הטכניון - מכון טכנולוגי לישראל
הפקולטה להנדסת חשמל

אותות אקראיים 044202

משה זכאי, משה סידי, אדם שורץ, עפר זיתוני

מהדורת תש"ע 2008/9

© 2003, 2006, 2007, 2008, 2009 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.

תוכן ענינים

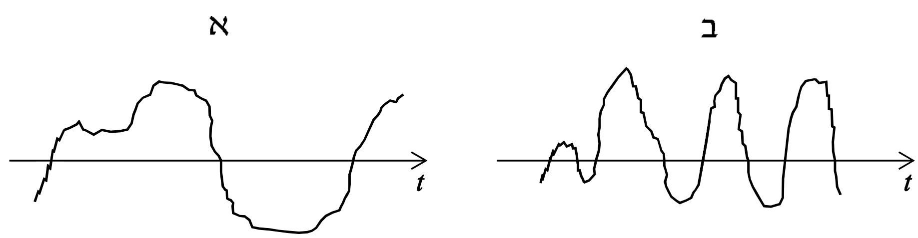
1 מבוא וחזרה על הסתברות

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, ורעש במעגלים חשמליים. נושאים אלו הם מרכזיים במגוון תחומים וישומים בהנדסת חשמל, כולל תקשורת, רשתות מחשבים, עיבוד אותות ועבוד תמונות, בקרה ומיכשור והתקנים אלקטרוניים. ואכן, מספר לא מבוטל של קורסים בשטחים שונים 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 (אותות אקראיים) 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 הים, 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.

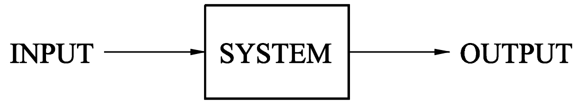
עבור צורת גל דטרמיניסטית אנו יכולים לדבר על רוחב סרט ולהגיד (לפחות במקרים מסויימים) איזו משתי צורות גל "מהירה" יותר ע"י השואת התמרת פוריה שלהן. במקרה ההסתברותי, נעיין בשתי צורות גל "טיפוסיות" המופיעות באיור 1.1. כאשר א' צורת גל טיפוסית של מכונה עם בולם זעזועים מסוג א', ו-ב' צורת גל טיפוסית עבור אותה מכונה



איור 1.1: צורות גל

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:

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

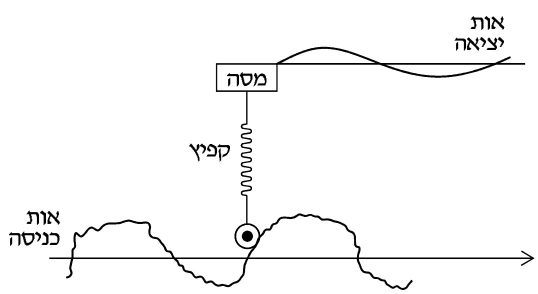


איור 1.2: מערכת

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 מעברת במערכת כזו. בהמשך כמובן נגדיר את מושגים אינטואיטיביים אלו בצורה מדוייקת.

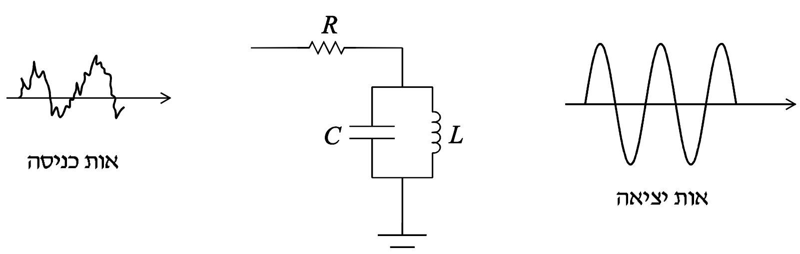
<u>דוגמאות:</u>

א. בולם זעזועים



איור 1.3: בולם זעזועים

ב. מעגל RLC



איור 1.4: מעגל RLC

<u>הערה:</u> 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. ל"יצור" 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 לגבו 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 באפשר-ות 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!). איפיון 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 בעטיים. 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 $X(t,\omega) = A(\omega)\sin(2\pi ft + \phi(\omega))$. 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.

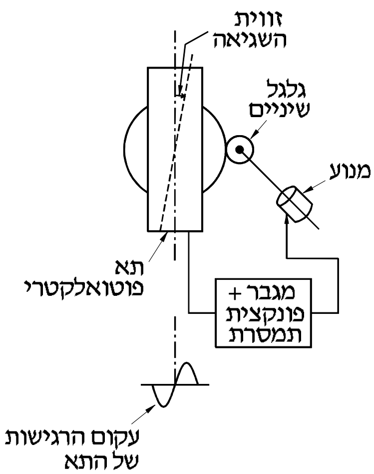
- יהיה N שלם חיובי קבוע או אקראי, ויהיו X_n משתנים אקראיים. נגדיר

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

אזי 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.

<u>דוגמא נוספת:</u> מערכת לעקיבה אחרי כוכב (או רובוט העוקב אחרי משהו) בנויה כמצויר:

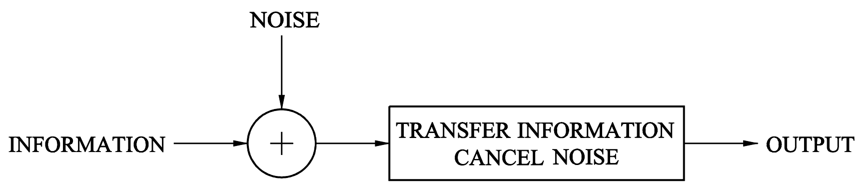


איור 1.5: מערכת עקיבה

אור הכוכב נופל על התא הפוטואלקטרי. באם אור הכוכב אינו נופל על מרכז התא (שגיאה אפס), נוצר אות שגיאה חיובי או שלילי בהתאם לכוון השגיאה. אות השגיאה מוגבר ומפעיל את המנוע בכוון לאפוס השגיאה. מערכת שבאופן עקרוני דומה למערכת כזו מופיעה במערכות עקיבה של מכ"ם בטווח ובזוית 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 לזוית השגיאה בין ציר הטלסקופ והכוכב (בכוון אפוס השגיאה). נניח שהמערכת יציבה: האם רצוי הגבר גדול או קטן? יש כאן דרישות מנוגדות 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

<u>חלק א׳:</u> מבוא וחזרה על הסתברות.

<u>חלק ב׳:</u> בחלק זה נעסוק בבעיות כגון הבעיה הבאה: נתונות המדידות Y_1 ו- Y_2,

$$Y_1 = X + n_1 \qquad ; \qquad Y_2 = X + n_2$$

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_2, Y_1 (וחוקי ההסתברות של X, n_2, n_1). בעיות מסוג זה נקראות שערוך. אפשר היה לשער כי הניחוש המיטבי הוא הממוצע בין שתי המדידות. בהמשך נראה כי שלא תמיד זהו ניחוש מוצלח. פתרון בעיה זו מהווה בסיס לשאלה קשה יותר---בעיית הסינון---בה אנו משערכים את ערך התהליך בזמן מסויים, על סמך ערכי כל המדידות בעבר. שערוך כזה מהווה בסיס לפתרון בעיית התקשורת ובעיית העקיבה אשר הזכרנו.

<u>חלק ג׳:</u> בחלק השלישי (והעיקרי) של הקורס נעסוק בתהליכים אקראיים, אפיונם ומעברם דרך מערכות לינאריות. נתחיל בתהליכים בזמן בדיד: פילוגים, מומנטים ותכונות. נרחיב בנושא שרשרות 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 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. לדוגמא 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. איחוד ניתן להמנות של מאורעות אף הוא מאורע, כלומר, אם $A_i \in F$ $i = 1, 2, \cdots$ גם $\cup A_i \in F$.

מדרישות אלו נובע כי F סגור 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 $F \doteq \Big\{\Omega, \emptyset, \{1,2,3\}, \{4,5,6\}\Big\}$ הוא מרחב מאורעות לגיטימי (בדוק!).

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

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

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

כאמור, פונקצית 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 $\omega \in \Omega, X(\omega)$ על מרחב הדגמים כך שעבור כל a ממשי $\{\omega : X(\omega) \le a\}$ 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. $n(0.25) = n(t)|_{t=0.25}$ (הרעש ברגע $t = 0.25$).
2. $X = \int_0^1 n^2(t)dt$ (האנרגיה של אות הרעש בתחום הזמן $[0, 1]$).
3. $Y = \max_{t\in[0,1]} n(t)$ (הערך המכסימלי של אות הרעש בתחום הזמן $[0, 1]$).

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

פונקצית פילוג

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

$$F_X(a) = \mathrm{Prob}\{X(\omega) \le a\}$$

מוגדרת תמיד כפונקציה של a (X הוא רק אינדקס), היא מונוטונית (לא יורדת), וניתן להוכיח שהיא רציפה מימין. בגבולות מקבלת פונקצית הפילוג את הערכים הבאים: $F_X(-\infty) = 0, F_X(+\infty) = 1$. עבור מ"א בדיד $F_X(a)$ עולה בקפיצות (צייר את $F_X(\alpha)$ 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 פונקצית 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 פונקצית הפילוג נובע מיד שעבור $a_1 < a_2$, מתקיים $\mathrm{Prob}\{a_1 < X \le a_2\} = F_X(a_2) - F_X(a_1)$, (כי המאורע $\{X \le a_2\}$ הוא איחוד של המאורעות הזרים $\{a_1 < X \le a_2\}$ ו- $\{X \le a_1\}$). מכאן גם נובעת המונוטוניות של $F_X(a)$.

תוחלת

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

$$m_x = \overline{X} = E[X] = \sum_i \alpha_i \mathrm{Prob}\{X = \alpha_i\} \ ,$$

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

עבור מ"א עם פ"ס:

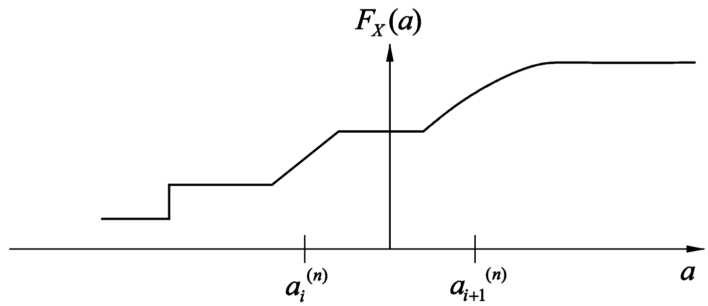
$$m_X = \overline{X} = E[X] = \int_{-\infty}^{\infty} \alpha f_X(\alpha) d\alpha$$

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

נגדיר:

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

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



איור 1.7:

יהיה $Y = g(X)$: אזי לפי ההגדרה $E[Y] = E[g(X)] = \int_{-\infty}^{\infty} \alpha dF_Y(\alpha)$.

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

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

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

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

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

$$f_X(\alpha) = \frac{1}{\pi} \frac{1}{1+\alpha^2} \tag{1.1}$$

אזי התוחלת אינה מוגדרת שכן האינטגרל עבור ערכים חיוביים הוא ∞ ואילו האינטגרל עבור ערכים שליליים הוא $(-\infty)$.

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 פונקצית 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; b, a 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: $E[(X - \overline{X})^n]$ (ובפרט $E\left[X - \overline{X}\right] = 0$)

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

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

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

עבור משתנה אקראי בודד, חוק ההסתברות $(F_X(a),\ a \in \mathbb{R})$, הוא "כל מה שאפשר להגיד על משתנה אקראי זה", בעקבות זאת אפשר לדבר על $EX, EX^2 \cdots$. בכוון ההפוך, אם נתונים $EX, EX^2, EX^3, \cdots$ 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 לא ידוע ולא מעניין; 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,

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

ומכאן:

$$P(|X| \geq \varepsilon) \leq \frac{EX^2}{\varepsilon^2} \, .$$

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

$$P(|Y - EY| \geq \varepsilon) \leq \frac{\mathrm{Var}\, Y}{\varepsilon^2}$$

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 מהממוצע. 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 מהממוצע) 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:

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

(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 פונקצית 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 האפיינית, 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:

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

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

<u>הוכחה:</u>

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

<u>שני משתנים אקראיים</u>

אם Y, X מ"א

$$F_{X,Y}(a,b) = \mathbb{P}\{X \le a, Y \le b\}$$

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

$$E[g(X,Y)] = \iint_{-\infty}^{\infty} g(\alpha,\beta) f_{X,Y}(\alpha,\beta) d\alpha d\beta$$

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

שני מ"א Y, X נקראים <u>בלתי תלויים</u> (ב"ת) סטטיסטית אם עבור כל a, b ממשיים מתקיים

$$\mathbb{P}\{X \le a, Y \le b\} = \mathbb{P}\{X \le a\} \cdot \mathbb{P}\{Y \le b\}$$

או

$$F_{X,Y}(a,b) = F_X(a) F_Y(b)\,.$$

אם למשתנים יש צפיפות אזי שוויון זה שקול ל-

$$f_{X,Y}(a,b) = f_X(a) \cdot f_Y(b).$$

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

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

<u>הקווריאנס</u> של זוג מ"א Y, X מוגדר ע"י:

$$\mathrm{Cov}(X,Y) = E(X - \overline{X})(Y - \overline{Y})$$

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

<u>תרגיל:</u> הוכח שאי תלות סטטיסטית מחייבת 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 $X \doteq \sin\Theta,\ Y \doteq \cos\Theta$, כמובן ש- X, Y תלויים סטטיסטית; אך קל לבדוק שהם חסרי קורלציה.

מקדם קורלציה

עבור שני משתנים אקראיים X_2 ו-X_1 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"י גזירה לפי λ נקבל ש - $\lambda^* = \frac{\mathrm{Cov}(X_1,X_2)}{\mathrm{Var}(X_2)}$ הוא האופטימלי. אם נציג את λ^* לביטוי עליו עשינו את המינימיזציה נקבל

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

ומכאן קבלנו את אי שוויון קושי שוורץ

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

למקדם

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

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

<u>סימונים:</u> $\underline{X}, \underline{a}$ יסמנו וקטורי עמודה. $\underline{\underline{B}}, \underline{\underline{A}}$ 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: $\underline{\underline{A}}, \underline{\underline{B}}$ מתאפשר להכפיל: $\underline{\underline{A}} \cdot \underline{\underline{B}}$ כאשר $\underline{\underline{A}}$ היא מטריצה $m \times n$ (m שורות n עמודות) ו- $\underline{\underline{B}}$ היא מטריצה $n \times k$. המטריצה המתקבלת היא $m \times k$. לדוגמא:

$$\begin{pmatrix} a_1 & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_1 & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix} = \begin{pmatrix} a_1 b_1 + a_{12} b_{21} + a_{13} b_{31} & a_1 b_{12} + a_{12} b_{22} + a_{13} b_{32} \\ a_{21} b_1 + a_{22} b_{21} + a_{23} b_{31} & a_{21} b_{12} + a_{22} b_{22} + a_{23} b_{32} \end{pmatrix}$$

יהי $\underline{X}$ וקטור אקראי בעל n רכיבים

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

פונקצית הפילוג של הווקטור 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 ווקטור על ידי

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

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

* $i = 1 \cdots n, EX_i$ נקראים המומנטים מסדר ראשון ו- $E\underline{X}$ הוא וקטור הממוצע.
* $i = 1, 2, \cdots, n, E|X_i|$ נקראים המומנטים המוחלטים מסדר ראשון.
* $1 \le i \le n, 1 \le j \le n, EX_iX_j$ נקראים המומנטים מסדר שני. בצורה מטריצית $E\left[\underline{X} \cdot \underline{X}^T\right] = \{E\left[X_iX_j\right]\}_{i,j=1}^n$ היא מטריצת המומנטים מסדר שני.
* $E(X_i - EX_i)(X_j - EX_j)$ עבור $1 \le i \le n, 1 \le j \le n$ נקראים המומנטים המרכזיים מסדר שני, כלומר הקווריאנסים. בסימון מטריצי

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

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

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

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

(או בכתיב וקטורי $F_{\underline{X}}(\underline{\alpha})$) יכול להיות מורכב ביותר ולא ידוע. אולם 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:

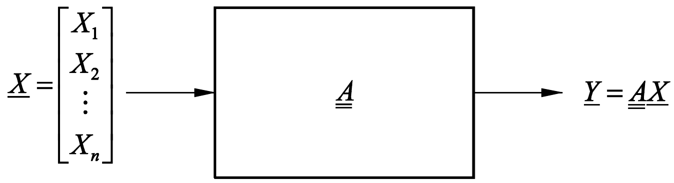
$$\underline{Y} = \underline{\underline{A}}\,\underline{X} + \underline{b}$$

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

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.

<u>תוחלות</u>

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



איור 1.8:

כלומר תוחלת היא פעולה לינארית.

<u>הוכחה</u>: התוצאה נובעת מהחשבון הישיר הבא:

$$E\left\{\begin{pmatrix} a_1 & a_{12} & \cdots & a_{1n} \\ a_{21} & & & \\ \vdots & & & \vdots \\ a_{m1} & & \cdots & a_{mn} \end{pmatrix}\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}\right\} = E\left\{\begin{pmatrix} a_1X_1 + a_{12}X_2 + & \cdots & +a_{1n}X_n + b_1 \\ & \vdots & \\ a_{m1}X_1 + a_{m2}X_2 + & \cdots & +a_{mn}X_n + b_m \end{pmatrix}\right\}$$

$$= \left\{\begin{pmatrix} a_1EX_1 + a_{12}EX_2 + & \cdots & +a_{1n}EX_n + b_1 \\ & \vdots & \\ a_{m1}EX_1 + a_{m2}EX_2 + & \cdots & +a_{mn}EX_n + b_m \end{pmatrix}\right\} = \underline{\underline{A}}\begin{pmatrix} EX_1 \\ \vdots \\ EX_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

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

$$\underline{\underline{Y}} = \begin{pmatrix} Y_1 & Y_{12} & \cdots & Y_{1n} \\ Y_{21} & & & \\ \vdots & & & \vdots \\ Y_{m1} & & \cdots & Y_{mn} \end{pmatrix} = [Y_{ij}]$$

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

<u>מומנטים מסדר שני</u>

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

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

שים לב, מתקבלת מטריצה $n \times n$ סימטרית; מטריצה זו נקראת <u>מטריצת המומנטים מסדר שני</u>. בצורה דומה נגדיר את <u>מטריצת הקווריאנס</u>:

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

<u>הערה:</u> עבור $\det(\underline{X}\cdot\underline{X}^T), n > 1$ 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 לינארית; 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 והדטרמיננט שווה 1.

