הוכחנו שהתהליך סטציונרי (וההוכחה ניתנת להרחבה ל- A אקראי בלתי תלוי ב- ϕ). נוכיח שהתהליך סטציונרי במובן הרחב. היות ו- ϕ מפולג אחיד בתחום $(0, 2\pi]$:

$$\begin{aligned} \mathbb{E}[X_t] &= \mathbb{E}[A(\omega) \cos(2\pi f_0 t + \phi(\omega))] = \mathbb{E}[A(\omega)] \mathbb{E}[\cos(2\pi f_0 t + \phi(\omega))] \\ &= \mathbb{E}[A(\omega)] \cdot \frac{1}{2\pi} \int_0^{2\pi} \cos(2\pi f_0 t + \theta) d\theta = 0 \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X_t X_{t+\tau}] &= \mathbb{E}[A^2(\omega) \cos(2\pi f_0(t + \tau) + \phi) \cdot \cos(2\pi f_0 t + \phi)] \\ &= \frac{1}{2} \mathbb{E}[A^2(\omega) \cos 2\pi f_0 \tau] + \frac{1}{2} \mathbb{E}[A^2 \cos(2\pi f_0(2t + \tau) + 2\phi)] \\ &= \frac{1}{2} \mathbb{E}[A^2(\omega)] \cdot \cos 2\pi f_0 \tau + 0 \end{aligned}$$

וכיוון ש- $\cos x = \cos(-x)$ התהליך סטציונרי במובן הרחב. כדאי לזכור שעבור תהליך פשוט זה

$$R_X(\tau) = \frac{\mathbb{E}[A^2]}{2} \cos(2\pi f_0 \tau)$$

הערה: חזרו על דוגמא זאת כאשר כמקודם, ϕ ו- A בלתי תלויים, אולם הפעם $\phi = 0, \pi/2, \pi, 3\pi/2$ בהסתברות 1/4. האם תהליך זה סטציונרי במובן הרחב? האם תהליך זה סטציונרי?

תכונות פונקציית האוטוקורלציה:

1. אם X_t סטציונרי במובן הרחב אזי $\mathbb{E}[(X_{t+\tau} - X_t)^2] = 2[R_X(0) - R_X(\tau)]$.

2. $R_X(0) \geq 0$

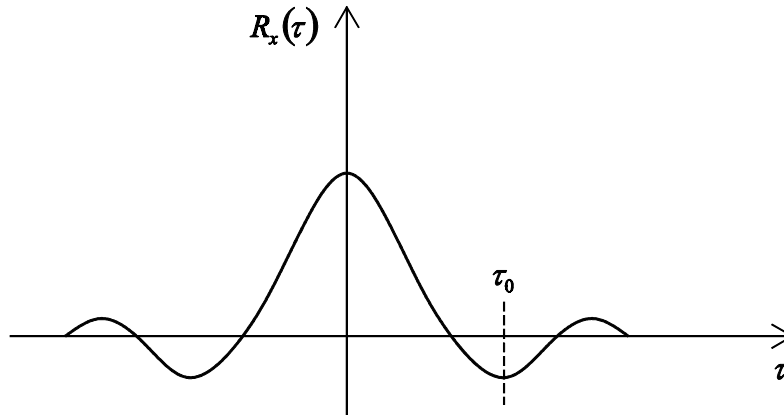
3. $R_X(\tau) = R_X(-\tau)$

4. $\mathbb{E}[X_{t+\tau} \pm X_t]^2 \geq 0$ (הוכח בעזרת $R_X(0) \geq |R_X(\tau)|$)

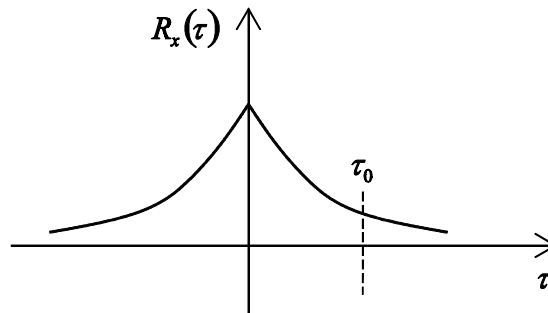
מדוע אנחנו מעוניינים בפונקציית האוטוקורלציה? היא מסכמת את המומנטים מסדר שני של התהליך וכפי שנראה אח"כ היא מאפשרת לענות על שאלות מעניינות. הסבר (חלקי מאוד) על משמעות פונקציית האוטוקורלציה ניתן לראות מהשיקול הבא: נניח סטציונריות, $\mu_X(t) = 0$, ונניח שעבור τ "גדול מאוד", X_t ו- $X_{t+\tau}$ "כמעט" בלתי תלויים סטטיסטית. במקרה זה

$$R_X(\tau) \xrightarrow{|\tau| \rightarrow \infty} 0$$

כמו כן ידוע לנו ש- $R_X(0) \geq |R_X(\tau)|$ ולכן מהלך $R_X(\tau)$ יהיה משהו כמתואר באיור 5.12 או כמתואר באיור 5.13. נוכל



איור 5.12: פונקציית אוטוקורלציה דועכת עם תנודות



איור 5.13: פונקציית אוטוקורלציה דועכת מונוטונית

איפוא לומר שבקרב גס פונקציית האוטוקורלציה "כמעט" מתאפסת עבור $\tau > \tau_0$ ולכן τ_0 מתאר במקרים אלה את "הזכרון האפקטיבי של התהליך". כדי לראות שפונקציית האוטוקורלציה נותנת מעין תאור גס של "הזכרון הטבעי" של התהליך נניח שהקלטנו את $X(t)$ על סרט מגנטי ואנו מריצים את ההקלטה במהירות שהיא α פעמים המהירות המקורית נקבל:

$$Y(t) = X(\alpha t)$$

$$R_Y(\tau) = R_X(\alpha \tau)$$

ואם α גדול אזי $R_Y(\tau)$ מתכווץ והזכרון מתקצר. ברור שישנם גם תהליכים עבורם $R_X(\tau)$ אינו דועך כאשר $|\tau| \rightarrow \infty$: למשל האות הסינוסי עם פאזה אקראית.

חשוב לזכור שמומנטים נותנים תאור חלקי בלבד של התהליך. יתכן מצב בו לשני תהליכים יש את אותה פונקציית תוחלת ואוטוקורלציה, אך הפילוגים וכן פונקציית המדגם שונים בתכלית. בפרט, ראינו שיתכן תהליך סטציונרי במובן הרחב, אשר הפילוגים שלו בזמנים שונים הם שונים וכן גם פונקציית המדגם.

דוגמה 5.8 ת"א סטציונרי במובן הרחב שאינו סטציונרי במובן הצר

נגדיר תהליך בדומה לדוגמה 5 דלעיל:

$$U \sim \mathcal{U}[0, T], \quad X_t = \sum_{n=-\infty}^{\infty} A_n p(t - nT - U)$$

כאשר $p(t)$ היא פונקציית חלון כפי שראינו בעבר (איור 5.21),

$$p(t) = \begin{cases} 1, & 0 \leq t \leq T \\ 0, & \text{otherwise,} \end{cases}$$

ו- $\{A_n\}$ ת"א ב"ת ב- U שדגימותיו בת"ס, אך הפעם:

$$A_n \sim \begin{cases} \mathcal{N}(0, 1), & n < 0 \\ \{\pm 1\} \text{ w.p. } \frac{1}{2}, & n \geq 0 \end{cases}$$

ברור ש- $\{X_t\}$ לא סטציונרי: יש זמנים בהם הוא בינארי ויש זמנים בהם פילוגו השולי הוא גאوسی (הוכיחו!). לתהליך תוחלת אפס, והוא בעל אוטוקורלציה כדלהלן:

$$\begin{aligned} \mathbb{E}[X_t X_s] &= \mathbb{E} \left[\sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} A_n A_m p(t - nT - U) p(s - mT - U) \right] \\ &= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \mathbb{E}[A_n A_m] \mathbb{E}[p(t - nT - U) p(s - mT - U)] \\ &= \sum_{n=-\infty}^{\infty} \mathbb{E}[p(t - nT - U) p(s - nT - U)] \end{aligned}$$

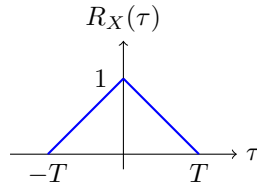
אנטגרל זה שונה מאפס רק עבור n_0 יחיד כלשהו (המתאים לאינטרוול $[0, T]$), ולכן:

$$\mathbb{E}[X_t X_s] = \mathbb{E}[p(t - n_0 T - U) p(s - n_0 T - U)]$$

קעת נשים לב כי התוחלת היא לגבי המשתנה U והפונקציית חלון, ולכן לאחר מעט אלגברה נקבל (איור 5.14):

$$\mathbb{E}[X_t X_s] = \begin{cases} 0, & |t - s| > T \\ 1 - \frac{|t - s|}{T}, & |t - s| \leq T \end{cases}$$

ואם נסמן $\tau = t - s$ נקבל $E[X_t X_s] = R_X(\tau)$ ולכן התהליך הינו סמ"ר.



איור 5.14: האוטוקורלציה של X_t

קורלציה מצטלבת

לצורך ההמשך נזדקק למושגים הבאים (הקשורים במומנטים מסדר שני של זוגות של תהליכים אקראיים). עבור זוג תהליכים אקראיים $\{X_t\}$ ו- $\{Y_t\}$ נגדיר את הקורלציה המצטלבת (קרוסקורלציה) כדלקמן:

$$\mathbf{R}_{X,Y}(t_1, t_2) \triangleq \mathbb{E}[X_{t_1} Y_{t_2}]$$

ולכן,

$$\mathbf{R}_{X,Y}(t_1, t_2) = \mathbf{R}_{Y,X}(t_2, t_1)$$

התהליכים נקראים חסרי קורלציה אם $\mathbf{R}_{X,Y}(t_1, t_2) = 0$ לכל t_1, t_2 . כאשר $Z_t = X_t + Y_t$ מתקבל

$$\begin{aligned} \mathbf{R}_Z(t_1, t_2) &= \mathbb{E}[(X_{t_1} + Y_{t_1})(X_{t_2} + Y_{t_2})] \\ &= \mathbf{R}_X(t_1, t_2) + \mathbf{R}_Y(t_1, t_2) + \mathbf{R}_{X,Y}(t_1, t_2) + \mathbf{R}_{Y,X}(t_1, t_2) \end{aligned}$$

כך ש- $\mathbf{R}_{X,Y}$ מופיע באופן טבעי כאשר אנו רוצים לחשב את \mathbf{R}_Z . כמובן שאם $\{X_t\}$ ו- $\{Y_t\}$ חסרי קורלציה אזי $\mathbf{R}_Z(t_1, t_2) = \mathbf{R}_X(t_1, t_2) + \mathbf{R}_Y(t_1, t_2)$

הערה: אם זוג התהליכים אקראיים $\{X_t, Y_t, -\infty < t < \infty\}$ הוא סטציונרי (=סטציונרי במשותף) אזי

$$\mathbf{R}_{X,Y}(t_1, t_2) = \mathbf{R}_{X,Y}(t_1 + \tau, t_2 + \tau) = \mathbf{R}_{X,Y}(0, t_2 - t_1)$$

ונגדיר במקרה זה

$$\mathbf{R}_{X,Y}(t_1, t_2) = R_{X,Y}(t_2 - t_1)$$

או

$$\mathbf{R}_{X,Y}(t_1, t_1 + \tau) = R_{X,Y}(\tau) = R_{Y,X}(-\tau)$$

הגדרה: זוג התהליכים האקראיים $\{X_t, Y_t, -\infty < t < \infty\}$ נקראים סטציונריים במשותף (או - התהליך הווקטורי נקרא סטציונרי) במובן הרחב אם מתקיים עבור כל t וכל τ :

א. $\mathbb{E}[Y_t] = \mathbb{E}[Y_0]; \quad \mathbb{E}[X_t] = \mathbb{E}[X_0]$

ב. $\mathbb{E}[Y_{t+\tau} Y_t], \mathbb{E}[X_{t+\tau} X_t]$ ו- $\mathbb{E}[Y_{t+\tau} X_t]$ הם פונקציות של τ בלבד.

כמו במקרה של סטציונריות במובן הצר, יתכן מצב בו כל אחד משני תהליכים הוא סטציונרי במובן הרחב, אך הם אינם סטציונריים במשותף במובן הרחב. לדוגמה נזכר בתהליך האפנון על ידי פולס מרובע. יהי $T > 1$ ו- U מ"א המפולג אחיד על $[0, T]$. יהי $\{a_n\}, \{b_n\}$ אוסף של משתנים בת"ס המקבלים ערכים ± 1 בהסתברות $1/2$ כ"א. אזי ראינו כי כל אחד מהתהליכים

$$(5.7) \quad X(t) \doteq \sum_{n=-\infty}^{\infty} a_n p(t - nT - U)$$

$$(5.8) \quad Y(t) \doteq \sum_{n=-\infty}^0 a_n p(t - nT - U)$$

$$(5.9) \quad + \sum_{n=1}^{\infty} b_n p(t - nT - U)$$

הוא סטציונרי (במובן הצר!). אולם

$$(5.10) \quad \mathbb{E}[X_1 Y_1] = \mathbb{E}[a_n^2 p(1 - U)] = 1$$

$$(5.11) \quad \mathbb{E}[X_{-1} Y_{-1}] = \mathbb{E}[a_{-1} b_{-1} p(1 - U)] = 0.$$

כלומר התהליך הווקטורי אינו סטציונרי במובן הרחב, או במילים אחרות זוג התהליכים אינו סטציונרי במשותף במובן הרחב.

דוגמה 5.9 דוגמא נוספת, ופשוטה יותר, היא הבאה:

$$X_t = \cos(2\pi f_x t + \phi), \quad Y_t = \cos(2\pi f_y t + \phi)$$

עם פאזה מפולגת אחידות, $\phi \sim \mathcal{U}[0, 2\pi]$. מדובר בשני תהליכים סמ"ר כל אחד, בעלי פאזה אקראית וזוהה, הנבדלים זה בתדר. כעת,

$$\mathbb{E}[X_t Y_t] = \frac{1}{2} \cos(2\pi(f_x - f_y)t)$$

ולכן הם אינם סמ"ר במשותף.

5.3 תהליך אקראי גאوسی

הגדרה: ת"א $\{X_t, t \in [a, b]\}$ נקרא גאوسی אם עבור כל n וכל $t_1, \dots, t_n \in [a, b]$ הווקטור האקראי $(X_{t_1}, \dots, X_{t_n})^T$ הוא ו"א גאوسی.

טענה: אם X_t ת"א גאوسی ב- $[a, b]$ אזי חוק ההסתברות שלו נקבע חד משמעית ע"י פונקציות התוחלת והאוטוקורלציה שלו

$$t, t_1, t_2 \in [a, b], \quad \mathbf{R}_X(t_1, t_2), \mu_X(t)$$

הוכחה: עבור כל n וכל המקיימים $t_1, \dots, t_n \in [a, b]$ הווקטור האקראי

$$\underline{Z} = (X_{t_1}, X_{t_2}, \dots, X_{t_n})^T$$

הוא וקטור אקראי גאוס (היות והנחנו ש- $\{X_t, t \in [a, b]\}$ הוא ת"א גאוס). עבור המומנטים מסדר ראשון ושני של \underline{Z} מתקיים:

$$\mathbb{E}[\underline{Z}] = \left(\mu_X(t_1), \mu_X(t_2), \dots, \mu_X(t_n) \right)^T$$

ו-

$$\mathbb{E}[\underline{Z}\underline{Z}^T] = \{\mathbb{E}[X_{t_i}X_{t_j}]\} = \{\mathbf{R}_X(t_i, t_j)\}$$

לכן עבור כל (ν_1, \dots, ν_n) , הפונקציה האופינית של הוקטור \underline{Z} נתונה ע"י:

$$\phi_Z(\nu_1, \dots, \nu_n) = \exp \left\{ i \sum \nu_i \mu_X(t_i) - \frac{1}{2} \sum_i \sum_j \nu_i \nu_j \mathbf{K}_X(t_i, t_j) \right\}$$

כאשר כזכור $\mathbf{K}_X(t_1, t_2) = \mathbf{R}_X(t_1, t_2) - \mu_X(t_1)\mu_X(t_2)$. הפונקציה האופינית של וקטור אקראי מגדירה חד משמעית את חוק ההסתברות של הוקטור האקראי, לכן $\mu_X(t)$ ו- $\mathbf{R}_X(t_1, t_2)$ מגדירים את חוק ההסתברות של התהליך.

מסקנה: תהליך אקראי גאוס סטציונרי במובן הרחב הוא סטציונרי.

טענה: אם $\{X_t, -\infty < t < \infty\}$ ת"א גאוס, גם התהליכים

$$a(t)X_t$$

$$a(t)X_t + b(t)X_{t+c}$$

תהליכים גאוסיים עבור כל $a(\cdot)$, $b(\cdot)$, c דטרמיניסטיים (מדוע? הוכח). לכן נצפה גם שאם הגבול

$$\frac{dX_t}{dt} = \lim_{\varepsilon \rightarrow 0} \frac{X_{t+\varepsilon} - X_t}{\varepsilon}$$

קיים, אזי (בהנחות מתאימות) הגבול גם הוא ת"א גאוס.

יהיה X_t ת"א גאוס; אזי

$$\int_a^b X_s ds \approx \sum_i X_{s_i} (s_{i+1} - s_i)$$

הוא מ"א גאוס והתהליכים

$$\int_{-\infty}^{\infty} h(t-\theta)X(\theta)d\theta, \quad \int_{-\infty}^{\infty} g(t,\theta)X(\theta)d\theta$$

הם ת"א גאוסיים כאשר $g(\cdot, \cdot)$, $h(\cdot)$ דטרמיניסטיים.

מסקנה: פעולה לינארית (לא אקראית) על תהליך אקראי גאוס נותנת תהליך אקראי גאוס.

הערה: תחת תנאים טכניים ניתן להחליף סדר תוחלת ואנטגרל מהשיקול הבא:

$$\mathbb{E} \int_0^t X_s ds \approx \mathbb{E} \sum X_{t_i} (t_{i+1} - t_i) = \sum \mathbb{E} X_{t_i} (t_{i+1} - t_i) \approx \int_0^t \mathbb{E} X_s ds$$

התנאי המרכזי הדרוש לשם החלפת הסדר הוא

$$\mathbb{E} \int_0^t |X_s| ds < \infty \quad \text{או התנאי השקול} \quad \int_0^t \mathbb{E} |X_s| ds < \infty \quad (5.12)$$

5.4 מעבר תהליכים אקראיים דרך מערכות לינאריות

נתונה מערכת כבאיר 5.15. תהליך הכניסה הוא $X(t)$ ותהליך היציאה הוא $Y(t)$.

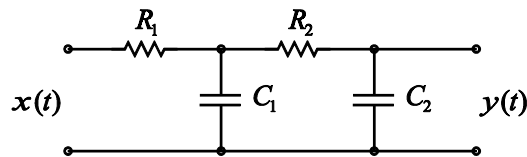


איור 5.15: מעבר אות במערכת דטרמיניסטית

שאלה: ידוע חוק ההסתברות של התהליך $\{X(t), -\infty < t < \infty\}$, נתונה מערכת (לא אקראית, לינארית או לא לינארית) שאפיונה ידוע. מהו חוק ההסתברות של תהליך היציאה $\{Y(t), -\infty < t < \infty\}$? בדרך כלל התשובה אינה ידועה אפילו אם מדובר במערכת לינארית קבועה בזמן. באופן מעשי נתעניין הרבה פעמים לא בחוק ההסתברות של $Y(t)$ אלא ב- $\mathbb{E}[Y^2(t)]$ או $\mathbb{E}[(Y(t) - \mathbb{E}[Y(t)])^2]$. למען הקיצור נקרא לגודל זה "הספק היציאה" (כאילו היה מדובר במתח על נגד של אוהם אחד, ואז $\mathbb{E}[Y^2(t)]$ הוא ההספק הממוצע הנמסר לנגד). לדוגמה, אם נתון מקלט הקולט אות רועש, והוא תוכנן כדי להפחית את השפעת הרעש, נרצה לנתח את פילוג אות המוצא, או לפחות לנתח את השפעת הרעש על ידי חישוב ההספק של הרעש במוצא המערכת. לבעיה הכללית אין פתרון, ולכן נסתפק בפתרונה (חישוב $\mathbb{E}[Y^2(t)]$) כאשר המערכת לינארית.

דוגמאות:

(א) מסנן R-C:

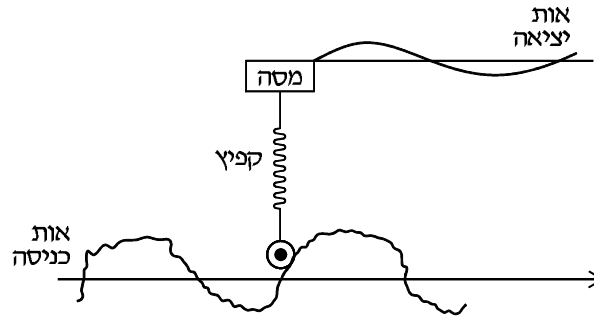


איור 5.16: מסנן

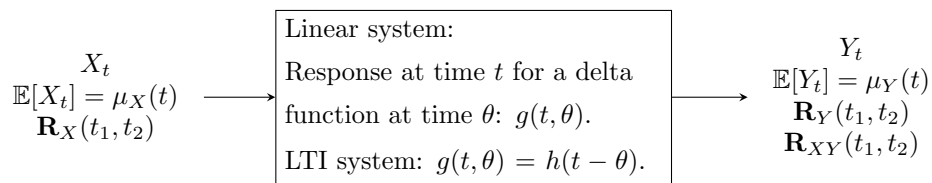
(ב) בולם זעזועים:

ננסח את השאלה הבאה: נתונה מערכת לינארית לא אקראית שהאפיון שלה ידוע; מה צריך לדעת על $\{X(t), -\infty < t < \infty\}$, על מנת שנוכל לחשב את $\mathbb{E}[Y^2(t)]$? בהמשך נקבל תשובה מלאה לשאלה זו. ברור שידיעת $\mathbb{E}[X^2(t)]$ אינה מספיקה (מדוע?). את הבעיה נתאר בצורה ציורית כדלקמן:

כבר עסקנו בשאלה דומה בהקשר של ו"א וטרנספורמציות לינאריות.



איור 5.17: בולם זעזועים



איור 5.18:

הקשר בין כניסת המערכת ויציאתה נתון ע"י:

$$Y(t) = \int_{-\infty}^{\infty} X(\theta)g(t, \theta)d\theta \cong \sum_i X(\theta_i)g(t, \theta_i)(\theta_{i+1} - \theta_i) \quad (5.13)$$

לא נתיחס כאן לבעיות של דיוק מתמטי ונרשום:

$$\mathbb{E}[Y(t)] \cong \sum_i \mathbb{E}[X(\theta_i)g(t, \theta_i)(\theta_{i+1} - \theta_i)] \rightarrow \int_{-\infty}^{\infty} \mu_X(\theta)g(t, \theta)d\theta$$

במילים אחרות, בצענו תוחלת על (5.13) והחלפנו את סדר האינטגרציה והתוחלת. כך קבלנו את המומנט מסדר ראשון של $Y(t)$ נעיין כעת במומנטים המצטלבים:

$$\begin{aligned} \mathbf{R}_{X,Y}(t_1, t_2) &= \mathbb{E} \left[X(t_1) \int_{-\infty}^{\infty} X(\theta)g(t_2, \theta)d\theta \right] \\ &= \mathbb{E} \left[\int_{-\infty}^{\infty} X(t_1)X(\theta)g(t_2, \theta)d\theta \right] \end{aligned}$$

ושוב נחליף סדר האינטגרציה והתוחלת ונקבל

$$\mathbf{R}_{X,Y}(t_1, t_2) = \int_{-\infty}^{\infty} \mathbf{R}_X(t_1, \theta)g(t_2, \theta)d\theta$$

לבסוף, נעיין ב- $\mathbf{R}_Y(t_1, t_2)$. שים לב שידעת $\mathbf{R}_Y(t_1, t_2)$ כוללת ידיעת "הספק היציאה הממוצע"

$$\mathbb{E}[Y^2(t)] = \mathbf{R}_Y(t, t)$$

$$\begin{aligned} \mathbb{E}[Y(t_1)Y(t_2)] &= \mathbb{E}\left[\int_{-\infty}^{\infty} X(\theta)g(t_1, \theta)d\theta \int_{-\infty}^{\infty} X(\eta)g(t_2, \eta)d\eta\right] \\ &= \mathbb{E}\left[\iint_{-\infty}^{\infty} X(\theta)X(\eta)g(t_1, \theta)g(t_2, \eta)n\theta d\eta\right] \end{aligned}$$

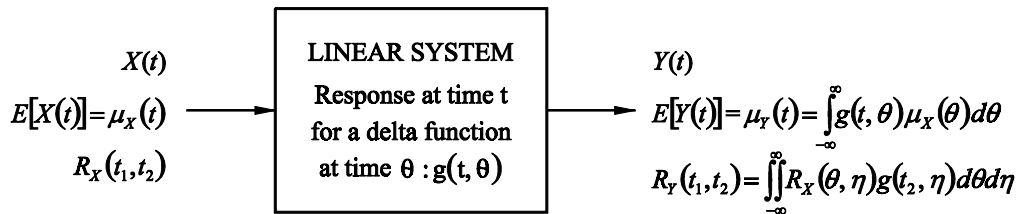
(5.14)

$$\mathbf{R}_Y(t_1, t_2) = \iint_{-\infty}^{\infty} \mathbf{R}_X(\theta, \eta)g(t_1, \theta)g(t_2, \eta)d\theta d\eta$$

מסקנות:

- (א) עבור מערכות לינאריות לא אקראיות (קבועות בזמן או משתנות בזמן, סיבתיות או לא סיבתיות), ידיעת $\mathbf{R}_X(t_1, t_2), \mu_X(t)$ עבור $-\infty < t, t_1, t_2 < \infty$ מאפשרת את קביעת $\mu_Y(t)$ ואת $\mathbf{R}_Y(t_a, t_b)$ עבור כל t_a, t_b .
- (ב) כאשר $X(t)$ ת"א גאוסי גם $Y(t)$ ת"א גאוסי ולכן ידיעת המומנטים מסדר ראשון ושני של $Y(t)$ מגדירה את חוק ההסתברות של תהליך היציאה.
- (ג) שים לב שעל מנת לדעת את $\mathbf{R}_Y(t, t)$ צריך לדעת את $\mathbf{R}_X(t_1, t_2)$ עבור כל t_1, t_2 .