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

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

נעיין בדוגמא:

$$\underline{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \quad ; \quad \underline{\underline{A}} = \begin{pmatrix} a_1 & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

במקרה זה:

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \underline{\underline{A}}\,\underline{X} = \begin{pmatrix} a_1X_1 + a_{12}X_2 \\ a_{21}X_1 + a_{22}X_2 \end{pmatrix}$$

וחשבון ישיר נותן:

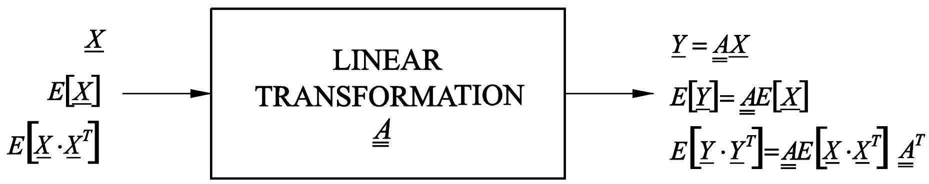
$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}^T = \begin{pmatrix} (a_1X_1 + a_{12}X_2)^2 & (a_1X_1 + a_{12}X_2)(a_{21}X_1 + a_{22}X_2) \\ (a_1X_1 + a_{12}X_2)(a_{21}X_1 + a_{22}X_2) & (a_{21}X_1 + a_{22}X_2)^2 \end{pmatrix}$$

מהסתכלות בתוצאה האחרונה ברור שאפשר לחשב את המטריצה $E[\underline{Y} \cdot \underline{Y}^T]$ מתוך ידיעת $E[\underline{X} \cdot \underline{X}^T]$ אולם החשבון נראה די מיגע. מושג המטריצות והכפל של מטריצות מאפשר לנו לסכם את החשבון בצורה מאוד פשוטה:

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

ובצורה דומה:

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



איור 1.9:

את התוצאות על המומנטים מסדר ראשון ושני כאשר מבצעים טרנספורמציה ליניארית על וקטור אקראי נסכם בציור 1.9.

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:

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

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

טענה: מטריצת המומנטים מסדר שני תמיד אי שלילית.

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:

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

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.

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

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 לווקטור האקראי $\underline{X}$ יש פונקצית צפיפות f_X אזי

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

ולכן, $\phi_{\underline{X}}(-\underline{\nu})$ היא בדיוק התמרת פוריה הרב-מימדית של פונקצית הצפיפות. חוק ההסתברות של $\underline{X}$ מגדיר את $\phi_{\underline{X}}(\nu)$. בכוון ההפוך, ללא הוכחה, ערכי הפונקציה $\phi_{\underline{X}}(\nu)$ 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) תתן 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

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

ותחת אי תלות, תוחלת המכפלה היא מכפלת התוחלות. ולהפך, אפשר להראות שאם הפונקציה האופינית 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 מכפלת 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 ν) ובנוסף עבור פונקציות מיוחדות (אקספוננציאליות).

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

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

<u>תשובה:</u>

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

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, מעונינים 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 $\underline{Y} = \underline{\underline{A}}\,\underline{X}$ כאשר

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

ועבור הפונקציה האופינית 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 וקטור:

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

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-$a_1 = a_2 = \cdots = a_n = 1$ אזי

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

<u>מסקנה:</u> הפונקציה האופיינית של סכום מ"א ב"ת היא מכפלת הפונקציות האופייניות.

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

<u>הערה:</u> 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 EX^mY^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 האפיינית 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 האפיינית: 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

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

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. שנית, קל יחסית לבצע חישובים עם פילוגים גאוסיים, ולכן משתמשים בו כקירוב נוח, וכן כאשר אנו מתעניינים רק במומנט ראשון ושני. לבסוף, משפט הגבול המרכזי מראה לנו כי הפילוג הגאוסי מהווה קירוב טוב לאוסף גדול של משתנים קטנים.

<u>המקרה הסקלרי</u>

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

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

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

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

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

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

<u>המקרה הוקטורי</u>

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

אם ו"א $\underline{X}$ הוא גאוסי אזי כל רכיב שלו הוא השלכה 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 מחייב שהוקטור הוא וקטור אקראי 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. חישוב ישיר מראה כי $\mathbb{E}Y = 0$, $\mathbb{E}[XY] = 0$ כלומר X, Y הם חסרי קורלציה.

תכונות של וקטורים אקראיים גאוסיים:

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

הוכחת (2):

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

ולכן כל השלכה של $\underline{Y}$ בכיוון מסוים היא בסופו של דבר השלכה של $\underline{X}$ (לא בהכרח באותו כיוון כמובן), ולכן $\underline{Y}$ ו"א גאוסי.

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

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

ואז מתקיים

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

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

הוכחה: נניח ש-$\underline{X}$ הוא ווקטור גאוסי. נסמן את ווקטור הממוצעים ואת מטריצת הקווריאנס ב-

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

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

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

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

ומחישוב קודם של קווריאנסים נקבל שהווריאנס של Y הוא

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

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

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

אולם מהגדרת Y,

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

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

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

מתכונות הפונקציה האפיינית,

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

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

ג. נניח ש-$\underline{\underline{\Lambda}}$ 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 $X_1, X_2, \cdots$ בלתי תלויים לינארית בזוגות, כלומר $\mathrm{Cov}(X_i, X_k) = 0, i \neq k$. 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-$\underline{\underline{\Lambda}}$ אלכסונית אם ורק אם קיימת אי תלות לינארית בזוגות אולם במקרה הגאוסי מתקיימת גם אי תלות סטטיסטית, ולכן עבור וקטור אקראי גאוסי אי תלות ליניארית 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:

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

ה. שאלה: אם $\underline{X}$ ו"א גאוסי, רכיבים ב"ת, $\underline{Y} = \underline{\underline{A}}\,\underline{X}$ אזי בדרך כלל הרכיבים של $\underline{Y}$ הם תלויים. מה בכוון ההפוך? האם ניתן להגיע לווקטור אשר רכיביו בת"ס על ידי טרנספורמציה לינארית? לכך יש חשיבות כי אם הרכיבים בת"ס 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 גאוסי 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 $\underline{\underline{D}}$ כך ש-$\underline{Y} = \underline{\underline{D}}\,\underline{X}$ 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 X. למשל, $\underline{X} = \binom{X_1}{X_1}$ כאשר X_1 מ"א גאוסי 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

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

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

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

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

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

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

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

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

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

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

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

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

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

וע"י השוואת עמודות נקבל $\underline{\underline{\Lambda}}\,\underline{d_i} = c_i\underline{d_i}$. לכן, הוקטורים $\underline{d_i}$ פותרים את המשואה $\underline{\underline{\Lambda}}\,\underline{d} = c\underline{d}$, במילים אחרות c_i הם הערכים העצמיים של $\underline{\underline{\Lambda}}$ ו-$\underline{d_i}$ הם הוקטורים העצמיים של $\underline{\underline{\Lambda}}$ (ולכן גם מתקיים $\underline{\underline{D}}^T = \underline{\underline{D}}^{-1}$).

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

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

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

שים לב כי כדי לרשום את הצפיפות בצורה מפורשת יש לחשב את ההופכי של המטריצה Λ. לצורך חישוב מפורש של הפונקציה האפיינית חישוב כזה אינו דרוש.

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

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

סיכום

א. תכונת הגאוסיות אינוריינטית לטרנספורמציות לינאריות.

ב. הפילוג של ו"א גאוסי $\underline{X}$ נקבע חד משמעית ע"י $\underline{m}\,,\underline{\underline{\Lambda}}$ (שני המומנטים הראשונים).

ג. אי תלות לינארית גוררת אי תלות סטטיסטית.

ד. ניתן לעבור למ"א ב"ת על ידי סבוב מערכת הצירים.

1.6 הסתברות מותנית

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

$$F_{Y|X}(\alpha|\beta) = \mathbb{P}\{Y \le \alpha | X = \beta\} = \frac{\mathbb{P}\{Y \le \alpha, X = \beta\}}{\mathbb{P}\{X = \beta\}}$$

ואם X מ"א רציף

$$F_{Y|X}(\alpha|\beta) = \lim_{\varepsilon \to 0} \frac{\mathbb{P}\{Y \le \alpha\,;\, \beta < X \le \beta + \varepsilon\}}{\mathbb{P}\{\beta < X \le \beta + \varepsilon\}}$$

פילוג סגולי מותנה

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

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

$$f_{Y|X}(\alpha|\beta) = f_Y(\alpha) \qquad f_{X|Y}(\beta|\alpha) = f_X(\beta).$$

תוחלת מותנית

$$E[Y|X = \beta] = \int_{-\infty}^{\infty} \alpha f_{Y|X}(\alpha|\beta) \cdot d\alpha$$

ובאופן כללי יותר

$$E[Y|X = \beta] = \int_{-\infty}^{\infty} \alpha F_{Y|X}(d\alpha|\beta)$$

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

$$E[g(Y)|X = \beta] = \int_{-\infty}^{\infty} g(\alpha) f_{Y|X}(\alpha|\beta) d\alpha$$

ובאופן כללי יותר

$$E[g(Y)|X = \beta] = \int_{-\infty}^{\infty} g(\alpha) F_{Y|X}(d\alpha|\beta)$$

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

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

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

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

מכאן שאם $Y=X+N$ כאשר X ו-N ב"ת אזי

$$E[Y|X=\beta]=E[X|X=\beta]+E[N|X=\beta]=\beta+E[N]$$

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

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

תכונות נוספות של תוחלת מותנית:

א) $E[X|X=\beta]=\beta$

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

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

ג) $E[Y]=E[E[Y|X]]$

<u>הערה:</u> תכונות (ב) ו-(ג) נקראות תכונות ההחלקה. מהן גם נובע הנסוח הבא של תכונת ההחלקה:

$$E\Big[h(X,Y)g(X)\Big]=E\Big[g(X)E[h(X,Y)|X]\Big] \qquad (+)$$

"מעין הוכחה" של (+):

$$\begin{aligned}E\Big[g(X)E[h(X,Y)|X]\Big] &= \int_{-\infty}^{\infty} g(\beta)\Big[\int_{-\infty}^{\infty} h(\beta,\alpha)f_{Y|X}(\alpha|\beta)d\alpha\Big]f_X(\beta)d\beta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(\beta)h(\beta,\alpha)f_{Y|X}(\alpha|\beta)f_X(\beta)d\alpha d\beta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(\beta)h(\beta,\alpha)f_{Y,X}(\alpha,\beta)d\alpha d\beta = E\Big[g(X)h(X,Y)\Big]\end{aligned}$$

<u>תוצאה חשובה ללא הוכחה:</u> 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:

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

<u>הערה:</u> **שכנע את עצמך 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 לך להשתכנע, נסה תחילה לחשוב על המקרה בו $\underline{Y}$ וקטור חד-מימדי).

דוגמה 1.8 נתונים זוג מ"א גאוסיים במשותף עם צפיפות

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

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:

$$f_{X_1}(x_1) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x_1^2}{2\sigma^2}}$$

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

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 (המדידה) של x_1! המצב דומה גם במקרה הכללי.

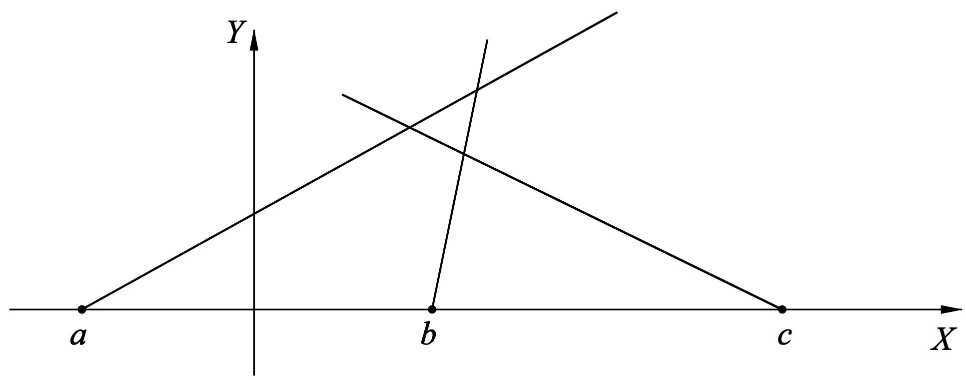
2 שערוך

2.1 מבוא

כיצד פועל מכשיר GPS (Global Positioning System)? המכשיר קולט שידורים מלווינים, ועל ידי השוואת זמן הקליטה (ביחס לזמן השידור) מפענח את המרחק מכל לוין. 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 נותן 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 השערוך.

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 (השערוך), ובנוסף, מה יהיה סדר הגודל של שגיאת השערוך.

"שערוך"---הערכה מחדש---עוסק בהערכה או השערה לגבי גודל רצוי. בסעיף 1.1 ראינו דוגמאות בהן רצינו "לנקות" אות מרעשים (לסנן), להעריך מרחק למטרה על פי מדידות רועשות וכד'. בפרק זה נניח את היסודות לנושא זה.

<u>ניסוח כללי של בעית השערוך:</u>

א. מצוי 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 כל שהוא (טוב או רע) הוא פונקציה של המדידה אשר נסמן ב- $\hat{\underline{X}} = \phi(\underline{Y})$.
לשם פשטות נעסוק תחילה 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 $E[g(X,\phi(\underline{Y}))] \triangleq$ השגיאה הממוצעת (לפי הקריטריון g). לדוגמה, עבור $g(a,b)=(a-b)^2$ השגיאה הממוצעת היא $E[(X-\phi(\underline{Y}))^2]$. דוגמה אחרת (תוחלת הערך המוחלט של השגיאה) $g(a,b)=|a-b|$ השגיאה הממוצעת היא $E[|X-\phi(\underline{Y})|]$. בתקשורת ספרתית, למשל, האות הרצוי X מקבל את הערכים $0,1$ וכדי לקבל סיכוי מרבי לשערוך נכון בוחרים $g(x,y)=0$ אם $x=y$ ו-$g(x,y)=1$ אם $x\neq y$.

ג. הבעיה: לאחר שהחלטת 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 (a-priori) $\mathbb{P}_X$ 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 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 שערוך אופטימלי

נחזור לבעית השערוך שהוצגה במבוא (נתרכז במקרה בו Y הוא מ"א ולא ו"א); X רצוי, Y מצוי, קיים פילוג הסתברות משותף ל-Y ו-X. המשערך ל-X כאשר נמדד Y מסומן ב-$\hat{X}=\phi(Y)$. קובעים סיפרת טיב $g(\cdot,\cdot)$ ובעזרתה מגדירים את השגיאה הממוצעת $E[g(X,\phi(Y))]$. עבור קריטריון השגיאה הריבועית הממוצעת $g(a,b)=(a-b)^2$ הגודל ε המוגדר כ-$\varepsilon \triangleq X-\phi(Y)$ נקרא שגיאת השערוך, ε^2 הוא השגיאה הריבועית ו-$E[\varepsilon^2]$ הוא השגיאה הריבועית 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 $E[(X-\phi_0(Y))^2]$ הוא מינימלי (על פני כל המשערכים $\phi(\cdot)$) הוא:

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

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

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

<u>דוגמא א'</u> (סתם דוגמא טכנית):

$$f_{X,Y}(\alpha,\beta) = \alpha+\beta \quad ; \quad 0 \le \alpha,\beta \le 1$$

$$f_Y(\beta) = \int_0^1 (\alpha+\beta) d\,\alpha = \frac{1}{2}+\beta$$

$$E(X|Y=\beta) = \int_0^1 \alpha \frac{\alpha+\beta}{\frac{1}{2}+\beta} d\,\alpha = \frac{\frac{1}{3}+\frac{\beta}{2}}{\frac{1}{2}+\beta}$$

<u>דוגמא ב'</u>: X,Y משתנים אקראיים גאוסים במשולב. 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-$E[ZY_c]=0$. 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:

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

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

מכאן:

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

ולכן המשערך האופטימלי של X מתוך Y נתון על ידי:

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

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

שים לב לכך שהמשערך האופטימלי במקרה זה, הוא פונקציה לינארית של המדידות. (השווה עם השורות האחרונות של סעיף 1.6). לכן $\hat{X}_{opt}$ הוא מ"א גאוסי, עם ממוצע $E[X]$. בנוסף, במקרה (החשוב!) בו $Y = X + N$ כאשר X, N חסרי קורלציה, קל לראות כי הווריאנס של $\hat{X}_{opt}$ תלוי רק בווריאנס של X.