את התוצאות (5.13) ו (5.14) נסכם בציור הבא:



איור 5.19: מומנטים של תהליך העובר במערכת לינארית

5.4.1 מעבר תהליכים אקראיים סטציונריים במובן הרחב דרך מערכות קבועות בזמן

כאשר המערכת קבועה בזמן היא מאופינת ע"י התגובה להלם $h(t)$ (כלומר $g(t, \theta) = h(t - \theta)$). נוכל לכן לקבל את התוצאות המבוקשות עבור מקרה זה ע"י החלפת $g(t, \theta)$ ב $h(t - \theta)$ בתוצאות הקודמות. אך נפתח מחדש. עבור תהליך סטציונרי במובן הרחב $\mu_X(t) = \mu_X$ ו- $g(t, \theta) = h(t - \theta)$. כיוון שהקשר בין כניסה ויציאה במקרה זה הוא

$$Y(t) = \int_{-\infty}^{\infty} X(t - \theta)h(\theta)d\theta$$

נקבל, בהחלפת סדר התוחלת והאינטגרציה

$$\mu_Y(t) \doteq \mathbb{E}[Y(t)] = \int_{-\infty}^{\infty} \mathbb{E}[X(t-\theta)]h(\theta)d\theta = \mu_X \int_{-\infty}^{\infty} h(\theta)d\theta$$

שים לב שהאינטגרל האחרון הוא ההגבר DC של המערכת הלינארית. קיבלנו שממוצע התגובה גם הוא אינו תלוי בזמן. באשר לאוטוקורלציה,

$$\begin{aligned} \mathbf{R}_Y(t, t + \tau) &= \mathbb{E}[Y(t)Y(t + \tau)] \\ &= \mathbb{E}\left[\int_{-\infty}^{\infty} X(t-\theta)h(\theta)d\theta \int_{-\infty}^{\infty} X(t + \tau - \eta)h(\eta)d\eta\right] \\ &= \iint_{-\infty}^{\infty} \mathbb{E}[X(t-\theta)X(t + \tau - \eta)]h(\theta)h(\eta)d\theta d\eta \\ &= \iint_{-\infty}^{\infty} R_X(\tau - \eta + \theta)h(\theta)h(\eta)d\theta d\eta \\ &= [R_X * h * \check{h}](\tau) \end{aligned}$$

כאשר הגדרנו $\check{h}(\tau) = h(-\tau)$ והמשוואה האחרונה היא סימון נכון יותר ל- $R_X(\tau) * h(\tau) * h(-\tau)$. כמו כן

$$(5.15) \quad \mathbf{R}_{X,Y}(t, t + \tau) = \mathbb{E}[X(t)Y(t + \tau)]$$

$$(5.16) \quad = \mathbb{E}\left[\int_{-\infty}^{\infty} X(t)X(t + \tau - \theta)h(\theta)d\theta\right]$$

$$(5.17) \quad = \int_{-\infty}^{\infty} R_X(\tau - \theta)h(\theta)d\theta$$

$$(5.18) \quad = [R_X * h](\tau)$$

$$(5.19) \quad = R_{X,Y}(\tau) .$$

בפיתוחים שעשינו החלפנו סדר בין תוחלת ואינטגרציה. כפי שהזכרנו החלפה זו מותרת אם מתקיים תנאי (5.12). אולם תנאי זה שקול ליציבות BIBO.

מסקנה: אם $\{X(t), -\infty < t < \infty\}$ סטציונרי במובן הרחב, המערכת לינארית קבועה בזמן ויציבה BIBO אזי $\{X(t), Y(t), -\infty < t < \infty\}$ סטציונריים במובן הרחב ו-

$$(5.20) \quad \mu_Y = \mu_X \int_{-\infty}^{\infty} h(\theta)d\theta$$

$$(5.21) \quad R_Y(\tau) = \iint_{-\infty}^{\infty} R_X(\tau - \eta + \theta)h(\theta)h(\eta)d\theta d\eta$$

בפרט, אם $\{X(t)\}$ ת"א גאוסי סטציונרי במובן הרחב אזי הוא גם סטציונרי, ונובע כי גם $\{Y(t)\}$ גאוסי וסטציונרי וכן $\{X(t), Y(t)\}$ גאויסיים במשותף וסטציונריים במשותף, ואז ידיעת הפילוג מסדר ראשון ושני שקולה לידיעת חוק ההסתברות המלא

אם המערכת סיבתית אזי $h(\theta) = 0$ עבור $\theta < 0$ ואז

$$R_Y(\tau) = \int_0^\infty \int_0^\infty R_X(\tau - \eta + \theta) h(\theta) h(\eta) d\theta d\eta \quad (5.22)$$

השימוש העיקרי של התוצאה עבור $R_Y(\tau)$ הוא "הספק היציאה" ואז

$$R_Y(0) = \iint R_X(\theta - \eta) h(\theta) h(\eta) d\theta d\eta$$

(כאשר גבולות האינטגרציה הם מ $-\infty$ עד ∞ או מאפס ועד ∞).

מעתה והלאה נעסוק רק במערכות קבועות בזמן ובתהליכי כניסה שהם סטציונריים במובן הרחב.

באותות ומערכות למדנו שיש שתי גישות לאפיון מערכות לינאריות קבועות בזמן (LTI) כאשר מדובר באותות דטרמיניסטיים:

1. בתחום הזמן: אפיון המערכת ע"י תגובתה להלם; היות ו- $X_t = \int_{-\infty}^\infty X_\tau \delta(t-\tau) d\tau$, והתגובה לכל $\delta(t-\tau)$ היא $h(t-\tau)$, אזי היציאה היא סופרפוזיציה: $Y_t = \int_{-\infty}^\infty X_\tau h(t-\tau) d\tau$ וזו הקונוולוציה.

2. בתחום התדר: אפיון התגובה לאות הרמוני, $e^{2\pi jft}$, אות הכניסה נתון ע"י סופרפוזיציה של הרמוניות: $X(t) = \int_{-\infty}^\infty X(f) e^{2\pi jft} df$. התגובה לכל הרמוניה, $e^{j2\pi ft}$, הינה $H(f) e^{2\pi jft}$, ולכן בסך-הכל היציאה הינה $Y(t) = \int_{-\infty}^\infty \underbrace{H(f) X(f)}_{Y(f)} e^{2\pi jft} df$

עד עתה ראינו את האנלוג לתחום הזמן במקרה של תהליכים אקראיים: $R_Y(\tau) = (R_X * h * h(-\cdot))(\tau)$ או באופן פחות פורמלי, $R_X(\tau) * h(\tau) * h(-\tau)$.

האם אפשר, בהקבלה המתאימה, לתת מובן לתחום התדר כאשר מדובר בתהליכים אקראיים? לאנליזה בתחום התדר יש פעמים רבות יתרון; משפט הקונוולוציה יתכן ויהיה שימושי גם פה, מכיוון שבתחום הזמן עלינו לבצע קונוולוציה לחישוב האוטוקורלציה במוצא המערכת.

חזרה: התמרות פוריה

נתחיל ב- $X(t)$ לא אקראי ונניח ש- $\int_{-\infty}^\infty |X(t)| dt < \infty$. התמרת פוריה של $X(t)$ מוגדרת ע"י:

$$\hat{X}(f) = \mathcal{F}\{X(t)\} = \int_{-\infty}^\infty X(t) e^{-2\pi ift} dt$$

והתמרת פוריה ההפוכה נתונה ע"י

$$X(t) = \mathcal{F}^{-1}\{\hat{X}(f)\} = \int_{-\infty}^\infty \hat{X}(f) e^{2\pi ift} df$$

קיימת נקודה עדינה לגבי קיום האינטגרל האחרון אולם נתעלם מבעיה זו). משפט פרסוול בהתמרות פוריה אומר שאם עבור $i = 1, 2$ אזי $\int_{-\infty}^{\infty} |X_i(t)|^2 dt < \infty$

$$\int_{-\infty}^{\infty} X_1(t)X_2^*(t) dt = \int_{-\infty}^{\infty} \hat{X}_1(f)\hat{X}_2^*(f) df$$

ל- $\int_{-\infty}^{\infty} X^2(t) dt$ נקרא האנרגיה של הפונקציה $X(\cdot)$ (מה האנרגיה של דגם טפוסי של ת"א סטציונרי?)

אם $\mathcal{F}\{X(t)\} = \hat{X}(f)$ אזי בתנאים מתאימים קיים

$$\mathcal{F}\left\{\frac{dX(t)}{dt}\right\} = 2\pi i f \hat{X}(f)$$

אם $X(\cdot)$ ממשי אזי

$$\mathcal{F}\{X(-t)\} = \hat{X}^*(f)$$

$$\mathcal{F}\{X(t + \tau)\} = e^{2\pi i f \tau} \hat{X}(f)$$

$$\mathcal{F}\{X * h(t)\} = \mathcal{F}\left\{\int_{-\infty}^{\infty} X(t - \theta)h(\theta) d\theta\right\} = \hat{X}(f) \cdot \hat{h}(f)$$

צפיפות ספקטרלית

כאשר $(X(t), -\infty < t < \infty)$ ת"א סטציונרי במובן הרחב, האנרגיה של פונקציה מדגם טיפוסית היא אינסופית ולכן לא ברור אם בכלל אפשר לבצע התמרת פוריה על פונקציה כזו, ואמנם, אנו לא נעשה זאת. אנו נתעניין בהתמרת פוריה של פונקציה האוטוקורלציה: $S_X(f) \triangleq \mathcal{F}\{R_X(\tau)\}$. זו נקראת פונקציה הצפיפות הספקטרלית (צה"ס, או ספקטרום) של $\{X(t)\}$. בהמשך נבין מה מקור השם.

בנוסף, נגדיר עבור ת"א סטציונריים במובן הרחב במשותף, $\{Y(t), X(t)\}$ את פונקציה הצפיפות הספקטרלית המצטלבת:

$$S_{X,Y}(f) \triangleq \mathcal{F}\{R_{X,Y}(\tau)\}$$

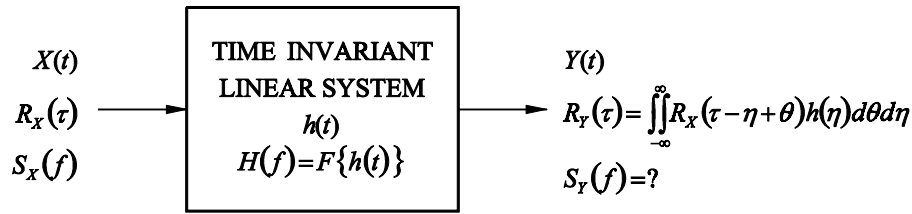
וזאת בהנחה שלפונקציות $R_X(\cdot), R_{X,Y}(\cdot)$ יש התמרת פוריה.

נניח עכשיו $\{X(t), -\infty < t < \infty\}$ סטציונרי במובן הרחב ונניח $\mathbb{E}[X(t)] = 0$. נתעניין בצפיפות הספקטרלית ביציאה של מערכת לינארית המתוארת באיור 5.20. אנו נניח כי h ממשית. במקרה הנוכחי:

$$R_{X,Y}(\tau) = \mathbb{E}[X(t)Y(t + \tau)] = \int_{-\infty}^{\infty} R_X(\tau - \theta)h(\theta) d\theta = (R_X * h)(\tau)$$

ולכן (מדוע?)

$$S_{X,Y}(f) = S_X(f) \cdot H(f)$$



איור 5.20: מעבר אות סמ"ר במערכת לק"ב יציבה

הקונוולוציה הפכה למכפלה. לגבי $S_Y(f)$:

$$R_Y(\tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_X(\tau - \theta + \eta) h(\theta) h(\eta) d\theta d\eta = R_X(\tau) * h(\tau) * h(-\tau)$$

על פי ההגדרה $S_Y(f) = \int_{-\infty}^{\infty} R_Y(\tau) e^{-2\pi i f \tau} d\tau$ ולכן בשינוי סדר האינטגרציה

$$\begin{aligned} \int_{\eta=-\infty}^{\infty} \int_{\theta=-\infty}^{\infty} \int_{\tau=-\infty}^{\infty} e^{-2\pi i f \tau} R_X(\tau + \theta - \eta) h(\theta) h(\eta) d\tau d\theta d\eta &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} S_X(f) e^{2\pi i f(\theta - \eta)} h(\theta) h(\eta) d\theta d\eta \\ &= S_X(f) H(f) \cdot H^*(f) \\ &= S_X(f) |H(f)|^2 \end{aligned}$$

כאשר בשוויון הלפני אחרון השתמשנו בעובדה כי h ממשי. כלומר

$$S_Y(f) = S_X(f) |H(f)|^2$$

וקבלנו שספקטרום אות המוצא ממערכת ממשית תלוי רק באמפליטודה (ולא בפאזה) של תגובת התדר. בפרט, הספק היציאה הממוצע הכולל הוא:

$$\mathbb{E}[Y^2(t)] = R_Y(0) = \int_{-\infty}^{\infty} S_Y(f) e^{j2\pi f \cdot 0} df = \int_{-\infty}^{\infty} S_X(f) |H(f)|^2 df$$

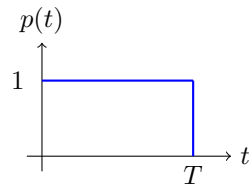
וזו נוסחה שחשוב לזכור.

דוגמא: נזכר בשני התהליכים שראינו בעבר:

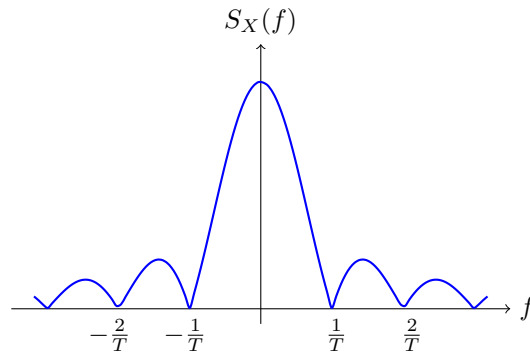
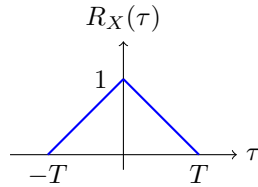
$$Y_t = \sum_n A_n p(t - nT - U)$$

$$X_t = N_t - N_{t-T} - \lambda T$$

כאשר הפולס $p(t)$ מוצג באיור 5.21. המשתנה U מפולג אחידות, $U \sim \mathcal{U}[0, T]$, התהליך $\{A_n\}$ הינו i.i.d ו- $\mathbb{E}[A_n^2] = 1$ וכן



איור 5.21:



איור 5.22: אוטוקורלציה וספקטרום של הת"א הנדון.

$\{N_t\}$ הינו תהליך פואסון.

לשני תהליכים אלו יש פונקציות אוטוקורלציה משולשת (איור 5.22), ולכן $S_X(f)$ פרופורציונלית ל- $\frac{\sin^2(\pi f T)}{(\pi f)^2}$.

עבור פולס $p(t)$ כללי הינו מקבלים $S_X(f) \propto |p(j2\pi f)|^2$, ולכן אחד מהשיקולים בבחירת פולסים בבעיות תקשורת הוא רוחב סרט קטן ככל האפשר, כדי לא לבזבז רוחב סרט וכדי להקטין חדירות לרעש.

המצב, נכון לעכשיו, הוא כדלקמן: הגדרנו, בצורה שרירותית למדי, את התמרת פוריה של $R_X(\tau)$ וראינו שחוקי המעבר דרך מערכת לינארית קבועה בזמן מקבלים צורה נוחה.

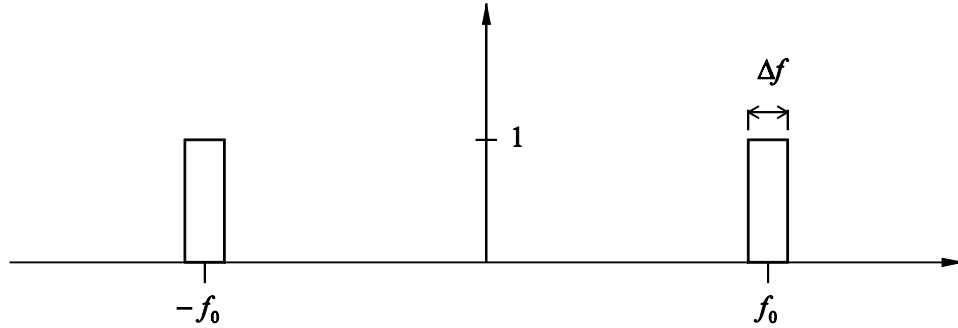
כדי להבין את משמעות המושג צפיפות ספקטרלית $S_X(f)$ נשים לב לתכונותיה:

(א) $S_X(-f) = S_X(f)$ וכן $S_X(f)$ ממשית (וזאת מפני ש- $R_X(\tau)$ ממשית וסימטרית)

(ב) $\int_{-\infty}^{\infty} S_X(f) df \geq 0$ (כי הביטוי משמאל הוא $R_X(0)$).

(ג) טענה: $S_X(f) \geq 0$.

הוכחה: נבחר ב- $H(f)$ להיות מסנן צר סרט אידיאלי סביב תדר שרירותי f_0 :



איור 5.23:

הספק היציאה במקרה זה יהיה (בערך)

$$0 \leq \mathbb{E}[Y^2(t)] = \int_{-\infty}^{\infty} S_X(f) |H(f)|^2 df \approx 2S_X(f_0) \Delta f$$

$$S_X(f_0) \approx \frac{\mathbb{E}[Y^2(t)]}{2\Delta f}$$

ולכן עבור כל f_0 , $S_X(f_0) \geq 0$.

$\mathbb{E}[Y^2(t)]$ מודד למעשה את ההספק של התהליך X בתחום תדרים צר (כפי שמכתיב המסנן) סביב f_0 . גודל זה הוא בקירוב לינארי ב- Δf כאשר תחום התדרים הוא צר מספיק, וקבוע הפרופורציה הוא בדיוק צפיפות ההספק הספקטרלית. מסיבה זו S_X נמדד ביחידות של $(\text{Volt})^2/\text{Hz}$ או $(\text{Amp})^2/\text{Hz}$, כתלות בהאם $X(t)$ הוא מתח או זרם.

ישנם מקרים בהם הספק היציאה, $\mathbb{E}[Y^2(t)]$, אינו שואף לאפס כאשר $\Delta f \rightarrow 0$. לדוגמא, אות סינוסי בתדר f_0 :

$X(t) = A \cos(2\pi f_0 t + \phi)$ כאשר A, ϕ ב"ת ו- ϕ מפולג אחיד בתחום $(0, 2\pi)$. כזכור

$$(5.23) \quad R_X(\tau) = \frac{1}{2} \mathbb{E}[A^2] \cos 2\pi f_0 \tau$$

$$(5.24) \quad S_X(f) = \frac{1}{4} \mathbb{E}[A^2] [\delta(f - f_0) + \delta(f + f_0)]$$

בדוגמה זו ה"צפיפות" היא פונקציה מוכללת, וההספק במוצא מסנן צר סרט אינו תלוי ברוחב הסרט, אלא רק בשאלה אם התדר f_0 נחסם על ידי המסנן או לא.

נשים לב כי מתכוונות אלו נובע שלא כל פונקציה זוגית יכולה להיות אוטוקורלציה: כדי להיות כזאת הפונקציה צריכה להיות בעלת התמרת פוריה אי-שלילית.

לסיכום: מושג הצפיפות הספקטרלית נותן תאור של תכולת ההספק בתדרים השונים, דהיינו, גם אם לא ניתן לדבר על התמרת פוריה של פונקציות מדגם של התהליך $\{X(t), -\infty < t < \infty\}$, פונקציות הצפיפות הספקטרלית של התהליך (הסטציונרי

במובן הרחב) נותנת את פילוג ההספק לפי ציר התדר. לפיכך, אם $S_X(f)$ הצפיפות הספקטרלית של התהליך, אזי

$$2 \int_{f_0}^{f_0+\Delta} S_X(f) df$$

הוא ההספק הממוצע ביציאה ממסננת bandpass אידאלית בעלת רחב סרט Δ בתחום $(f_0, f_0 + \Delta)$.

שים לב: כאשר $\phi(t)$ פונקציה דטרמיניסטית בעלת אנרגיה סופית, אנו יכולים לדעת את תכולת האנרגיה בתחום תדרים מסוים ע"י עיון באינטגרל של $|\Phi(f)|^2$ על תחום התדרים המעניין. כאן אנו עוסקים בתכולת ההספק!

כדי להוכיח כי $S_X(f)$ אכן מתאר את תכולת ההספק של פונקציות המדגם, נעביר את התהליך X דרך מסנן מעביר סרט $(-f_0, f_0)$, כאשר $S_X(f) = 0$ מחוץ לסרט. אם נגדיר תהליך שגיאה כ- $\varepsilon(t) = X(t) - X * h(t)$ אזי

$$\mathbb{E}\varepsilon^2 = \int S_X(f) |1 - H(f)|^2 df = 0.$$

הערות:

(א) גוזר $Y(t) = dX(t)/dt$ מבליט תדרים גבוהים ומתקיים עבורו $F\{Y(t)\} = 2\pi i f F\{X(t)\}$.

(ב) מערכת שהקשר בין הכניסה $X(t)$ והיציאה $Y(t)$ נתון ע"י $Y(t) = \int_{t-\Delta}^t X(\theta) d\theta$ מתאימה לתגובה להלם שהיא 1 בין הזמנים אפס ו- Δ ואפס בזמנים אחרים. מתוך עיון בהתמרת פוריה של התגובה להלם נובע מיד שמערכת זו מחליקה (מרסנת תדרים גבוהים).

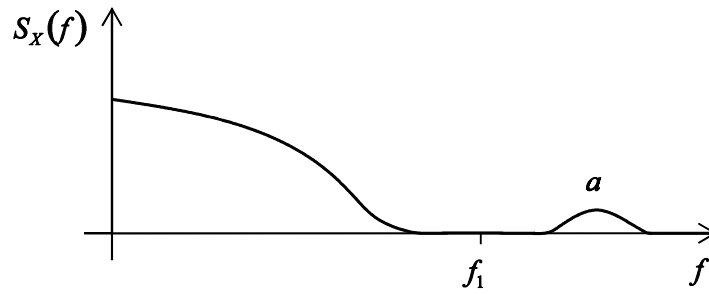
(ג) בכל מקרה של ציור העוסק בבעיות מהסוג שאנו מטפלים בהם יש לבדוק את משמעות הציור האופקי על מנת לברר אם הוא מבטא זמן או תדר.

(ד) כזכור הגדרנו את הצפיפות הספקטרלית עבור תהליך אקראי בעל תוחלת אפס. אפשר ליחס צפיפות ספקטרלית גם לתהליכים (סטציונריים במובן הרחב) בעלי תוחלת שונה מאפס. במקרה זה תהיה לצפיפות הספקטרלית פונקציות דירק בתדר אפס. דהיינו, אם לת"א $X(t)$ נוסף מ"א C בלתי תלוי ב $X(t)$, לקבלת $Y(t) = X(t) + C$, אזי: $S_Y(f) = S_X(f) + \mathbb{E}[C^2]\delta(f)$. כלומר, ת"א שיש לו רכיב DC הוא בעל ספקטרום המכיל דיראק בראשית.

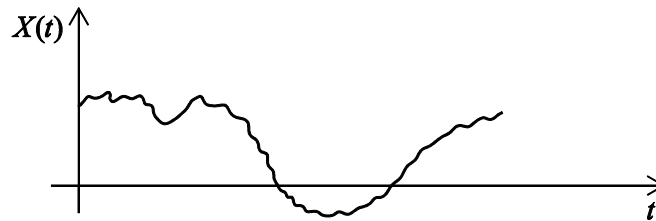
דוגמה: נעסוק באות אקראי $X(t)$ סטציונרי במובן הרחב. תהיה $S_X(f)$ הצפיפות הספקטרלית—ראה איור 5.24—של אות זה (כולל החלק α).

אנו מניחים שה"תוספת" המסומנת ב- α נובעת מרעש שאיננו מעוניינים בו. בהנחה ש- $S_X(f)$ (כולל החלק α) מתיחס לתהליך אקראי גאוס, דגם טיפוסי יראה כמתואר באיור 5.25.

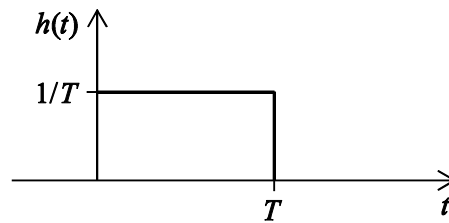
ה"רעידות" של $X(t)$ באות מהרעש, היינו, מהחלק המסומן במישור התדר ב- α . אנו מתבקשים לבצע גזירה. במרחב התדר עלינו להכפיל את $S_X(f)$ ב- $(2\pi f)^2 = |H(f)|^2$ והגזירה תגביר את אפקט הרעש בצורה ניכרת, זה ברור הן במרחב הזמן והן במרחב התדר. בכדי להתגבר על אפקט הרעש מוצע לבצע החלקה ע"י העברת האות $X(t)$ דרך מערכת ליניארית קבועה בזמן שתגובתה להלם ניתנת ע"י $h(t)$ מאיור 5.26 ואז $H(f) = F\{h(t)\}$ נראה כמתואר באיור 5.27. אם נבחר ב- T גדול



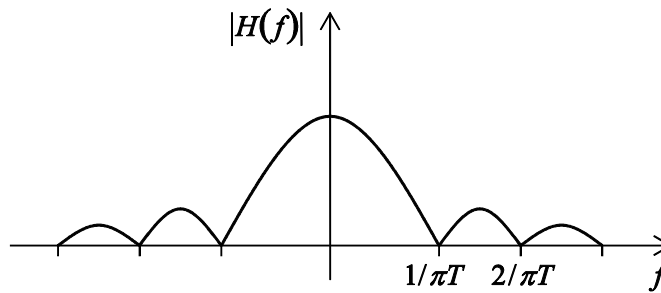
איור 5.24: צפיפות ספקטרלית של אות הכולל רעש



איור 5.25: פונקצית מדגם של האות הכולל רעש



איור 5.26: תגובת הלבם של המסנן



איור 5.27: תגובת תדר של המסנן

מאוד, נפגע בסיגנל עצמו, אם נבחר ב- T קטן מאוד לא נחליק את הרעש. מתוך שיקולים ענייניים נראה שכדאי לבחור את T כך ש- $f_1 = 1/\pi T$ כאשר f_1 מופיע בציור לעיל.

שאלה: האם בבעיה זו יש קודם לבצע החלקה ואח"כ גזירה או להפך?

המשפט הבא עוזר לנו להבין את המושג של צפיפות ספקטרלית, ומקשר בינה לבין התמרת פוריה של האות עצמו.

משפט: נניח $\int_{-\infty}^{\infty} |\tau R(\tau)| d\tau < \infty$, נסמן

$$X_T(t) = \begin{cases} X(t) & \text{if } |t| < T \\ 0 & \text{if } |t| \geq T \end{cases}$$

$$\hat{X}_T(f) = F\{X_T(t)\}$$

אזי

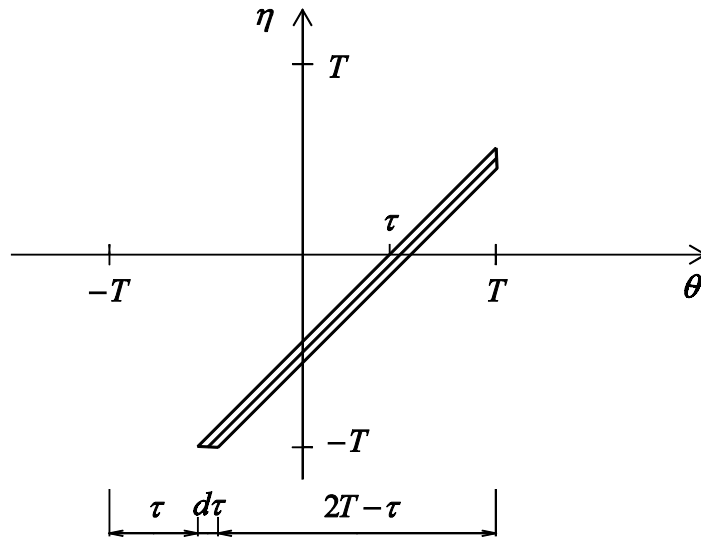
$$\mathbb{E} \left[\frac{1}{2T} |\hat{X}_T(f)|^2 \right] \xrightarrow{T \rightarrow \infty} S_X(f)$$

כאן רואים את הקשר בין התמרת פוריה של האות עצמו לצפיפות הספקטרלית. החלוקה באורך אינטרוול הזמן מעבירה אותנו מאנרגיה להספק.