$\underline{Y} = (Y_1, \cdots, Y_n)^T$, $(\underline{Y}, X)$ ו"א גאוסי, ונניח כי $\underline{\underline{\Lambda}}_Y = \mathrm{Cov}\{\underline{Y}\}$ היא מטריצה לא סינגולרית. נחשב את $E(X|\underline{Y})$.
נגדיר

$$\underline{Y}_c = \underline{Y} - E\,\underline{Y}$$
$$X_c = X - E\,X$$
$$Z \doteq X_c - \underline{Y}_c^T \Lambda_Y^{-1} E\left[X_c\,\underline{Y}_c\right].$$

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

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

בצורה מפורשת יותר

$$E[X|\underline{Y}] = EX + (\underline{Y} - E\underline{Y})^T \Lambda_Y^{-1} \begin{pmatrix} \mathrm{Cov}(Y_1, X) \\ \vdots \\ \mathrm{Cov}(Y_n, X) \end{pmatrix} \tag{2.5}$$

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 $\mathbb{E}\left[\,\underline{X} \mid \underline{Y}\,\right]$ ניתן לחשב על ידי

$$\mathbb{E}\left[\,\underline{X} \mid \underline{Y}\,\right] = \mathbb{E}[\underline{X}] + \left[(\mathrm{Var}\,\underline{Y})^{-1}\,\mathrm{Cov}(\underline{Y}, \underline{X})\right]^T \left[\underline{Y} - \mathbb{E}\,\underline{Y}\right] \tag{2.6}$$

<u>דוגמא ג':</u> דוגמא פשוטה אולם חשובה וחשוב לזכור אותה כולל התוצאות המסומנות ב-*:

$$Y = X + N \tag{2.7}$$
$$E[X] = E[N] = 0 \tag{2.8}$$
$$E[X^2] = 1, \quad E[N^2] = \sigma_n^2 \tag{2.9}$$

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). לצורך דוגמה זאת לא נשתמש בתוצאות דוגמא ב', אלא במשפט שהופיע ללא הוכחה בסוף סעיף 1.6. לכן נוכל לרשום:

$$E[X|Y] = c_0 Y$$

עבור קבוע לא ידוע c_0. על מנת למצוא את c_0, נבצע מינימיזציה על השגיאה הרבועית הממוצעת:

$$E\Big[(X-cY)^2\Big] = E\Big[(X-cX-cN)^2\Big] = (1-c)^2 + c^2\sigma_n^2$$

כיוון ש- $\hat{X} = c_0(X+N)$ הרי שהמינימיזציה מחפשת פשרה (על ידי התאמת c_0) בין אי-שינוי X (על ידי בחירה $c_0 = 1$) מחד לבין סילוק הרעש (על ידי בחירה $c_0 = 0$) מאידך. מינימיזציה נותנת (ע"י גזירה והשוואה ל-0),

$$c_0 = \frac{1}{1+\sigma_n^2} \qquad (*)$$

במקרה הכללי יותר בו הווריאנס של X הוא σ_x^2 נקבל

$$\frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2} \qquad (2.10)$$

שהוא מספר בין 0 ל-1 התלוי ב"עוצמות" היחסיות של האות והרעש. במבט ראשון על (2.7) נראה שהשערוך 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 הנותרת (במקרה של ווריאנס כללי של X) היא:

$$E\Big[(X-c_0Y)^2\Big] = E\Big[(X-c_0[X+N])^2\Big] \qquad (2.11)$$

$$= \left(1-\frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2}\right)^2\sigma_x^2 + \left(\frac{\sigma_x^2}{\sigma_x^2+\sigma_n^2}\right)^2\sigma_n^2 \qquad (2.12)$$

$$= \left(\frac{\sigma_x^2\sigma_n^2}{\sigma_x^2+\sigma_n^2}\right) \qquad (*)$$

נחזור להוכחת טענה 2.1.

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

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

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

$$E\Big[(X-\phi_0(Y))(\phi_0(Y)-\phi(Y))\Big] = E\Big[(\phi_o(Y)-\phi(Y))E[X-\phi_0(Y)|Y]\Big]$$

ומתקיים

$$E\Big[X-\phi_0(Y)|Y\Big]=E[X|Y]-\phi_0(Y)=\phi_0(Y)-\phi_0(Y)=0$$

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

$$E\Big[(X-\phi(Y))^2\Big]\geq E\Big[(X-\phi_0(Y))^2\Big]$$

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

בכך השלמנו את הוכחת הטענה. שים לב שהתוצאה נשארת ללא שינוי גם אם במקום מ"א Y יהיה ו"א $\underline{Y}$.

הערה: תהי g פונקציה כלשהי: אזי

$$E\Big[(X-\phi_0(Y))g(Y)\Big]=0\,. \tag{2.13}$$

המשמעות היא: שגיאת השערוך חסרת קורלציה עם כל פונקציה של המדידות. ההוכחה דומה להוכחה לעיל: מתכונות (ב) ו (ג) נקבל

$$E\Big[(X-\phi_0(Y))g(Y)\Big]=E\Big[(E[X-\phi_0(Y)|Y])g(Y)\Big]=0$$

כאשר השוויון האחרון הוא מהגדרת $\phi_0(Y)$.

בעזרת ההערה נוכל להוכיח כי המשערך האופטימלי הוא יחיד.

טענה 2.3 המשערך האופטימלי הוא יחיד, במובן הבא. נסמן ב-ϕ_0 את משערך התוחלת המותנית ויהי ϕ_1 משערך אחר אשר השגיאה שלו שווה לזו של ϕ_0. אזי $E\Big[(\phi_0(Y)-\phi_1(Y))^2\Big]=0$.

הוכחה:

$$E\Big[(X-\phi_1(Y))^2\Big]=E\Big[(X-\phi_0(Y)+[\phi_0(Y)-\phi_1(Y)])^2\Big] \tag{2.14}$$

$$=E\Big[(X-\phi_0(Y))^2\Big]+E\Big[(\phi_0(Y)-\phi_1(Y))^2\Big]+2E\Big[(X-\phi_0(Y))(\phi_0(Y)-\phi_1(Y))\Big]. \tag{2.15}$$

בגלל ההערה לעיל האיבר האחרון שווה לאפס. כיוון שלפי ההנחה לשני המשערכים אותה שגיאה, קיבלנו כי

$$E\Big[(\phi_0(Y)-\phi_1(Y))^2\Big]=0$$

ולכן המשערכים הם זהים, במובן שהם נותנים את אותה התוצאה עבור כל מדידה.

ההוכחה עבור מדידה ווקטורית זהה לחלוטין.

נשים לב כי היחידות היא באשר לתוצאה של השערוך, ולא באשר למבנה של המשערך. כדי להבין נקודה זו נתבונן במקרה פשוט ביותר של שערוך 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 השערוך היא X למרות שהנוסחה עבור המשערך שונה.

נעבור כעת לדוגמא נוספת לחישוב מפורש של המשערך האופטימלי. דוגמא זו עוסקת בשערוך של תוצאת זריקת קוביה "הוגנת" $p_i = 1/6$; $i = 1, \cdots, 6$ לפי קריטריון השגיאה הרבועית המינימלית.

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

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

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:

$$Y = 0 \quad \text{if} \quad X = 4 \quad \text{or} \quad 5 \quad \text{or} \quad 6$$
$$Y = 1 \quad \text{if} \quad X = 1 \quad \text{or} \quad 2 \quad \text{or} \quad 3$$

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:

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

שגיאה קטנה משמעותית מהשגיאה אשר קיבלנו כאשר השערוך 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 שערוך לינארי

במקרים רבים חישוב התוחלת המותנית קשה ביותר. בכל מקרה, דורש חישוב זה 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 פונקציה לינארית של המדידות. בסעיף זה איננו מניחים ש- $X, \underline{Y}$ הוא ווקטור גאוסי, 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 השערוך 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 בעית מציאת המשערך הלינארי האופטימלי נגדיר כדלקמן:

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

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

יהיה מינימלי. שים לב שזו בעיה הרבה יותר פשוטה ממציאת המשערך 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:

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

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:

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

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:

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

ולכן המשערך הלינארי האופטימלי הוא:

$$\hat{X}^l_{\mathrm{opt}} = E[X] + \frac{\mathrm{Cov}(Y, X)}{\mathrm{Var}(Y)}\Big[Y - E[Y]\Big]$$

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

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

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

הערות:

1. שים לב שלצורך חישוב המשערך הלינארי האופטימלי יש צורך לדעת רק מומנטים מסדר ראשון ושני.
2. משערך אופטימלי לעולם אינו גרוע יותר ממשערך לינארי אופטימלי; משערך לינארי אופטימלי לעולם אינו גרוע יותר ממשערך ללא מדידה כלל.
3. כאשר $\mathrm{Cov}(Y,X)=0$ 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 בשערוך לינארי של X.
4. במקרים מסוימים המשערך הליניארי האופטימלי הוא המשערך האופטימלי. לדוגמא, כאשר X ו-Y בלתי תלויים. דוגמא אחרת היא המקרה הגאוסי (וזאת ומאחר וכבר צוין, ללא הוכחה שבמקרה הגאוסי $E[X|Y]=c_0Y$ עבור קבוע c_0 מסוים). דוגמאות נוספות יובאו בתרגילים.

המקרה הוקטורי

נעבור למקרה בו השערוך מתבצע על סמך מספר מדידות $Y_1, Y_2, \cdots Y_n$. לשם פשטות נניח ש-$E[X]=0$ וכן $E[\underline{Y}]=0$ כאשר $\underline{Y}$ הוא הוקטור האקראי שרכיביו הם $Y_1, Y_2, \ldots, Y_n$. 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 $\underline{a}^T = (a_1, a_2, \ldots, a_n)$ 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 השערוך הריבועית הממוצעת תהיה מינימלית. במילים אחרות, מצא קבועים $(a_1, a_2, \ldots, a_n)$ כך שהביטוי

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

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 השערוך הריבועית הממוצעת:

$$E[\varepsilon^2] = E\Big[(X - \sum a_j Y_j)^2\Big] \ . \qquad (2.16)$$

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:

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

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

$$\varepsilon = X - \sum_{j=1}^n a_j^* Y_j \tag{2.17}$$

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:

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

ובצורה מקוצרת:

$$E[\underline{Y}X] = E[\underline{Y}\underline{Y}^T] \cdot \underline{a}^* \tag{2.18}$$

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 $\mathrm{Cov}(\underline{Y}, X) = \mathrm{Var}\,\underline{Y} \cdot \underline{a}^*$. בהנחה שמטריצת הקווריאנס $E[\underline{Y}\underline{Y}^T]$ אינה סינגולרית, יש פתרון (והוא יחיד) עבור $\underline{a}^*$ (במקרה הסינגולרי נטפל אח"כ). פתרון זה נתון ע"י

$$\underline{a}^* = \left(E[\underline{Y}\underline{Y}^T]\right)^{-1} E[\underline{Y}X] \tag{2.19}$$

והמשערך הלינארי האופטימלי הוא

$$\hat{X}^l_{\mathrm{opt}} = (\underline{a}^*)^T \underline{Y} = \sum_{i=j}^n a_j^* Y_j \tag{2.20}$$

מה השגיאה הנותרת $E[(X - (\underline{a}^*)^T \cdot \underline{Y})^2] = ?$. דרך אחת לחשב את השגיאה הנותרת היא לרשום את (2.18) בצורת $\underline{b} = \underline{\underline{C}} \cdot \underline{a}^*$ כאשר $\underline{b} = E[X\underline{Y}]$ ו-$\underline{\underline{C}} = E[\underline{Y} \cdot \underline{Y}^T]$ ואז להציב לתוך (2.16). דרך אחרת, אשר נחזור אליה בצורה כללית בהמשך, היא הבאה: מהגדרת השגיאה,

$$E[X^2] = E\left[\sum_{i=j}^n a_j^* Y_j + \varepsilon\right]^2 \tag{2.21}$$

$$= E\left[\sum_{i=j}^n a_j^* Y_j\right]^2 + E[\varepsilon^2] + 2\left[\sum_{i=j}^n a_j^* E[Y_j\varepsilon]\right] . \tag{2.22}$$

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

$$(E[\varepsilon^2])_{\min} = E[X^2] - 2\underline{b}^T \underline{a}^* + (\underline{a}^*)^T \underline{\underline{C}}\, \underline{a}^* = E[X^2] - (\underline{a}^*)^T \underline{\underline{C}}\, \underline{a}^* \tag{2.23}$$

או

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

(2.19), (2.20) ו- (2.23) נותנים את התשובה המלאה לבעית השערוך הלינארי במקרה הוקטורי כאשר המומנטים מסדר ראשון מתאפסים (לפחות כאשר $\underline{\underline{C}}$ הפיכה).

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

הוכחה חלופית לנוסחת המשערך האופטימלי היא הבאה. נשים לב כי מנוסחת השגיאה (2.16) נובע כי השגיאה תלויה אך ורק במומנטים עד סדר שני, כלומר ב-$\mathbb{E}\, X, \mathbb{E}\, \underline{Y}, \mathbb{E}[Y_i X], \mathrm{Var}\, X, \mathrm{Var}\, \underline{Y}$. לכן אם נחליף את המ"א במשתנים $\tilde{X}, \underline{\tilde{Y}}$ בעלי אותם מומנטים, השגיאה תשאר זהה, ולכן גם כל התנאים אשר קובעים את נוסחת המשערך האופטימלי. אם כן, נבחר משתנים חדשים כאלו להיות גאוסיים במשותף. במקרה זה אנו יודעים את נוסחת המשערך האופטימלי---אשר הוא לינארי---משוואה (2.5). זו בדיוק משוואת המשערך הלינארי האופטימלי.

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

$$\hat{X}^l_{\mathrm{opt}} = E[X] + \sum_{i=1}^{n} a_i^* (Y_i - E[Y_i])$$

כאשר ה-a_i^* נתונים ע"י פתרון משוואה (2.19) עבור המשתנים הממורכזים (ממוצע אפס) והשגיאה הנותרת תהיה:

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

(הוכח תוצאות אלה).

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

דוגמה 2.5 נתונים

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

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

אזי

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

ומכאן

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

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

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

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

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

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

$$E[Y_1Y_2] = E[(X + N_1)(X + N_2)] = E[X^2] = 1$$

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

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

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

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

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

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

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

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

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

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

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

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

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 $\underline{\underline{C}} = [\mathrm{Cov}(Y_i, Y_j)]$ היא סינגולרית. במקרה זה קיים וקטור קבוע n מימדי, $\underline{h}$ שאינו ווקטור האפס, כך ש $\underline{\underline{C}}\,\underline{h} = 0$ כי השורות של $\underline{\underline{C}}$ תלויות. לכן:

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

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

$$E\left[(Y_n + h_{n-1}Y_{n-1} + \cdots + h_1 Y_1)^2\right] = 0$$

ולכן

$$Y_n = -h_{n-1}Y_{n-1} - h_{n-2}Y_{n-2} - \cdots - h_1 Y_1$$

דהיינו, אם נתונים $Y_1, \ldots, Y_{n-1}$ אזי נוכל לנחש את Y_n ללא שגיאה. במלים אחרות, באם ידועים $Y_1, \ldots, Y_{n-1}$ אזי אין צורך בידיעת Y_n. כלומר יש כאן יתירות: כל פונקציה לינארית של $Y_1, \ldots, Y_n$ אפשר לייצג כפונקציה לינארית של $Y_1, \ldots, Y_{n-1}$: המשתנה Y_n אינו מוסיף מידע לצורך השערוך הלינארי. לכן, במקום לשערך את X כשנתון $Y_1, \ldots, Y_n$ מספיק לשערך את X כשנתון $Y_1, \ldots, Y_{n-1}$. אם $h_n = 0$ אבל $h_i \neq 0$ עבור i כלשהוא יהיה החשבון בדיוק אותו דבר. מוסר השכל: אם $\underline{\underline{C}}$ סינגולרית אזי קיים i_0 (לפחות אחד) כך שאפשר לוותר על Y_{i_0} מבלי לקלקל את השערוך. אם המטריצה החדשה $(n-1)\times(n-1)$ שמתקבלת ע"י מחיקת העמודה והשורה ה-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.

2.4 עקרון ההשלכה (עקרון האורתוגונליות)

בחלק זה נתרכז במקרה בו $EX = 0; E\underline{Y} = 0$. עבור שערוך אופטימלי ראינו כי שגיאת השערוך תמיד חסרת קורלציה עם כל פונקציה של המדידות (משוואה (2.13)), כלומר

$$E\Big[(X - \phi_0(Y))g(Y)\Big] = 0\,. \tag{2.26}$$

כמו כן מצאנו שהמשערך הלינארי האופטימלי של X כשנמדד $\underline{Y}$ הוא $\hat{X}^l_{\text{opt}} = (\underline{a}^*)^T\underline{Y}$ כאשר הקבועים $\underline{a}^*$ מקיימים:

$$E[\underline{Y} \cdot X] = E[\underline{Y} \cdot \underline{Y}^T] \cdot \underline{a}^* \tag{2.27}$$

או בצורה מפורטת יותר

$$E\Big[(X - (\underline{a}^*)^T\underline{Y}) \cdot Y_i\Big] = 0 \quad \text{עבור כל } i \tag{2.28}$$

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

אנו נראה דרך נוספת למציאת תוצאה זו, הפעם דרך נימוקים "גיאומטריים". הסתכלות גאומטרית כזו מסייעת מאד, במקרים רבים, לפתור בעיות שערוך: לעיתים קל יותר לתקוף את בעיית השערוך מזווית זו, ובנוסף היא מסייעת להבין את תכונות הפתרון.

מדוע יש לצפות כי שגיאת השערוך של המשערך האופטימלי תהיה חסרת קורלציה עם המדידות? ומדוע תכונה דומה קיימת במקרה הלינארי? כדי לקבל תובנה, 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 ווקטור מקדמים $\underline{a}$ עבורו השגיאה הריבועית $E[(X - \underline{a}^T\underline{Y})^2]$ מינימלית, כך נגדיר בעית הטלה גאומטרית, בה מחפשים ווקטור מקדמים כך שהשגיאה הריבועית $X - \underline{a}^T\underline{Y}$ היא מינימלית, כאן המדידות $\underline{Y}$ הן אוסף של ווקטורים: נפרט כעת את הבעיה.

יהיה $\underline{X}$ וקטור דטרמיניסטי במרחב האוקלידי ה-m ממדי $\mathbb{R}^m$. יהיו $\underline{Y}_1, \ldots, \underline{Y}_n$ וקטורים דטרמיניסטיים באותו מרחב m ממדי. נעיין בתת המרחב $M(\underline{Y}_1, \ldots, \underline{Y}_n)$ 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-$\underline{\hat{X}}$ 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).

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

כדי להבין מדוע הטענה נכונה-הפעל את משפט פיתגורס לחישוב השגיאה ושים לב כי אם אין ניצבות אזי ניתן להקטין את השגיאה.

<u>שאלה ב-</u> $\mathbb{R}^m$: נתונים $\underline{X}, \underline{Y}_1, \ldots, \underline{Y}_n$ וקטורים ב-$\mathbb{R}^m$. מצא ההשלכה $\underline{\hat{X}}$ של $\underline{X}$ על תת המרחב הנפרש על ידי $\underline{Y}_1, \cdots, \underline{Y}_n$.