הוכחה:

$$\mathbb{E} \left[\frac{1}{2T} |\hat{X}_T(f)|^2 \right] = \frac{1}{2T} \mathbb{E} \left[\iint_{-T}^T X(\theta) X(\eta) e^{-2\pi i f(\theta - \eta)} d\theta d\eta \right] = \frac{1}{2T} \iint_{-T}^T R_X(\theta - \eta) e^{-2\pi i f(\theta - \eta)} d\theta d\eta$$

נסמן $\tau = \theta - \eta$ ונעיין באיור 5.28: נבצע קודם אינטגרציה על הפס הצר שעבורו $\tau \leq \theta - \eta \leq \tau + d\tau$ לכן



איור 5.28: שינוי סדר אינטגרציה

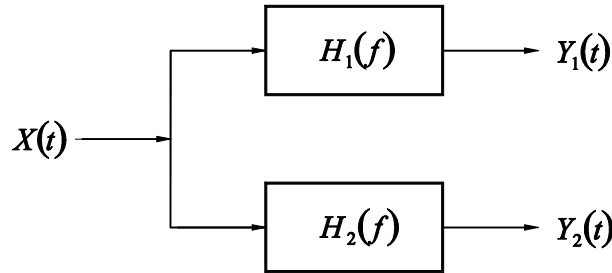
$$\begin{aligned}
\mathbb{E} \left[\frac{1}{2T} |\hat{X}_T(f)|^2 \right] &= \frac{1}{2T} \int_{-2T}^{2T} R_X(\tau) e^{-2\pi i f \tau} \underbrace{(2T - |\tau|) \cdot \sqrt{2}}_{\text{שטח הטרפז}} \frac{d\tau}{\sqrt{2}} \\
&= \int_{-2T}^{2T} R_X(\tau) e^{-2\pi i f \tau} \left(1 - \frac{|\tau|}{2T} \right) d\tau \\
&= \int_{-2T}^{2T} R_X(\tau) e^{-2\pi i f \tau} d\tau - \frac{1}{2T} \underbrace{\int_{-2T}^{2T} R_X(\tau) e^{-2\pi i f \tau} |\tau| d\tau}_{< \infty} \\
&\xrightarrow{T \rightarrow \infty} \mathcal{F}\{R_X(\tau)\} = S_X(f)
\end{aligned}$$

מ.ש.ל.

נחזור למושג הצפיפות הספקטרלית המצטלבת

$$S_{X,Y}(f) = \mathcal{F}\{R_{X,Y}(\tau)\} = \mathcal{F}\{\mathbb{E}[X(t)Y(t+\tau)]\}$$

נאמר כי X, Y חסרי קורלציה אם $\mathbb{E} X(t_1)Y(t_2) \equiv 0$ (כלומר שווה לאפס לכל (t_1, t_2) או, באופן שקול: $S_{XY}(f) \equiv 0$).
נעיין במיוחד במקרה המתואר באיור 5.29.



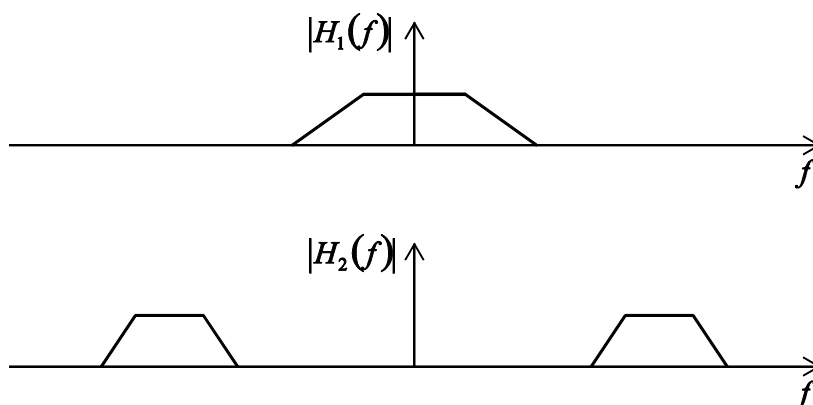
איור 5.29: כניסה משותפת לשתי מערכות במקביל

$$\begin{aligned}
Y_i(t) &= \int_{-\infty}^{\infty} h_i(\theta) X(t - \theta) d\theta \\
R_{Y_1, Y_2}(\tau) &= \mathbb{E} [Y_1(t)Y_2(t + \tau)] = \mathbb{E} \left[\iint_{-\infty}^{\infty} X(t - \theta)X(t + \tau - \eta)h_1(\theta)h_2(\eta) d\theta d\eta \right] \\
&= \iint_{-\infty}^{\infty} R_X(\tau + \theta - \eta)h_1(\theta)h_2(\eta) d\theta d\eta = [R_X * \check{h}_1 * h_2](\tau)
\end{aligned}$$

כאשר השתמשנו בסימון $\check{h}_1(\tau) = h_1(-\tau)$ לכן עבור h ממשיה

$$\begin{aligned}
S_{Y_1, Y_2}(f) &= \iiint R_X(\tau + \theta - \eta) e^{-2\pi i f \tau} h_1(\theta)h_2(\eta) d\theta d\eta d\tau \\
&= \iint S_X(f) e^{2\pi i f(\theta - \eta)} h_1(\theta)h_2(\eta) d\theta d\eta = S_X(f) H_1^*(f) H_2(f)
\end{aligned}$$

במיוחד אם $H_2(f)H_1(f) \equiv 0$, כגון כנתון באיור 5.30. אזי התהליכים האקראיים $Y_1(t)$ ו- $Y_2(t)$ חסרי קורלציה (ולכן



איור 5.30: שני מסננים ללא חפיפה בתגובת התדר

במקרה של תהליכים גאוסיים במשותף הת"א $\{Y_1(t), -\infty < t < \infty\}$ ו- $\{Y_2(t), -\infty < t < \infty\}$ בלתי תלויים).

הערות:

- הוכיחו שהמשתנים האקראיים $X(t)$ ו- $dX(t)/dt$ חסרי קורלציה עבור אותו t , אולם כתהליכים אקראיים הם לא חסרי קורלציה.

- הזכרו בדוגמא 5.9, בה הצגנו שני תהליכים אקראיים סטציונריים במובן הרחב:

$$X_t = \cos(2\pi f_x t + \phi), \quad Y_t = \cos(2\pi f_y t + \phi)$$

צפיפות ההספק הספקטרלית של תהליכים אלו מכילה שני דיראקים בתדרים שונים, ואי לכך הם אינם חופפים בתדר, אך כפי שראינו בעבר, התהליכים תלויים. זאת מכיוון שהם אינם סמ"ר במשותף.

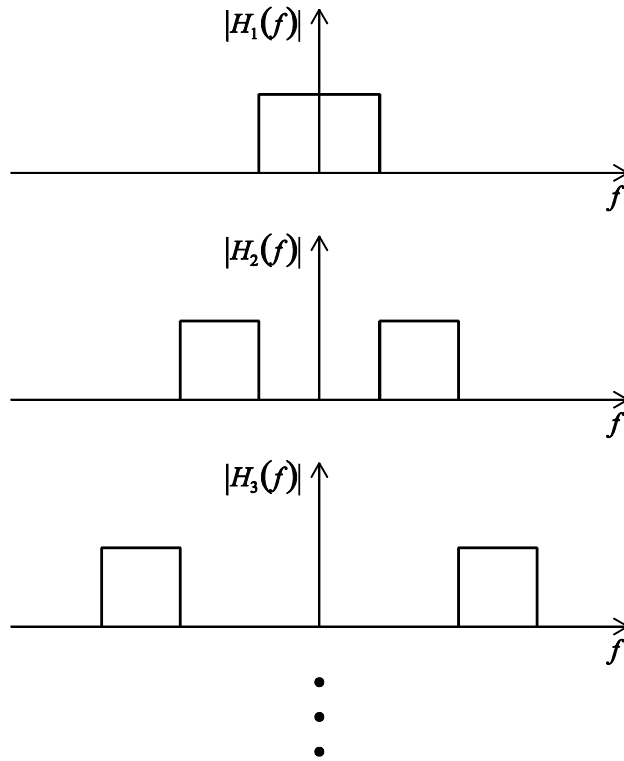
נגדיר אוסף של מסננים $H_1(f), H_2(f), \dots$ ע"י איור 5.31. מסננים אלו מקימים:

$$\forall i \neq j, \quad H_i(f) \cdot H_j(f) \equiv 0; \quad \sum_i H_i(f) = 1$$

כלומר המסננים מחלקים את האות במישור התדר כך שכל מסנן מעביר תחום תדר אחר, וכל תדר מועבר על ידי מסנן כלשהו. נחשוב כעת על אות $X(t)$ הנכנס במקביל לכל המסננים: המסננים מפרידים את האות לפי תדרים. מהדוגמה הקודמת (איור 5.30 והדיון שאחריו) נסיק כי אות המוצא ממסנן i חסר קורלציה ביחס לאות המוצא ממסנן j עבור $u \neq j$. במילים אחרות, בין פסי התדר של אות אין קורלציה.

זוהי טכניקה שימושית במגוון בעיות בעבוד אותות ותקשורת, בדחיסת אינפורמציה למשל, ובפרט בדחיסת אותות שמע; "מפרקים" את האות (דיבור, מוזיקה וכו') לפסי תדר ודוחסים את המידע בכל פס תדר לחוד, באופן מקבילי.

בעזרת מושג הצפיפות הספקטרלית אנו יכולים לחשב את ההספק ביציאת כל אחד מהמסננים. ההספק ביציאת מסנן ברוחב Δf הוא בערך $2 \cdot \Delta f \cdot S_X(f_0)$, כאשר f_0 הוא התדר המרכזי של המסנן. הנקודה החשובה היא, שאנו יכולים גם ללכת הפוך



איור 5.31: אוסף מסננים ללא חפיפה במישור התדר

בשיטה זאת ולשערך את $S_X(f)$ של תהליך $X(t)$ לא ידוע, מתוך מדידת הספקי היציאה של מערך מסננים כנ"ל. בצורה כזו אנו מפרקים את האות לאוסף של אותות צרי סרט, כל אחד בתחום תדר נפרד, כאשר האותות הם חסרי קורלציה (ולכן, במקרה הגאוס, בת"ס), ומטפלים בכל אות בנפרד. זוהי שיטה שמושית בבעיות מעשיות רבות, למשל בדחיסת אינפורמציה.

5.5 הזתת ספקטרום

בסעיף זה נעסוק בבעיה הבאה: מה קורה כשמכפילים ת"א סטציונרי במובן הרחב $X(t)$ ב- $\cos(2\pi f_0 t + \phi)$ כאשר ϕ מ"א בלתי תלוי בתהליך $X(t)$ ומפולג אחיד בתחום $[0, 2\pi]$. נחשוב על X כאות הבסיסי אותו רוצים לשדר, ומאפננים אותו בעזרת ה- \cos . לפעולה כזו חשיבות רבה במערכות תקשורת. נעיין ב-

$$Y(t) = X(t) \cos(2\pi f_0 t + \phi)$$

אזי $Y(t)$ סטציונרי במובן הרחב, וחשבון פשוט נותן

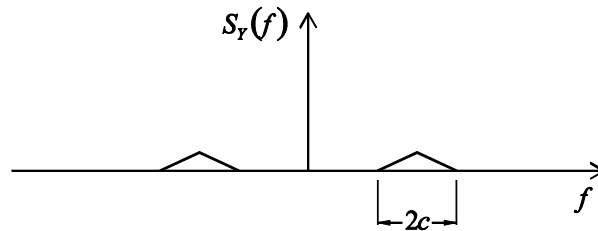
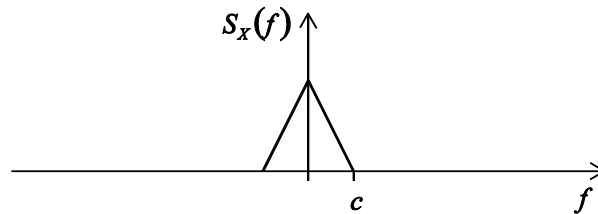
$$R_Y(\tau) = \frac{1}{2} R_X(\tau) \cos 2\pi f_0 \tau$$

כידוע, עבור פונקציה דטרמיניסטית כלשהי g מתקיים

$$\mathcal{F}\{g(t) \cos 2\pi f_0 t\} = \frac{1}{2} [G(f + f_0) + G(f - f_0)]$$

$$S_Y(f) = \frac{1}{4} [S_X(f + f_0) + S_X(f - f_0)]$$

מקרה א': אות צר סרט מאופנן על ידי אות בתדר גבוה כמתואר באיור 5.32. מקרה ב': נזכר באות שבו נתקלנו בעבר,



איור 5.32: ספקטרום של אות צר סרט ושל האות המאופנן

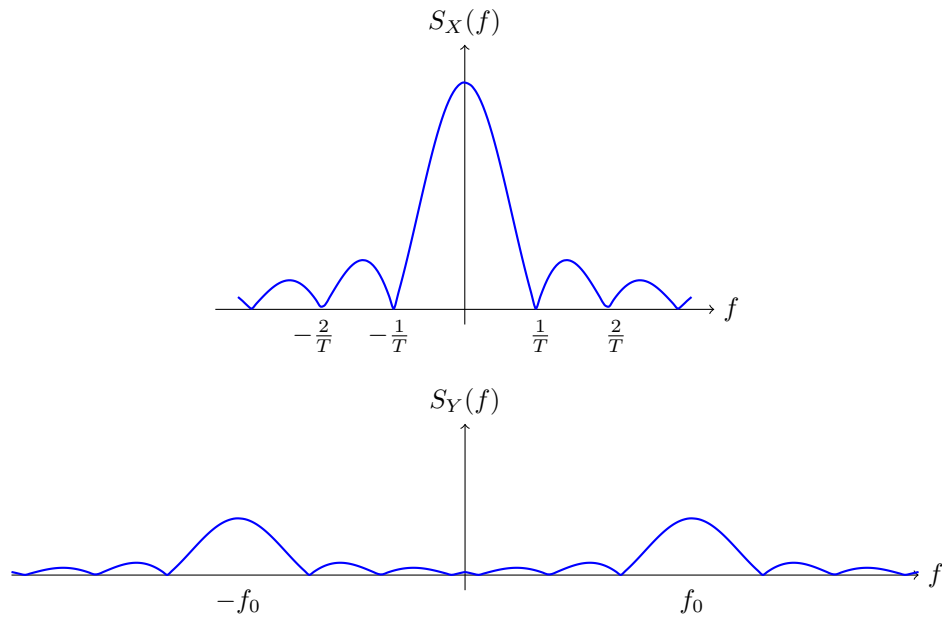
$$X_t = \sum_n A_n p(t - nT - U)$$

לאות זה אוטוקורלציה משולשת (איור 5.22) ולכן הוא בעל תמך אינסופי בתדר. האות המאופנן מתואר באיור 5.33.

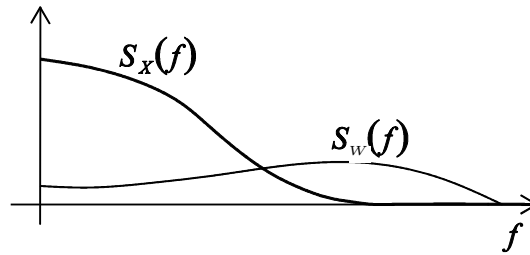
5.6 סינון לינארי אופטימלי

נעבור כעת לבעיה כללית של סינון לינארי. זוהי הרחבה (לאותות) של בעית השערוך הלינארי, כאשר כעת המדידה היא תהליך. נניח שנתונים שני ת"א $\{X(t), -\infty < t < \infty\}$ ו- $\{W(t), -\infty < t < \infty\}$, לכל אחד מהם תוחלת אפס וכל אחד מהם סטציונרי במובן הרחב. כן נניח שהתהליכים חסרי קורלציה ($\mathbb{E}[X(t_1)W(t_2)] = 0$ לכל t_1 ו- t_2) ולכן הם סטציונריים במשותף במובן הרחב. נקלט האות $Y(t)$ שהוא סכום האות הרצוי $X(t)$ והרעש $W(t)$, כלומר, $Y(t) = X(t) + W(t)$. מתוך האות הנקלט $Y(t)$ רוצים לסלק את הרעש ולקבל את האות הרצוי $X(t)$ ע"י העברת $Y(t)$ דרך מסנן לינארי שתגובתו להלם היא $h(t)$ וקבלת יציאה נאמנה ככל האפשר ל- $X(t)$. לפעולה זו קוראים סינון (של הרעש) והיא מתוארת באיור 5.35. האות המצוי אינו זהה לאות הרצוי היות ויש שגיאה הנובעת הן מהרעש והן מעיוות האות הנגרם ע"י $h(t)$. על מנת להבהיר זאת נצייר את איור 5.35 בצורה השקולה, המופיעה באיור 5.36. בגלל אי התלות הלינארית בין $X(\cdot)$ לבין $W(\cdot)$ נוכל לרשום

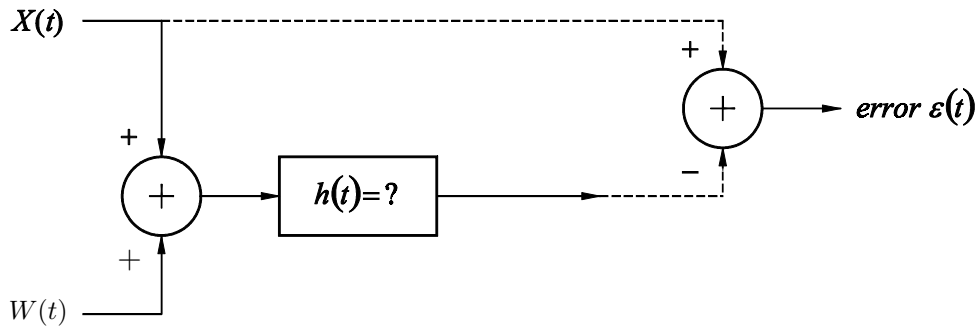
$$\mathbb{E}[\varepsilon^2(t)] = \mathbb{E}[z_1^2(t)] + \mathbb{E}[z_2^2(t)]$$



איור 5.33: ספקטרום של אות ושל האות המאופנן



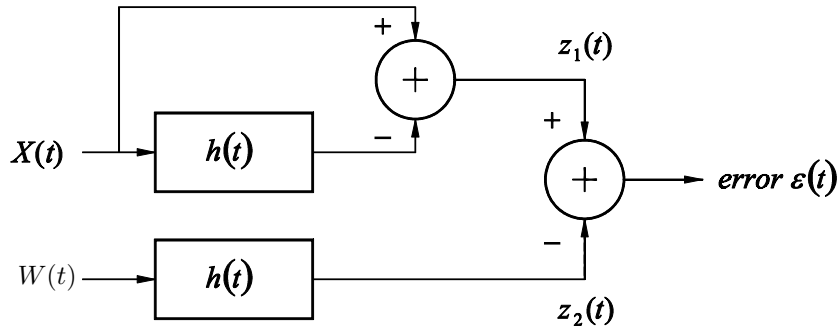
איור 5.34: צפיפות ספקטרלית טיפוסית של אות ושל רעש



איור 5.35: אות בתוספת רעש עוברים במערכת לינארית

ולכן מאיור 5.36:

$$\mathbb{E}[\varepsilon^2(t)] = \int_{-\infty}^{\infty} S_W(f)|H(f)|^2 df + \int_{-\infty}^{\infty} S_X(f)|1 - H(f)|^2 df$$



איור 5.36: הצגה שקולה של אות בתוספת רעש העוברים במערכת לינארית

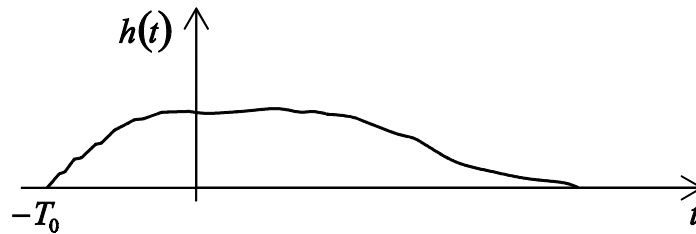
קיבלנו ביטוי מפורש עבור השגיאה הריבועית הממוצעת הכוללת. עכשיו נוכל להשוות הצעות שונות ל- $H(f)$. למשל, עבור $H(f) = (1 + if/f_0)^{-1}$ (דהיינו, מסננת R-C) נוכל לחשב את השגיאה כפונקציה של הפרמטר f_0 , לבצע אופטימיזציה ולמצוא את הפרמטר f_0 הטוב ביותר שיגרום לשגיאה ריבועית ממוצעת מינימלית עבור מסנן כזה.

גישה נועזת יותר היא לשאול מה הוא $h(t)$ האופטימלי שיביא את השגיאה הריבועית הממוצעת למינימום. במקרה זה יש להבחין בין שני מקרים של מציאת $h(t)$ האופטימלי:

א. אין הגבלות על $h(t)$, דהיינו, אין דרישה ש- $h(t)$ יהיה סיבתי.

ב. קיימת מגבלה ש- $h(t)$ יהיה סיבתי.

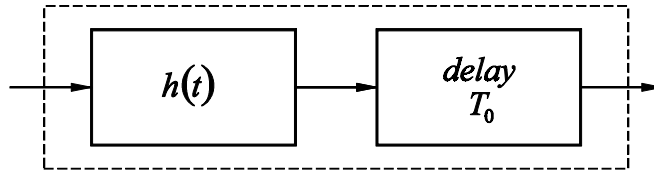
אם הפתרון הוא סיבתי אזי ניתן לממש אותו (או קרוב שלו) אם באופן דיגיטלי ואם באופן אנלוגי. בעית הסינון האופטימלי עם מגבלת הסיבתיות נפתרה ע"י Wiener⁹ וידועה כמסננת וינר אולם הפתרון הוא קשה ולא נביא אותו בקורס זה. הפתרון ללא מגבלת הסיבתיות הוא קל יחסית ונתרכז בו בהמשך. השאלה הנשאלת מיד היא: האם לפתרון ללא מגבלת הסיבתיות יש מובן פיזיקלי, דהיינו, האם ניתן לממש אותו (או קרוב טוב שלו) באופן אנלוגי או דיגיטלי? אם קבלנו פתרון מהצורה המופיעה באיור 5.37 כאשר עבור $h(t) = 0, t < -T_0$ (או זניח) אזי $h(t)$ אינה סיבתית אולם $h(t - T_0)$ סיבתית. כלומר את



איור 5.37: תגובת הלם לא סבתית עם חלק שלילי סופי

המערכת 5.38 נוכל לממש. לכן, בבעיות בהן אפשר לסבול השהייה (כגון בעיות תקשורת בניגוד לבעיות בקרה) יש לפתרון ללא סיבתיות משמעות פיזיקלית. לפתרון זה קוראים "מסננת וינר עם השהייה אינסופית".

⁹נוברט וינר (Norbert Wiener), 1894–1964, היה מתמטיקאי ופילוסוף אמריקאי. הוא היה פרופסור באוניברסיטת MIT ונחשב לאבי הקיברנטיקה, ושל הקונספט של משוב במערכות. מסנן וינר פותח על ידו בשנת 1949.



$$h(t-T_0)$$

איור 5.38: קרוב מעשי למערכת לא סבתית על ידי השהיה

נחזור לבעיית מציאת המסננת האופטימלית המביאה את השגיאה הריבועית הממוצעת למינימום. השגיאה הריבועית הממוצעת ניתנת כאמור ע"י:

$$\mathbb{E}[\varepsilon^2(t)] = \int_{-\infty}^{\infty} S_W(f)|H(f)|^2 df + \int_{-\infty}^{\infty} S_X(f)|1-H(f)|^2 df$$

והבעיה—מצא מסננת $H(f)$ כך ש- $\mathbb{E}[\varepsilon^2(t)]$ יהיה מינימום על פני כל המסננות $H(f)$.

הבעיה הכללית של מציאת פונקציה המביאה אינטגרל למינימום היא מורכבת, ופתרון כללי מצוי בתחום הנקרא חשבון ווריאציות. למזלנו, הבעיה שלפנינו פשוטה יחסית, ונפתור אותה בשלושה שלבים.

ראשית, נרשום את הביטוי לשגיאה כאינטגרל אחד:

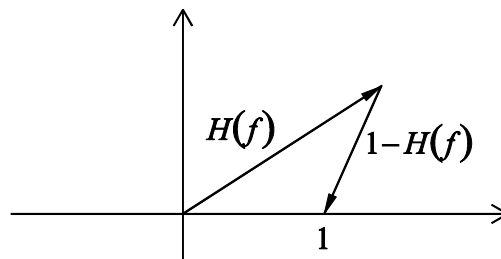
$$\mathbb{E}[\varepsilon^2(t)] = \int_{-\infty}^{\infty} [S_W(f)|H(f)|^2 + S_X(f)|1-H(f)|^2] df \quad (5.25)$$

ונשים לב כי כל הביטויים תחת האינטגרל חיוביים. אם כך, אנו מחפשים $H(f)$ כך שהשטח מתחת לפונקציה

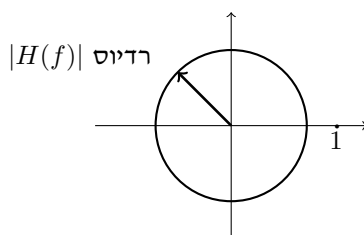
$$S_W(f)|H(f)|^2 + S_X(f)|1-H(f)|^2$$

יהיה מינימלי. אולם כדי להקטין שטח זה עלינו להקטין את הפונקציה, בכל ערך של f בנפרד—אין בבעיה שלפנינו אילוצים הקושרים את ערכי $H(f)$ בתדרים שונים. כלומר, פישטנו את הבעיה ממצאת מינימום של אינטגרל למציאת מינימום של פונקציה.

שלב שני: עבור f מסוים נעניין ב- $|1-H(f)|$. נשרטט זאת באיור 5.39.



איור 5.39: בחירת הפאזה של H



איור 5.40: הנקודה הקרובה ביותר ל-1 על פני המעגל ברדיוס $|H(f)|$ היא הנקודה הממשית $|H(f)|$. זה נכון גם אם רדיוס המעגל גדול מ-1.

שים לב שבין כל הנקודות על המעגל (סביב הראשית) $|H(f)| = \text{const}$, הקרובה ביותר לנקודה 1 היא הנקודה (הממשית) $|H(f)|$ (איור 5.40). לכן אם עבור אותו f נחליף את $H(f)$ ב- $|H(f)|$ נגרום להקטנת $|1 - H(f)|$. באופן אלגברי:

$$1 = H(f) + 1 - H(f) \leq |H(f)| + |1 - H(f)| \Rightarrow |1 - H(f)| \geq 1 - |H(f)|$$

$$|H(f)| = |H(f) - 1 + 1| \leq |H(f) - 1| + 1 \Rightarrow |1 - H(f)| \geq |H(f)| - 1$$

ולכן

$$|1 - H(f)| \geq |1 - |H(f)||$$

$$|1 - H(f)|^2 \geq |1 - |H(f)||^2$$

עם שוויון כאשר $H(f)$ ממשי ואי-שלילי. ע"י החלפה זו לא נשנה את הביטוי $|H(f)|^2 S_W(f)$ אבל נקטין את האיבר השני $|1 - H(f)|^2 S_X(f)$, ולכן מותר להגביל את החיפוש לפונקציות $H(f)$ שהן ממשיות ולא שליליות עבור כל f .