תשובה: $\underline{\hat{X}}$ הוא צרוף לינארי של הוקטורים $\underline{Y}_i$:

$$\underline{\hat{X}} = h_1^*\underline{Y}_1 + h_2^*\underline{Y}_2 + \cdots + h_n^*\underline{Y}_n$$

ועקרון ההשלכה קובע שעבור כל i:

$$\Big((\underline{X}-\hat{\underline{X}}),\underline{Y}_i\Big)=0$$

לכן $\underline{h}^*$ מוגדר ע"י האוסף של n משוואות לינאריות

$$(\underline{X},\underline{Y}_i)=\sum_{j=1}^{n}h_j^*(\underline{Y}_i,\underline{Y}_j)$$

את התוצאה ניסחנו במרחב m ממדי $\mathbb{R}^m$ אבל זה ניתן להרחבה למרחבים "אין סוף ממדיים" בתנאי שנגדיר כהלכה מרחבים כאלה ויהיה לרשותנו מושג של מכפלה סקלרית במרחבים אלה. באופן מדוייק יותר:

משפט 2.8 נתבונן באוסף אלמנטים במרחב לינארי עבורו מוגדרת מכפלה סקלרית (x,y), כך שהמרחק בין זוג נקודות x,y נתון על ידי $||x-y||^2=(x-y,x-y)$; כלומר המרחק נגזר מהמכפלה הסקלרית. נתון תת-מרחב לינארי S. נתונה נקודה x שאיננה ב- S ואנו מחפשים נקודה $\hat{x}$ ב- S שתהייה הקרובה ביותר ל- x. אזי בהכרח השגיאה $\varepsilon\doteq x-\hat{x}$ ניצבת לכל איבר ב- S כלומר $(\varepsilon,z)=0$ לכל $z\in S$.

הוכחה: נניח שהטענה אינה נכונה: כלומר קיים z ב- S כך ש- $(\varepsilon,z)=\alpha\neq 0$. על ידי חלוקה ב- $||z||$ (וכפל ב- (-1) במידת הצורך) אפשר להניח כי $\alpha>0$ וכן $(z,z)=1$. נגדיר כעת נקודה חדשה $\tilde{x}$ ב- S ונוכיח שהיא ממש קרובה יותר ל x: זאת תהיה סתירה, ובכך נוכיח את הטענה. נגדיר $\tilde{x}=\hat{x}+\alpha z$. אזי כיוון ש- S הוא מרחב לינארי וכיוון שגם $\hat{x}$ וגם z הם ב- S, הצירוף הלינארי גם הוא ב- S. נשים לב כי השגיאה עבור $\tilde{x}$ מאונכת ל- z כי מהלינאריות של המכפלה הסקלרית

$$(x-\tilde{x},z)=(x-\hat{x}-\alpha z,z)=(\varepsilon,z)-\alpha(z,z)=\alpha-\alpha=0. \tag{2.29}$$

כעת נחשב את המרחק בין $\tilde{x}$ ו- x:

$$||x-\tilde{x}||^2=(x-\tilde{x},x-\tilde{x}) \tag{2.30}$$
$$=(x-\hat{x}-\alpha z,x-\hat{x}-\alpha z) \tag{2.31}$$
$$=(x-\hat{x},x-\hat{x})+\alpha^2(z,z)-2\alpha(x-\hat{x},z) \tag{2.32}$$
$$=||x-\hat{x}||^2+\alpha^2-2\alpha^2 \tag{2.33}$$
$$=||x-\hat{x}||^2-\alpha^2 \tag{2.34}$$

ואכן $\tilde{x}$ קרוב יותר ל- x. בכך הוכחנו כי כל משערך אופטימלי חייב לקיים תכונת ניצבות (כמובן בתנאי שקיימת מכפלה פנימית).

לסיכום, ניזכר בהקבלה בין בעית השערוך הלינארי של מ"א לבין בעית ההיטל עם שגיאה ריבועית מינימלית. נבדוק

את האנלוגיה הבאה:

Y_1, Y_2 מ"א סקלריים $E[Y_i]=0 \longleftrightarrow \underline{Y}_1, \underline{Y}_2$ וקטורים דטרמיניסטיים ב- $\mathbb{R}^m$

$\mathrm{Cov}(Y_1,Y_2) \longleftrightarrow$ מכפלה סקלרית $(\underline{Y}_1, \underline{Y}_2)$

$\sqrt{\mathrm{Var}(Y)} \longleftrightarrow$ אורך וקטור $= (\underline{Y},\underline{Y})^{\frac{1}{2}} = \|\underline{Y}\|$

$\mathrm{Cov}(Y_1,Y_2)=0 \longleftrightarrow$ ניצבות $(\underline{Y}_1, \underline{Y}_2)=0$

תת מרחב במקרה הלינארי-מ"א $Y=\sum_{i=1}^{n}\theta_i Y_i \longleftrightarrow$ תת מרחב-וקטורים $\underline{Y}=\sum_{i=1}^{n}\theta_i \underline{Y}_i$

תת מרחב במקרה האופטימלי-מ"א-כל g $Y=g(\underline{Y}) \longleftrightarrow$

והאנלוגיה: ההטלה מאופיינת על ידי כך שהשגיאה ניצבת לכל וקטור בתת המרחב.

נעבור לבעית השערוך הלינארי של המ"א X מתוך $Y_1,\ldots,Y_n$; נפעיל את טבלת האנלוגיה לתוצאה ב-$\mathbb{R}^m$ ונקבל לכל i:

$$\mathrm{Cov}(Y_i,X)=\sum_{j=1}^{n} h_j^* \,\mathrm{Cov}(Y_i,Y_j)$$

וזה בדיוק (2.28)! ולכן האנלוגיה ועקרון ההשלכה נותנים תוצאות נכונות. בצורה זהה נקבל את עקרון הנצבות עבור שערוך אופטימלי---המשעך האופטימלי מאופיין על ידי ניצבות השגיאה.

<u>נסכם:</u> השגיאה של המשערך הלינארי האופטימלי $X-\hat{X}=X-\sum a_i^* Y_i$ ניצבת לתת המרחב שעליו אנו משליכים, דהיינו:

$$E\left[\left(X-\sum_{i=1}^{n} a_i^* Y_i\right)\cdot Y_j\right]=0; \quad \text{לכל } j$$

לגבי השגיאה הנותרת: ב-$\mathbb{R}^m$ אנו יודעים בגלל הנצבות שמתקיים

$$\|\underline{X}\|^2=\|\underline{\hat{X}}\|^2+\|\underline{X}-\underline{\hat{X}}\|^2$$

נראה זאת בחישוב מפורש עבור בעית השערוך:

$$\begin{aligned}E[\varepsilon^2] &= E\left[(X-\hat{X})^2\right]\\ &= E\left[X^2\right]+E\left[\hat{X}^2\right]-2E\left[X\hat{X}\right]\\ &= E\left[X^2\right]+E\left[\hat{X}^2\right]-2E\left[\hat{X}^2\right]\\ &= E[X^2]-E[\hat{X}^2]\end{aligned}$$

כאשר השויון הלפני אחרון נובע מהנצבות: $E\left[X\hat{X}\right]=E\left[\hat{X}^2\right]$. לסיכום, ניתן לחשוב על בעית השערוך הלינארי של מ"א X מתוך אוסף מ"א $Y_1,\ldots,Y_n$ כעל בעית היטל של X על המרחב הנפרש על ידי ה- Y_i, כלומר על ידי כל

הצירופים הלנאריים שלהם. המשערך האופטימלי הוא היחיד עבורו השגיאה ניצבת לכל מדידה, ותכונה זו מאפיינת את המשערך הלינארי האופטימלי.

<u>שים לב</u>: $E[(X-\sum a_i^* Y_i)\cdot Y_j]=0,\ j=1,\ldots,N$ מאפיין את המשערך הלינארי האופטימלי ואילו $E[(X-\hat{X}(\underline{Y}))\cdot g(Y_1,\ldots,Y_n)]=0$ מאפיין את המשערך האופטימלי (התוחלת המותנית). אפשר לכן לנסח עקרון השלכה גם עבור השערוך האופטימלי.

3 תהליכים אקראיים בזמן בדיד

"וקטור אקראי" (הגדרה 8.14) הוא אוסף סופי של משתנים אקראיים: זוהי הרחבה של מושג ה"משתנה אקראי" למספר מימדים. המושג "תהליך אקראי" הוא הרחבה נוספת, כאשר מספר המשתנים יכול להיות אין סופי. בנוסף, אנו מתייחסים לאוסף המשתנים כאילו הופיעו בנקודות זמן עוקבות.

דוגמה 3.1 נניח שמשדרים אות סיפרתי: 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 סידרה 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 סידרה של משתנים אקראיים $\{X(t),\ T_1 \le t \le T_2\}$ (כלומר t מקבל ערכים בדידים). פונקציית מדגם של התהליך האקראי היא כל פונקציה של משתנה הזמן t עבור $\omega = \omega_0$ קבוע. כלומר, זוהי כל פונקציה מהצורה

$$X(t,\omega_0),\ T_1 \le t \le T_2, \qquad \omega_0 \text{ קבוע}$$

בדרך כלל t מקבל ערכים שלמים, T_2 יהיה סופי או $+\infty$, ו-T_1 יקבל את הערך 0 או $(-\infty)$. 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) $X(t)$ או X_t וכו'.

3.1 חוק ההסתברות של תהליך אקראי

עבור וקטור אקראי, ניתן להגדיר פונקצית פילוג (הגדרה 8.15). אולם בבואנו להרחיב הגדרה זו לתהליך אקראי, אנו נתקלים בקשיים: כוון שתהליך אקראי יכול להיות מוגדר עבור אין-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 האין-סופי"

$$\mathbb{P}\{X_1 \le a_1, X_2 \le a_2, \dots\} \tag{3.1}$$

לכל סידרה אין-סופית $a_1, a_2, \dots$ אם ורק אם מתקיים התנאי הבא. לכל N, לכל סידרת זמנים (אינדקסים) $t_1 < t_2 < \cdots < t_N$ ולכל סידרה $a_1, a_2, \dots, a_N$ באורך N אנו יודעים לחשב את הפילוג של הווקטור

$$F_{X(t_1),X(t_2),\dots,X(t_N)}(a_1,a_2,\dots,a_N) = \mathbb{P}\{X(t_1)\le a_1, X(t_2)\le a_2,\dots,X(t_N)\le a_N\}$$

רמז: בדוק בניפרד את המקרה בו מספר הפעמים ש-$a_i < 1$ הוא אין-סופי (ואז ההיסתברות 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 אין סופי של תנאים, בדרך כלל הסתברות זו תהייה 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 סידרת זמנים סופית. כלומר, זהו האוסף של כל הפילוגים במימד סופי

$$F_{X(t_1),X(t_2),\dots,X(t_N)}(a_1,a_2,\dots,a_N) = \mathbb{P}\left\{X(t_1) \leq a_1, X(t_2) \leq a_2, \dots, X(t_N) \leq a_N\right\}$$

עבור כל N, כל סידרת זמנים (אינדקסים) $t_1 < t_2 < \cdots < t_N$ וכל סידרה $a_1,a_2,\dots,a_N$. הגודל $F_{X(t)}$, שהוא הפילוג של המשתנה האקראי $X(t)$ נקרא הפילוג החד-מימדי, או הפילוג השולי של התהליך בזמן t.

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

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

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

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

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

הגדרה 3.5 דרישת עיקביות: לכל $N \geq 2$ ו- $1 \leq k \leq N$, לכל סידרת זמנים (אינדקסים) $t_1 < t_2 < \cdots < t_N$ ולכל סידרה $a_1,a_2,\dots,a_N$ מתקיים התנאי

$$\begin{aligned}&\mathbb{P}\left\{X(t_1) \leq a_1, \dots, X(t_{k-1}) \leq a_{k-1}, X(t_{k+1}) \leq a_{k+1}, \dots, X(t_N) \leq a_N\right\} \\ &\qquad = \mathbb{P}\left\{X(t_1) \leq a_1, \dots, X(t_{k-1}) \leq a_{k-1}, X(t_k) < \infty, X(t_{k+1}) \leq a_{k+1}, \dots, X(t_N) \leq a_N\right\}\end{aligned}$$

או, בסימון של פונקציות פילוג,

$$\begin{aligned}&F_{X(t_1),\dots,X(t_{k-1}),X(t_{k+1}),\dots,X(t_N)}(a_1,\dots,a_{k-1},a_{k+1},\dots,a_N) \\ &\qquad = F_{X(t_1),\dots,X(t_N)}(a_1,\dots,a_{k-1},\infty,a_{k+1},\dots,a_N)\end{aligned}$$

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 $T_1, T_1+1, \dots, T_1+N$.

מספר דוגמאות של אותות אקראיים בזמן בדיד.

דוגמה 3.7 יהיה $\{X(n),\ n = 1, 2, \ldots\}$ אוסף של משתנים בלתי תלויים סטטיסטית ושווי פילוג (בקצרה i.i.d = independent, identically distributed) אפשר להתייחס לאוסף המסודר כאל תהליך אקראי בזמן בדיד, כאשר הזמנים הם שלמים $t = 1, 2, \ldots$. תהליך כזה נקרא רעש לבן. אם המ"א הם גאוסיים אזי התהליך נקרא רעש לבן גאוסי. יש לשים לב כי לעיתים משתמשים במונח "רעש לבן" ומתכוונים לרעש לבן גאוסי.

תהליך רעש לבן מופיע למשל כסידרת הזכיות בהפעלות חוזרות של מכונת הימורים. בנוסף, במקרים רבים מודל הרעש במדידות הוא כזה. בפרט, במקרים רבים משתמשים ברעש לבן גאוסי כמודל לרעש פיזיקלי.
הבה נראה כי הפילוג של המ"א $X(1)$ קובע את חוק הפילוג של תהליך הרעש הלבן, ונחשב חוק פילוג זה. נסמן ב-F_X את חוק הפילוג של המ"א $X(1)$ (שהוא גם חוק הפילוג של כל אחד מהמשתנים האחרים). יהיו $\{a_i,\ i = 1, 2, \ldots\}$ מספרים ממשיים כלשהם. אזי בגלל אי התלות, אם $\{t_1 < t_2 < \ldots < t_k$ נקבל

$$F_{X(t_1),X(t_2),\ldots,X(t_k)}(a_1, a_2, \ldots, a_k) = F_{X(t_1)}(a_1) \cdot F_{X(t_2)}(a_2) \cdots F_{X(t_k)}(a_k) = \prod_{i=1}^{k} F_X(a_i)$$

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

$$Y(0) = 0, \qquad Y(n) = Y(n-1) + X(n) \qquad n > 0$$

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

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

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

דוגמאות מעשיות 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,

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

בגלל אי התלות. בצורה דומה נמשיך ונפתח

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

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

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

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 $a_1,\ldots,a_n$ ניתן לחשב את הפילוג $\mathbb{P}\left\{A\cos(\omega_0 k+\phi)\le a_k,\ 1\le k\le n\right\}$ בדרך הבאה. עבור כל k הדרישה $A\cos(\omega_0 k+\phi)\le a_k$ קובעת אילוץ על הערכים של A,ϕ. ההסתברות שכל האילוצים מתקיימים היא בדיוק הפילוג המבוקש.

הגדרה 3.10 ת"א אקראי $\{X(n)\}$ יקרא גאוסי אם לכל $N,t_1,\ldots,t_N$ הווקטור האקראי $(X(t_1),\ldots,X(t_N))$ הוא וקטור גאוסי.

למשל רעש לבן עבורו הפילוג של כל אחד מהמשתנים האקראיים הוא גאוסי, הוא ת"א גאוסי (מדוע?).

3.2 תוחלת ומומנטים של תהליך אקראי

הגדלים הבסיסיים ביותר המתארים תהליך אקראי הם פונקציית התוחלת, הווריאנס ופונקצית אוטוקורלציה. עבור משתנים אקראיים, התוחלת (הגדרה 8.17), המומנטים, השונות (הגדרה 8.22), והקווריאנס (הגדרה 8.23), הם מספר-ים. בעזרתם נוכל להגדיר תכונות מקבילות של תהליך אקראי, הכוללות מומנטים עד (כולל) סדר שני, ע"י התבוננות במשתנה האקראי $X(t_1)$ או בזוג המשתנים $(X(t_1),X(t_2))$. הגדלים יהיו כמובן תלויים בזמנים $t_1,\ t_2$.

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

1. פונקציית התוחלת $\mu_X(t)=\mathbb{E}[X(t)]$
2. פונקצית האוטוקורלציה $\mathbf{R}_X(t_1,t_2)=\mathbb{E}\left[X(t_1)\cdot X(t_2)\right]$

3. פונקצית הקווריאנס

$$\mathbf{K}_X(t_1,t_2) = \mathbb{E}\left[(X(t_1)-\mu_X(t_1))\cdot(X(t_2)-\mu_X(t_2))\right]$$
$$= \mathbf{R}_X(t_1,t_2) - \mu_X(t_1)\cdot\mu_X(t_2)$$

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 רעש לבן (דוגמה 3.7) והילוך שיכור (דוגמה 3.8).

דוגמה 3.12 עבור התהליך $X(n)$,

$$\mu_X(n) = \mathbb{E}[X(n)]$$
$$= \mathbb{E}[X(1)]$$

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

$$\mathbf{R}_X(t_1,t_2) = \mathbb{E}[X(t_1)X(t_2)]$$
$$= \begin{cases} \mathbb{E}[(X(1))^2] & t_1 = t_2 \text{ אם} \\ (\mu_X)^2 & t_1 \neq t_2 \text{ אם.} \end{cases}$$

בגלל אי התלות בזמן, ובגלל אי התלות הסטטיסטית בין $X(t_1)$ לבין $X(t_2)$ כאשר הזמנים הם שונים. כיוון ש-

$$\mathbb{E}[(X(1))^2] = \mathbb{E}[((X(1)-\mu_x)+\mu_x)^2] = \sigma_x^2 + 2\,\mathbb{E}[\mu_x(X(1)-\mu_x)] + \mu_x^2 = \sigma_x^2 + \mu_x^2 \tag{3.4}$$

קיבלנו את הנוסחה הכללית

$$\mathbf{R}_X(t_1,t_2) = \mu_x^2 + \sigma_x^2\delta(t_1-t_2)\ . \tag{3.5}$$

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

$$\mathbf{K}_X(t_1,t_2) = \sigma_x^2\delta(t_1-t_2)\ . \tag{3.6}$$

נשים לב כי $\mathbf{K}_X(t_1,t_2)$ ו-$\mathbf{R}_X(t_1,t_2)$ תלויים במשתני הזמן רק דרך ההפרש t_1-t_2.