שלב שלישי: הבעיה מצטמצמת, לכן, למציאת $H(f)$ ממשית ולא שלילית שעבורה הביטוי

$$S_W(f) [H(f)]^2 + S_X(f) [1 - H(f)]^2$$

הוא מינימלי. נסמן (עבור f נתון) ב- $a = H(f)$ את אותו $H(f)$ עבורו האינטגרנד מינימלי. נמצא את המינימום על ידי

$$\frac{\partial}{\partial a} \{ S_W(f) a^2 + S_X(f) (1 - a)^2 \} = 0$$

ונקבל

$$a = \frac{S_X(f)}{S_X(f) + S_W(f)}$$

ולכן

$$H_{\text{opt}}(f) = \frac{S_X(f)}{S_X(f) + S_W(f)} \quad (5.26)$$

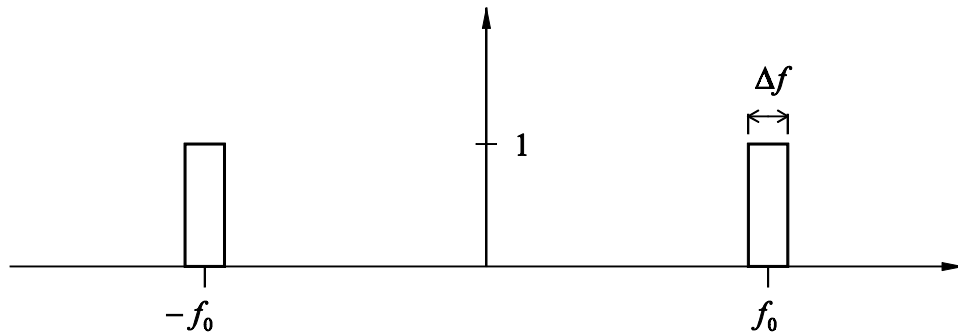
נשים לב כי זהו מסנן לא סיבתי. ניתן לממשו באופן מעשי ע"י השהיה, ויש גם פתרון סיבתי הנלמד בקורס אחר.

השגיאה המינימלית מתקבלת ע"י הצגת $H_{opt}(f)$ לתוך נוסחת השגיאה הרבועית הממוצעת:

$$\begin{aligned} \mathbb{E}[\varepsilon_{\min}^2(t)] &= \int_{-\infty}^{\infty} \left(\frac{S_W(f)S_X^2(f)}{(S_X(f) + S_W(f))^2} + \frac{S_X(f)S_W^2(f)}{(S_X(f) + S_W(f))^2} \right) df \\ &= \int_{-\infty}^{\infty} \frac{S_X(f)S_W(f)}{S_X(f) + S_W(f)} df \end{aligned}$$

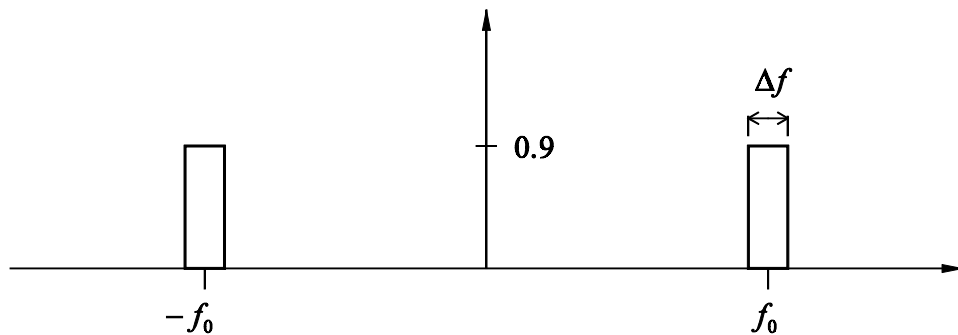
נשים לב כי מאיורים 5.35-5.36. ברור שהבחירה של ערך של a היא פשרה בין $a = 0$ הגורם לניחות מוחלט של הרעש אך גם מאפס את האות הרצוי, לבין $a = 1$ אשר אינו מעוות את האות כלל אך גם אינו מנחית את הרעש. השווה לדוגמה של שערך לינארי (גאוסני אופטימלי שהוא לינארי) של משתנה אקראי כל סמך מדידה רועשת שלו—ראה (2.5) והמקדם האופטימלי (2.8). על הבעיה שפתרנו כאן אפשר לחשוב כאילו פרקנו את האות הרועש Y לפסי תדר. כיוון שבין פסי התדר אין קורלציה, פתרנו עבור כל פס לחוד, ואז הבעיה היא של שערך של מ"א.

דוגמה: יהיה $X(t)$ תהליך אקראי בעל צפיפות הספק כנתון באיור 5.41. אליו מתווסף רעש $W(t)$ בעל צפיפות הספק



איור 5.41: תהליך צר סרט סביב תדר f_0

המסננת האופטימלית $H_{opt}(f)$ לשחזור $X(t)$ מתוך $Y(t) = X(t) + W(t)$, נתונה ע"י (5.26), ונראית כנתון באיור 5.42.

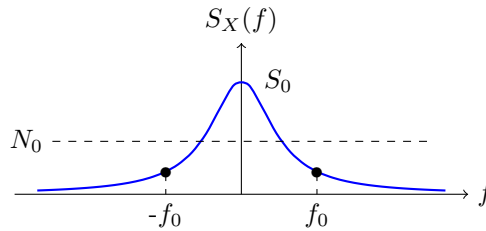


איור 5.42: מסנן אופטימלי

שים לב ש- $H_{opt}(f)$ אינה מעבירה אותות באותם תדרים בהם תכולת התדר של האות $X(t)$ אינה קיימת. מדוע יש הנחתה של 0.9 בתדרים האחרים? (ראה דוגמה ג' בסעיף 2.2).

דוגמה 5.10 בהנתן אות $Y_t = X_t + W_t$ כאשר $\{X_t\}$ ו- $\{W_t\}$ סמ"ר, חסרי קורלציה, בעלי תוחלת אפס וצפיפויות ההספק הספקטרלית הבאות (איור 5.43):

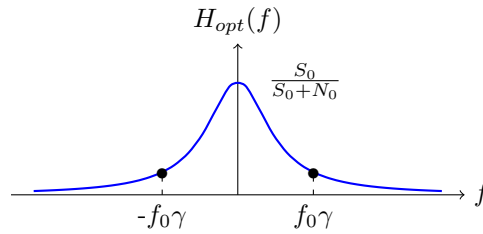
$$S_X(f) = \frac{S_0}{1 + (f/f_0)^2}; \quad S_W(f) = N_0$$



איור 5.43: $S_X(f)$ ו- $S_W(f) = N_0$

המסנן האופטימלי נתון ע"י (איור 5.44):

$$\begin{aligned} H_{opt}(f) &= \frac{S_X(f)}{S_X(f) + S_W(f)} = \frac{\frac{S_0}{1+(f/f_0)^2}}{\frac{S_0}{1+(f/f_0)^2} + N_0} \\ &= \frac{S_0}{S_0 + N_0} \cdot \frac{1}{1 + \left(\frac{f}{f_0 \cdot \gamma}\right)^2}; \quad \gamma \triangleq \sqrt{1 + \frac{S_0}{N_0}} \end{aligned}$$



איור 5.44: $H_{opt}(f)$

כאשר $\frac{S_0}{N_0} \rightarrow \infty$, רוחב הסרט גדל והולך, והניחות שואף ליחידה:

$$H_{opt}(f) = \frac{S_0}{S_0 + N_0} \rightarrow 1$$

כאשר $\frac{S_0}{N_0} \rightarrow 0$, $H_{opt}(f)$ שואף לאפס. השגיאה נתונה על ידי:

$$\begin{aligned} \mathbb{E}[\varepsilon^2(t)] &= \int_{-\infty}^{\infty} \frac{S_X(f) S_W(f)}{S_X(f) + S_W(f)} df = \int_{-\infty}^{\infty} H_{opt}(f) S_W(f) df \\ &= \frac{S_0 N_0}{S_0 + N_0} \int_{-\infty}^{\infty} \frac{1}{1 + \left(\frac{f}{f_0 \gamma}\right)^2} df = \frac{\pi S_0 N_0 f_0 \gamma}{S_0 + N_0} = \frac{\pi S_0 f_0}{\gamma} \end{aligned}$$

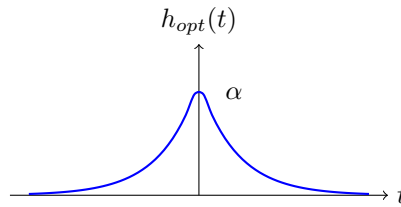
נחשב את יחס האות לרעש במוצא המסנן:

$$\text{SNR}_{out} = \frac{\mathbb{E}[X^2(t)]}{\mathbb{E}[\varepsilon^2(t)]} = \frac{\int_{-\infty}^{\infty} S_X(f) df}{\pi f_0 S_0 / \gamma} = \frac{\pi S_0 f_0}{\pi S_0 f_0 / \gamma} = \gamma$$

כלומר, הפרמטר γ מקבל את המובן של SNR במוצא המסנן.

בתחום הזמן (איור 5.45):

$$h_{opt}(t) = \mathcal{F}^{-1}\{H_{opt}(f)\} = \underbrace{\frac{S_0}{S_0 + N_0}}_{\alpha} \cdot \pi f_0 \gamma e^{-2\pi f_0 \gamma |t|}$$

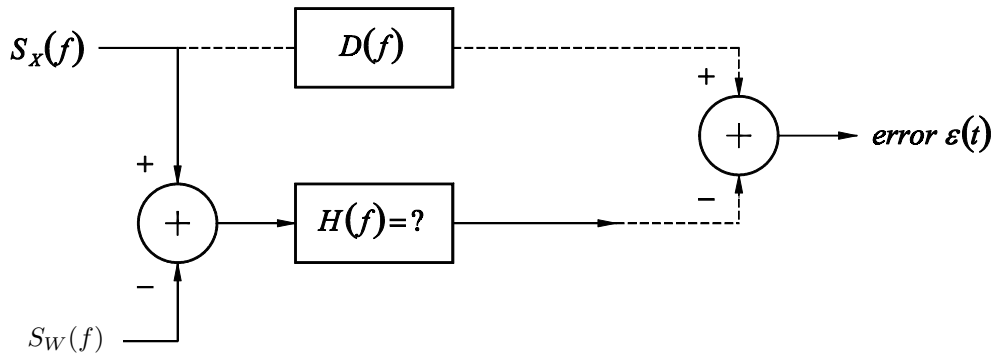


איור 5.45: $h_{opt}(t)$

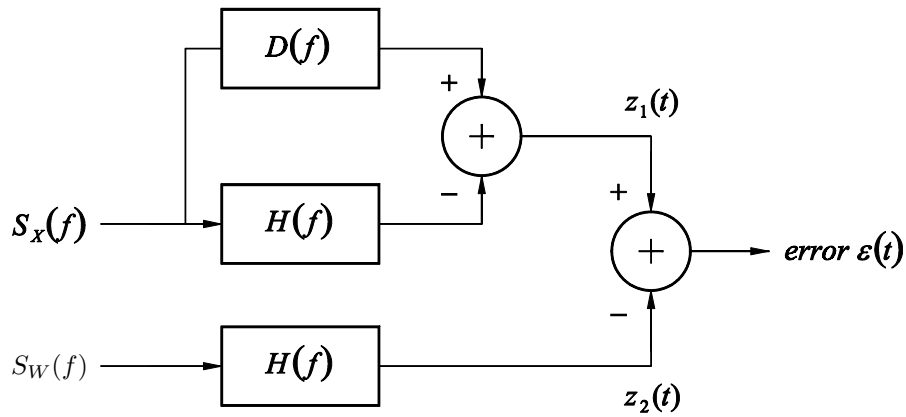
כאשר $\frac{S_0}{N_0} \rightarrow \infty$ ולכן גם $\gamma \rightarrow \infty$, $h_{opt}(t)$ מתקרב והולך ל- $\delta(t)$, כלומר – מסנן היחידה. זו תגובה לא סיבתית שאפשר לקרבה עם השהיה מתאימה, $T \gg (f_0 \gamma)^{-1}$, עם קיצוץ הזנב הלא-סיבתי. שימו לב: היות ו- $H_{opt}(f)$ לעולם ממשי וסימטרי, גם $h_{opt}(t)$ יהיה כזה, ולכן בעית הסיבתיות תהיה קיימת בדרך-כלל.

ניתן להרחיב במקצת את בעיית מציאת המסננת האופטימלית. נמשיך להניח $\mathbb{E}[W(t)] = \mathbb{E}[X(t)] = 0$ וכן ש- $X(t_1), W(t_2)$ חסרי קורלציה. ההרחבה תהיה בכך שבמקום שהאות הרצוי יהיה $X(t)$ נחפש כאות רצוי את $Y(t)$ כאשר $Y(t)$ הוא היציאה של מערכת לינארית עם תגובה להלם $d(t)$ כאשר הכניסה היא $X(t)$. לדוגמה, $Y(t) = dX(t)/dt$ כאשר $d(t)$ הוא גזור. מוטיבציה אפשרית לשימוש בהרחבה זו היא כאשר אנו מעוניינים מראש רק בתחום תדר מסוים בתהליך המסונן; אם, למשל, בידינו מדידה של אות שמע שבו אנו מעוניינים לשערך רק את התדרים הנמוכים, המסנן $D(f)$ יהיה מסנן מעביר-נמוכים (LPF).

המערכת מתוארת (במרחב התדר) באיור 5.46. כאשר $D(f) = F\{d(t)\} = 1$ נקבל את המקרה הקודם. ע"י סופרפוזיציה



איור 5.46: אות תגובה מעובד ומסונן



איור 5.47: אות תגובה מעובד ומסונן

נוכל להחליף את איור 5.46 באיור 5.47 ולכן

$$\mathbb{E}[\varepsilon^2(t)] = \int_{-\infty}^{\infty} S_W(f)|H(f)|^2 df + \int_{-\infty}^{\infty} S_X(f)|D(f) - H(f)|^2 df$$

את המשואה לעיל נרשום בצורה הבאה:

$$\mathbb{E}[\varepsilon^2(t)] = \int_{-\infty}^{\infty} \left| \frac{H(f)}{D(f)} \right|^2 S_W(f)|D(f)|^2 df + \int_{-\infty}^{\infty} S_X(f)|D(f)|^2 \left| 1 - \frac{H(f)}{D(f)} \right|^2 df$$

אם במקום לחפש $H(f)$ אופטימלי נחפש $(H(f)/D(f))$ אופטימלי (ואח"כ נכפיל את הפתרון ב- $D(f)$) אזי יש לנו בדיוק אותה בעיה כמו במקרה $D(f) = 1$. ולכן:

$$\left(\frac{H(f)}{D(f)} \right)_{\text{opt}} = \frac{S_X(f)|D(f)|^2}{|D(f)|^2[S_X(f) + S_W(f)]} = \frac{S_X(f)}{S_X(f) + S_W(f)}$$

ולכן

$$H_{\text{opt}}(f) = \frac{D(f)S_X(f)}{S_X(f) + S_W(f)} \quad (5.27)$$

והשגיאה הריבועית הממוצעת המינימלית המתקבלת היא:

$$\mathbb{E}[\varepsilon_{\min}^2(t)] = \int_{-\infty}^{\infty} \frac{|D(f)|^2 S_X(f) S_W(f)}{S_X(f) + S_W(f)} df$$

עקרון האורתוגונליות

לסיכום סעיף זה על סינון לינארי אופטימלי, נחשב בחשבון ישיר וקצר את המסנן האופטימלי (כולל המקרה של קורלציה בין האות לרעש) בעזרת עקרון האורתוגונליות.

התהליכים $\{X(t)\}$ ו- $\{Y(t)\}$ סמ"ר במשותף, עם ספקטרא $S_X(f)$ ו- $S_Y(f)$ בהתאמה, ועם קרוס-ספקטרום (צפיפות הספק ספקטרלית מצטלבת) $S_{XY}(f)$. נסמן ב- $d(t)$ את התמרת פוריה ההפוכה של $D(f)$ אזי

$$\mathbb{E}[\varepsilon^2(t)] = \mathbb{E} \left[\left(X(t) - \int_{-\infty}^{\infty} Y(t-\theta) h(\theta) d\theta \right)^2 \right].$$

ע"י עקרון האורתוגונליות נקבל שהמסננת האופטימלית חייבת לקיים:

$$\mathbb{E} \left[\left(X(t) - \int_{-\infty}^{\infty} Y(t-\theta) h_{\text{opt}}(\theta) d\theta \right) Y(\eta) \right] = 0$$

עבור כל $\eta, -\infty < \eta < \infty$ ולכן:

$$R_{Y,X}(t-\eta) = \int_{-\infty}^{\infty} R_Y(t-\theta-\eta) h_{\text{opt}}(\theta) d\theta$$

$$R_{Y,X}(\tau) = \int_{-\infty}^{\infty} R_Y(\tau-\theta) h_{\text{opt}}(\theta) d\theta = (R_Y * h_{\text{opt}})(\tau)$$

כאשר הצבנו $\tau = t - \eta$. כעצ נבצע התמרת פוריה על שני אגפי המשוואה ונקבל:

$$S_{Y,X}(f) = S_Y(f) H_{\text{opt}}(f)$$

ולכן:

$$H_{\text{opt}}(f) = \frac{S_{Y,X}(f)}{S_Y(f)}$$

שימו לב כי כאן לא השתמשנו למעשה בהנחה הקודמת, ש- $Y(t) = X(t) + W(t)$ ו- $\{X(t)\}$ ו- $\{W(t)\}$ חסרי קורלציה. ההנחה היחידה כאן היתה שהתהליכים $\{X(t)\}$ ו- $\{Y(t)\}$ סמ"ר במשותף. עבור המקרה הפרטי שבו $Y(t) = X(t) + W(t)$, והתהליכים $\{X(t)\}$ ו- $\{W(t)\}$ חסרי קורלציה, נקבל:

$$R_{YX}(\tau) = \mathbb{E}[Y(t)X(t+\tau)] = \mathbb{E}\{[X(t) + W(t)]X(t+\tau)\}$$

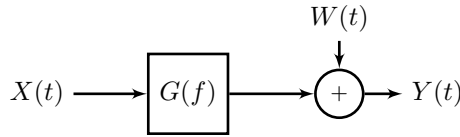
$$= \mathbb{E}[X(t)X(t+\tau)] + \mathbb{E}[W(t)X(t+\tau)] = R_X(\tau) + 0$$

ולכן $S_{YX}(f) = S_X(f)$ בנוסף,

$$\begin{aligned} R_Y(\tau) &= \mathbb{E}[Y(t)Y(t+\tau)] = \mathbb{E}\{[X(t) + W(t)][X(t+\tau) + W(t+\tau)]\} \\ &= \mathbb{E}[X(t)X(t+\tau)] + \mathbb{E}[W(t)W(t+\tau)] + 0 + 0 = R_X(\tau) + R_W(\tau) \end{aligned}$$

ולכן $S_Y(f) = S_X(f) + S_W(f)$. נעיר כי, ככלל, הספקטרום של סכום שני ת"א חסרי קורלציה היא סכום הספקטרא. דבר זה נובע מהחישוב שעשינו כעת שבו חישבנו את הספקטרום של שני תהליכים חסרי קורלציה כלליים.

דוגמה 5.11 נרחיב מעט את המקרה הפרטי הנ"ל. נניח שנמדובר במצב המתואר באיור 5.48. זוהי בעיה עם חשיבות מעשית



איור 5.48:

גדולה. נתעניין בסינון של $X(t)$, ובמקרה זה:

$$S_Y(f) = S_X(f) \cdot |G(f)|^2 + S_W(f)$$

$$S_{YX}(f) = S_X(f) \cdot G^*(f)$$

ולכן

$$H_{opt}(f) = \frac{S_{YX}(f)}{S_Y(f)} = \frac{S_X(f) \cdot G^*(f)}{S_X \cdot |G(f)|^2 + S_W(f)}$$

כך שבהיעדר רעש, $S_W(f) \equiv 0$, מקבלים, כצפוי, $H_{opt}(f) = \frac{1}{G(f)}$. התוספת של $S_W(f)$ במכנה "מרסנת" הגברים אפשריים גבוהים של הרעש, היכן שממילא $G(f) \approx 0$, ואין לנו גישה לאות המקורי.

היתרון שיש לאנליזה הקודמת שעשינו היא שקל באמצעותה לקחת בחשבון אילווצים בתחום התדר, אם יש. למשל, אם $H(f)$ מוגבל להיות מסיבות כלשהן בעל תמך מסוים, $[-W, W]$, אז עושים אופטימיזציה לאנטגרנד $S_\epsilon(f)$ רק בתחום הרלוונטי, ובשאר התחום אין לנו שליטה. דוגמא נוספת היא כאשר רוצים של- $H(f)$ יהיה מבנה מסוים וקל לביטוי בתחום התדר.

השיטה של עקרון האורתוגונליות, לעומת זאת, היא כללית יותר, ומתאימה לאילווצים בתחום הזמן, כגון סיבתיות; במקרה הסיבתי משתמשים בעקרון האורתוגונליות בו השתמשנו קודם רק עבור $t \leq \eta$, ולא לכל η . זה הופך את הבעיה לקשה יותר!

הערה מסכמת לגבי ת"א בזמן בדיד: לנושאים בהם עסקנו בפגישות האחרונות – מעבר ת"א סמ"ר דרך מערכות לינאריות קבועות בזמן, וכן סינון אופטימלי, יש אנלוגיות ישירות בעולם של ת"א בזמן בדיד. במקרה זה אנטגרלים הופכים לסכומים, קונוולוציות רציפות הופכות לקונוולוציות דיסקרטיות, התמרות פוריה רציפות הופכות להתמרות פוריה בזמן בדיד, וכו'. כך, למשל, אם $\{X(t)\}$ ת"א סמ"ר בזמן בדיד עם סדרת אוטוקורלציה $R_X(k)$ והוא מזין מערכת עם תגובה לדגם יחידה $\{h(n)\}$,

אזי במוצא נקבל $R_Y(k) = (R_X * h * h(\cdot))(k)$ ובמישור התדר

$$S_X(e^{j\theta}) \triangleq \mathcal{F}\{R_X(n)\} = \sum_{k=-\infty}^{\infty} R_X(k)e^{-j\theta k}$$

$$S_Y(e^{j\theta}) = S_X(e^{j\theta}) \cdot |H(e^{j\theta})|^2$$

$$S_{XY}(e^{j\theta}) = S_X(e^{j\theta}) \cdot H(e^{j\theta})$$

וכו'. סינון לינארי אופטימלי במקרה הכללי נתון ע"י המסנן: $H_{opt}(e^{j\theta}) = \frac{S_{YX}(e^{j\theta})}{S_Y(e^{j\theta})}$

5.7 כמה מילים על ארגודיות

תהיה $\{X_1(\omega), X_2(\omega), \dots\}$ סדרה של מ"א בלתי תלויים ובעלי פילוג זהה. במקרה זה, חוק המספרים הגדולים אומר שעבור כל פונקציה $f(x)$, $-\infty < x < \infty$ שעבורה $\mathbb{E}|f(X_1)| < \infty$ מתקיים

$$\frac{1}{N} \sum_{i=1}^N f(X_i(\omega)) \xrightarrow{N \rightarrow \infty} \mathbb{E}[f(X_1)] \quad (5.28)$$

בכדי לתת לתוצאה זו מובן יש צורך להגדיר מה אנחנו מבינים כאשר אומרים שסדרה של משתנים אקראיים מתכנסת. נבטא זאת בצורה הבאה.

נניח ש- $\mathbb{E}[|X_1(\omega)|^2] < \infty$ ונרשום $X_i(\omega) = \mathbb{E}[X_i(\omega)] + Y_i(\omega)$ אזי ה- $Y_i(\omega)$ בעלי תוחלת אפס, ב"ת ובעלי פילוג זהה.

טענה:

$$\mathbb{E} \left(\frac{1}{N} \sum_1^N X_i(\omega) - \mathbb{E}[X_1] \right)^2 \xrightarrow{N \rightarrow \infty} 0$$

הוכחה: צ"ל

$$\mathbb{E} \left(\frac{1}{N} \sum_1^N Y_i \right)^2 \xrightarrow{N \rightarrow \infty} 0$$

אולם

$$\mathbb{E} \left(\frac{1}{N} \sum_1^N Y_i \right)^2 = \frac{1}{N^2} \sum_1^N \mathbb{E}[Y_i^2] = \frac{1}{N} \mathbb{E}[Y_1^2] \rightarrow 0$$

מכאן גם נובע (ראה אי שוויון צ'ביצ'ב):

$$\mathbb{P} \left\{ \frac{1}{N} \sum_{i=1}^N |X_i - \mathbb{E}[X_1]| > \delta \right\} \leq \frac{\mathbb{E}[Y_1^2]}{N\delta^2}$$

ולכן הביטוי הולך ל-0 כאשר $N \rightarrow \infty$.

שים לב שאם במקום לטפל ב- X_i נטפל ב- $f(X_i)$, ונדרוש $\mathbb{E}[|f(X_1)|^2] < \infty$, נקבל:

$$\frac{1}{N} \sum_{i=1}^N (\mathbb{E}[f(X_i)] - \mathbb{E}[f(X_1)]) \rightarrow 0$$

תוצאה זאת מהווה "פרוש" אפשרי ל- (5.28).

מתבקשת השאלה אם (5.28) נשאר נכון כאשר יש תלות בין האיברים השונים בסדרה האקראית.

הגדרה: יהיה $\{X(t), -\infty < t < \infty\}$ ת"א סטציונרי. התהליך נקרא ארגודי אם עבור כל K , כל פונקציה (חסומה) f של K משתנים וכל סדרה t_1, t_2, \dots, t_k מתקיים (בהסתברות 1):

$$\frac{1}{2T} \int_{-T}^T f(X(t_1 + s, \omega), \dots, X(t_k + s, \omega)) ds \xrightarrow{T \rightarrow \infty} \mathbb{E} f(X(t_1), \dots, X(t_k))$$

שוב אנו מתעלמים מהשאלה באיזה מובן מתכנסים המ"א שבאגף שמאל למשתנה האקראי המנוון בצד ימין (משפט ידוע וקשה - המשפט הארגודי - אומר שלאגף שמאל יש אכן גבול אבל המשפט לא אומר מתי הגבול אכן שווה לתוחלת). את התכונה הארגודית נבטא ע"י הסיסמה: עבור תהליכים ארגודיים, ממוצע הזמן שווה לממוצע האנסמבל (=התוחלת).

דוגמאות לתהליכים לא ארגודיים:

1. משתנה אקראי $X_t(\omega) = C(\omega)$

2. $X_t(\omega) = A(\omega) \cos(2\pi f_0 t + \phi(\omega))$ כאשר ϕ בלתי תלוי ב- A ומפולג אחיד בתחום $[0, 2\pi]$. אם $A(\omega)$ מ"א לא מנוון אזי

$$\frac{1}{2T} \int_{-T}^T X_t^2 dt \rightarrow \frac{A^2(\omega)}{2} \neq \mathbb{E} \left[\frac{A^2(\omega)}{2} \right]$$

אפשר להראות שעבור $A(\omega) \equiv \text{const}$ התהליך אכן ארגודי.

3. יהיו

$$X_t(\omega) = \cos(2\pi f_0 t + \phi)$$

$$Y_t(\omega) = \cos(2\pi f_0 t + \psi)$$

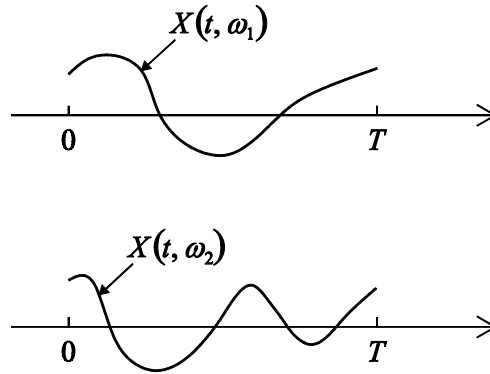
כאשר ϕ, ψ בת"ס ומפולגים אחיד בתחום $[0, 2\pi]$.

נגדיר

$$Z_t(\omega) \triangleq X_t(\omega) Y_t(\omega)$$

נקל לראות ש- $Z_t(\omega)$ איננו ת"א ארגודי.

4. תערובת של שני תהליכים שונים שכל אחד ארגודי לא חייבת להיות ארגודית.