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

$$\mu_Y(n) = \mathbb{E}[Y(n)]$$
$$= \left[\sum_{i=1}^{n}\mathbb{E}[X(i)]\right]$$
$$= \sum_{i=1}^{n}\mu_X(i)$$

כוון שההפרשים הם שווי פילוג, נקבל את הנוסחה הפשוטה

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

כעת, עבור $k \leq n$,

$$\begin{aligned}
\mathbf{R}_Y(k,n) =& \mathbb{E}[Y(n) \cdot Y(k)] \\
=& \mathbb{E}\left[\sum_{i=1}^{n} X(i)\right] \cdot \left[\sum_{j=1}^{k} X(j)\right] \\
=& \mathbb{E}\left[\sum_{j=1}^{k}\sum_{i=1}^{n} X(i)X(j)\right] \\
=& \mathbb{E}\left[\sum_{j=1}^{k}\sum_{i=1,\ i\neq j}^{n} X(i)X(j) + \sum_{i=1}^{k}\big(X(i)\big)^2\right]
\end{aligned}$$

נסמן $\mu = \mathbb{E}[X(1)]$ ו-$\sigma^2 = \mathbb{E}[(X(i)-\mu)^2]$, ונשתמש באי התלות, כך ש-$\mathbb{E}[X(i)X(j)] = \mu^2$ עבור $i \neq j$. חישבנו כבר ש-$\mathbb{E}[(X(i))^2] = \sigma^2 + \mu^2$ ומכוון שלכל $j \leq k$ ישנו בדיוק ערך אחד של $i \leq n$ כך ש-$i = j$, נקבל

$$\begin{aligned}
\mathbb{E}[Y(n) \cdot Y(k)] =& k(n-1)\mu^2 + k(\sigma^2 + \mu^2) \\
=& kn\mu^2 + k\sigma^2 \\
=& \mathbf{R}_Y(k,n)
\end{aligned}$$

וכן

$$\mathbf{K}_Y(k,n) = \mathbb{E}[Y(n) \cdot Y(k)] - \mathbb{E}[Y(n)]\,\mathbb{E}[Y(k)] \tag{3.7}$$

$$= kn\mu^2 + k\sigma^2 - (n\mu)(k\mu) \tag{3.8}$$

$$= k\sigma^2\ . \tag{3.9}$$

לסיכום, עבור הילוך שיכור

$$\boxed{\mathbf{R}_Y(k,n) = k\sigma^2 + nk\mu^2 \qquad \text{כאשר } k \leq n:} \tag{3.10}$$

$$\boxed{\mathbf{K}_Y(k,n) = k\sigma^2 \qquad \text{כאשר } k \leq n:} \tag{3.11}$$

נשים לב כי אם $\{Y(n)\}$ הוא ת"א גאוסי אזי מההגדרות נובע כי $\mu_Y(n)$ ו- $\mathbf{R}_Y(k,n)$ קובעים את הפילוג של התהליך.

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

דוגמה 3.13 יהיה $\{X(n),\ n=0,1,\dots\}$ רעש לבן (תהליך אקראי המורכב ממ"א בלתי תלויים סטטיסטית). תהי g פונקציה נתונה, ונקבע את $Y(0)=y_0$. נגדיר תהליך $\{Y(n),\ n=0,1,\dots\}$ ע"י

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

זהו ניסוח כללי של אלגוריתם רקורסיבי "בנוכחות רעש". נוסחה זו היא כללית מדי מכדי שנוכל לנתח אותה. אולם ניתן לנתח מיקרים 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-$\mu_Y(t)$, ופונקציות האוטוקורלציה $\mathbf{R}_X(t_1,t_2)$ ו-$\mathbf{R}_Y(t_1,t_2)$. שרטט (רצוי בעזרת MATLAB) את פונקציות התוחלת, מספר פונקציות מדגם של התהליך, ואת הממוצע של פונקציות המדגם (אם פונקציות המדגם הן $X(t,\omega_1),X(t,\omega_2),\dots,X(t,\omega_k)$ אזי הממוצע של פונקציות המדגם הוא

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

שהיא פונקציה של משתנה הזמן t). שרטט את פונקציות האוטוקורלציה $\mathbf{R}_X(1,t)$ ו-$\mathbf{R}_Y(1,t)$ וכן את הפונקציות

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

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-

$$Y(n+1)=aY(n)+X(n)\ ,\ n\geq 0, \qquad Y(0)=0 \tag{3.12}$$

כאשר $|a|<1$. נחשב את המומנטים של $\{Y(n)\}$: מתוך משוואת ההפרשים,

$$\mu_y(n+1)\doteq \mathbb{E}\,Y(n+1) \tag{3.13}$$

$$=\mathbb{E}[aY(n)+X(n)] \tag{3.14}$$

$$=a\mu_y(n)+\mu_x(n)\ . \tag{3.15}$$

נשים לב שאפשר לחשוב על סידרת הממוצעים (שהיא סידרה דטרמיניסטית) $\{\mu_y(n)\}$ 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

$$\mu_y=a\mu_y+\mu_x \tag{3.16}$$

כלומר ל- $\mu_y = \mu_x/(1-a)$.

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

$$\sigma_y^2(n) = \mathbb{E}\, Y^2(n) \tag{3.17}$$

$$= \mathbb{E}\left[aY(n-1) + X(n)\right]^2 \tag{3.18}$$

$$= a^2\sigma_y^2(n-1) + 2a\,\mathbb{E}[Y(n-1)X(n)] + \sigma_x^2(n)\ . \tag{3.19}$$

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

$$\sigma_y^2(n) = a^2\sigma_y^2(n-1) + \sigma_x^2(n)\ . \tag{3.20}$$

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

$$\sigma_y^2(n) \to \frac{\sigma_x^2}{1-a^2}\ . \tag{3.21}$$

התהליך המתואר בדוגמה זו הוא מקרה פרטי של תהליך AR: Auto-Regressive.

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 0. אנו מפעילים את האלגוריתם הבא לחיפוש הערך של a:

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

שם 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).

כאשר מעבירים אות קבוע בזמן, או אות מחזורי, דרך מערכת לינארית קבועה בזמן ויציבה, מקבלים תגובה המורכבת מתופעת מעבר וממצב יציב. תופעת מעבר קשורה לתנאי ההתחלה, והתאמתם לאות הכניסה. אפשר לראות תופעה דומה בתהליכים אקראיים. תהליך קבוע (אולי פרט לגודל אקראי) הוא המקביל לאות קבוע, אך הוא אינו מעניין... משפחה מעניינת יותר של אותות בעלי תכונה חלשה יותר של "קביעות בזמן" הם האותות הסטציונריים. לאותות אלו יש התנהגות סטטיסטית שהיא קבועה בזמן במובן שנגדיר להלן, ונוכל לשאול האם כאשר אות כזה עובר במערכת לינארית וקבועה בזמן, גם אות המוצא הוא בעל תכונות סטטיסטיות קבועות בזמן.

הגדרה 3.17 **סטציונריות.** תהליך אקראי בזמן בדיד $\{X(n),\ -\infty < n < \infty\}$ נקרא **סטציונרי** אם לכל קבוע τ ולכל $t_1 < t_2$ הפילוג של $\{X(t_1), X(t_1+1), \ldots, X(t_2)\}$ זהה לפילוג של $\{X(t_1+\tau), X(t_1+1+\tau), \ldots, X(t_2+\tau)\}$.
כלומר, הפילוג אינו משתנה תחת הזזות זמן.
נאמר שהתהליך סטציונרי החל מזמן T אם התנאי מתקיים עבור כל $t_1 \geq T$ וכל $\tau > 0$.

דוגמה 3.18 יהיה Y מ"א כלשהוא, ויהיה Z מ"א המתאר זריקת 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

$$B(n) \doteq B \cdot (-1)^n \tag{3.22}$$

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 $\tilde{Z} = [Z+\tau] \bmod 6 + 1$ ($[m] \bmod 6$ 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, מכוון ש- $\sin[(\pi/3)n]$ היא פונקציה מחזורית עם מחזור באורך של 6, מתקיים

$$\sin[(\pi/3)(n+Z-1+\tau)] = \sin[(\pi/3)(n+[Z+\tau-1] \bmod 6)] \tag{3.23}$$

$$= \sin[(\pi/3)(n+\tilde{Z}-1)] \tag{3.24}$$

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. עבור רעש לבן, נסמן ב-F_x את הפילוג של המשתנה $X(n)$. נקבל

$$\mathbb{P}\left\{X(t_1)\le a_0, X(t_1+1)\le a_1,\dots,X(t_2)\le a_{t_2-t_1}\right\}=\prod_{i=0}^{t_2-t_1}F_x(a_i) \tag{3.25}$$

וברור שהפילוג אינו משתנה תחת הזזת זמן.

7. עבור הילוך שיכור ראינו כי הווריאנס תלוי בזמן, ולכן אפילו הפילוג של $Y(n)$ תלוי בזמן. לכן תהליך זה אינו סטציונרי.

ברור למדי כי סטציונריות היא תכונה אשר קשה לבדוק, גם כאשר כל המידע על הפילוגים מצוי בידנו. אם אנו מוכנים להסתפק בתאור התהליך על ידי מומנטים ראשון ושני, אזי ניתן להגדיר סוג חלש יותר של סטציונריות, אשר מספיקה ברוב יישומים:

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

1. המומנט הראשון אינו תלוי בזמן: $\mathbb{E}\,X(n)=\mu_x(n)=\mu_x$,
2. $\mathbf{R}(n_1,n_2)$ ולכן גם $\mathbf{K}(n_1,n_2)$ תלויים רק בהפרש n_2-n_1.

כאשר יש צורך לחדד את ההבדלים נקרא לסטציונריות במובן המקורי סטציונריות במובן הצר.

בהתייחס לדוגמה 3.18 כמובן שכל אות סטציונרי הוא גם סטציונרי במובן הרחב. ברור גם שהבדיקה קלה יותר---למשל חישבנו מומנט ראשון ושני עבור רעש לבן, והם מקיימים את התנאים. לעומת זאת הילוך שיכור אינו סטציונרי במובן הרחב כיוון שהממוצע תלוי בזמן, גם 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:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}g(X_n,\dots,X_{n+k})=Eg(X_0,\dots,X_k)\qquad\forall\omega \tag{3.26}$$

(למען הדיוק, נדרוש שהשוויון יתקיים בהסתברות 1, כלומר שלמאורע שיש שוויון תהייה הסתברות 1).

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

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

כיוון ש-

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

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

$$F_{X(0)}(a) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbf{1}_{(-\infty,a]}(X_n) = \lim_{N\to\infty} \frac{\{\text{מספר ה- } n\text{-ים, } 1 \le n \le N\text{, עבורם } X_n \le a\}}{N} \tag{3.27}$$

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

$$F_{X(0),X(k)}(a,b) = \lim_{N\to\infty} \frac{\{\text{מספר ה- } n\text{-ים, } 1 \le n \le N\text{, עבורם } X(n) \le a \text{ וכן } X(n+k) \le b\}}{N} \tag{3.28}$$

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

דוגמה 3.21 רעש לבן (דוגמה 3.7) הוא ארגודי. עובדה זו נובעת מחוק המספרים הגדולים. לעומת זאת, התהליך $\{X(n) = X,\ n = \ldots, -1, 0, 1, \ldots\}$ (אינו תלוי בזמן) אינו ארגודי, שכן הממוצע האריתמטי תמיד שווה ל-X.

דוגמה 3.22 התהליך

$$B(n) \doteq B \cdot (-1)^n \tag{3.29}$$

(משוואה (3.22)) הוא ארגודי, כיוון שלכל k ופונקציה g

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g(B(n), \ldots, B(n+k)) = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g(B(n), -B(n), \ldots, (-1)^{k-1}B(n), (-1)^k B(n)) \tag{3.30}$$

$$= \frac{1}{2} g(B(1), -B(1), \ldots, (-1)^{k-1}B(1), (-1)^k B(1)) \tag{3.31}$$

$$+ \frac{1}{2} g(B(2), -B(2), \ldots, (-1)^{k-1}B(2), (-1)^k B(2)) \tag{3.32}$$

$$= \mathbb{E}\left[g(B(0), \ldots, B(k))\right]. \tag{3.33}$$

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

$$A(n) \doteq (B(2n), B(2n+1)). \tag{3.34}$$

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 (3.26) יתקיימו בקירוב עבור T סופי. אנו גם מניחים ש"תופעות מעבר", אם ישנן, כבר דעכו.

דוגמה 3.23 **יהי $\{X(n), n = 0, 1, \dots\}$ רעש לבן. נגדיר משתנה אקראי נוסף $Y(0)$ ונדרוש שיהיה בת"ס ב-$\{X(n), n = 0, 1, \dots\}$. נגדיר כעת תהליך $\{Y(n),\ n = 0, 1, \dots\}$ 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). 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 הארגו-דיות (3.26):

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

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

4 שרשרות מרקוב

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

שרשרות מרקוב הן משפחה של תהליכים 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, לשרשרות מרקוב תאוריה וכלים מתמטיים מפותחים ונוחים לשימוש.

4.1 דוגמאות והגדרות

אחד המודלים הבסיסיים של רשתות מחשבים הוא תור עם שרת בודד.

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

התור עם שרת בודד קרוב מאד 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 $(1-\lambda-\mu)+2\cdot\lambda\cdot\mu$ התור נשאר ללא שינוי: אם בגלל שלא קרה דבר (בהסתברות $(1-\mu)(1-\lambda)$), 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 $A(n),\ D(n)$ ת"א בינריים i.i.d. ובת"ס זה בזה המייצגים הגעות וסיום שירות בהתאמה:

$$\mathbb{E}\,A(n)=\lambda,\quad \mathbb{E}\,D(n)=\mu\ . \tag{4.1}$$

אזי אורך התור מקיים את המשוואה

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

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

מודל כללי יותר עם תכונת מרקוביות הוא האלגוריתם הרקורסיבי המתואר בדוגמה 3.13. גם עבור אלגוריתם זה, בהנתן שהערך הוא $Y(n)$ ברגע n, אפשר לחשב את הפילוג של $Y(n+1)$. אם למשל ידוע ש-$Y(n)=y$, אזי ההסתברות

של המאורע $\{Y(n+1) \leq \alpha\}$ היא ההסתברות ש-$\{y + g(y, X(n)) \leq \alpha\}$. ברור כי הסתברות זו אינה תלוייה כלל בעבר הרחוק: ברגע שידוע הערך הנוכחי, נקבע פילוג הצעד הבא.

בפרק זה נעסוק בשרשרות מרקוביות, כלומר בתהליכים מרקוביים המקבלים ערכים בדידים. הדבר ייקל מאד על ההבנה. נזכר בהגדרה 8.28 של הסתברות מותנית.

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-$\{i, j_{n-k}, \dots, j_n\}$,

$$\mathbb{P}\{X(n+1) = i \mid X(n) = j_n, X(n-1) = j_{n-1}, \dots, X(n-k) = j_{n-k}\}$$
$$= \mathbb{P}\{X(n+1) = i \mid X(n) = j_n\}$$

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.

הביטוי $\mathbb{P}\{X(n+1) = i \mid X(n) = j_n\}$ 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 ולאו דוקא חיוביים. לשם נוחיות בסימונים נניח בדרך כלל כי ערכי התהליך הם שלמים.

שים לב שגם עבור שרשרת מרקובית,

$$\mathbb{P}\{X(n+1) = i \mid X(n) = j_n, X(n+2) = l\} \neq \mathbb{P}\{X(n+1) = i \mid X(n) = j_n\} \tag{4.3}$$

תרגיל 4.3 יהי $\{Z(n)\}$ רעש לבן 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 רעש לבן 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

$$\mathbb{P}\{X(n+1) = k \mid X(n) = i, X(0) = j\} = \mathbb{P}\{X(n+1) = k \mid X(n) = i\} \tag{4.4}$$

כלומר התלות היא במידע האחרון בלבד. ואכן

$$\mathbb{P}\{X(n+1)=k \mid X(n)=i, X(0)=j\} = \frac{\mathbb{P}\{X(n+1)=k, X(n)=i, X(0)=j\}}{\mathbb{P}\{X(n)=i, X(0)=j\}} \tag{4.5}$$

$$\mathbb{P}\{X(n+1)=k, X(n)=i, X(0)=j\} \tag{4.6}$$

$$= \sum_{i_{n-1}} \cdots \sum_{i_1} \mathbb{P}\{X(n+1)=k, X(n)=i, X(n-1)=i_{n-1}, \ldots, X(1)=i_1, X(0)=j\} \tag{4.7}$$

$$= \sum_{i_{n-1}} \cdots \sum_{i_1} \mathbb{P}\{X(n+1)=k \mid X(n)=i, \ldots, X(0)=j\} \, \mathbb{P}\{X(n)=i, \ldots, X(0)=j\} \tag{4.8}$$

$$= \sum_{i_{n-1}} \cdots \sum_{i_1} \mathbb{P}\{X(n+1)=k \mid X(n)=i\} \, \mathbb{P}\{X(n)=i, \ldots, X(0)=j\} \tag{4.9}$$

$$= \mathbb{P}\{X(n+1)=k \mid X(n)=i\} \, \mathbb{P}\{X(n)=i, X(0)=j\} \tag{4.10}$$

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

$$\mathbb{P}\{X(n+1)=k \mid X(n)=i, X(0)=j\} = \mathbb{P}\{X(n+1)=k \mid X(n)=i\} \tag{4.11}$$

באותה דרך ניתן להראות גם

תרגיל 4.4 עבור שרשרת מרקוב הראה כי

$$\mathbb{P}\{Y(n+m)=i_m, \ldots, Y(n+1)=i_1 \mid Y(n)=j_n, \ldots, Y(0)=j_0\} \tag{4.12}$$

$$= \mathbb{P}\{Y(n+m)=i_m, \ldots, Y(n+1)=i_1 \mid Y(n)=j_n\} \ .$$

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. כדי לתת לעובדה זו מובן מדויק, נתבונן בשרשרת מרקוב $\{X(n)\}$ ונגדיר תהליך חדש על ידי ההגדרה $Y(n) = X(-n)$. נוכיח כי גם $\{Y(n)\}$ הוא שרשרת מרקוב. מהגדרת תוחלת מותנית,

$$\mathbb{P}\{Y(n+1)=i \mid Y(n)=j_n, Y(n-1)=j_{n-1}, \ldots, Y(n-k)=j_{n-k}\} \tag{4.13}$$

$$= \frac{\mathbb{P}\{Y(n+1)=i, Y(n)=j_n, Y(n-1)=j_{n-1}, \ldots, Y(n-k)=j_{n-k}\}}{\mathbb{P}\{Y(n)=j_n, Y(n-1)=j_{n-1}, \ldots, Y(n-k)=j_{n-k}\}}$$

$$= \frac{\mathbb{P}\{Y(n-1)=j_{n-1}, \ldots, Y(n-k)=j_{n-k} \mid Y(n+1)=i, Y(n)=j_n\} \, \mathbb{P}\{Y(n+1)=i, Y(n)=j_n\}}{\mathbb{P}\{Y(n-1)=j_{n-1}, \ldots, Y(n-k)=j_{n-k} \mid Y(n)=j_n\} \, \mathbb{P}\{Y(n)=j_n\}}$$