איור 5.49: שתי פונקציות מדגם

בהרגשה ברור (פרט אולי למקרים מנוונים) שעל מנת שת"א יהיה ארגודי צריך לדרוש שבמהלך ההתפתחות בזמן של כל דגם, הוא יקבל את כל צורות הגל האפשריות. דהיינו: אם נמדדים אותות כמתואר באיור 5.49, אזי אם נסתכל על הדגם עם ω_1 עבור זמן מספיק ארוך, נמצא $(t_0, t_0 + T)$ שעליו התהליך דומה מאוד לתהליך $X(t, \omega_2)$ בקטע $(0, T)$.

כאמור, הבעיה לקבוע אם ת"א נתון הוא ארגודי היא בעיה קשה. נסתפק לכן בפתרון חלקי (מאד) של בעיית הארגודיות:

הגדרה: ת"א $\{X(t), -\infty < t < \infty\}$ סטציונרי (או אפילו סטציונרי במובן הרחב) ייקרא ארגודי לגבי הממוצע אם

$$\mathbb{E} \left(\frac{1}{2T} \int_{-T}^T X(t) dt - \mu \right)^2 \xrightarrow{T \rightarrow \infty} 0$$

כאשר $\mathbb{E}[X(t)] = \mu$, דהיינו, אם

$$\mathbb{E} \left(\frac{1}{2T} \int_{-T}^T (X(t) - \mu) dt \right)^2 \xrightarrow{T \rightarrow \infty} 0.$$

משפט: אם הקוריאנס

$$\mathbf{K}_X(t_1, t_2) = \mathbb{E}[(X(t_1) - \mu)(X(t_2) - \mu)]$$

מקיים:

$$\frac{1}{4T^2} \int_{-T}^T \int_{-T}^T \mathbf{K}_X(t_1, t_2) dt_1 dt_2 \xrightarrow{T \rightarrow \infty} 0$$

אזי התהליך ארגודי לגבי הממוצע.

הוכחה: מידית.

מתוך משפט זה נוכל לקבל את התוצאה הבאה (לא ניתן כאן את ההוכחה, ראה הערה בהמשך): אם $X(t)$ סטציונרי במובן

הרחב ו- $\mathbf{K}_X(t_1, t_1 + \tau) = \psi(\tau)$ ואם $\psi(\tau)$ מקיימת

$$\frac{1}{2T} \int_{-2T}^{2T} \psi(\tau) \left(1 - \frac{|\tau|}{2T}\right) d\tau \xrightarrow{T \rightarrow \infty} 0 \quad (5.29)$$

אזי התהליך ארגודי לגבי הממוצע. ההוכחה מתבססת על הקשר

$$\int_{-T}^T \int_{-T}^T Q(t_1 - t_2) dt_1 dt_2 = \int_{-2T}^{2T} Q(\theta)(2T - |\theta|) d\theta$$

הערה: ההוכחה לקשר זה מצויה סביב איור 5.28.

ברוח ההגדרה של ארגודיות לגבי הממוצע נגדיר גם:

הגדרה: יהיה $X(t)$ ת"א סטציונרי. נסמן

$$z_\lambda(t) = X(t + \lambda)X(t)$$

אם עבור כל λ , $z_\lambda(t)$ ארגודי לגבי הממוצע אזי נגיד ש- $X(t)$ ארגודי לגבי הקורלציה. לכן, על מנת ש- $X(t)$ יהיה ארגודי לגבי הקורלציה צריך להתקיים

$$\mathbb{E} \left(\frac{1}{2T} \int_{-T}^T X(t + \tau)X(t) dt - R_X(\tau) \right)^2 \xrightarrow{T \rightarrow \infty} 0$$

נסמן ב- $R_Z(\tau)$ את פונקציית האוטוקורלציה של $z_\lambda(t)$:

$$R_Z(\tau) = \mathbb{E} \left[X(t + \tau + \lambda)X(t + \tau)X(t + \lambda)X(t) \right]$$

אזי

$$\psi(\tau) = R_Z(\tau) - R_X^2(\lambda)$$

צריך לקיים את התנאי (5.29) לעיל.

הערה: על מנת לדעת את $\psi(\tau)$ עלינו לדעת לא רק את המומנטים מסדר שני של $X(t)$ אלא גם את המומנטים מסדר רביעי. במקרה הגאוסני יש קשר בין השניים: מתקיים (כאשר $X(t)$ גאוסני עם תוחלת אפס)

$$\psi(\tau) = R_X(\lambda + \tau)R_X(\lambda - \tau) + R_X^2(\tau)$$

ולכן במקרה זה אם $R_X(\tau)$ יורד לאפס מספיק מהר אזי התהליך ארגודי לגבי הקורלציה.

משמעות הארגודיות לגבי מדידות

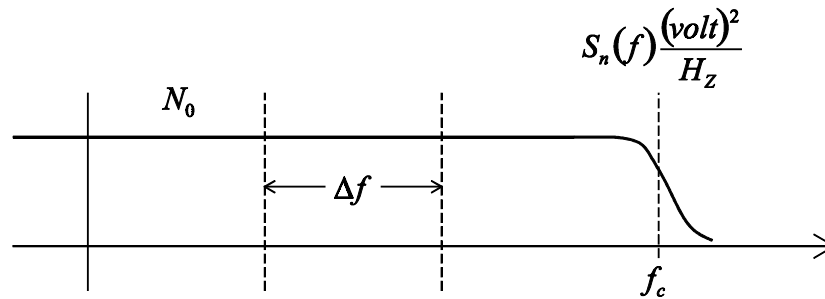
נתונה מערכת כגון מקלט המיוצר בכמויות. על מנת לבדוק את תגובת המקלטים לאותות המשודרים אליהם בתנאי רעש חיצוני, אנו בונים מערכת מדידה המכילה מקור המייצר את האותות המשודרים. המקור מורכב מגנרטור אות וגנרטור רעש. למערכת זו נחבר את אחד המקלטים המיוצרים ונמדוד את התנהגות המערכת ונוכל מתוך המדידות לקבל תוצאות עבור טיבו של אותו מקלט. אם נחזור על המדידות ונמצע (בצורה נאותה) על פני מספר גדול של מקלטים נקבל הערכה על טיב אותו סוג מקלטים. בכל אותן מדידות, בין אם על מקלט בודד ובין אם על מספר גדול של מקלטים, נשאר (במציאות) אותו גנרטור רעש. בעצם אולי היינו צריכים לבדוק כל מקלט על מספר גנרטורי רעש ולמצע אולם לא עושים זאת. מדוע? התשובה היא שאנו מניחים שהרעש הנוצר מגנרטור הרעש הוא ארגודי ולכן אפשר להסתפק בגנרטור רעש אחד היות וכפי שהוסבר כבר, בהרגשה, על מנת שתהליך יהיה ארגודי כל דגם בודד חייב (בהסתברות 1) לקבל את כל צורות הגל האפשריות שהתהליך יכול לקבל. אותו שיקול תופס גם לגבי האות המשודר. אם המקור המשודר הוא אות אקראי, היינו צריכים לחבר כל מקלט

להרבה גנרטורי אות ולמצע. שוב, בהנחה שאות המקור האקראי הוא ארגודי, נוכל להסתפק בגנרטור אות יחיד. שאות המקור האקראי הוא ארגודי, נוכל להסתפק בגנרטור אות יחיד.

6 רעשים במערכות ובמעגלים חשמליים

6.1 רעש לבן

בבעיות מעשיות עוסקים, רבות במקרה הבא: קיים מקור רעש בעל תוחלת אפס, שהצפיפות הספקטרלית שלו מתחילה בתדרים נמוכים מאוד ונמשכת עד לתדרים גבוהים מאוד ("מקור רחב סרט") כמצויר:



איור 6.1: מקור רחב סרט

רעש זה מועבר דרך מגבר בעל רוחב סרט ΔF והגברת מתח K . המגבר אולי רחב סרט אולם רוחב הסרט שלו קטן בהרבה מזה של הרעש. במילים אחרות, בתחום התדרים שבו המגבר פועל כמגבר, הצפיפות הספקטרלית למעשה קבועה. השאלה היא מה עוצמת הרעש האפקטיבית ביציאה מהמגבר.



איור 6.2: השפעת מגבר על רעש

עוצמת הרעש האפקטיבית ביציאה היא

$$\mathbb{E}[V_{out}^2(t)] = R_{out}(0) = \int_{-W}^W N_0 |H(f)|^2 df = K^2 2W N_0$$

שים לב ש- f_c לא מופיעה ביציאה. מה שחשוב זה N_0 . אם במקום $S_n(f)$ הנתון היינו בונים מודל מקורב עבור הבעיה ומניחים $S_n(f) \equiv N_0$ בכל התדרים היינו מקבלים אותה תוצאה עבור $R_{opt}(0)$. מקור רעש שהצפיפות הספקטרלית שלו היא N_0 בכל התדרים נקרא "רעש לבן", באנלוגיה לאור לבן, המכיל את כל התדרים באותה עצמה.

שים לב שאם

$$S_n(f) = N_0$$

$$R_n(\tau) = N_0 \delta(\tau)$$

פירושו של דבר הוא ש- $n(t)$ חסר קורלציה עם $n(s)$ לכל $t \neq s$. נקבל, אפוא, $R_n(0) \simeq \infty$. בעיה זו מופיעה כבר ב- $S_n(f) = N_0$ כי ההספק נתון ע"י $\int_{-\infty}^{\infty} N_0 df = \infty$.

במלים אחרות: רעש לבן הוא אידיאליזציה של תהליך אקראי פיזיקלי, הוא אינו קיים כתהליך עם הספק ממוצע סופי. יש לראותו כקרוב של תהליך פיזיקלי כפי שתארנו בדוגמה. במציאות לעיתים קרובות לא ידועים הערכים של $S_n(f)$ בתדרים גבוהים ואפילו f_c אינו ידוע (רק ידוע שהוא מעל לתדר מסויים). כל שידוע הוא שבתדרים המעניינים אותנו הצפיפות הספקטרלית קבועה ברמה N_0 (אגב, במקרה הדטרמיניסטי האם אפשר ליחס אנרגיה סופית לפונקציה $\delta(t)$?). לכן במקרים אלה הקרוב נוח מאוד. עבור מקור רעש לבן

$$\text{הספק היציאה} = \int_{-\infty}^{\infty} N_0 |H(f)|^2 df = \int_0^{\infty} 2N_0 |H(f)|^2 df$$

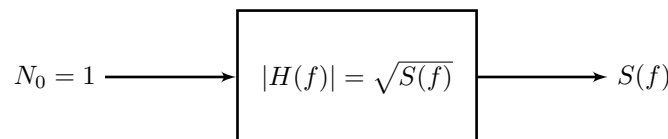
יש המגדירים צפיפות ספקטרלית חד-צדדית:

$$S(f) = \begin{cases} 0, & f < 0 \\ 2S(f), & f > 0 \end{cases}$$

$$\text{הספק היציאה} = N_0 \cdot \Delta f \cdot K^2$$

בחשבונות תיאורטיים נוח להגדיר את $S(f)$ כפי שהגדרנו (צפיפות ספקטרלית דו-צדדית) ואילו בעבודה הנדסית נוח להשתמש בצפיפות ספקטרלית חד-צדדית. בכל מקרה של עיון בספר או מאמר צריך לודא באיזו הגדרה הם משתמשים. אנחנו נמשיך להשתמש בהגדרה הדו-צדדית.

אפשר לראות כל רעש לא לבן "כאילו נוצר ע"י רעש לבן" כדלקמן:



איור 6.3: ייצור רעש עם ספקטרום מוכתב

למשל, אם $S(f) = \frac{1}{1+(2\pi f)^2}$ אזי אפשר לבחור באחת האפשרויות

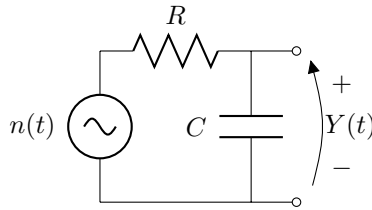
$$H(f) = \begin{cases} \frac{1}{1+i2\pi f} \\ \frac{1}{1-i2\pi f} \\ \frac{1}{\sqrt{1+(2\pi f)^2}} \end{cases}$$

וכל אחת מהן תיתן את $S(f) = \frac{1}{1+(2\pi f)^2}$. שים לב שרק אחת מהשלוש היא סיבתית (איזו?). אפשר להראות (הקריטריון של Paley-Wiener) שאם

$$\int_{-\infty}^{\infty} \frac{|\log S(f)|}{1+f^2} df < \infty$$

אזי קיימת $H(f)$ סיבתית הנותנת ביציאה צפיפות ספקטרלית $S(f)$ כאשר בכניסה רעש לבן. מדוע זה מעניין? בבעיות סימולציה לעתים נתעניין להפיק ת"א עם ספקטרום נתון.

דוגמה 6.1 נתבונן במעגל באיור 6.4. מתקיים:



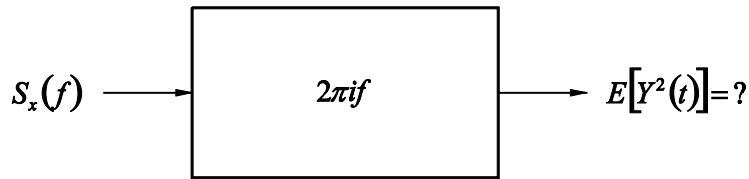
איור 6.4:

$$S_Y(f) = N_0 \left| \frac{\frac{1}{2\pi j f C}}{R + \frac{1}{2\pi j f C}} \right|^2 = \frac{N_0}{1 + (2\pi f RC)^2}$$

$$R_Y(\tau) = \frac{N_0}{2RC} e^{-\frac{|\tau|}{RC}}$$

כלומר, באמצעות מעגל זה ניתן לקבל אוטוקורלציה זהה לזו של תהליך טלגרף אקראי (בדומה לאיור 5.13) אך צורות הגל שונות לחלוטין.

נעיין בגוזר



איור 6.5: גזירה של רעש

נעיין במקרים הבאים עבור $S(f)$

$$S(f) = \begin{cases} N_0 \\ \frac{1}{1+f^2} \\ \frac{1}{1+f^4} \end{cases}$$

אילו ממקורות אלה אפשר לגזור?

מוסר ההשכל הוא שאם מדובר בגוזר הרי שהקרוב של רעש לבן איננו תופס.

אינטגרלים של רעש לבן יהיה $\{n(t), -\infty < t < \infty\}$ רעש עם צפיפות ספקטרלית N_0 ותהינה $h_1(t), h_2(t), \dots$ פונקציות דטרמיניסטיות בעלות אנרגיה סופית $\int_{-\infty}^{\infty} h_i^2(t) dt < \infty, i = 1, 2, \dots$.

נסמן

$$X_i = \int_{-\infty}^{\infty} n(t)h_i(t)dt$$

אזי, ברוח ההגדרה של רעש לבן ($n(t)$ "תהליך בעל רחב סרט גדול מאוד")

$$E[X_i] = 0$$

$$\begin{aligned} E[X_i, X_j] &= E \left[\int_{-\infty}^{\infty} n(t)h_i(t)dt \cdot \int_{-\infty}^{\infty} n(s)h_j(s)ds \right] \\ &= \iint_{-\infty}^{\infty} h_i(t)h_j(s)E[n(t)n(s)]dt ds \\ &= N_0 \iint_{-\infty}^{\infty} h_i(t)h_j(s)\delta(t-s)dt ds \\ &= N_0 \int_{-\infty}^{\infty} h_i(t)h_j(t)dt \end{aligned}$$

ובמיוחד, אם $n(\cdot)$ רעש לבן גאוס (X_1, X_2) וקטור אקראי גאוס עם תוחלת אפס ומטריצת קורלציה

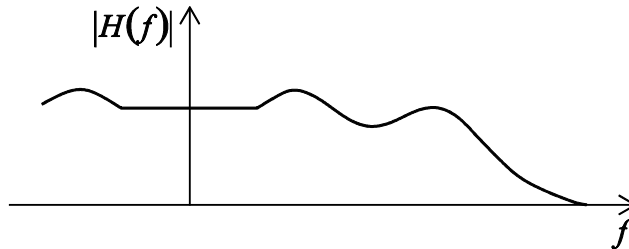
$$\begin{aligned} E[X_1^2] &= N_0 \int_{-\infty}^{\infty} h_1^2(s)ds \\ E[X_2^2] &= N_0 \int_{-\infty}^{\infty} h_2^2(s)ds \\ E[X_1X_2] &= N_0 \int_{-\infty}^{\infty} h_1(s)h_2(s)ds \end{aligned}$$

מה צריך להיות הקשר בין $h_1(\cdot)$ ל- $h_2(\cdot)$ במקרה זה על מנת ש- X_1, X_2 יהיו מ"א בלתי תלויים?

הערה: כאן מדובר בקונספט שנתקלים בו במידה רבה בקורס "מבוא לתקשורת ספרתית" בהקשר של גילוי אות רועש העובר דרך קורלטור, ו- X_i מסמל את תרומת הרעש לתוצאת הקורלציה. בבעיות דטקציה בסיסיות (למשל, בעית מכ"מ) קולטים אות $Y(t)$ ורוצים לבחון האם זהו רעש טהור: $H_0: Y(t) = N(t)$, או אות ידוע כלשהו הטבול ברעש: $H_1: Y(t) = X(t) + N(t)$. פתרון מקובל הוא לחשב את הקורלציה $\int Y(t)X(t)dt$ ולהשוות לסף. מהו סיכוי הטעות? במקרה הגאוס התוצאה תלויה במומנטים.

6.2 רוחב סרט אפקטיבי לרעש

נתונה פונקציית תמסורת $H(f)$ מסוג Low-Pass דהיינו $H(f)$ כמעט אפס בתדרים גבוהים ושונה מאפס בתדרים נמוכים כמתואר באיור 6.3.



איור 6.6: פונקציית תמסורת מעבירה נמוכים מעשית

רוצים להגדיר "רוחב סרט אפקטיבי" לפונקציית תמסורת כזו. הגדרה מקובלת היא ההגדרה הבאה:

$$\Delta F \triangleq \frac{\int_{-\infty}^{\infty} |H(f)|^2 df}{2|H(0)|^2}$$

כלומר רוחב הסרט האפקטיבי לרעש, ΔF , הוא היחס שבין הספק היציאה כאשר בכניסה רעש לבן עם $N_0 = 1$ לבין $2|H(0)|^2$. הגדרה זו נוחה למדי להרבה בעיות, אך כדאי לזכור שיש גם הגדרות אחרות.

רוחב סרט אפקטיבי לרעש עונה על השאלה הבאה: מהוא רוחב הסרט של מסנן LPF אידאלי, עם הגבר $H(0)$ בכל תחום ההעברה, אשר יעביר אותו הספק כמו המסנן הנתון כאשר הוא מוזן ברעש לבן?

דוגמא:

$$H(f) = \frac{1}{1 + i \frac{f}{f_0}}$$

אזי f_0 הוא רוחב הסרט של 3 db. $H(0) = 1$ ולכן

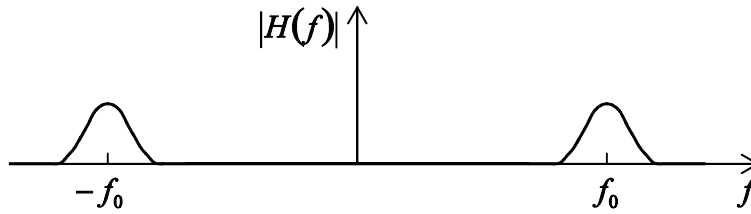
$$\Delta F = \frac{1}{2} \int_{-\infty}^{\infty} |H(f)|^2 df = \frac{f_0}{2} \int_{-\infty}^{\infty} \frac{df/f_0}{1 + (f/f_0)^2} = \frac{\pi}{2} f_0$$

ולכן רוחב הסרט האפקטיבי לרעש הוא $\frac{\pi}{2} f_0$.

במקום להגדיר את רוחב הסרט האפקטיבי ל- $H(0)$ אפשר גם להתיחס לתדר אחר:

$$\Delta F = \frac{\int_{-\infty}^{\infty} |H(f)|^2 df}{2|H(f_0)|^2}$$

הגדרה זו נוחה במיוחד לפונקציות תמסורת (כולל כמובן מגברים) מטפוס Band Pass, דהיינו לפונקציות תמסורת מהטיפוס:



איור 6.7: מסנן מעביר סרט

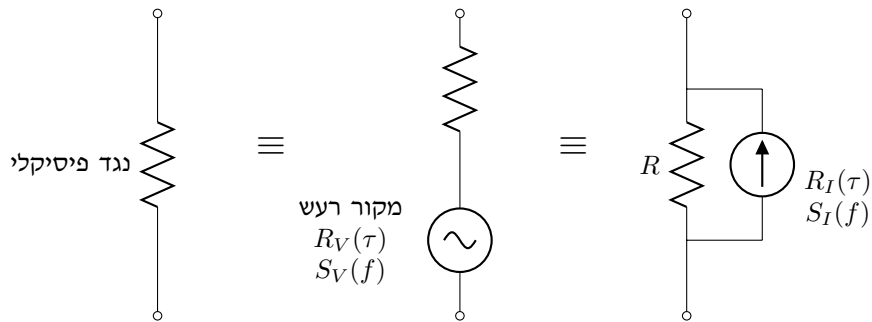
הערה: לפעמים יותר נח לחשב בתחום הזמן: $\int_{-\infty}^{\infty} |H(f)|^2 df = \int_{-\infty}^{\infty} h^2(t) dt$, למשל, אם התגובה להלם הינה:

$$h(t) = \begin{cases} 1, & |t| < T \\ 0, & \text{אחרת} \end{cases}$$

6.3 רעש תרמי (רעש הנגד, רעש Nyquist)

רעש תרמי בנגד הוא תולדה של פלקטואציות אקראיות הקשורות לתנועות האלקטרונים החפשיים שבמוליך. תנועות אלו יוצרות זרם אקראי רגעי, $I(t)$. ישנה אינטראקציה של כל אלקטרון עם האטומים ועם אלקטרונים אחרים, ומשך הזמן בין שתי התנגשויות עוקבות של כל אלקטרון הוא, בממוצע, קצר ביותר. מאחר והתנגשויות אלו יכולות להביא לידי שינוי כיוון, אז גם "זמן הקורלציה" של $I(t)$ הוא קצר ביותר, בסדר גודל של 10^{-14} שניות, ועל כן זהו בקירוב מצוין (לכל צורך מעשי) רעש לבן.

לכן, עלינו לראות כל נגד פיזיקלי כמורכב מנגד אידיאלי בצרוף מקור רעש:



איור 6.8: מודל לנגד רועש

בתנאי "שיווי משקל" עם הסביבה נוכל להניח שרעש הנגד הוא תהליך אקראי סטציונרי והבעיה היא למצוא את $S_V(f)$. באופן כללי $S_V(f)$ יהיה אולי פונקציה של התדר, של ההתנגדות הכוללת R , של הטמפרטורה T , של החומרים מהם מורכב הנגד ושל הגיאומטריה שלו. נדגיש שאנו עוסקים בנגד (או רשת נגדים, סלילים וקבלים) ללא מקורות הספק חיצוניים. לפני שנמשיך נשאל את השאלה הבאה: האם גם אלמנטים חסרי הפסדים - קבלים, סלילים, טרנספורמטורים אידיאליים יוצרים רעש על ההדקים שלהם כמו נגד? התשובה היא:

משפט א': רשת חסרת הפסדים איננה יוצרת רעש.

הוכחה: נעיין ברשת חסרת הפסדים המחוברת לנגד:



איור 6.9: רשת חסרת הפסדים

ונניח שהמערכת נמצאת בשיווי משקל תרמודינמי. אם הרשת חסרת ההפסדים היתה יוצרת מתחי רעש, מתחים אלו היו מחממים את הנגד. באותו זמן מתחי הרעש של הנגד אינם יכולים לחמם את הרשת חסרת ההפסדים מאחר ורשת חסרת הפסדים איננה יכולה לקבל הספק. התוצאה: הנגד היה מתחמם ואילו הרשת היתה מתקררת וזאת בניגוד לחוק השני של התרמודינמיקה. לכן מצב כזה בלתי אפשרי ולכן רשת חסרת הפסדים איננה יכולה לייצר רעש.

משפט ב': עבור מקור הרעש הטורי עם הנגד:

$$S_V(f) = 2kTR$$

עד לתדירויות "גבוהות מאוד" כאשר k היא הקונסטנטה של בולצמן $1.38 \cdot 10^{23} \text{ Joule}/^\circ K$. (הערה: "הנוסחה המעשית" היא $S_n(f) = 4kTR$; מכל מקום, תמיד נכון ש- $4kTR\Delta f = \{ \text{ממוצע המתח בריבוע בתחום התדרים } \Delta f \}$.)

שים לב שהתוצאה היא אוניברסלית ואיננה תלויה בחומרים או בגיאומטריה של הנגד. לפי תוצאה זו אין "נגד רועש" ו-"נגד שקט" - כל הנגדים טובים או רעים באותה מידה מנקודת המבט של הרעש שהם יוצרים. יש להדגיש שתוצאה זו בהחלט נכונה כאשר מדובר במערכת מבודדת, אולם היא איננה נכונה כשמדובר בנגד שמזרימים דרכו זרמים חיצוניים. נקודה זו תתבהר בהוכחה (החלקית) למשפט זה שנביא.

הוכחה חלקית: נעיין בשני נגדים R_a ו- R_b נחבר אותם זה אל זה באמצעות מעגלי תאום אימפדנסים ומסננות (כאשר מעגלי התאום והמסננות חסרי הפסדים) ונניח שהמערכת נמצאת בשיווי משקל תרמודינמי בינה לבין עצמה ועם הסביבה



איור 6.10: מעגל לחישוב רעש נגד

נשרטט מערכת שקולה פשוטה יותר, כדלקמן:



איור 6.11: מעגל מפושט לחישוב רעש נגד

כאשר $n_b(t), n_a(t)$ הם הרעשים הרגועים של הנגד אחרי מעבר דרך מסננת צרת סרט, דהיינו:

$$E[n_a^2(t)] \approx S_n^a(f_0) \cdot 2\Delta f$$

$$E[n_b^2(t)] \approx S_n^b(f_0) \cdot 2\Delta f$$

בהנחה של תאום אימפדנסים, הזרם שיוצא מהמקור השמאלי הוא $n_a(t)/2R_a$ ומהמקור הימני הוא $n_b(t)/2R_b$. ההספק שהמקור השמאלי מוסר לעומס (הנגד) הימני הוא (נזכור שקיים תאום אימפדנסים):

$$\frac{1}{2} n_a(t) \frac{n_b(t)}{2R_b}$$

בצורה דומה, ההספק שהמקור הימני מוסר לעומס השמאלי הוא:

$$\frac{n_b^2(t)}{4R_b}$$

ולכן, מתוך שיקול של שיווי משקל תרמודינמי, חייב להתקיים איזון ממוצע של הספקים

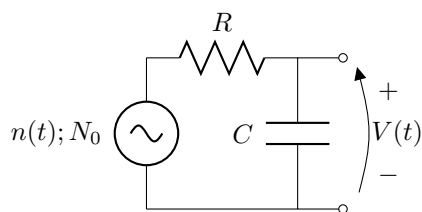
$$E\left[\frac{n_a^2(t)}{4R_a}\right] = E\left[\frac{n_b^2(t)}{4R_b}\right]$$

או

$$\frac{S_n^a(f_0)}{R_a} = \frac{S_n^b(f_0)}{R_b}$$

מסקנה: היחס $S_n(f)/R$ הוא יחס אוניברסלי שיכול להיות תלוי בתדר f או בטמפרטורה T אבל איננו יכול להיות תלוי בסוג הנגד ז.א. לא תלוי בחומר ובגיאומטריה).

בכדי לקבל ביטוי לגובה הצפיפות הספקטרלית של הזרם, $S_I(f)$, שהיא קבועה ושווה ל- N_0 כלשהו לכל f ברוחב סרט גדול מאד, נתבונן במעגל באיור 6.12. במעגל זה $n(t)$ הינו מקור מתח טורי המתאר את מקור רעש המתח האקראי שמייצר הנגד.