נשתמש כעת בהגדרה של $Y(n)$, במרקוביות של $\{X(n)\}$ ובקשר (4.12) ונקבל

$$\begin{aligned}
\mathbb{P}\{Y(n-1)=j_{n-1},\dots,Y(n-k)=j_{n-k}\mid Y(n+1)=i,Y(n)=j_n\} \\
&= \mathbb{P}\{X(-n+1)=j_{n-1},\dots,X(-n+k)=j_{n-k}\mid X(-n-1)=i,X(-n)=j_n\} \\
&= \mathbb{P}\{X(-n+1)=j_{n-1},\dots,X(-n+k)=j_{n-k}\mid X(-n)=j_n\} \\
&= \mathbb{P}\{Y(n-1)=j_{n-1},\dots,Y(n-k)=j_{n-k}\mid Y(n)=j_n\}
\end{aligned} \tag{4.14}$$

כאשר בשורה האחרונה ניצלנו את המרקוביות של $\{X(n)\}$. נציב זאת ב-(4.13) ונקבל

$$\begin{aligned}
\mathbb{P}\{Y(n+1)=i\mid Y(n)=j_n,Y(n-1)=j_{n-1},\dots,Y(n-k)=j_{n-k}\} \\
&= \frac{\mathbb{P}\{Y(n-1)=j_{n-1},\dots,Y(n-k)=j_{n-k}\mid Y(n)=j_n\}\,\mathbb{P}\{Y(n+1)=i,Y(n)=j_n\}}{\mathbb{P}\{Y(n-1)=j_{n-1},\dots,Y(n-k)=j_{n-k}\mid Y(n)=j_n\}\,\mathbb{P}\{Y(n)=j_n\}} \\
&= \frac{\mathbb{P}\{Y(n+1)=i,Y(n)=j_n\}}{\mathbb{P}\{Y(n)=j_n\}} \\
&= \mathbb{P}\{Y(n+1)=i\mid Y(n)=j_n\}
\end{aligned} \tag{4.15}$$

כלומר $\{Y(n)\}$ גם הוא שרשרת מרקובית. נסכם זאת:

טענה 4.5 אם $\{X(n)\}$ היא שרשרת מרקובית אזי גם $\{Y(n)\}\doteq\{X(-n)\}$ היא שרשרת מרקובית. הסתברויות המעבר של $\{Y(n)\}$ נתונות על ידי

$$\mathbb{P}\{Y(n+1)=j\mid Y(n)=i\}=\mathbb{P}\{X(-n)=i\mid X(-n-1)=j\}\,\frac{\mathbb{P}\{X(-n-1)=j\}}{\mathbb{P}\{X(-n)=i\}}\ . \tag{4.16}$$

הביטוי הראשון בצד שמאל הוא בדיוק הסתברות המעבר של X. אנו רואים כי גם במקרה שהסתברויות המעבר של X אינן תלויות במפורש בזמן, הסתברויות המעבר של Y תלויות בזמן.

הוכחה: נובע מיידית מהגדרת הסתברות מותנית.

תרגיל 4.6 יהי $\{Z(n)\}$ רעש לבן 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

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

(דוגמה 3.13) מייצר תהליך מרקובי $\{Y(n),n\ge 0\}$. 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 רעש לבן 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 שרשרות הומוגניות

נתרכז מעתה במקרה בו מספר המצבים סופי וכן הסתברויות המעבר אינן תלויות במפורש 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

$$\mathbb{P}\left\{X(n+1)=j \mid X(n)=i\right\}=\mathbb{P}\left\{X(n)=j \mid X(n-1)=i\right\}$$

כלומר אם הסתברויות המעבר אינן תלויות בזמן.

במקרה זה נשתמש באחד מהסימונים המקובלים הבאים:

$$\boxed{\mathbb{P}\left\{X(n+1)=j \mid X(n)=i\right\}=\mathrm{p}_{ij}=\mathrm{p}(j \mid i)}$$

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

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

כאשר מספר המצבים כאן הוא N ובחרנו לקרוא להם $\{1,2,\ldots,N\}$. כלומר $S=\{1,2,\ldots,N\}$. 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 שרשרת הומוגנית אינה בהכרח תהליך אקראי סטציונרי! מסיבה זו אנו נקפיד להשתמש במינוח "שרשרת הומוגנית".

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

דוגמה 4.8 בהמשך לדוגמה 4.1, נניח שבתור יש מספר סופי K של מקומות. במקרה זה, אם מגיע משתמש לתור כאשר התור מלא, הוא נחסם (ולכן אינו משפיע על אורך התור). את אורך התור נוכל כעת לייצג בעזרת המשוואה

$$Y(n+1) = Y(n) + A(n)I[Y(n) < K] - D(n)I[Y(n) > 0] \; . \tag{4.17}$$

לכן, הסתברויות המעבר הן:

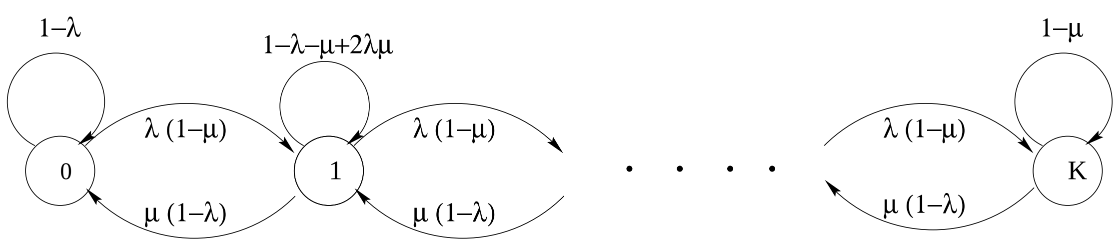
$$\mathrm{p}_{ij} = 0 \qquad \text{אם } |i-j| > 1$$

$$\mathrm{p}_{i(i+1)} = \begin{cases} \lambda & \text{אם } i = 0 \\ \lambda(1-\mu) & \text{אם } 0 < i < K \\ 0 & \text{אם } i = K \end{cases}$$

$$\mathrm{p}_{ii} = \begin{cases} 1-\lambda & \text{אם } i = 0 \\ 1-\mu & \text{אם } i = K \\ (1-\lambda)(1-\mu) + \lambda \cdot \mu & \text{אחרת} \end{cases}$$

$$\mathrm{p}_{i(i-1)} = \begin{cases} 0 & \text{אם } i = 0 \\ \mu(1-\lambda) & \text{אם } 0 < i < K \\ \mu & \text{אם } i = K \end{cases}$$

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



איור 4.1: שרשרת מרקוב: תור סופי

בדוגמה זו, מצבי השרשרת הם $\{0, 1, \ldots, K\}$. 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 אינה 0. השרשרת מדוגמה 4.8 נקראת גם "תהליך לידה מיתה".

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

$$0 \le \mathrm{p}_{ij} \le 1$$
$$\sum_j \mathrm{p}_{ij} = 1 \qquad \forall i \in S\ .$$

כלומר סכום אברי 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-

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

כלומר ווקטור העמודה $(1,1,\dots,1)^T$ הוא 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 נקבע שרשרת מרקוב מסויימת $\{X(n),\ n=0,1,\dots\}$, ונסמן את מצבי השרשרת ב- $S=\{1,2,\dots,K\}$. נסמן את ההסתברות שהשרשרת נמצאת בזמן n במצב k ב-$\nu_k(n)$ כלומר $\nu_k(n) \doteq \mathbb{P}\{X(n)=k\}$. 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-

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

תרגיל 4.11 הראה כי עבור $m>n$ שרשרת הומוגנית מקיימת

$$\mathbb{P}\{X(m)=k \mid X(n)=i, X(0)=j\} = \mathbb{P}\{X(m)=k \mid X(n)=i\} \tag{4.18}$$

רמז: השתמש בשיטת ההוכחה של משוואה (4.4).

טענה 4.12 (ללא הוכחה; ראה תרגיל *4.11*). עבור שרשרת מרקוב, לכל סדרת זמנים (שלמים) $n > n_k > n_{k-1} > \cdots > n_1$ מתקיים

$$\mathbb{P}\left\{X(n) = j \mid X(n_k) = i_k, \ldots, X(n_1) = i_1\right\} = \mathbb{P}\left\{X(n) = j \mid X(n_k) = i_k\right\}. \tag{4.19}$$

טענה 4.13 חישוב הפילוג והסתברויות המעבר של שרשרת הומוגנית וסופית.

1. ההסתברות למעבר ב-n צעדים מקיימת את המשוואה הרקורסיבית

$$\mathrm{p}_{ij}^{(n)} = \sum_k \mathrm{p}_{ik}^{(n-1)}\, \mathrm{p}_{kj}\ , \qquad \mathrm{p}_{ij}^{(1)} = \mathrm{p}_{ij}$$

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

$$\mathrm{p}_{ij}^{(n+m)} = \sum_k \mathrm{p}_{ik}^{(n)} \cdot \mathrm{p}_{kj}^{(m)}$$
$$\mathrm{P}^{(n+m)} = \mathrm{P}^n \cdot \mathrm{P}^m$$

3. את פונקציית ההסתברות של התהליך ברגע n ניתן לחשב על ידי

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

כאשר הביטוי האחרון הוא בייצוג ווקטורי.

הוכחת הטענה. נרשום מנוסחת ההסתברות השלמה

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

מהגדרת הסתברות מותנית

$$\begin{aligned}&\mathbb{P}\left\{X(n) = j,\ X(n-1) = k \mid X(0) = i\right\} \\ &\qquad = \mathbb{P}\left\{X(n) = j \mid X(n-1) = k,\ X(0) = i\right\} \cdot \mathbb{P}\left\{X(n-1) = k \mid X(0) = i\right\}\end{aligned}$$

ובגלל המרקוביות,

$$\mathbb{P}\left\{X(n)=j \mid X(n-1)=k,\ X(0)=i\right\}=\mathbb{P}\left\{X(n)=j \mid X(n-1)=k\right\}$$

לכן נקבל

$$\begin{aligned}
\mathrm{p}_{ij}^{(n)} &= \mathbb{P}\left\{X(n)=j \mid X(0)=i\right\} \\
&= \sum_k \mathbb{P}\left\{X(n)=j,\ X(n-1)=k \mid X(0)=i\right\} \\
&= \sum_k \mathbb{P}\left\{X(n)=j \mid X(n-1)=k,\ X(0)=i\right\} \cdot \mathbb{P}\left\{X(n-1)=k \mid X(0)=i\right\} \\
&= \sum_k \mathbb{P}\left\{X(n)=j \mid X(n-1)=k\right\} \cdot \mathbb{P}\left\{X(n-1)=k \mid X(0)=i\right\} \\
&= \sum_k \mathrm{p}_{ik}^{(n-1)}\,\mathrm{p}_{kj}
\end{aligned}$$

ובכך סיימנו להוכיח את טענה 1. כיוון שלפי ההגדרה

$$\mathrm{P}=\{\mathrm{p}_{ij}\}_{ij}$$

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

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

או בסימון וקטורי

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

ובכך סיימנו את ההוכחה.

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:

$$
\begin{aligned}
&\mathbb{P}\left\{X(5)=i,\ X(100)=j\right\} \\
&\qquad = \sum_k \mathbb{P}\left\{X(100)=j,\ X(5)=i,\ X(0)=k\right\} \\
&\qquad = \sum_k \mathbb{P}\left\{X(100)=j \mid X(5)=i,\ X(0)=k\right\}\cdot \mathbb{P}\left\{X(5)=i,\ X(0)=k\right\} \\
&\qquad = \sum_k \mathbb{P}\left\{X(100)=j \mid X(5)=i\right\}\cdot \mathbb{P}\left\{X(5)=i \mid X(0)=k\right\}\cdot \mathbb{P}\left\{X(0)=k\right\} \\
&\qquad = \mathrm{p}_{ij}^{(95)}\cdot \sum_k \mathrm{p}_{ki}^{(5)}\,\nu_k(0) = \left(\mathrm{P}^{95}\right)_{ij}\sum_k \left(\mathrm{P}^{5}\right)_{ki}\nu_k(0)\ .
\end{aligned}
$$

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

טענה 4.14 **חוק הפילוג של שרשרת מרקובית הומוגנית $\{X(n),\ n\geq 0\}$ נקבע באופן חד משמעי על ידי הפילוג בזמן 0, ומטריצת הסתברויות המעבר.**

הוכחה: נבחר סדרת זמנים $\{t_1<t_2<\ldots<t_k\}$ וסדרת מצבים $\{j_1,j_2,\ldots,j_k\}$ כלשהם. נחשב את הפילוג המשותף:

$$
\begin{aligned}
&\mathbb{P}\left\{X(t_k)=j_k, X(t_{k-1})=j_{k-1},\ldots,X(t_1)=j_1\right\} \\
&\qquad = \mathbb{P}\left\{X(t_k)=j_k \mid X(t_{k-1})=j_{k-1},\ldots,X(t_1)=j_1\right\} \\
&\qquad\quad \times \mathbb{P}\left\{X(t_{k-1})=j_{k-1},\ldots,X(t_1)=j_1\right\} \\
&\qquad = \mathbb{P}\left\{X(t_k)=j_k \mid X(t_{k-1})=j_{k-1}\right\}\cdot \mathbb{P}\left\{X(t_{k-1})=j_{k-1},\ldots,X(t_1)=j_1\right\}
\end{aligned}
$$

בגלל המרקוביות. נמשיך בשיטה זהה ונקבל

$$
\begin{aligned}
&= \mathbb{P}\left\{X(t_k)=j_k \mid X(t_{k-1})=j_{k-1}\right\}\cdot \mathbb{P}\left\{X(t_{k-1})=j_{k-1} \mid X(t_{k-2})=j_{k-2}\right\} \\
&\quad \times\cdots\ \mathbb{P}\left\{X(t_2)=j_2 \mid X(t_1)=j_1\right\}\cdot \mathbb{P}\left\{X(t_1)=j_1\right\} \\
&= \left(\prod_{l=2}^{k}\mathbb{P}\left\{X(t_l)=j_l \mid X(t_{l-1}=j_{l-1})\right\}\right)\cdot \mathbb{P}\left\{X(t_1)=j_1\right\} \\
&= \left(\prod_{l=2}^{k}\mathrm{p}_{j_{l-1}\,j_l}^{(t_l-t_{l-1})}\right)\cdot \nu_{j_1}(t_1) = \left(\prod_{l=2}^{k}\mathrm{P}^{t_l-t_{l-1}}\right)_{j_{l-1}\,j_l}\cdot \nu_{j_1}(t_1)
\end{aligned}
$$

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

משפט 4.15 יהי $\underline{\nu}$ פילוג שהוא וקטור עצמי שמאלי של מטריצת המעברים P עם ערך עצמי 1 כלומר $\underline{\nu}\,\mathrm{P}=\underline{\nu}$, אזי השרשרת עם פילוג התחלתי $\underline{\nu}(0)=\underline{\nu}$ היא תהליך אקראי סטציונרי החל מזמן 0 (ראה הגדרה *3.17*).

הוכחה: ממשוואת צ'פמן קולמוגורוב נובע כי $\underline{\nu}(n)=\underline{\nu}(n-1)$. סטציונרית נובעת לכן מיידית מהוכחת משפט 4.14.

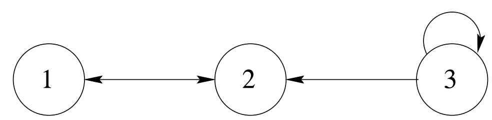
<u>הערה:</u> את החישובים שעשינו ניתן להרחיב למקרה של שרשרות שאינן סופיות. הקושי היחידי הוא בסימון המטריצי. אולם נחשוב על "מטריצות אין-סופיות", אשר מקיימות את כללי החיבור והכפל הרגילים בין מטריצות. אזי כל הפיתוחים שעשינו תקפים, ולכן המסקנות נכונות גם עבור שרשרות לא סופיות.

4.4 מצבים נשנים וחולפים

על מנת להבין את הדינמיקה וההתנהגות של שרשרות מרקוב בזמנים ארוכים יש לעמוד על כמה מושגי יסוד ותכונות בסיסיות של שרשרות מרקוב.

ניתן לסווג את המצבים של שרשרת מרקוב, לפי הקריטריון: האם נחזור למצב שוב ושוב, או שמא נבקר בו רק מספר סופי של פעמים?

דוגמה 4.16 בדוגמה שלפנינו לא רשומות הסתברויות המעבר. אולם כל קשת מתארת הסתברות חיובית.



איור 4.2: מצבים חולפים ונשנים

אם נתחיל במצב 3, אזי בכל צעד יש סיכוי לעבור למצב 2, ולכן מעבר זה מובטח (אם כי לא ברור מתי). מרגע שעברנו, לא ניתן לחזור: לכן מספר הפעמים שנהיה במצב זה הוא סופי (אם כי אקראי). לעומת זאת, למצבים 1 ו-2 נחזור שוב ושוב.

דוגמה 4.17 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. האיום יתכן שהמשחק ימשך לנצח? אם לא, פירוש הדבר כי מספר הפעמים שנהיה בכל מצב שאינו 0 הוא סופי (אקראי, כמובן).

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

$$\begin{aligned}\rho_{ji} &= \mathbb{P}\left\{X(n)=i,\ \text{כלשהו } n>0 \text{ עבור} \mid X(0)=j\right\}\\ &= \sum_{n=1}^{\infty}\mathbb{P}\left\{X(n)=i, X(n-1)\neq i,\ldots,X(2)\neq i, X(1)\neq i \mid X(0)=j\right\}\end{aligned}$$

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

דוגמה 4.19 אם נתחיל במצב 1, לא נוכל לעבור למצב 2, ולהיפך. לפי ההגדרה, שני המצבים נשנים.

איור 4.3: מצבים נשנים

מצב 3 באיור 4.2 הוא מצב חולף שכן יש הסתברות (שאינה אפס) שנעזוב מצב זה בצעד ראשון - ובמקרה זה לא נחזור אליו לעולם.

הגדרה 4.20 תהליך מרקובי נקרא <u>תהליך חולף</u> 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.

לכאורה, לא הגבלנו את מספר הביקורים במצב חולף אם התחלנו במצב אחר. אולם בגלל המרקוביות, הדברים קשורים.
נסמן ב-N_i את מספר הביקורים של התהליך במצב i: זהו משתנה אקראי המקיים $N_i = \sum_{n=1}^{\infty} I\{x(n) = i\}$.

טענה 4.21 מצב i הוא חולף אם ורק אם לכל j,

$$\mathbb{E}\left[N_i \mid X(0) = j\right] < \infty \tag{4.20}$$

הוכחה: לצורך חישוב התוחלת נשתמש בקשר $\mathbb{E}\, N = \sum_{k=1}^{\infty} \mathbb{P}\left\{N \geq k\right\}$ (המתקיים למשתנה חיובי שערכיו שלמים - ראה טענה 8.21). ראשית, $N_i \geq 1$ משמעותו לפחות ביקור אחד במצב i ולכן

$$\mathbb{P}\left\{N_i \geq 1 \mid X(0) = j\right\} = \rho_{ji} \tag{4.21}$$

כעת $N_i \geq 2$ משמעו ביקור ב-i, נאמר בזמן m, ואחריו ביקור נוסף ב-i כלומר

$$\begin{aligned}
&\mathbb{P}\left\{N_i \geq 2 \mid X(0)=j\right\} \\
&= \sum_{m=1}^{\infty} \sum_{n=m+1}^{\infty} \mathbb{P}\left\{X(n)=i, X(n-1) \neq i, \ldots, X(m+1) \neq i, X(m)=i, X(m-1) \neq i, \ldots, X(1) \neq i \mid X(0)=j\right\} \\
&= \sum_{m=1}^{\infty} \sum_{n=m+1}^{\infty} \mathbb{P}\left\{X(n)=i, X(n-1) \neq i, \ldots, X(m+1) \neq i \mid X(m)=i, X(m-1) \neq i, \ldots, X(1) \neq i, X(0)=j\right\} \\
&\quad \times \mathbb{P}\left\{X(m)=i, X(m-1) \neq i, \ldots, X(1) \neq i \mid X(0)=j\right\} \\
&= \sum_{m=1}^{\infty} \sum_{n=m+1}^{\infty} \mathbb{P}\left\{X(n)=i, X(n-1) \neq i, \ldots, X(m+1) \neq i \mid X(m)=i\right\} \\
&\quad \times \mathbb{P}\left\{X(m)=i, X(m-1) \neq i, \ldots, X(1) \neq i \mid X(0)=j\right\} \\
&= \sum_{n=1}^{\infty} \mathbb{P}\left\{X(n)=i, X(n-1) \neq i, \ldots, X(1) \neq i \mid X(0)=i\right\} \\
&\quad \times \sum_{m=1}^{\infty} \mathbb{P}\left\{X(m)=i, X(m-1) \neq i, \ldots, X(1) \neq i \mid X(0)=j\right\} \\
&= \rho_{ii} \rho_{ji}.
\end{aligned}$$

כאשר במעבר הלפני אחרון השתמשנו בהומוגניות, ובאחרון בהגדרת ρ. באותה צורה נקבל גם

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

כעת, עבור מצב חולף, לפי ההגדרה $\rho_{ii} < 1$. נחשב את התוחלת בעזרת טענה 8.21:

$$\begin{aligned}
\mathbb{E}[N_i \mid X(0)=j] &= \sum_{k=1}^{\infty} \mathbb{P}\left\{N_i \geq k \mid X(0)=j\right\} \\
&= \sum_{k=1}^{\infty} \rho_{ji} \cdot \rho_{ii}^{k-1} \\
&= \rho_{ji} \frac{1}{1-\rho_{ii}} < \infty
\end{aligned} \tag{4.22}$$

שכן כאמור $\rho_{ii} < 1$. הוכחנו כי עבור מצב חולף, $\mathbb{E}[N_i \mid X(0)=j] < \infty$. לעומת זאת, אם המצב נשנה $\rho_{ii} = 1$ ואז לפחות עבור $j = i$ הסכום במשוואה (4.22) יתן ∞. בכך הוכחנו את הטענה.

בדוגמה 4.16 מצב 3 הוא חולף: נראה כי $\rho_{33} < 1$. כיוון שלא ניתן לחזור ממצב 1 או 2 למצב 3, נובע שאם $X(k) \neq 3$ אזי בהכרח $X(n) \neq 3$ לכל $n > k$. לכן מהגדרת ρ_{33} נובע כי

$$\rho_{33} = \mathbb{P}\left\{X(1)=3 | X(0)=3\right\} + 0 < 1 \tag{4.23}$$

ואכן זהו מצב חולף.

הערה: מספיק לבדוק את סופיות הביטוי במשוואה (4.20) עבור $j = i$.

כעת נשתמש בתוצאה זו כדי להראות את התוצאה (האינטואיטיבית) הבאה.

טענה 4.22 שרשרת מרקובית בעלת מספר סופי של מצבים אינה חולפת, כלומר יש לה לפחות מצב נשנה אחד.

הוכחה: נניח שקבענו מצב התחלתי, ונשמיט אותו מהסימונים. נשים לב כי מספר הביקורים במצב כלשהוא 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

$$\infty = \mathbb{E}\left[\sum_i N_i\right] = \sum_i \mathbb{E}\left[N_i\right]$$

כיוון שמספר המצבים הוא סופי, לא יתכן שיתקיים בו זמנית

$$\mathbb{E}\left[\sum_i N_i\right] = \infty, \qquad \mathbb{E}\left[N_i\right] < \infty \text{ לכל } i$$

ולכן לפחות מצב אחד אינו חולף.

טענה זו נותנת גם שיטה נוספת לבדוק אם מצב הוא חולף, על ידי חישוב תוחלת מספר הביקורים בו.

טענה 4.23 $\mathbb{E}\left[N_i \mid X(0) = i\right] = \sum_{n=1}^{\infty} \mathrm{p}_{ii}^{(n)}$

הטענה נותנת שיטה נוספת לבדיקה אם מצב i הוא חולף - אם הביטויים שווים ∞.

הוכחה: נתחיל במצב i ונזכר כי לפי הטענה הקודמת, מצב הוא נישנה אם ורק אם $\mathbb{E}\left[N_i \mid X(0) = i\right] = \infty$. נזכר בסימון של פונקציה מציינת, ונחשב

$$\begin{aligned}
\mathbb{E}\left[N_i \mid X(0) = i\right] &= \mathbb{E}\left[\sum_{n=1}^{\infty} 1_i[X(n)] \mid X(0) = i\right] \\
&= \sum_{n=1}^{\infty} \mathbb{E}\left[1_i[X(n)] \mid X(0) = i\right] \\
&= \sum_{n=1}^{\infty} \mathbb{P}\left\{X(n) = i \mid X(0) = i\right\} \\
&= \sum_{n=1}^{\infty} \mathrm{p}_{ii}^{(n)}
\end{aligned}$$

כדרוש.

דוגמה 4.24 נניח שבכל רגע יכולים להגיע למחשב לכל היותר משימה אחת מסוג א' ומשימה אחת מסוג ב'. נניח שההגעות הן בלתי תלויות סטטיסטית. נניח שבכל רגע המחשב מטפל בדיוק במשימה אחת, כאשר למשימות מסוג א' יש עדיפות. כלומר, בכל רגע תצא משימה מסוג א' מהמערכת, אלא אם כן אין במערכת משימות כאילו: במקרה זה יופנה המחשב למשימות מסוג ב'. המצב כאן הוא הווקטור

$$\underline{X} = \begin{pmatrix} \text{מספר משימות ממתינות מסוג א'} \\ \text{מספר משימות ממתינות מסוג ב'} \end{pmatrix}$$

ברור לפי תאור המערכת כי אם ברגע כלשהוא לא נותרו משימות מסוג א' במערכת, אזי בעתיד תוכל להיות במערכת לכל היותר משימה אחד מסוג זה. אולם יתכן שהתחלנו את פעולת המערכת עם מספר משימות גדול. לכן יתכן שבפרק זמן התחלתי תהיינה במערכת מספר משימות מסוג א', אולם כל מצב מהצורה

$$\underline{X} = \begin{pmatrix} k \\ m \end{pmatrix}$$

כאשר $k > 1$ הוא מצב חולף.

4.5 פרוק מרחב המצב

נניח שמצב i הוא מצב נשנה, ונניח כי $\mathrm{p}_{ij}^{(n)} = 1$ וכן $\mathrm{p}_{ji}^{(m)} = 1$ עבור n, m כלשהם. אזי ברור כי מצב j גם הוא נשנה. אולם יש תוצאות מדויקות יותר בנושא. בפרק זה נראה כי ניתן לסווג את המצבים לקבוצות בהתאם לתכונות הנישנות שלהם.

הגדרה 4.25 אם $\mathrm{p}_{ij}^{(n)} > 0$ עבור n כלשהוא או ש-$i = j$ אזי נאמר ש-i מוביל ל-j, ונסמן $i \to j$. אם $j \to i$ וגם $i \to j$ נאמר ש-i ו-j מקושרים (communicating), ונסמן $i \leftrightarrow j$.

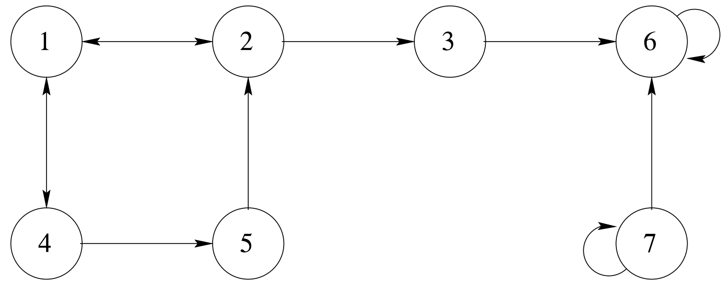
נשים לב כי לפי הגדרה, כל מצב מקושר לעצמו גם אם הסתברות המעבר מהמצב לעצמו היא 0.

בסימון הקודם, קיים n כך ש-$\mathrm{p}_{ij}^{(n)} > 0$ אם ורק אם $\rho_{ij} > 0$. מבחינת דיאגרמת המצבים, $i \to j$ אם (ורק אם) קיימת מסילה בדיאגרמה המובילה מ-i ל-j. כתוצאה מכך, $i \leftrightarrow j$ אם ורק אם קיימת בדיאגרמה מסילה סגורה מ-i לעצמו, העוברת דרך j.

דוגמה 4.26 בדוגמה זו, מצבים $1, 2$ מקושרים, וגם מצבים $1, 4$ מקושרים. פחות מיידי לראות כי $1, 5$ מקושרים: 1 מוביל ל-5 דרך $1 \to 4 \to 5$ 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).

טענה 4.27 <u>תכונות יחס הקשירות</u>

1. $i \leftrightarrow i$ רפלקסיביות reflexive relation



איור 4.4: מצבים קשורים

2. $i \leftrightarrow j$ אם ורק אם $j \leftrightarrow i$ סימטריה symmetry

3. אם $i \to j$ וגם $j \to k$ אזי $i \to k$ טרנזיטיביות transitivity

4. אם $i \leftrightarrow j$ אזי שניהם נשנים או שניהם חולפים.

<u>הוכחה:</u> שלושת הטענות הראשונות נובעות מיידית מההגדרות (ראה דוגמה). נוכיח את 4. כיוון ש-$i \leftrightarrow j$ קיימים m, k כך ש-

$$\mathrm{p}_{ij}^{(m)} > 0 \qquad \text{וגם} \qquad \mathrm{p}_{ji}^{(k)} > 0$$

נניח ש-i נשנה ונבדוק את הקריטריון עבור j

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathrm{p}_{jj}^{(n)} &\geq \sum_{n=1}^{\infty} \mathrm{p}_{jj}^{(k+n+m)} = \sum_{n=1} \sum_{i_1, i_2} \mathrm{p}_{ji_1}^{(k)} \, \mathrm{p}_{i_1 i_2}^{(n)} \, \mathrm{p}_{i_2 j}^{(m)} \\
&\geq \sum_{n=1}^{\infty} \mathrm{p}_{ji}^{(k)} \, \mathrm{p}_{ii}^{(n)} \, \mathrm{p}_{ij}^{(m)} = \mathrm{p}_{ji}^{(k)} \left(\sum_{n=1}^{\infty} \mathrm{p}_{ii}^{(n)} \right) \mathrm{p}_{ij}^{(m)}
\end{aligned}$$

אם i נשנה אזי ראינו כי

$$\sum_{n=1}^{\infty} \mathrm{p}_{ii}^{(n)} = \infty$$

ולכן גם

$$\sum_{n=1}^{\infty} \mathrm{p}_{jj}^{(n)} = \infty$$

וקיבלנו כי j נשנה. מצד שני, אם j חולף אזי

$$\sum_{n=1}^{\infty} \mathrm{p}_{jj}^{(n)} < \infty$$

ואזי גם

$$\sum_{n=1}^{\infty} \mathrm{p}_{ii}^{(n)} < \infty$$

וגם i חולף. בכך הראינו את תכונה 4.

מסקנה מתכונות 1-3: 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.

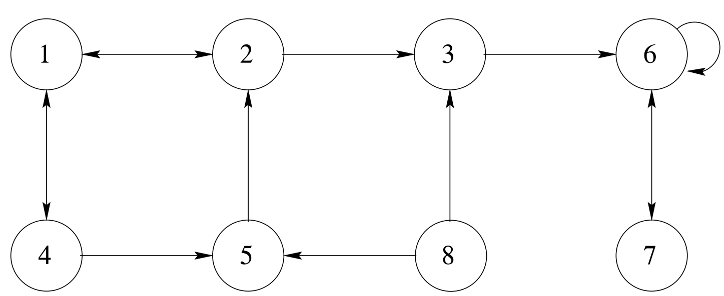
הגדרה 4.28 אוסף מצבים נקרא קבוצה סגורה אם ההסתברות לצאת מהקבוצה היא אפס. כלומר קבוצה היא סגורה אם מתקיים התנאי הבא. לכל מצב i בקבוצה ולכל מצב j שאינו בקבוצה, $\mathrm{p}_{ij}^{(n)} = 0$ לכל n.

מהגדרה זו נובע כי בדיאגרמת המעברים, אין קשתות היוצאות מקבוצה סגורה. למשל בדוגמה של איור 4.4 מצבים $6,7$ מהווים קבוצה סגורה. בנוסף אוסף כל המצבים הוא (כמובן) קבוצה סגורה.

הגדרה 4.29 קבוצת הקשירות של מצב i היא קבוצת השקילות שלו ביחס ליחס הקשירות. כלומר זוהי קבוצת כל המצבים המקושרים ל-i. קבוצת קשירות A נקראת קבוצה סגורה אם כל מצב ב-A מוביל רק למצבים ב-A.

קבוצת נקראת נישנית אם כל מצביה נישנים.

דוגמה 4.30 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 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.



איור 4.5: סיווג מצבים

בקבוצת קשירות סופית וסגורה כל האברים נישנים. אם קבוצת קשירות אינה סגורה אזי כל אבריה חולפים כי יש מעבר חד סיטרי ממנה והחוצה.

טענה 4.31 (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 4.32 שרשרת מרקובית נקראת שרשרת פריקה אם יש לה יותר מאשר קבוצה סגורה אחת. אחרת היא נקראת לא פריקה indecomposeable. קבוצה נקראת irreducible אם אין לה תת קבוצה (פרט לעצמה) סגורה.

נשים לב כי שרשרת סופית שיש לה מצב חולף אחד אינה irreducible אך היא לא פריקה. בספרות אין אחידות מלאה בהגדרות ויש המשתמשים במושג irreducible כאשר הכוונה היא ל-indecomposeable.

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 4.33 (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, בציור 4.5 רק מצבים $6, 7$ הם נישנים, ושאר המצבים הם חולפים.

4.6 סטציונריות ושרשרות מרקוביות

ראינו בהמחשות כי הפילוג של שרשרות מרקוב נוטה בדרך כלל להתייצב לאחר מספר צעדים מספיק גדול. כיצד לחבר עובדה זו עם המושג של סטציונריות (הגדרה 3.17)? נראה מייד כי אם מתחילים שרשרת מרקובית עם פילוג מתאים, אזי היא תהיה תהליך סטציונרי.

הגדרה 4.34 פילוג $\underline{\nu}$ ניקרא פילוג סטציונרי או פילוג אינווריאנטי Stationary, invariant עבור שרשרת מרקובית אם $\underline{\nu}(0) = \underline{\nu}$ גורר $\underline{\nu}(n) = \underline{\nu}$ 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 4.35 עבור שרשרת מרקוב הומוגנית,

1. הפילוג $\underline{\nu}$ הוא פילוג **סטציונרי** אם ורק אם הוא מקיים $\underline{\nu} \cdot \mathrm{P} = \underline{\nu}$, כלומר הוא ווקטור עצמי שמאלי של מטריצת המעברים, עם ערך עצמי 1.