איור 6.12:

נתבונן במתח, $V(t)$, המתפתח על הקבל, ונניח שהמעגל שרוי בשווי משקל תרמי. מתרמודינמיקה ופיסיקה סטטיסטית, אנו יודעים כי צפיפות ההסתברות למצוא מערכת הנמצאת בשו"מ תרמי, במצב עם אנרגיה E , מתנהגת לפי פילוג בולצמן, כלומר פרופורציונלי ל- $e^{-E/(kT)}$. במקרה שלנו, כל האנרגיה אגורה בקבל, ולכן $E = \frac{1}{2}CV^2$. כלומר - צפיפות ההסתברות של המתח, $V(t)$ (כל אימת שהוא במצב מתמיד), מתנהגת באופן פרופורציונלי ל-

$$e^{-\frac{CV^2}{2kT}}.$$

מכאן נובע כי המתח מפולג גאוסית עם שונות $\mathbb{E}[V^2(t)] = kT/C$. כעת נשתמש בתכונות שלמדנו לגבי הקשר בין השונות התהליך לבין הספקטרום, ונקבל:

$$\begin{aligned} \frac{kT}{C} &= \mathbb{E}[V^2(t)] = R_V(0) = \int_{-\infty}^{\infty} S_V(f) df \\ &= \int_{-\infty}^{\infty} N_0 \left| \frac{1}{1 + 2\pi j f RC} \right|^2 df = N_0 \int_{-\infty}^{\infty} \frac{1}{1 + (2\pi f RC)^2} df \\ &= \frac{N_0}{2\pi RC} \int_{-\infty}^{\infty} \frac{1}{1 + u^2} du = N_0 \frac{\pi}{2\pi RC} = \frac{N_0}{2RC} \\ \Rightarrow N_0 &= 2kTR \end{aligned}$$

ללא תלות ב- C .

באופן דומה נקבל לגבי הזרם

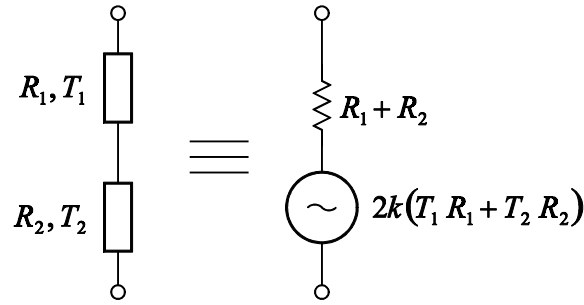
$$S_I(f) = \frac{N_0}{R^2} = \frac{2kT}{R} = 2kTG$$

ברור שנוסחה זו אינה נכונה בתדרים (מאד) גבוהים, בסדר ג" של אלפי GHz, בגלל אפקטים קוונטיים. ובגלל השיקול שהבאנו, התוצאה מתקיימת עבור כל הנגדים.

הפיתוח הנ"ל ממחיש קשר בין רעש לפיסיקה סטטיסטית. המתעניינים בנושא מוזמנים להרשם לקורס "פיסיקה סטטיסטית, תנודות ורעש" (046242) העוסק בהרחבה בתופעות אלו ועוד.

הערות:

א. שים לב שמתחי הרעש מסתכמים "על בסיס של הספקים" דהיינו:

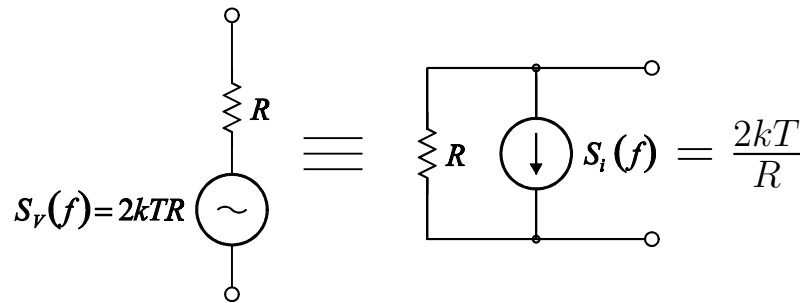


איור 6.13: חיבור נגדים בטור-מעגל שקול

ולכן, נגדיר טמפרטורה אפקטיבית של שני הנגדים כך שהן ההתנגדות השקולה והן הספק הרעש יקבלו את הערכים הנכונים, כלומר $2k(T_1 R_1 + T_2 R_2) = 2kT_{\text{eff}}(R_1 + R_2)$ או

$$T_{\text{eff}} = \frac{T_1 R_1 + T_2 R_2}{R_1 + R_2}$$

ב. מעגל תמורה מקבילי במקום מעגל תמורה טורי



איור 6.14: מעגל תמורה מקבילי לרעש נגד

מעגל תמורה עם הרעש כמקור מתח טורי, אפשר כידוע, להחליף במעגל תמורה עם הרעש כמקור זרם מקבילי. על מנת למצוא את הקשר בין הצפיפות הספקטרלית של מקור המתח $S_v(f)$ עם זה של הזרם $S_i(f)$, נשווה את המתח ריקם על פני ההדקים: $i(t)R = v(t)$

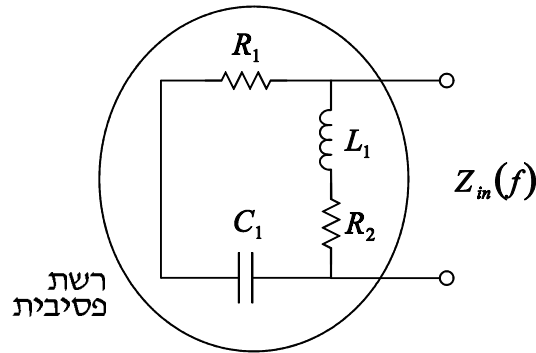
ולכן

$$R^2 E[i(t + \tau)i(t)] = E[v(t + \tau)v(t)]$$

ומכאן

$$S_i(f) = \frac{S_v(f)}{R^2} = 2ktTG$$

(ועבור $S_v(f) = 4kTR$ נקבל $S_i(f) = 4kTG$ כאשר $G = R^{-1}$.)



איור 6.15: רעש של מעגל

ציור 6.12

נעבור למשפט ג': נעיין ברשת פסיבית

הרשת נמצאת בשיווי משקל תרמודינמי עם סביבתה. מה ספקטרום הרעש על פני הדקי הכניסה של הרשת?

נוכל כמובן להוסיף לכל נגד את מקור הרעש שלו ולחפש את סה"כ הרעש על פני ההדקים, אולם:

משפט ג': אם אימפדנס הכניסה של רשת (פסיבית הנמצאת בשיווי משקל תרמודינמי) הוא

$$Z_{in}(f) = R_{in}(f) + iX_{in}(f)$$

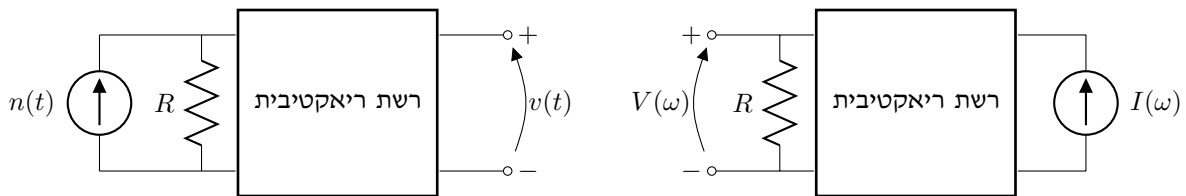
אזי הצפיפות הספקטרלית של מתח הרעש על פני הדקי המעגל יהיה:

$$S(f) = 2kT R_{in}(f)$$

$$\bar{v}^2 = \int_{-\infty}^{\infty} 2kT R_{in}(f) df$$

וסה"כ הספק הרעש יהיה

הוכחה חלקית: נניח, ראשית, כי יש רק נגד אחד ברשת, ואילו שאר האלמנטים הם ריאקטיביים טהורים (איור 6.16). במעגל



איור 6.16:

הימני, ההספק הנכנס לרשת הוא $|I(\omega)|^2 \cdot \text{RE}\{Z(\omega)\}$ והוא נמסר כולו לנגד. מצד שני, ההספק הנמסר לנגד הוא $\frac{V(\omega)^2}{R}$. לכן:

$$\frac{|V(\omega)|^2}{R} = |I(\omega)|^2 \cdot \text{RE}\{Z(\omega)\} \Rightarrow \left| \frac{V(\omega)}{I(\omega)} \right|^2 = R \cdot \text{RE}\{Z(\omega)\}.$$

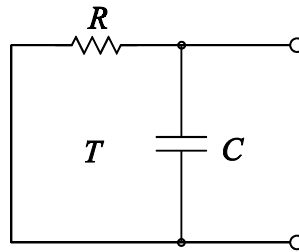
מצד שני, במעגל השמאלי מתקיים:

$$S_V(f) = S_I(f) \cdot |H(\omega)|^2 = \frac{2kT}{R} \cdot R \cdot \text{RE}\{Z(\omega)\} \\ = 2kT \cdot \text{RE}\{Z(\omega)\}$$

כאשר השתמשנו במשפט ההיפוך (reciprocity theorem), הקובע כי תמסורת מהזרם בזוג ההדקים השמאלי, למתח בזוג הימני (טרנס-אימפדנס) שווה לזה שבכוון ההפוך. ההרחבה ליותר נגדים היא ישירה:

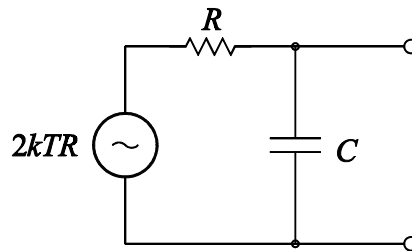
$$|I(\omega)|^2 \cdot \text{RE}\{Z(\omega)\} = \sum_{i=1}^K \frac{|V_i(\omega)|^2}{R_i} \Rightarrow \text{RE}\{Z(\omega)\} = \sum_{i=1}^K \frac{1}{R_i} \left| \frac{V_i(\omega)}{I(\omega)} \right|^2 \\ S_V(f) = \sum_{i=1}^K 2kT \frac{1}{R_i} \left| \frac{V_i(\omega)}{I(\omega)} \right|^2 = 2kT \text{RE}\{Z(\omega)\}.$$

דוגמא: נעיין במעגל



איור 6.17: רעש במעגל פשוט

נוכל לחשב את הרעש על פני ההדקים ע"י מעגל התמורה:



איור 6.18: מעגל תמורה - רעש במעגל פשוט

וכאשר $H(f)$ פונקציית התמסורת ממקור מתח הרעש להדקי היציאה נקבל:

$$S(f) = |H(f)|^2 \cdot 2kTR = \left(\frac{\frac{1}{2\pi i f C}}{R + \frac{1}{2\pi i f C}} \right)^2 2kTR = \frac{2kTR}{(2\pi)^2 f^2 R^2 C^2 + 1}$$

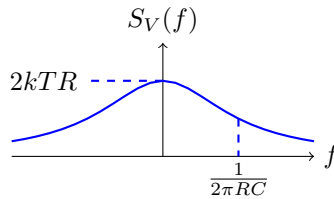
לפי המשפט שקבלנו

$$Z_{in}(f) = \frac{R \cdot \frac{1}{i2\pi fC}}{R + \frac{1}{i2\pi fC}} = \frac{R(1 - i2\pi fRC)}{1 + (2\pi fRC)^2}$$

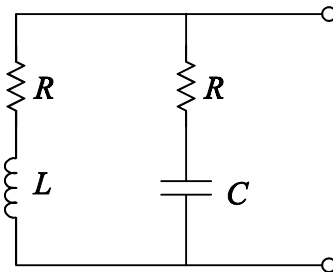
ולכן

$$R_{in}(f) = \frac{R}{1 + (2\pi fRC)^2}$$

ומתקבלת, כמובן אותה תוצאה.



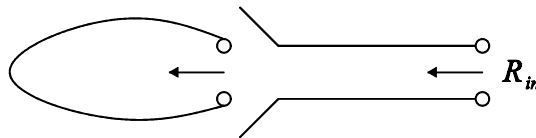
תרגיל:



איור 6.19: מעגל טורי-מקבילי

עבור $R = \sqrt{\frac{L}{C}}$ המעגל מתנהג כמו נגד טהור בעל התנגדות R . מצא את הרעש על פני ההדקים בשתי שיטות: (א) ע"י שמוש במשפט ג'. (ב) ע"י הוספת מקור רעש לכל אחד מהנגדים וחישוב השפעת מקורות אלה על הרעש שעל פני הדקי המעגל.

הערת סיום לסעיף זה: נעיין באנטנה כיוונית שכאנטנת שידור היא קורנת את כל ההספק הנמסר לה לכוון אחד (ראה הציור) ושהיא חסרת הפסדים, (דהיינו, כל ההספק הנמסר לה מוקרן החוצה)



איור 6.20: אנטנה כיוונית

נניח שאימפדנס הכניסה לאנטנה הוא אוהמי טהור, R_{in} . נניח שהאנטנה צופה אל אחד הכוונים הבאים:

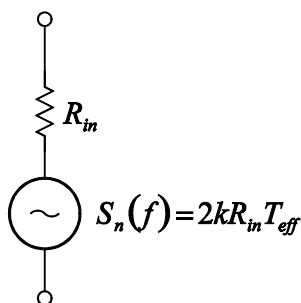
(א) החלל (בניצב לשביל החלב) - (בסדר גודל של 1000°K ב- 100 MHz , יורד לכ- 5°K ב- 1000 MHz).

(ב) שביל החלב - (בסדר גודל של 4000°K ב- 100 MHz , יורד לכ- 15°K ב- 1000 MHz).

(ג) כדור הארץ - (כ- 300°K)

(ד) השמש (נניח אלומת קרינה צרה מספיק כך שתקרין לשמש בלבד) - (בסדר גודל של אלפי מעלות קלווין).

מנקודת המבט של רעש תהיה סכימת התמורה של האנטנה



איור 6.21: מודל של אנטנה כיוונית

כאשר בכל אחד מהמקרים יש להציג את הטמפרטורות המתאימות.

6.4 רעש הדיודה (Shot noise)

לפני שנתייחס לרעש הדיודה, נביא תוספת לתהליכי פואסון. כזכור, תהליך פואסון הוא תהליך מניה עם הפרשים בלתי תלויים ופילוג שולי

$$\mathbb{P}\{N_t = k\} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

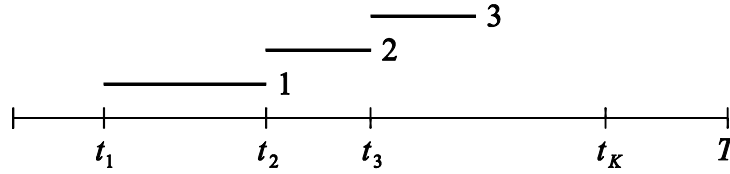
כאשר λ הוא קצב התהליך.

נבצע סימולציה של תהליך אקראי (על מחשב) כדלקמן: נקבע T מסוים, נגדיל מספר אקראי K לפי פילוג פואסון עם תוחלת λT . אחרי שהגרלנו את K נבחר באקראי K זמנים ב- $[0, T]$, בלתי תלויים ומפולגים בצורה אחידה בתחום הזה. נסמן את הזמנים האלה $(\theta_1, \dots, \theta_K)$, נסדר את הזמנים בסדר עולה ונסמנם מחדש (t_1, \dots, t_K) (כעת $t_1 \leq t_2 \leq \dots \leq t_K$) ונבנה מהם פונקצית מדרגות עולה כדלקמן:

טענה: חוק ההסתברות של התהליך האקראי שהמצאנו הוא חוק ההסתברות של תהליך פואסון (טענה זו הגיונית מאוד: אם נחשוב על תהליך פואסון כעל התהליך האקראי שמכונית מגיעה לצומת ואם אנו יודעים ש- K מכוניות הגיעו לצומת בקטע הזמן $[0, T]$, כל אחת מהמכוניות שיגיעו לצומת לא יודעת על האחרת וסביר להניח שפילוג הגעתה לצומת יהיה פילוג אחיד ב- $(0, T)$).

הוכחה: נשאל את עצמנו, תחילה, מה ההסתברות של- N_t יהיו בדיוק k קפיצות בזמנים סביב (t_1, \dots, t_k) , שהם זמנים

————— K



איור 6.22: תהליך פואסון

מסודרים בסדר עולה:

$$\begin{aligned} & \mathbb{P}\{N_T = k, t_1 \leq \tau_1 \leq t_1 + \Delta, \dots, t_k \leq \tau_k \leq t_k + \Delta\} \\ &= e^{-\lambda t_1} \lambda \Delta e^{-\lambda \Delta} e^{-\lambda(t_2 - t_1 - \Delta)} \lambda \Delta e^{-\lambda \Delta} \dots e^{-\lambda(t_k - t_1 - \Delta)} \lambda \Delta e^{-\lambda \Delta} e^{-\lambda(T - t_k - \Delta)} \\ &= e^{-\lambda t_1} \lambda \Delta e^{-\lambda \Delta} \prod_{i=2}^k [e^{-\lambda(t_i - t_{i-1} - \Delta)} \cdot e^{-\lambda \Delta} \lambda \Delta] = e^{-\lambda T} (\lambda \Delta)^k. \end{aligned}$$

ההסתברות המותנית המתאימה:

$$\begin{aligned} & \mathbb{P}\{t_1 \leq \tau_1 \leq t_1 + \Delta, \dots, t_k \leq \tau_k \leq t_k + \Delta | N_T = k\} \\ &= \frac{e^{-\lambda T} (\lambda \Delta)^k}{e^{-\lambda T} \frac{(\lambda T)^k}{k!}} = k! \cdot \left(\frac{\Delta}{T}\right)^k. \\ f(t_1, \dots, t_k | N_T = k) &= \lim_{\Delta \rightarrow 0} \frac{\mathbb{P}\{t_1 \leq \tau_1 \leq t_1 + \Delta, \dots, t_k \leq \tau_k \leq t_k + \Delta | N_T = k\}}{\Delta^k} = k! \cdot \left(\frac{1}{T}\right)^k. \end{aligned}$$

זוהי צפיפות ההסתברות של הזמנים כשהם מסודרים בסדר עולה. לכל קטור זמנים מסודר יש $k!$ "מקורות" של וקטורים זמניים "לא מסודרים", $(\theta_1, \dots, \theta_k)$, ולכן:

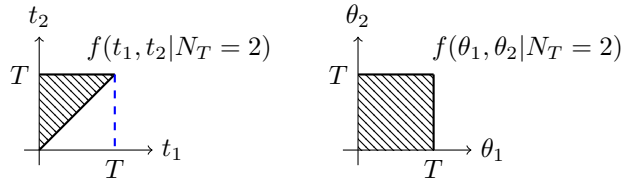
$$f(\theta_1, \dots, \theta_k | N_T = k) = \frac{1}{k!} f(t_1, \dots, t_k | N_T = k) = \left(\frac{1}{T}\right)^k,$$

שפירושו ש- k הזמנים $\{\theta_i\}$ מפולגים לפי פילוג אחיד, $\frac{1}{T}$, ללא תלות זה בזה (איור 6.23).

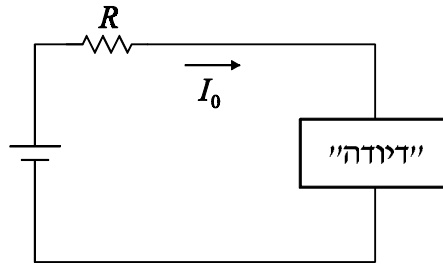
נעבור כעת לזרם בדיודה: נסמן ב- q את מטען האלקטרון ונניח שאנו מזרימים זרם I_0 (קבוע) דרך הדיודה, משמעות הדבר היא שבכל שניה I_0/q אלקטרונים (בממוצע) עוברים מהקטודה לאנודה.

אם מטען האלקטרון היה "אפס" (ו"אינסוף" אלקטרונים בשניה) היה עובר דרך הדיודה והנגד זרם D.C. טהור I_0 והצפיפות הספקטרלית של הזרם היתה

$$S(f) = I_0^2 \delta(f)$$



איור 6.23: דוגמא עבור $k = 2$. בצד שמאל ניתן לראות פילוג אחיד רק במשולש המקווקו, ואילו בצד ימין מדובר בפילוג אחיד ובלתי תלוי של שני המשתנים.



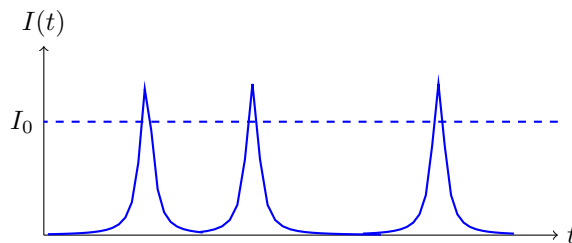
איור 6.24: מקור רחב סרט

ציור 6.17

מאחר והזרם נגרם ע"י אלקטרונים בעלי מטען סופי נצפה לצפיפות ספקטרלית מהצורה

$$S(f) = \underbrace{I_0^2 \delta(f)}_{\text{טוהור DC}} + \underbrace{S_n(f)}_{\text{פלקטואציות}}$$

מטרתנו היא לחשב את $S_n(f)$.



במודל שלנו, אלקטרון הנפלט ברגע $t = 0$ משרה במעגל זרם $i_e(t)$ ולכן:

$$\text{מטען האלקטרון} = q = \int_{-\infty}^{\infty} i_e(t) dt$$

בקטע הזמן T יפלטו בממוצע $I_0 T / q$ אלקטרונים. מאחר והאלקטרונים נפטים מאיזורים שונים בקטודה סביר להניח שיש אי תלות בין זמני הפליטה. נוכל, לכן, להניח שתהליך פליטת האלקטרונים הוא תהליך פואסון. לכן (בהזנחת אפקט הקצוות,

ז.א. T גדול)

$$I(t) = \sum_{k=1}^K i_e(t - t_k)$$

כאשר K מ"א המפולג לפי פילוג פואסון; $\lambda = I_0 T / q$ ו- t_k מפולגים אחיד בתחום $[0, T]$. לכן:

$$E[I(t)] = E \left[E \left[\sum_{k=1}^K i_e(t - t_k) \middle| K \right] \right] = E \left[K \frac{1}{T} \int_0^T i_e(t - \theta) d\theta \right] = \frac{q}{T} E[K] = \frac{q}{T} I_0 T = I_0$$

השוויון הראשון הוא מתכונת ההחלקה, בשוויון השני משתמשים בעובדה: $Eg(X) = \int f_X(\theta)g(\theta)d\theta$ (עבור מ"א X ופונקציה g), בשוויון השלישי באה לידי ביטוי הזנחת הקצוות ובשוויון הרביעי מופיעה התוחלת של תהליך פואסון. נפנה כעת למומנט השני:

$$\begin{aligned} R(\tau) &= E[I(t+\tau)I(t)] = E \left[\sum_{k=1}^K i_e(t+\tau - t_k) \sum_{j=1}^K i_e(t - t_j) \right] \\ &= E \left[\sum_{k=1}^K i_e(t+\tau - t_k) i_e(t - t_k) \right] + E \left[\sum_{k \neq j} i_e(t+\tau - t_k) i_e(t - t_j) \right] \end{aligned}$$

נניח כעת שה- t_k מפולגים אחיד בתחום $[-T/2, T/2]$. נמצע תחילה על פני ה- t_k ואח"כ על פני K . לכן, בהזנחת אפקט הקצוות, (E יסמן מצוע על K)

$$\begin{aligned} R(\tau) &= E \left[\sum_1^K \frac{1}{T} \int_{-T/2}^{T/2} i_e(t+\tau - \theta) i_e(t - \theta) d\theta \right] + E \left[\sum_{k \neq j} \frac{1}{T^2} \int_{-T/2}^{T/2} i_e(t+\tau - \theta) d\theta \int_{-T/2}^{T/2} i_e(t - \eta) d\eta \right] \\ &\cong \frac{E[K]}{T} \int_{-\infty}^{\infty} i_e(\tau + \theta) i_e(\theta) d(\theta) + \frac{E[K^2 - K]}{T^2} q^2 \end{aligned}$$

עבור תהליך פואסון

$$E[K^2 - K] = (\lambda T)^2 \quad ; \quad E[K] = \lambda T$$

לכן

$$R(\tau) = \frac{I_0}{q} \int_{-\infty}^{\infty} i_e(t+\tau) i_e(t) dt + I_0^2$$

נסמן ב- $G_e(f) = F\{i_e(t)\}$, ונקבל ע"י התמרת פוריה

$$S(f) = I_0^2 \delta(f) + \frac{I_0}{q} G_e(f) G_e^*(f)$$

יהיה τ זמן המעבר של אלקטרון מהקטודה לאנודה, עבור תדרים f המקיימים $\tau f \ll 1$ יתקיים $G_e(f) \cong G_e(0)$ (במלים אחרות $i_e(t) \cong q\delta(t)$ ובמקרה שלנו $G_e(0) = q$ ולכן

$$S(f) \cong I_0^2 \delta(f) + I_0 \cdot q$$

או

$$S_n(f) = I_0 q$$

התוצאה שקבלנו היא נוסחת רעש הדיודה.

הערות:

(א) שים לב שבעזרת התוצאה שקבלנו וע"י מדידת רעש הדיודה אפשר למצוא באופן נסיוני את מטען האלקטרון q .

(ב) בספרי עזר מוצאים הנוסחה $2I_0 q$ מדוע?

(ג) התוצאה שקבלנו היא עבור זרמי דיודה לא גדולים מדי. עבור זרמי דיודה גדולים מופיע אפקט נוסף הנקרא אפקט המטען המרחבי והוא כדלקמן: אם ברגע מסוים עוזבים הרבה אלקטרונים את הקטודה הם יוצרים מטען מרחבי (בין הקטודה לאנודה) המעכב את הפליטה של אלקטרונים נוספים. אפקט זה גורם לרעש דיודה נמוך מהנתון ע"י הנוסחה שקבלנו אולם בזרמי דיודה לא גבוהים מדי הוא זניח.

6.5 רעש טרמי - פתוח מתוך מודל רעש הדיודה

נזכיר כי בדיודה,

$$I = I_0(e^{qV/kT} - 1)$$

ולמעשה, ניתן למדל את הזרמים בדיודה כך:

| | |
|-------------------------------|-------------|
| $\xrightarrow{I_0 e^{qV/kT}}$ | זרם סחיפה |
| $\xleftarrow{-I_0}$ | זרם דיפוזיה |

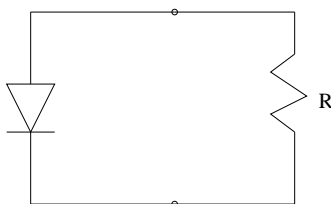
זרם הסחיפה וזרם הדיפוזיה הם זרמים פואסוניים, בת"ס. לכן,

$$S_n(I) = qI_0(1 + e^{qV/kT})$$

ועבור $V = 0$,

$$\boxed{S_n(I) = 2qI_0} \quad (6.1)$$

נחבר כעת את המעגל ראשית, נשים לב כי עבור הדיודה,



$$\frac{\partial I}{\partial V} = \frac{qI_0}{kT} \cdot e^{qV/kT}$$

ולכן, סביב מתח 0, מתנהגת הדיודה כמו נגד:

$$\left. \frac{\partial I}{\partial V} \right|_{V=0} = \frac{qI_0}{kT} = \frac{1}{R} \quad (6.2)$$

לכן, עם בחירת R כזה, מתקיים תאום אמפדנסים. לכן,

$$2qI_0 = S_n^{R,I}$$

ומכאן

$$S_n^{R,I} = \frac{2kT}{R}$$

ומכאן, במעבר לרעש מתח,

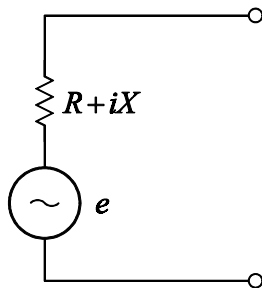
$$S_n^{R,V} = 2kTR \quad (6.3)$$

6.6 אפיון רעש מגבר

נניח שנתונים שני מגברים, האחד עם הגבר גבוה ורעש גבוה ביציאתו והשני עם הגבר נמוך ורעש נמוך ביציאתו. איזה משניהם עדיף לשים לפני השני? בבעיות מסוג זה נעסוק בסעיף הנוכחי.

הספק מצוי, הגבר הספק

נעיין בתדר מסוים ונגדיר: הספק מצוי של המקור = הספק מכסימלי שאפשר לקבל מהמקור. הספק זה יתקבל בתנאי התאמת אימפדנסים ולכן עבור איור 6.25 יתקבל ההספק המצוי:



איור 6.25: מעגל הרעש שהוא מייצר

עבור רעש טרמי בתחום תדרים Δf , נקבל שהספק הרעש המצוי נתון ע"י:

$$E \left[\frac{e^2(t)}{4R(f)} \right] = \frac{2kTR(f)\Delta f \cdot 2}{4R(f)}$$

$$= kT \Delta f$$

וזה הספק הרעש המצוי במקור פיזיקלי פסיבי.

הגדרות:

א. הטמפרטורה האפקטיבית של מקור כלשהוא (פסיבי או אקטיבי)

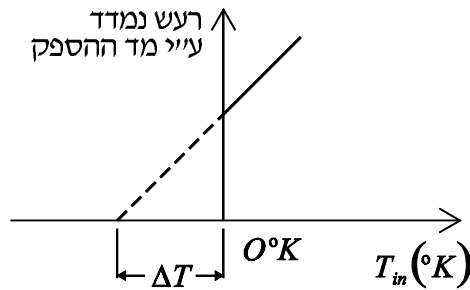
$$T_{\text{eff}} = \frac{1}{k \cdot \Delta f} \cdot [\text{הספק הרעש המצוי ברחב סרט } \Delta f]$$

ב. הגבר הספק: $G_a = \frac{\text{הספק מצוי ביציאה}}{\text{הספק מצוי בכניסה}}$

להלן נביא סקירה מקוצרת על המושגים "טמפרטורת הרעש האפקטיבית של מגבר" ו"ספרת רעש". נעיין במערכת המצוירת; נשנה את הטמפרטורה של המקור (T_{in}) ונצייר את הספק הרעש ביציאה כפונקציה של טמפרטורת המקור. (שים לב, המקור R_0 הוא לא חלק מהמגבר). טמפרטורת הרעש האפקטיבית ΔT מוגדרת כמצויר:



איור 6.26: רעש המגבר



איור 6.27: רעש המגבר - מדידה

במילים אחרות נחליף המגבר הרועש במגבר זהה אבל ללא רעש ונעלה את טמפרטורת המקור כך שסה"כ לא יהיה שינוי. כמה צריך להוסיף? תשובה ΔT .

לאור ההגדרה, הספק הרעש המצוי ביציאת המגבר ברוב סרט Δf והגבר G יהיה:

$$G \cdot k \cdot (T_{in} + \Delta T) \cdot \Delta f$$

לכן: אם המגבר מחובר דרך קו תמסורת ואנטנה חסרת הפסדים והאנטנה צופה ל- $T_{(מקור)}$, אזי,

אם $\Delta T \ll T_{(מקור)}$ אזי שיפור ב- ΔT ישפר את הקליטה.

אם $\Delta T \ll T_{(מקור)}$ אזי שיפור ΔT לא ישפר למעשה.

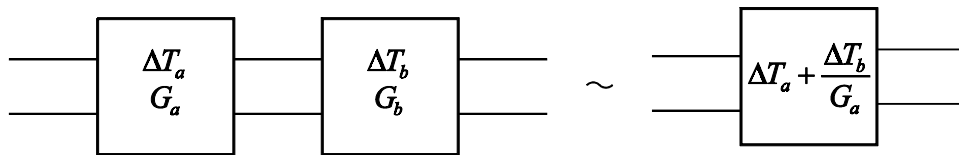
מספרי ΔT טיפוסיים:

מקלט מיקרוגלים לא מעולה $\Delta T \cong 10 \times T_0$ $T_0 = 300^\circ \text{K}$

מקלט מיקרוגלים מעולה $4 \times T_0$

מגבר פרמטרי $\Delta T \sim 80^\circ \text{K}$

טענה: הטמפרטורה האפקטיבית הכוללת של שני מגברים בזה אחר זה (קסקדה):



איור 6.28: חיבור מגברים בטור

הוכחה: ראשית כפי שכבר הערנו, ההספק המצוי של הרעש ביציאה של מגבר (בודד), הוא $Gk\Delta f(T_{מקור} + \Delta T)$, (כאשר G - ההגבר המצוי ו- $T_{מקור}$ טמפרטורת הרעש של המקור). עבור קסקדה, הספק הרעש המצוי ביציאת מגבר a נתון ע"י $G_a k \Delta f (T_{מקור} + \Delta T_a)$, ולכן ביציאה של מגבר b יהיה הספק רעש מצוי:

$$G_b G_a k \Delta f (T_{מקור} + \Delta T_a) + G_b k \Delta f \Delta T_b = G_a G_b k \Delta f \left(T_{מקור} + \Delta T_a + \frac{\Delta T_b}{G_a} \right)$$

ולכן:

$$\Delta T_{ab} = \Delta T_a + \frac{\Delta T_b}{G_a}$$

שים לב שמהנוסחה נוכל להחליט אם עדיף לשים את a לפני b או את b לפני a .

מאפיין נוסף לרעש המגבר הוא ספרת הרעש F המוגדרת כדלקמן: נחבר לכניסת המגבר נגד בטמפרטורת הסביבה (300°K). ההספק המצוי של המקור יהיה כמובן $300k\Delta fG$. נגדיר כעת:

$$F = \frac{\text{הספק הרעש המצוי ביציאת המגבר (כולל רעש המגבר)}}{300k\Delta fG}$$

ולכן F קשור ל ΔT ע"י

$$F = 1 + \frac{\Delta T}{300}$$

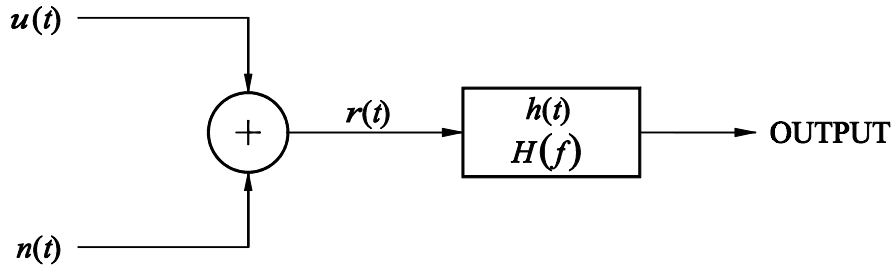
עבור מגבר אידיאלי שאינו יוצר רעש $\Delta T = 0$ ו- $F = 1$. שים לב ש- ΔT מאפיין את רעש המגבר ע"י תוספת הרעש שגורם המגבר, ואילו F מאפיין את רעש המגבר ע"י מקדם כפל לרעש המקור, כאשר המקור בטמפרטורה של 300°K . מכיון ש- F מוגדר ביחס לטמפרטורה שרירותית ואילו הגדרת ΔT אינה דורשת נקודת יחוס שרירותית, ΔT הוא מושג בסיסי יותר).

תרגיל: הוכח שעבור חיבור שני מגברים בקסקדה

$$F_{ab} = F_a + \frac{F_b - 1}{G_a}$$

6.7 מסננת מתואמת Matched Filter

אות הכניסה: אות דטרמיניסטי בעל אנרגיה סופית, $u(t)$. למען הפשטות נניח ש- $u(t) = 0$ עבור $t > 0$. נסמן $U(f) = F\{u(t)\}$. עיין באיור 6.29



איור 6.29: אות ורעש

רעש הכניסה: רעש לבן, תוחלת אפס צפיפות ספקטרלית N_0 .

היציאה: כאשר בכניסה מצוי האות $u(t)$ בלבד (ללא רעש):

$$Y(t) = \int_{-\infty}^{\infty} u(t - \theta)h(\theta)d\theta$$

כיוון שהאות הרצוי מתאפס בזמנים חיוביים, ערך המוצא בזמן 0 כולל את כל המידע על אות זה. ברגע $t = 0$ תהיה היציאה

$$Y(0) = \int_0^{\infty} u(-\theta)h(\theta)d\theta$$

נגדיר יחס אות לרעש ביציאה ברגע $t = 0$ כדלקמן:

$$\left(\frac{S}{N}\right)_{\text{out}}^{t=0} = \frac{(Y(0))^2}{\text{הספק רעש ממוצע ביציאה}}$$

ונרצה שיחס זה יהיה גדול ככל האפשר. לפי משפט פרסוול נוכל לרשום את $Y(0)$ במרחב התדר

$$Y(0) = \int_{-\infty}^{\infty} H(f)U^*(f)df$$

הספק הרעש ביציאה (הספק ממוצע):

$$\int_{-\infty}^{\infty} |H(f)|^2 \cdot N_0 df$$

ובמרחב הזמן, ממשפט פרסוול נובע

$$\int_{-\infty}^{\infty} |H(f)|^2 \cdot N_0 df = N_0 \int_{-\infty}^{\infty} h^2(\theta) d\theta$$

ולכן

$$\left(\frac{S}{N}\right)_{\frac{\text{out}}{t=0}} = \frac{\left[\int_0^{\infty} u(-\theta)h(\theta)d\theta\right]^2}{N_0 \int_0^{\infty} h^2(\theta)d\theta} = \frac{\left[\int_0^{\infty} U^*(f)H(f)df\right]^2}{N_0 \int_0^{\infty} |H(f)|^2 df}$$

הבעיה: מצא מערכת לינארית (מאופיינת ע"י $h(\theta)$ או $H(f)$) כך שהיחס $\left(\frac{S}{N}\right)_{\frac{\text{out}}{t=0}}$ יהיה מכסימלי.

תזכורת: אי השיוויון של שוורץ (לפונקציות קומפלקסיות)

$$\left(\int_{T_1}^{T_2} f_1(t)f_2^*(t)dt\right)^2 \leq \int |f_1(t)|^2 dt \cdot \int |f_2(t)|^2 dt$$

ואי השיוויון הופך לשיוויון כאשר $f_2(t) = \alpha f_1(t)$ מקדם פרופורציה שרירותי. לכן, נקבל מאי השיוויון של שוורץ ומהביטויים לעיל

$$\left(\frac{S}{N}\right)_{\frac{\text{out}}{t=0}} \leq \frac{\int_0^{\infty} (u(-\theta))^2 d\theta \int_0^{\infty} h^2(\theta) d\theta}{N_0 \int_0^{\infty} h^2(\theta) d\theta} = \frac{1}{N_0} \int_0^{\infty} (u(-\theta))^2 d\theta$$

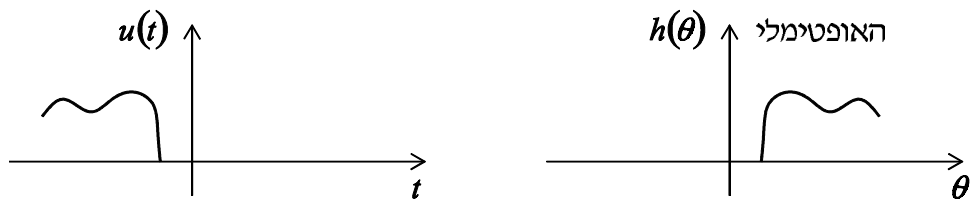
כמו כן נובע מאי השיוויון של שוורץ שאי השיוויון שקבלנו זה עתה יתקיים כשיוויון כאשר

$$h(\theta) = \alpha u(-\theta)$$

מכאן המסקנה: $h(t)$ האופטימלי הוא תמונת הראי של $u(t)$ לגבי הציר האנכי.

$h(t)$ האופטימלי נקרא: המסננת המתואמת.

הערה: בעמוד הקודם מופיע ביטוי במרחב הזמן וביטוי במרחב התדר. במקום להשתמש בביטוי במרחב הזמן, השתמש בביטוי במרחב התדר על מנת לקבל (בעזרת אי השיוויון של שוורץ) ש- $H(f)$ האופטימלי הוא $\alpha U^*(f)$ וע"י התמרת פוריה הפוכה נקבל כמובן שנית את התוצאה ש- $h(t)$ האופטימלי הוא $\alpha u(-t)$.



איור 6.30: אות ומסננת מתואמת

נספח 2

8 חזרה נוספת על "מבוא להסתברות"

בפרק זה נאסוף הגדרות ותזכורות

הקשורות לתורת ההסתברות, וכן לאלגברה לינארית.

8.1 הסתברות

הגדרה 8.1 מרחב המדגם Ω הוא אוסף של "נקודות" $\{\omega\}$.

אינטואיטיבית נחשוב על מרחב המדגם כעל אוסף כל התוצאות האפשריות של נסוי.

דוגמה 8.2 נניח שרוצים לבנות מודל של זריקת קוביה. כוון שיש שש תוצאות אפשריות, נבחר אוסף (שרירותי!) של שישה עצמים. למשל: אוסף המספרים $\{1, 2, 3, 4, 5, 6\}$. נוכל גם לבחור את האוסף $\{a, b, c, d, e, f\}$.

ככל שרוצים לבנות מודל של תופעה מסובכת יותר, כך יהיה מרחב המדגם מסובך יותר. לפעמים בונים מרחב מדגם גדול יותר מאשר נחוץ: זאת כדי להשיג מרחב עם מבנה פשוט, או לאפשר הרחבות של המודל בשלב מאוחר יותר.

דוגמה 8.3 כדי לבנות מודל של רולטה, נוכל לבחור כמרחב מדגם את אוסף הנקודות על היקף מעגל היחידה. באוסף זה יותר נקודות מאשר יש תוצאות אפשריות ברולטה. ניתן לתאר כל מאורע (תוצאה ברולטה) כקטע על המעגל.

אם רוצים לבנות מודל עבור רעש יציאה ממגבר בקטע הזמן $[0, 1]$, מרחב המדגם יכול להיות אוסף כל הפונקציות הרציפות בקטע הזמן $[0, 1]$.

הגדרה 8.4 אוסף המאורעות \mathbb{F} הוא אוסף של תת קבוצות של Ω , המקיימות את התנאים הבאים.

1. $\Omega \in \mathbb{F}$, כלומר המקרה שתמיד קורה הוא מאורע, אורתת הקבוצה Ω של Ω היא מאורע.

2. אם A_1 וכן A_2 הם מאורעות, אזי גם $A_1 \cup A_2$ הוא מאורע, כלומר $A_1 \cup A_2 \in \mathbb{F}$.

3. אם A שייך ל- \mathbb{F} אזי גם המשלים שלו A^c שייך ל- \mathbb{F} .

4. איחוד ניתן להמנות של מאורעות אף הוא מאורע. בסימון מתמטי, אם $A_i \in \mathbb{F}$ עבור $i = 1, 2, \dots$ אזי גם $\bigcup_{i=1}^{\infty} A_i \in \mathbb{F}$.

אוסף תת קבוצות המקיים את כל התנאים לעיל נקרא סיגמה שדה (σ -field) או סיגמה אלגברה (σ -algebra): שני שמות לאותו מושג).

הגדרה 8.5 הקבוצה הריקה, כלומר הקבוצה שאינה מכילה דבר, תסומן ב- \emptyset . זוג מאורעות A, B נקראים זרים אם אין להם דבר משותף, כלומר אם $A \cap B = \emptyset$.

הגדרה 8.6 פונקציית הסתברות, או הסתברות $\mathbb{P}\{A\}$ היא פונקציה המוגדרת עבור כל מאורע (ב- \mathbb{F}), ומקיימת את התנאים הבאים:

$$1. \quad \mathbb{P}\{A\} \leq 1 \quad \text{לכל } A \in \mathbb{F}.$$

$$2. \quad \mathbb{P}\{\Omega\} = 1.$$

3. לכל אוסף (בן מניה) של מאורעות $\{A_i, i = 1, 2, \dots\}$ שהם זרים (כלומר $A_i \cap A_j = \emptyset$ אם $i \neq j$), מתקיים

$$\mathbb{P}\left\{\bigcup_{i=1}^{\infty} A_i\right\} = \sum_{i=1}^{\infty} \mathbb{P}\{A_i\}$$

מ-2 ו-3 נובע כי לכל מאורע A מתקיים $\mathbb{P}\{A\} + \mathbb{P}\{A^c\} = 1$.

הגדרה 8.7 לשלשה $\{\Omega, \mathbb{F}, \mathbb{P}\}$ קוראים מרחב הסתברות.

8.2 משתנה אקראי ופילוג

דוגמה 8.8 בדוגמאות 8.2–8.3 תארנו מאורעות פשוטים. עבור רעש היציאה מהמגבר, אפשר לשאול מהי ההסתברות שהאנרגיה תהיה לא גדולה מדי: למשל $\int_0^1 n^2(t) dt \leq 3$ כאשר $n(t) = n(t, \omega)$ הוא תהליך הרעש (את "משתנה המזל" ω נשמיט בדרך כלל, כדי שהסימון יהיה פשוט יותר: אך הוא קיים ומשפיע!). כאן דרושה יותר זהירות מתמטית כדי לוודא שאכן זהו מאורע (כי אם לא, אזי ההסתברות אינה מוגדרת!).

הגדרה 8.9 משתנה אקראי (בקיצור: מ"א) הוא פונקציה $X(\omega)$ על מרחב המדגם Ω כך שעבור כל מספר ממשי a הקבוצה $\{\omega : X(\omega) \leq a\}$ היא מאורע.

שים לב שמקובל להשמיט את המשתנה ω , כלומר מקובל לרשום X כאשר מתכוונים ל- $X(\omega)$. בצורה דומה, לפעמים נרשום בקיצור (למשל) $\{X(\omega) \leq a\}$, כאשר אנו מתכוונים לתאר את המאורע $\{\omega : X(\omega) \leq a\}$.

דוגמה 8.10 נסמן את המשתנה האקראי ב- X . נניח שהרעש ביציאה ממגבר נתון ע"י $n(t) = n(t, \omega)$ ונניח שהרעש רציף (כפונקציה של הזמן). אזי ניתן להגדיר את המ"א הבאים:

• הערך של רעש המגבר ברגע $t = 2$ למשל (מתח חשמלי, נמדד בוולט V), הוא $X(\omega) = n(2, \omega) = n(t, \omega)|_{t=2}$.

• אפשר להגדיר מ"א מורכבים יותר: למשל האנרגיה של אות הרעש בתחום בזמן $[0, 1]$ היא $X(\omega) = \int_0^1 n^2(t, \omega) dt$.

• הערך המקסימלי של אות הרעש בתחום $[0, 1]$ הוא $Y = \max_{0 \leq t \leq 1} n(t, \omega)$.

הערה מתמטית: ברור כי $\{\omega : X(\omega) \leq a\}$ היא קבוצה ב- Ω . אך כדי להראות ש- $X(\omega)$ הוא משתנה אקראי, יש לוודא כי זהו מאורע. לשם כך יש להתייחס להגדרה של \mathbb{F} . כל זה ניתן להעשות, אך לא במסגרת הנוכחית.

הגדרה 8.11 פונקציית הפילוג של משתנה אקראי $X(\omega)$ היא פונקציה של משתנה ממשי, המקבלת ערכים בן 0 ו-1, והגדרתה

$$F_X(a) = \mathbb{P}\{X(\omega) \leq a\}$$

בהגדרה זו, הסימן X מזהה שאנו עוסקים בפונקציית הפילוג של המשתנה האקראי $X(\omega)$. פונקציית הפילוג $F_X(a)$ היא פונקציה של המשתנה a בלבד.

טענה 8.12 תכונות פונקציית הפילוג $F_X(a)$ (ללא הוכחה):

- פונקציית הפילוג היא מונוטונית (לא יורדת), ורציפה מימין,
- (בסימון לא לגמרי מדויק מתמטית) $F_X(+\infty) = 1$ ו- $F_X(-\infty) = 0$. כלומר, משתנה אקראי מקבל רק ערכים סופיים.
- אם $a_1 < a_2$ אזי

$$\mathbb{P}\{a_1 < X \leq a_2\} = F_X(a_2) - F_X(a_1)$$

אם X הוא משתנה בדיד, אזי $F_X(a)$ קבועה על פני קטעים, ועולה בקפיצות (ציר את פונקציית הפילוג של קוביה).

הגדרה 8.13 אם קיימת פונקציה (אינטגרלית רימן), שנסמן ב- $f_X(\theta)$, כך שלכל a

$$(8.1) \quad F_X(a) = \int_{-\infty}^a f_X(\theta) d\theta$$

אזי f_X תקרא פונקציית הצפיפות של המ"א X , או הפילוג הסגולי של המ"א X .

8.3 ווקטור אקראי

הגדרה 8.14 ווקטור אקראי \underline{X} במימד N הוא אוסף של N משתנים אקראיים:

$$\underline{X}(\omega) = \{X_1(\omega), X_2(\omega), \dots, X_N(\omega)\}'$$

כמובן ש- ω הוא משותף לכל הרכיבים (אותו פרמטר מזל לכל רכיבי הווקטור $X_i(\omega)$). הפילוג של ווקטור אקראי הוא הרחבה של מושג הפילוג של משתנה אקראי

הגדרה 8.15 פונקציית הפילוג של ווקטור אקראי $\underline{X}(\omega) = \{X_1(\omega), X_2(\omega), \dots, X_N(\omega)\}'$ היא פונקציה של ווקטור ממשי $\underline{a} = \{a_1, a_2, \dots, a_N\}'$ המקבלת ערכים בן 0 ו-1, והגדרתה

$$(8.2) \quad F_{\underline{X}}(\underline{a}) = \mathbb{P}\{X_1(\omega) \leq a_1, X_2(\omega) \leq a_2, \dots, X_N(\omega) \leq a_N\}$$

לוקטור אקראי יש פונקציית צפיפות (ראה 8.13) אם קיימת פונקציה $f_{\underline{X}}(\underline{a})$ כך שלכל \underline{a}

$$F_{\underline{X}}(\underline{a}) = \int_{-\infty}^{a_1} \cdots \int_{-\infty}^{a_N} f_{\underline{X}}(\underline{a}) da_1 \cdots da_N$$

ניתן לחשב אותה ע"י

$$f_{\underline{X}}(\underline{a}) = \frac{\partial^N F_{\underline{X}}(\underline{a})}{\partial a_1 \partial a_2 \cdots \partial a_N}$$

8.4 אי תלות סטטיסטית

הגדרה 8.16 אי תלות סטטיסטית.

זוג מאורעות A, B נקראים בלתי תלויים סטטיסטית (בת"ס) אם $\mathbb{P}\{A \cap B\} = \mathbb{P}\{A\} \cdot \mathbb{P}\{B\}$.
 זוג משתנים אקראיים X, Y נקראים בלתי תלויים סטטיסטית (בת"ס) אם לכל זוג מספרים a, b המאורעות $\{\omega : X \leq a\}$ ו- $\{\omega : Y \leq b\}$ הם בת"ס. במילים אחרות, המשתנים האקראיים X, Y הם בת"ס אם לכל a, b מתקיים

$$\mathbb{P}\{X \leq a, Y \leq b\} = \mathbb{P}\{X \leq a\} \cdot \mathbb{P}\{Y \leq b\}$$

או, בסימון שונה,

$$F_{X,Y}(a, b) = F_X(a) \cdot F_Y(b)$$

בצורה זהה, שני וקטורים $\underline{X}, \underline{Y}$ נקראים בת"ס אם לכל $\underline{a}, \underline{b}$ מתקיים

$$F_{\underline{X}, \underline{Y}}(\underline{a}, \underline{b}) = F_{\underline{X}}(\underline{a}) \cdot F_{\underline{Y}}(\underline{b})$$

אם X, Y בת"ס ו- g היא פונקציה (דטרמיניסטית) כלשהיא (למדקדקים-פונקציית בורל) אזי $X, g(Y)$ גם הם בת"ס.

8.5 תוחלת

הגדרה 8.17 את התוחלת של משתנה אקראי X נסמן באחת מהצורות $\mathbb{E} X = \bar{X} = m_X$. התוחלת מוגדרת בצורה הבאה.

• יהי X משתנה אקראי המקבל ערכים בדידים, למשל $\{\alpha_i, i = 1, 2, \dots\}$. התוחלת מוגדרת בתנאי ש-

$$\sum_{i=1}^{\infty} |\alpha_i| \mathbb{P}\{X_i = \alpha_i\} < \infty$$

ואם תנאי זה מתקיים אזי התוחלת היא

$$\mathbb{E} X = \sum_{i=1}^{\infty} \alpha_i \mathbb{P}\{X_i = \alpha_i\}$$

- נניח שלמשתנה X יש צפיפות סגולית. אזי התוחלת מוגדרת בתנאי ש-

$$\int_{-\infty}^{\infty} |\alpha| f_X(\alpha) d\alpha < \infty$$

ואם תנאי זה מתקיים אזי התוחלת היא

$$\mathbb{E} X = \int_{-\infty}^{\infty} \alpha f_X(\alpha) d\alpha$$

- באופן כללי, מגדירים תוחלת על ידי קרובים. מקרבים את המשתנה על ידי משתנה בדיד, משתמשים בנוסחה של משתנה בדיד, ומשפרים את הקירוב. ביתר פירוט, לכל מספר n נבחר חלוקה (סופית) של הקטע $(-n, n)$ ל- $k(n)$ קטעים. נדרוש מהחלוקה שתקיים

$$-n = \alpha_1^{(n)} < \alpha_2^{(n)} < \dots < \alpha_i^{(n)} < \alpha_{i+1}^{(n)} < \dots < \alpha_{k(n)-1}^{(n)} < \alpha_{k(n)}^{(n)} = n$$

$$\max_i \alpha_{i+1}^{(n)} - \alpha_i^{(n)} \rightarrow 0$$

כאשר $n \rightarrow \infty$ (לשם כך צריך ש- $k(n)$ יגדל מהר יותר מ- n). התוחלת של משתנה אקראי X מוגדרת בתנאי שקיים קבוע B כך ש-

$$\sum_{i=1}^{k(n)} |\alpha_i^{(n)}| \left[F_X(\alpha_{i+1}^{(n)}) - F_X(\alpha_i^{(n)}) \right] < B$$

לכל n גדול מספיק. אם תנאי זה מתקיים, אזי התוחלת היא

$$(8.3) \quad \mathbb{E}[X] = \lim_{n \rightarrow \infty} \sum_{i=1}^{k(n)} \alpha_i^{(n)} \left[F_X(\alpha_{i+1}^{(n)}) - F_X(\alpha_i^{(n)}) \right]$$

בגלל צורת הסכום, מסמנים גבול זה כאינטגרל, באחת מהצורות הבאות:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} \alpha dF_X(\alpha) = \int_{-\infty}^{\infty} \alpha F_X(d\alpha)$$

אינטגרל כזה נקרא אינטגרל סטילצ'ס-לבג Stieltjes-Lebesgue.

תרגיל 8.18 בדוק שההגדרה הכללית של תוחלת נותנת תוצאה זהה להגדרות הקודמות כאשר המשתנה האקראי מקבל ערכים בדידים, וכאשר למשתנה האקראי יש צפיפות.

טענה 8.19 בהנתן משתנה אקראי X ופונקציה g נגדיר משתנה אקראי $Y = g(X)$. אזי

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{-\infty}^{\infty} \alpha dF_Y(\alpha) = \int_{-\infty}^{\infty} g(\alpha) dF_X(\alpha)$$

כלומר, כדי לחשב את התוחלת של המשתנה Y אין צורך לחשב את הפילוג שלו: אפשר לבצע את החישוב בעזרת פונקציית הפילוג של X .

לא לכל מ"א יש תוחלת סופית, ולא תמיד ניתן להגדיר תוחלת. לדוגמה, עבור מ"א עם פונקציות פילוג סגולי

$$f_X(\alpha) = \begin{cases} 0 & \alpha \geq 0 \\ \frac{2/\pi}{1+\alpha^2} & \alpha < 0 \end{cases}$$

$$f_Y(\alpha) = \frac{1/\pi}{1+\alpha^2}$$

מתקיים כי

$$\int_{-\infty}^{\infty} \alpha f_X(\alpha) d\alpha = \int_{-\infty}^0 \alpha f_Y(\alpha) d\alpha = -\infty$$

כך שאפשר להגדיר תוחלת עבור המשתנה X , אולם התוחלת היא אי-סופית. לעומת זאת, כיוון ש-

$$\int_0^{\infty} \alpha f_Y(\alpha) d\alpha = \infty$$

ולכן עבור Y לא ניתן כלל להגדיר תוחלת.

בהינתן מאורע A נסמן ב- A^c את המאורע המשלים (כלומר $\omega \in A^c$ אם ורק אם $\omega \notin A$). נסמן ב- I_A את הפונקציה המציינת של המאורע A . זהו משתנה אקראי המוגדר כך:

$$I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \in A^c \end{cases}$$

טענה 8.20 תכונות התוחלת.

1. לכל מאורע A ,

$$\mathbb{E}[I_A] = 0 \cdot \mathbb{P}\{A^c\} + 1 \cdot \mathbb{P}\{A\} = \mathbb{P}\{A\}$$

2. אם $X = C$ קבוע שאינו אקראי, אזי $\mathbb{E}X = C$.

3. לינאריות: נניח שלמשתנים X, Y יש תוחלת, ויהיו a, b זוג קבועים. אזי

$$\mathbb{E}[aX + bY] = a \cdot \mathbb{E}[X] + b \cdot \mathbb{E}[Y]$$

בין אם המשתנים תלויים סטטיסטית ובין אם לאו.

4. אם המשתנים X, Y בלתי תלויים סטטיסטית (ויש להם תוחלת) אזי

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

5. אם $X \geq Y$ (או $\mathbb{P}\{X \geq Y\} = 1$) אזי $\mathbb{E}[X] \geq \mathbb{E}[Y]$, כאשר שוויון יתכן רק אם $\mathbb{P}\{X = Y\} = 1$.

טענה 8.21 אם המשתנה X מקבל ערכים שלמים וחיוביים, אזי

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{P}\{X \geq k\}$$

הוכחה: לפי הגדרת התוחלת למקרה זה,

$$\begin{aligned} \mathbb{E}[X] &= \sum_{k=0}^{\infty} k \mathbb{P}\{X = k\} \\ &= \sum_{k=1}^{\infty} \mathbb{P}\{X = k\} + \sum_{k=1}^{\infty} (k-1) \mathbb{P}\{X = k\} \\ &= \mathbb{P}\{X \geq 1\} + \sum_{k=2}^{\infty} (k-1) \mathbb{P}\{X = k\} \\ &= \mathbb{P}\{X \geq 1\} + \sum_{k=2}^{\infty} \mathbb{P}\{X = k\} + \sum_{k=2}^{\infty} (k-2) \mathbb{P}\{X = k\} \\ &= \mathbb{P}\{X \geq 1\} + \mathbb{P}\{X \geq 2\} + \sum_{k=3}^{\infty} (k-2) \mathbb{P}\{X = k\} \\ &= \mathbb{P}\{X \geq 1\} + \mathbb{P}\{X \geq 2\} + \dots \\ &= \sum_{k=1}^{\infty} \mathbb{P}\{X \geq k\} \end{aligned}$$

8.6 מומנטים

לתוחלות של חזקות של משתנה אקראי קוראים מומנטים.

הגדרה 8.22 יהי X משתנה אקראי. נגדיר

- מומנט שני: $\mathbb{E}[X^2]$
- מומנט מסדר n : $\mathbb{E}[X^n]$.
- מומנט מוחלט מסדר n : $\mathbb{E}[|X|^n]$.
- מומנט מרכזי מסדר n : $\mathbb{E}\left[(X - \bar{X})^n\right]$

• שונות (ווריאנס): $\text{Var}(X) = \mathbb{E}[(X - \bar{X})^2]$. סטיית התקן $\sigma_X = \sqrt{\text{Var}(X)}$.

• מ"א X נקרא מנורמל אם $\mathbb{E}[X] = 0$ ו- $\mathbb{E}[X^2] = 1$.

לפי ההגדרה, אם X הוא מ"א כל שהוא בעל מומנט שני סופי, אזי המ"א

$$Z = \frac{X - \mathbb{E}[X]}{\sigma_X} \quad (8.4)$$

הוא מ"א מנורמל.

מתוך חוק ההסתברות של מ"א אפשר לחשב את כל המומנטים שלו. מצד שני, התוחלת (מומנט ראשון) היא קירוב דטרמיניסטי סביר למ"א, והמומנט המרכזי השני מתאר מהו הפיזור סביב הקירוב הדטרמיניסטי. מומנטים גבוהים יותר מספקים מידע נוסף על הפילוג.

הערה: באופן כללי, אם מתקיים התנאי הטכני ש- $\mathbb{E}[|X|^n]$ אינו עולה מהר מדי, אזי המומנטים מגדירים את חוק הפילוג באופן חד משמעי.

יהיו X, Y זוג משתנים אקראיים.

הגדרה 8.23 הקווריאנס של זוג מ"א מוגדר ע"י

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \bar{X})(Y - \bar{Y})]$$

אם $\text{Cov}(X, Y) = 0$ נאמר שהמ"א הם חסרי קורלציה, או בלתי מתואמים, או בלתי תלויים לינארית (להבדיל מבלתי תלויים סטטיסטית!). מקדם הקורלציה, או מקדם המיתאם ρ בין המ"א מוגדר ע"י

$$\rho = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}$$

אם X, Y בלתי תלויים לינארית אזי $\rho = 0$. הבה נראה כי $|\rho| \leq 1$. כיוון שמקדם המתאם אינו תלוי במוצג, נניח שלשני המשתנים ממוצע 0. לצורך החישוב נתבונן בביטוי החיובי $\mathbb{E}[(X - \lambda Y)^2]$ ונחפש את הערך של המספר λ עבורו הביטוי יהיה מינימלי. כדי למצוא מינימום, נגזור לפי λ ונשווה ל-0. מהגזירה נקבל כי המינימום מושג עבור

$$\lambda^* = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}$$

אם נציב ערך זה, נקבל

$$\begin{aligned} 0 &\leq \mathbb{E}[(X - \lambda^* Y)^2] \\ &= \mathbb{E}[X^2] - 2\lambda^* \text{Cov}(X, Y) + (\lambda^*)^2 \mathbb{E}[Y^2] \\ &= \mathbb{E}[X^2] - \frac{(\text{Cov}(X, Y))^2}{\text{Var}(Y)} \end{aligned}$$

ומכאן ש-

$$\text{Var}(X) \text{Var}(Y) \geq (\text{Cov}(X, Y))^2$$

או $|\rho| \leq 1$ כנדרש.

כסיכום בניים, נתאר קשר מפורסם בין מומנטים לבין הסתברויות. קשר זה מראה כיצד מומנטים יכולים לסייע לשם קירוב הסתברויות מסוימות.

משפט 8.24 יהי X משתנה אקראי כלשהוא ותהי g פונקציה חיובית ועולה. נדרוש בנוסף כי $\mathbb{E}[g(X)]$ מוגדר היטב וסופי. אזי לכל מספר α ,

$$(8.5) \quad \mathbb{P}\{X \geq \alpha\} \leq \frac{\mathbb{E}[g(X)]}{g(\alpha)}$$

הוכחה: נשתמש בהוכחה זו בסימון הכללי (ראה 3.8) לתוחלת כאינטגרל. לטובת האינטואיציה, אפשר לחשוב על הביטוי $dF_X(x)$ כקיצור ל- $f_X(x) dx$. נקבע את α ונחשב

$$\begin{aligned} \mathbb{E}[g(X)] &= \int_{-\infty}^{\infty} g(x) dF_X(x) \\ &\geq \int_{-\infty}^{\infty} I_{\{x \geq \alpha\}}(x) g(x) dF_X(x) \\ &\geq \int_{-\infty}^{\infty} I_{\{x \geq \alpha\}}(x) g(\alpha) dF_X(x) \end{aligned}$$

כאשר אי השוויון הראשון מתקיים כי g חיובית, והשני כי היא פונקציה עולה. כעת נשתמש בתכונות התוחלת ונקבל

$$\begin{aligned} \mathbb{E}[g(X)] &\geq g(\alpha) \int_{-\infty}^{\infty} I_{\{x \geq \alpha\}}(x) dF_X(x) \\ &= g(\alpha) \mathbb{P}\{X \geq \alpha\} \end{aligned}$$

ולכן, כפי שטעננו,

$$\mathbb{P}\{X \geq \alpha\} \leq \frac{\mathbb{E}[g(X)]}{g(\alpha)}$$

ממשפט זה אפשר להקיש מספר מסקנות מפורסמות. נביא תחילה את משפט מרקוב (Markov).

טענה 8.25 לכל משתנה אקראי חיובי X ולכל $\alpha > 0$ מתקיים

$$\mathbb{P}\{X \geq \alpha\} \leq \frac{\mathbb{E}[X]}{\alpha}$$

טענה זו נובעת מייד מהבחירה $g(x) = x$, כיוון ש- X חיובי. למדקדקים, שימו לב שכיוון ש- X חיובי, אפשר להגדיר לו תוחלת ללא כל תנאי. כמובן שאים התוחלת איך-סופית, הטענה היא חסרת תועלת...

טענה נוספת הנובעת ממשפט זה נקראת חסם צ'בישב (Chebyshev).

טענה 8.26 לכל משתנה אקראי X ולכל α חיובי מתקיים

$$\mathbb{P}\{|X| \geq \alpha\} \leq \frac{\mathbb{E}[X^2]}{\alpha^2}$$

$$\mathbb{P}\{|X - \mathbb{E}[X]| \geq \alpha\} \leq \frac{\text{Var}[X]}{\alpha^2}$$

השורה הראשונה נובעת מיידית מהמשפט, כיוון שלכל α חיובי,

$$\mathbb{P}\{|X| \geq \alpha\} = \mathbb{P}\{|X|^2 \geq \alpha^2\}$$

והטענה השנייה נובעת מההגדרה של ווריאנס, ע"י הפעלת המשפט על המשתנה האקראי $X - \mathbb{E}[X]$.

חסם חדש יחסית הנובע מאותו משפט הוא חסם צ'רנוב (Chernoff).

טענה 8.27 לכל משתנה אקראי X , לכל α ולכל $\theta \geq 0$ מתקיים

$$\mathbb{P}\{X \geq \alpha\} \leq \mathbb{E}\left[e^{\theta(X-\alpha)}\right]$$

בתנאי שהתוחלת בצד ימין קיימת.

נשים לב שהפונקציה $e^{\theta \cdot x}$ של המשתנה x היא פונקציה חיובית ועולה. לכן אפשר להפעיל את המשפט ולקבל

$$\mathbb{P}\{X \geq \alpha\} \leq \frac{\mathbb{E}\left[e^{\theta \cdot X}\right]}{e^{\theta \cdot \alpha}}$$

$$= \mathbb{E}\left[e^{\theta(X-\alpha)}\right]$$

8.7 פונקציה אפיינית

הפונקציה האפיינית של ϕ_X של מ"א X היא פונקציה של המשתנה הממשי ν . היא מוגדרת ע"י

$$\phi_X(\nu) = \mathbb{E}\left[e^{i\nu X}\right] = \int_{-\infty}^{\infty} e^{i\nu x} dF_X(x).$$

אם למשתנה יש פילוג סגולי (צפיפות) $f_X(\alpha)$

$$\phi_X(\nu) = \int_{-\infty}^{\infty} e^{i\nu \alpha} f_X(\alpha) d\alpha$$

כלומר, $\phi_x(\nu)$ הוא הצמוד המרוכב של התמרת פוריה של $f_x(\alpha)$. אפשר להראות כי בכל מקרה, הפונקציה האפיינית ϕ_x מגדירה חד משמעית את פונקציית הפילוג $F_x(\cdot)$.

שם לב כי הפונקציה האפיינית מוגדרת היטב (ולכן תמיד קיימת), כי

$$|e^{i\alpha}| = 1$$

לכל α . אם קיימים מומטים מכל סדר, אזי ניתן לייצג את הפונקציה האפיינית על ידי טור חזקות

$$\phi_x(\nu) = \sum_{k=0}^{\infty} \frac{(i\nu)^k}{k!} \mathbb{E}[X^k]$$

כאשר (תחת תנאים טכניים)

$$\phi_x(0) = 1$$

$$\left. \frac{\partial \phi_x(\nu)}{\partial \nu} \right|_{\nu=0} = i \mathbb{E}[X]$$

$$\left. \frac{\partial^k \phi_x(\nu)}{\partial \nu^k} \right|_{\nu=0} = (i)^k \mathbb{E}[X^k]$$

כך שבפרט, ידיעת הפונקציה האפיינית מאפשרת חישוב של המומנטים.

8.8 הסתברות ותוחלת מותנים

הגדרה 8.28 ההסתברות של מאורע A בהנתן מאורע B מוגדרת על ידי

$$\mathbb{P}\{A | B\} = \frac{\mathbb{P}\{A \cap B\}}{\mathbb{P}\{B\}}$$

שים לב כי עבור B קבוע, זוהי הסתברות (כפונקציה של A) המרוכזת בקבוצה B , ולכן יש לה את כל התכונות של הסתברות (הגדרה 8.6). הטענה הבאה שימושית מאד, ונובעת ישירות מההגדרה:

טענה 8.29 עבור מאורעות $\{A_k, k = 1, 2, \dots, K\}$ המקיימים $\mathbb{P}\{A_k\} \neq 0$ מתקיים

$$\mathbb{P}\{A_1 \cap A_2\} = \mathbb{P}\{A_1 | A_2\} \cdot \mathbb{P}\{A_2\}$$

$$\mathbb{P}\{A_1 \cap A_2 | A_3\} = \mathbb{P}\{A_1 | A_2 \cap A_3\} \cdot \mathbb{P}\{A_2 | A_3\}$$

$$\mathbb{P}\{\cap_{k=1}^K A_k\} = \mathbb{P}\{A_1 | \cap_{m=2}^K A_m\} \cdot \mathbb{P}\{A_2 | \cap_{m=3}^K A_m\} \times \dots$$

$$\times \mathbb{P}\{A_{K-2} | A_{K-1} \cap A_K\} \cdot \mathbb{P}\{A_{K-1} | A_K\} \cdot \mathbb{P}\{A_K\}$$

$$\mathbb{P}\{\cap_{k=1}^{K-1} A_k | A_K\} = \mathbb{P}\{A_1 | \cap_{m=2}^K A_m\} \cdot \mathbb{P}\{A_2 | \cap_{m=3}^K A_m\} \times \dots$$

$$\times \mathbb{P}\{A_{K-2} | A_{K-1} \cap A_K\} \cdot \mathbb{P}\{A_{K-1} | A_K\}$$

נניח כעת כי X, Y הם משתנים אקראיים, ונגדיר עבור β קבוע את המאורע $B = \{\omega : X(\omega) = \beta\}$. אזי ניתן לרשום

$$\mathbb{P}\{A | B\} = \mathbb{P}\{A | X(\omega) = \beta\} = \frac{\mathbb{P}\{A \cap B\}}{\mathbb{P}\{B\}}$$

אולם במיקרים רבים (למשל אם למשתנה X יש צפיפות), ההסתברות של המאורע B היא אפס: $\mathbb{P}\{B\} = 0$. כיצד נגדיר אז את ההסתברות המותנית?

הגדרה 8.30 עבור מאורע A ומשתנה אקראי X נגדיר

$$\mathbb{P}\{A | X(\omega) = \beta\} = \lim_{\epsilon \rightarrow 0} \mathbb{P}\{A | \beta \leq X(\omega) \leq \beta + \epsilon\}$$

אנו נניח שגבול זה תמיד קיים, ומגדיר הסתברות מותנית כך שכל התכונות מתקיימות.

הגדרה 8.31 הפילוג המותנה של משתנה Y בהנתן המשתנה X מוגדר כ-

$$F_{Y|X}(\alpha | \beta) = \mathbb{P}\{Y \leq \alpha | X = \beta\}$$

צד ימין הוגדר למעלה. אם X משתנה עם צפיפות, נגדיר אותו דרך הגבול.

אם X, Y בלתי תלויים סטטיסטית, אזי ההתניה אינה מוסיפה כל מידע.

טענה 8.32 אם המאורעות A, B הם בלתי תלויים אזי

$$\mathbb{P}\{A | B\} = \mathbb{P}\{A\} \quad (8.6)$$

לכן, אם המשתנים X, Y הם בלתי תלויים סטטיסטית, אזי

$$F_{Y|X}(\alpha | \beta) = F_Y(\alpha)$$

שתי הטענות נובעות מהגדרת הסתברות מותנית: אם A, B בת"ס אזי

$$\mathbb{P}\{A | B\} = \frac{\mathbb{P}\{A \cap B\}}{\mathbb{P}\{B\}} = \frac{\mathbb{P}\{A\} \mathbb{P}\{B\}}{\mathbb{P}\{B\}} = \mathbb{P}\{A\}$$

הטענה השנייה נובעת מכך ומהגדרת הפילוג המותנה.

הגדרה 8.33 אם קיימת פונקציה f כך ש-

$$F_{Y|X}(\alpha | \beta) = \int_{-\infty}^{\alpha} f_{Y|X}(\theta | \beta) d\theta$$

אזי f תקרא פילוג סגולי מותנה או צפיפות מותנית של Y בהנתן X .

מההגדרה של צפיפות מותנית נובע כי אם יש ל- X ו- Y צפיפות משותפת $f_{X,Y}(\alpha, \beta)$ אזי אפשר לחשב את הצפיפות המותנית כלהלן:

$$\begin{aligned} f_{Y|X}(\alpha | \beta) &= \lim_{\delta \rightarrow 0} \frac{F_{Y|X}(\alpha + \delta | \beta) - F_{Y|X}(\alpha | \beta)}{\delta} \\ &= \lim_{\delta \rightarrow 0} \frac{\mathbb{P}\{\alpha \leq Y \leq \alpha + \delta | X = \beta\}}{\delta} \\ &= \lim_{\delta \rightarrow 0} \left[\lim_{\epsilon \rightarrow 0} \frac{\mathbb{P}\{\alpha \leq Y \leq \alpha + \delta, \beta \leq X \leq \beta + \epsilon\}}{\delta \cdot \mathbb{P}\{\beta \leq X \leq \beta + \epsilon\}} \right] \\ &= \frac{\delta \cdot \epsilon \cdot f_{X,Y}(\alpha, \beta)}{\delta \cdot \epsilon \cdot f_X(\beta)} \\ &= \frac{f_{X,Y}(\alpha, \beta)}{f_X(\beta)} \end{aligned}$$

לסיכום, אם יש צפיפות משותפת אזי

$$(8.7) \quad \boxed{f_{Y|X}(\alpha | \beta) = \frac{f_{X,Y}(\alpha, \beta)}{f_X(\beta)}}$$

ההסתברות המותנית $\mathbb{P}\{Y \leq \alpha | X = \beta\}$ וכן גם הפילוג המותנה $F_{Y|X}(\alpha | \beta)$ והצפיפות המותנית $f_{Y|X}(\alpha | \beta)$ הם כולם פונקציות של המשתנה β . ניתן כעת להציב במקום המספר β , את המשתנה האקראי X . יתקבל, כמובן, משתנה אקראי חדש (תלוי ב- ω) המוגדר על ידי

$$\mathbb{P}\{Y \leq \alpha | X\} = \mathbb{P}\{Y \leq \alpha | X = \beta\}|_{\beta=X}$$

מהגדרה זו נובע שהשוויון

$$F_{Y|X}(\alpha | X) = \mathbb{P}\{Y \leq \alpha | X = \beta\}$$

יתקיים אם ורק אם ω מקיים $X(\omega) = \beta$. בצורה דומה יש להבין את המשמעות של $f_{Y|X}(\alpha | X)$.

כאמור, עבור β קבוע הפילוג המותנה הוא בעל כל התכונות של פילוג "רגיל", אפשר להגדיר תוחלת מותנית בדרך שהגדרנו תוחלת.

הגדרה 8.34 התוחלת המותנית של משתנה אקראי Y בהנתן שהמשתנה X מקבל את הערך β מוגדרת על ידי

$$\mathbb{E}[Y | X = \beta] = \int_{-\infty}^{\infty} \alpha dF_{Y|X}(\alpha | \beta) = \int_{-\infty}^{\infty} \alpha F_{Y|X}(d\alpha | \beta)$$

איים יש פילוג סגולי מותנה, אזי

$$\mathbb{E}[Y | X = \beta] = \int_{-\infty}^{\infty} \alpha f_{Y|X}(\alpha | \beta) d\alpha$$

עבור פונקציה כלשהי g מתקיים

$$\mathbb{E}[g(Y) | X = \beta] = \int_{-\infty}^{\infty} g(\alpha) F_{Y|X}(d\alpha | \beta) = \int_{-\infty}^{\infty} g(\alpha) f_{Y|X}(\alpha | \beta) d\alpha$$

כאשר השוויון האחרון מתקיים במידה וקיים פילוג סגולי מותנה.

כמו במקרה של תוחלת רגילה, התוחלת המותנית שהגדרנו היא פונקציה של משתנה ההתנייה β , ולכן כמו במקרה של הסתברות (פילוג או צפיפות) מותנים, אפשר להציב את המשתנה המתנה X במקום β .

$$\mathbb{E}[Y | X] = \mathbb{E}[Y | X = \beta]_{\beta=X}$$

הערה למדקדקים: ההגדרה המדויקת של תוחלת מותנית היא מופשטת יותר (ומדויקת יותר). התוחלת המותנית (כמישתנה אקראי) מוגדרת רק עד כדי מאורע שהסתברותו אפס: לכן כל שוויון שמופיעה בו תוחלת מותנית יהיה נכון פרט לקבוצת ω שהסתברותה אפס. בהמשך לא נתייחס לנקודה זו.

עבור β קבוע התוחלת המותנית מחושבת כמו תוחלת רגילה. לכן היא יורשת את תכונות התוחלת: אולם יש לתוחלת המותנית גם תכונות יחודיות.

טענה 8.35 תכונות התוחלת המותנית. התוחלת המותנית מוגדרת לכל משתנה שיש לו תוחלת סופית. בתאור התכונות לעיל נניח תמיד שמתקיימים התנאים הדרושים ככדי שהתוחלת המותנית תהיה קיימת. בדומה לתוחלת הרגילה,

1. לכל מאורע A ולכל משתנה אקראי X ,

$$\mathbb{E}[I_A | X] = \mathbb{P}\{A | X\}$$

2. אם $Y = C$ קבוע שאינו אקראי, אזי $\mathbb{E}[Y | X] = C$.

3. לינאריות: יהיו a, b זוג קבועים. אזי

$$\mathbb{E}[aZ + bY | X = \beta] = a \cdot \mathbb{E}[Z | X = \beta] + b \cdot \mathbb{E}[Y | X = \beta]$$

$$\mathbb{E}[aZ + bY | X] = a \cdot \mathbb{E}[Z | X] + b \cdot \mathbb{E}[Y | X]$$

4. אם $Z \geq Y$ (או $\mathbb{P}\{Z \geq Y | X = \beta\} = 1$) אזי $\mathbb{E}[Z | X = \beta] \geq \mathbb{E}[Y | X = \beta]$. התכונות הבאות הן מיוחדות לתוחלת המותנית.

5. $\mathbb{E}[X | X] = X$ ולכן גם $\mathbb{E}[X | X = \beta] = \beta$.

6. עבור פונקציות g, h כלשהן,

$$\mathbb{E}[g(X) \cdot h(Y) \mid X = \beta] = g(\beta) \mathbb{E}[h(Y) \mid X = \beta]$$

ובאופן כללי יותר, אם h תלוייה בשני המשתנים X, Y

$$\mathbb{E}[g(X) \cdot h(X, Y) \mid X = \beta] = g(\beta) \mathbb{E}[h(\beta, Y) \mid X = \beta]$$

לכן, לפי ההגדרות, מתקיים בשני המקרים בהתאמה

$$\mathbb{E}[g(X) \cdot h(Y) \mid X] = g(X) \mathbb{E}[h(Y) \mid X]$$

$$\mathbb{E}[g(X) \cdot h(X, Y) \mid X] = g(X) \mathbb{E}[h(X, Y) \mid X]$$

7.

$$\mathbb{E}[\mathbb{E}(Y \mid X)] = \mathbb{E}[Y]$$

ואם המשתנים X, Y בלתי תלויים סטטיסטית אזי

$$\mathbb{E}[Y \mid X] = \mathbb{E}[Y]$$

8.

$$\mathbb{E}[g(X) \cdot h(X, Y)] = \mathbb{E}[g(X) \cdot \mathbb{E}(h(X, Y) \mid X)]$$

הערות: תכונה 5 אומרת כי אים ערכו של X כמתנה נקבע לערך מסויים, אז אמנם אפשר להתייחס ל- X כקבוע (גם בתפקידו כמותנה, לא רק כמתנה). תכונות 7 ו-8 נקראות "החלקה".

תכונה 5 נובעת מההגדרות, שכן הפילוג המשותף של X עם עצמו הוא "דלתה". את תכונה 6 נראה בעזרת הפילוג המותנה. לצורך ההוכחה נגדיר משתנה חדש Z שהגדרתו היא $Z = X$. הגדרה זו תאפשר להפריד בין תפקידים שונים של אותו המשתנה X . בסימון החדש, עלינו להראות כי

$$\mathbb{E}[g(Z) \cdot h(Z, Y) \mid X = \beta] = g(\beta) \mathbb{E}[h(\beta, Y) \mid X = \beta]$$

לפי ההגדרה,

$$\begin{aligned} \mathbb{E}[g(Z) \cdot h(Z, Y) \mid X = \beta] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(z)h(z, y)F_{Y,Z|X}(dz dy \mid \beta) \\ &= \int_{-\infty}^{\infty} g(\beta)h(\beta, y)F_{Y,Z|X}(dy \mid \beta) \end{aligned}$$

מכיוון מהגדרת הפילוג המותנה המשותף (ראה בפרט את הנוסחה לצפיפות המותנית), הפילוג מתרכז ב- $X = Z$ כך שהצפיפות המותנית היא פונקציה דלתה. בשלב זה המשתנה Z אינו רלוונטי מבחינת פונקציית הפילוג המותנית: הוא אינו משפיע על הארגומנטים. בנוסף, $g(\beta)$ הוא קבוע מבחינת האינטגרל, ולכן אפשר להוציאו אל מחוץ לאינטגרל. אם כך

$$\begin{aligned}\mathbb{E}[g(Z) \cdot h(Z, Y) \mid X = \beta] &= g(\beta) \cdot \int_{-\infty}^{\infty} h(\beta, y) F_{Y|X}(dy \mid \beta) \\ &= g(\beta) \mathbb{E}[h(\beta, Y) \mid X = \beta]\end{aligned}$$

כאשר השוויון האחרון נובע מהגדרת התוחלת המותנית.

שאר השוויונים ב-6 נובעים מההגדרה. השוויון הראשון ב-7 נובע מתכונות הפילוג המותנה-ראה בפרט את נוסחה (8.7) עבור הצפיפות המותנית. השוויון השני נובע מטענה 8.32. לבסוף, 8 נובע מהטענות הקודמות.