2. אם $\underline{\nu}(T)$ 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 4.15. נחזור על ההוכחה ביתר פירוט. $\underline{\nu} = \underline{\nu}(0)$ ו- $\underline{\nu} \cdot \mathrm{P} = \underline{\nu}$ אזי

$$\begin{aligned}
\underline{\nu}(n) &= \underline{\nu}(0)\,\mathrm{P}^n \\
&= (\underline{\nu}(0)\,\mathrm{P})\,\mathrm{P}^{(n-1)} \\
&= \underline{\nu}(0)\,\mathrm{P}^{(n-1)} \\
&= (\underline{\nu}(0)\,\mathrm{P})\,\mathrm{P}^{(n-2)} \\
&= \cdots = \underline{\nu}(0)\,\mathrm{P} = \underline{\nu}(0)
\end{aligned}$$

ולכן לפי ההגדרה הפילוג הוא סטציונרי. מצד שני, אם הפילוג הוא סטציונרי אזי לפי ההגדרה

$$\begin{aligned}
\underline{\nu}\,\mathrm{P} &= \underline{\nu}(0)\,\mathrm{P} \\
&= \underline{\nu}(1) = \underline{\nu}
\end{aligned}$$

ובכך הוכחנו את טענה 1. כעת נזכר בהגדרת הסטציונריות 3.17 ונבדוק לפי ההגדרה. ראינו בהוכחת משפט 4.14 כי

$$\mathbb{P}\left\{X(n_j) = i_j, j = 1, \ldots, k\right\} = \prod_{j=2}^{k} \mathrm{p}_{i_{j-1}\,i_j}^{(n_j - n_{j-1})}\,\nu_{i_1}(n_1)\;. \tag{4.24}$$

ביטוי זה אינו משתנה אם נזיז את כל הזמנים ב-τ: זאת כיוון ש-$(n_j+\tau)-(n_{j-1}+\tau) = n_j - n_{j-1}$, והראנו לעיל כי $\nu_{i_1}(n_1)$ כלל אינו תלוי בערך של n_1 כלומר

$$\mathbb{P}\left\{X(t_1+\tau) = j_1\right\} = \nu_{j_1}(t_1+\tau) = \nu_{j_1}(t_1) = \mathbb{P}\left\{X(t_1) = j_1\right\}$$

ולכן ההסתברות לעיל אינה תלויה ב-τ ובפרט ערכה שווה ב-τ וב-T. 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 מנקודת זמן רחוקה בעבר ומכל פילוג, הפילוג החד ממדי מתכנס 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 4.36 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 $-1 > n$?
2. שרשרת עם מצבים $\{1,2\}$ עבורה הסתברויות המעבר הן $\mathrm{p}_{12}=\mathrm{p}_{22}=1$ והפילוג ב-$n=-1$ הוא $\mathbb{P}\{X(-1)=1\}=1$. האם שרשרת זו יכולה להיות מוגדרת עבור $-1 > n$? בהנתן $T>0$ מצא שרשרת בעלת מספר מצבים סופי המוגדרת עבור $0 \le n < \infty$ 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

$$\mathbb{P}\left\{x(n) \in T_r \mid x(n-N)=i\right\} < 1. \tag{4.25}$$

הסק מכך כי עבור שרשרת המוגדרת על אוסף הזמנים האינסופי כאמור לעיל מתקיים, לכל n

$$\mathbb{P}\left\{x(n) \in T_r\right\} = 0. \tag{4.26}$$

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 (האין-סופית!)

$$\underline{\nu} \cdot \mathrm{P} = \underline{\nu}$$

הוא הווקטור $\underline{\nu} = (\ldots, 1, 1, 1, \ldots)$ אשר אינו פילוג.

טענה 4.37 (ללא הוכחה). לכל שרשרת מרקוב סופית יש פילוג סטציונרי. השרשרת אינה פריקה אם ורק אם הפילוג הוא יחיד. הפילוג הסטציונרי מקבל ערכים שונים מ-0 רק על מצבים נשנים.

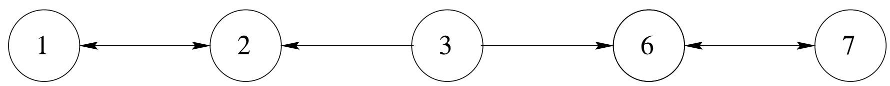
נשים לב כי שרשרת סופית לא פריקה, גם אם היא אינה irreducible עדיין יש לה פילוג סטציונרי יחיד. חוסר יחידות נגרם על ידי קיום שתי קבוצות סגורות (או יותר).

הערה: קיום פילוג סטציונרי נובע מתוצאות באלגברה לינארית: למשוואה

$$\mathrm{P}\,\underline{x}=\underline{x}$$

יש פיתרון (הווקטור $(\underline{x}=(\ldots,1,1,1,\ldots)^T$, כלומר למטריצה P יש וקטור עצמי (ימני) עם ערך עצמי 1. מכך נובע שלמטריצה זו יש גם ווקטור עצמי שמאלי עם ערך עצמי 1. משפט פרון-פרובניוס מתורת המטריצות מבטיח כי, כיוון שאברי המטריצה P הם חיוביים, אזי לווקטור העצמי השמאלי (השייך לערך העצמי הגדול ביותר של P) רכיבים חיוביים. כיוון שמדובר בווקטור סופי, ניתן לנרמל אותו כך שיהיה פילוג, ומטענה 4.35 זהו פילוג סטציונרי.

דוגמה 4.38 בדוגמה שלפנינו יש שתי קבוצות סגורות, ולכן הדוגמה אינה מקיימת את תכונת האי-פריקות. נניח שכל המעברים קורים בהסתברות $1/2$.



איור 4.6: פילוג סטציונרי

אזי קל לבדוק כי לכל $0\le\alpha\le 1$, הפילוג

$$\underline{\nu}=(\alpha/2,\alpha/2,0,(1-\alpha)/2,(1-\alpha)/2) \tag{4.27}$$

הוא פילוג סטציונרי. כלומר, הפילוג הסטציונרי אינו יחיד. באופן אינטואיטיבי, ניתן לחלק את ההסתברות הסטציונרית בין הקבוצות הסגורות כאוות נפשינו. החלוקה 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 4.39 מצב i נקרא מחזורי אם המחלק המשותף הגדול ביותר (שנסמן ב-d) של אוסף המספרים

$$\left\{n:\mathrm{p}_{ii}^{(n)}>0\right\} \tag{4.28}$$

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 4.40 עבור שרשרת עם מצבים $\{1,2,3,4,5\}$, אם הסתברויות המעבר הן

$$\mathrm{p}_{i(i+1)} = 1,\ i = 1,2,3,4, \quad \mathrm{p}_{51} = 1$$

אזי לשרשרת מחזור $d = 5$. לעומת זאת אם הסתברויות המעבר הן

$$\mathrm{p}_{12} = \mathrm{p}_{23} = \mathrm{p}_{34} = \mathrm{p}_{41} = 1, \quad \mathrm{p}_{55} = 1$$

אזי מצבים $1,2,3,4$ מחזוריים עם מחזור $d = 4$ ומצב 5 אינו מחזורי. לבסוף, אם

$$\mathrm{p}_{12} = 1,\ \mathrm{p}_{21} = 1/2,\ \mathrm{p}_{23} = 1/2,\ \mathrm{p}_{34} = \mathrm{p}_{42} = \mathrm{p}_{55} = 1$$

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 $n \to \infty$, לפילוג האינווריאנטי. קצב ההתכנסות הוא גאומטרי, כלומר לאחר t צעדים המרחק בין הפילוג לבין הפילוג האינווריאנטי 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

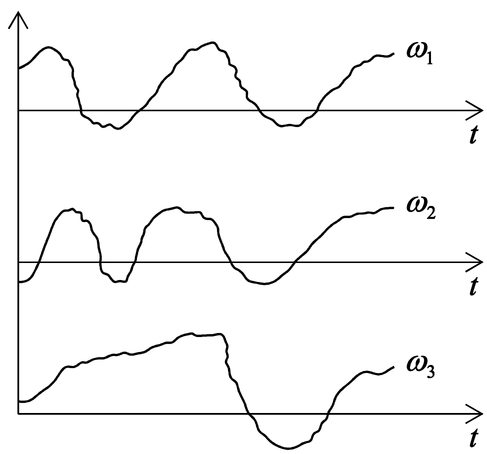
5.1 מבוא, הגדרות ודוגמאות

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 סדרת זמנים כלשהיא. בפרק זה נעסוק באינדקס זמן רציף. כלומר התהליך יוגדר עבור אינטרוול זמן $T_1 \le t \le T_2$, 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 $T_2 = \infty$). כמו בזמן בדיד, תהליך אקראי הוא אוסף של משתנים.

הגדרה 5.1 תהליך אקראי כל קטע $[a, b]$ הוא אוסף של מ"א $X(t, \omega)$, משתנה לכל זמן $a \le t \le b$.

גם כאן נשתמש לפי הנוחיות בסימונים (השקולים) $X(t,\omega),\ X_t(\omega)$ ולעיתים נשמיט את פרמטר המזל ונרשום את התהליך כ-$X(t), X_t$.

כפונקציה של פרמטר המזל, $X(t,\omega)$ הוא מ"א, לדוגמא, $X(2,\omega)$ הוא מ"א. עבור ω קבוע, בקטע $a \le t \le b$ $\quad X(t,\omega),$ 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:

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-$X(t,\omega) = A(\omega)\sin(2\pi f t + \phi(\omega))$ 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

$$p(t) \doteq \begin{cases} 1 & 0 \le t < 1 \\ 0 & \text{אחרת.} \end{cases} \tag{5.1}$$

אזי שידור של סידרה (אינסופית) של ביטים (משתנים אקראיים A_n המקבלים ערכים ± 1) המאפננים את הפולס המרובע נותנים את התהליך האקראי

$$X(t)=\sum_{n=-\infty}^{\infty}A_n p(t-nT)\ . \tag{5.2}$$

כאשר $T \ge 1$.

חוק ההסתברות של תהליך אקראי:

יהיה $\{X_t, T_1 \le t \le T_2\}$ תהליך אקראי. נעיין באוסף חוקי ההסתברות כדלקמן: עבור כל n וכל $t_1,t_2,\ldots,t_n$ $(T_1 \le t_i \le T_2)$

$$F_{X_{t_1},\ldots,X_{t_n}}(a_1,\ldots,a_n)=\mathbb{P}\{X_{t_1}\le a_1,\ldots,X_{t_n}\le a_n\}$$

אוסף זה של פונקציות פילוג נקרא חוק ההסתברות של התהליך האקראי X_t והוא חייב לקיים את חוק הקונסיסט-נטיות:

$$F_{X_{t_1},\ldots,X_{t_{i-1}},X_{t_{i+1}},\ldots,X_{t_n}}(a_1,\ldots,a_{i-1},a_{i+1},\ldots,a_n)=F_{X_{t_1},\ldots,X_{t_n}}(a_1,\ldots,a_{i-1},\infty,a_{i+1},\ldots,a_n)$$

תהליך פואסון Poisson process

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 מספר המאורעות המתרחשים בפרק הזמן $[0,t]$.

הגדרה 5.2 תהליך מנייה הוא תהליך המקבל ערכים שלמים, ואשר פונקציות המדגם שלו אינן יורדות. תהליך מניה פשוט הוא תהליך אשר כל קפיצה (עליה) היא בגודל 1 בדיוק.

התהליך N_t הוא תהליך מניה פשוט, דהיינו, הוא מונוטוני לא יורד ויכול לקבל ערכים שלמים $0,1,2,\ldots$ בלבד. נשתמש בסימון המתמטי הבא: $o(\Delta t)$ הוא גודל זניח לגבי Δt כאשר $\Delta t \to 0$ (באופן כללי $f(\varepsilon)$ הוא $o(\varepsilon)$ אם $f(\varepsilon)/\varepsilon \to 0$ כאשר $\varepsilon \downarrow 0$). לגבי חוק ההסתברות של התהליך נניח את ההנחות הבאות:

הגדרה 5.3 תהליך פואסון הוא תהליך מניה 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

$$\lim_{\Delta t\downarrow 0}\frac{\mathbb{P}\{N_{t+\Delta t}-N_t=1\}}{\Delta t}=\lambda\ . \tag{5.3}$$

(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.

מ (א) ו-(ב) נובע ש: $\mathbb{P}\{(N_{T+\Delta t}-N_t)\geq 2\}=\mathbb{P}\ \{\text{שני מאורעות או יותר}\}=o(\Delta t)$

<u>תזכורת:</u> 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

$$\binom{n}{k}p^kq^{n-k}=\frac{n!}{(n-k)!k!}\,p^kq^{n-k}\ .$$

נחזור לתהליך פואסון. התהליך יתקבל בגבול של חוק ההסתברות הבינומי לעיל, כאשר נחלק את ציר הזמן לקטעים קטנים באורך Δt 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 ("כשלון"), ואילו ב k קטעי זמן צריך להיות ארוע ("הצלחה"); לכן, עבור k קבוע, כאשר Δ קטן $\Delta\to 0$, כיוון ש-$n=T/\Delta$ נקבל $n\to\infty$ ולכן

$$\mathbb{P}\{N_T=k\}\cong\frac{n!}{(n-k)!k!}(\lambda\Delta)^k(1-\lambda\Delta)^{n-k}=\frac{n!}{(n-k)!k!}\left(\frac{\lambda T}{n}\right)^k\left(1-\lambda\frac{T}{n}\right)^{n-k} \tag{5.4}$$

$$=\frac{n!}{(n-k)!n^k}\left(1-\lambda\frac{T}{n}\right)^{-k}\frac{(\lambda T)^k}{k!}\left(1-\lambda\frac{T}{n}\right)^{n} \tag{5.5}$$

נשים לב כעת כי הביטוי הראשון בשורה האחרונה

$$\frac{n!}{(n-k)!n^k}=\frac{n(n-1)\cdots(n-k+1)(n-k)!}{nn\cdots n(n-k)!}=\frac{n(n-1)\cdots(n-k+1)}{nn\cdots n}\to 1 \tag{5.6}$$

כאשר $n \to \infty$. כלומר

$$\frac{n!}{(n-k)!k!} \to 1 \ .$$

בנוסף

$$\left(1-\lambda\frac{T}{n}\right)^{-k} \to 1 \ ; \quad \left(1-\lambda\frac{T}{n}\right)^{n} \to e^{-\lambda T}$$

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

$$\mathbb{P}\{N_{T_2} - N_{T_1} = k\} = \frac{[\lambda(T_2-T_1)]^k}{k!} e^{-\lambda(T_2-T_1)}$$

קל לראות שאכן מתקיימות הנחות (א)-(ב) לעיל. כאשר Δ קטן נקבל $\mathbb{P}\{N_{t+\Delta} - N_t = 1\} = \lambda\Delta e^{-\lambda\Delta}$ והביטוי השני שואף ל-1. לצורך טענה (ב) נשים לב כי $e^{-\lambda\Delta} \approx 1 - \lambda\Delta$ (קירוב טיילור).

אפשר להשתמש בתכונות מ"א פואסוני, או להוכיח ע"י חשבון ישיר ש-

$$\mathbb{E}[N_T] = \lambda T$$

$$\mathbb{E}[N_T^2] = (\lambda T)^2 + \lambda T$$

$$\mathrm{Var}(N_T) = (\lambda T)^2 + \lambda T - (\lambda T)^2 = \lambda T$$

השוויון הראשון מסביר מדוע λ נקרא פרמטר הקצב של התהליך. במקום להוכיח תוצאות אלה ישירות ע"י שימוש בתוצאות ידועות עבור משתנים אקראיים פואסונים, נראה כיצד ניתן להשתמש ישירות בהגדרה. נחשב כדלקמן: את הקטע $(0,T]$ נחלק לקטעים קטנים ונסמן את נקודות החלוקה ב-$t_0, t_1, \ldots, t_n$ ($t_0 = 0, t_n = T$):

$$\mathbb{E}[N_T] = \mathbb{E}\left[\sum_{i=1}^{n-1}(N_{t_{i+1}} - N_{t_i})\right] = \left[\sum_{i=1}^{n-1}\mathbb{E}(N_{t_{i+1}} - N_{t_i})\right] \cong \sum_{i=1}^{n-1}\lambda(t_{i+1}-t_i) = T\cdot\lambda$$

$$\begin{aligned}\mathbb{E}[N_T^2] &= \mathbb{E}\left[\left(\sum_{i=1}^{n-1}(N_{t_{i+1}} - N_{t_i})\right)^2\right] = \mathbb{E}\Big[\sum_{i=1}^{n-1}(N_{t_{i+1}} - N_{t_i})^2\Big] + \mathbb{E}\left[\sum\sum_{i\neq j}(N_{t_{i+1}} - N_{t_i})(N_{t_{j+1}} - N_{t_j})\right] \\ &= \lambda T + \sum\sum_{i\neq j}\Big(\lambda(t_{i+1}-t_i)\cdot\lambda(t_{j+1}-t_j)\Big) = \lambda T + \lambda^2 T^2\end{aligned}$$

אגב, עבור $t_2 > t_1$

$$\mathbb{E}[N_{t_2}N_{t_1}] = \mathbb{E}[N_{t_1}^2] + \mathbb{E}\Big[N_{t_1}(N_{t_2} - N_{t_1})\Big] = (\lambda t_1)^2 + \lambda t_1 + \lambda^2 t_1(t_2-t_1) = \lambda t_1 + \lambda^2 t_1 t_2$$

השווה תוצאה זו לתוצאה (3.10) שקיבלנו עבור הילוך שיכור---גם הוא תהליך הפרשים בת"ס.

בגלל הנחה (ג), תהליך פואסון הוא תהליך של "תוספות בלתי תלויות", דהיינו: אם $t_1 < t_2 \le t_3 < t_4$, 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 $0 < t_1 < \cdots < t_n$ נגדיר $t_0 = 0, k_0 = 0$ ונקבל

$$\begin{aligned}\mathbb{P}\{N_{t_1} = k_1, \ldots, N_{t_n} = k_n\} &= \mathbb{P}\{N_{t_1} = k_1\}\,\mathbb{P}\{N_{t_2} - N_{t_1} = k_2 - k_1\} \cdots \mathbb{P}\{N_{t_n} - N_{t_{n-1}} = k_n - k_{n-1}\}\\ &= \prod_{i=1}^{n} e^{-\lambda(t_i - t_{i-1})} \frac{[\lambda(t_i - t_{i-1})]^{k_i - k_{i-1}}}{(k_i - k_{i-1})!}\\ &= e^{-\lambda t_n} \prod_{i=1}^{n} \frac{[\lambda(t_i - t_{i-1})]^{k_i - k_{i-1}}}{(k_i - k_{i-1})!}\ .\end{aligned}$$

בצורה כזו אנו יכולים לקבל את חוק ההסתברות של $N_{t_1}, \ldots, N_{t_n}$ לכל n ולכל $t_1, \ldots, t_n$ ולכן חוק ההסתברות של תהליך פואסון ידוע. בגלל ההפרשים 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 גאוסי).

תרגיל 5.4 הוכח 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: בדוק את ההנחות.

תרגיל 5.5 הוכח את הכיוון ההפוך: יהי Q_t תהליך פואסון עם קצב $\lambda + \nu$. יהי $X(n)$ אוסף משתנים בת"ס ושווי פילוג אשר כולם בת"ס ב-Q_t עם פילוג $\mathbb{P}\{X(1) = 1\} = \lambda/(\lambda + \nu)$ ו-$\mathbb{P}\{X(1) = 0\} = \nu/(\lambda + \nu)$. נגדיר תהליכים N_t ו-M_t בצורה הבאה. $N_0 = M_0 = 0$. נניח שבזמן t_n התהליך Q עלה מ-$n - 1$ ל-n. אם $X(n) = 1$ אזי נגדיר כי N יגדל ב-1 ברגע t_n. אם $X(n) = 0$ אזי M יגדל.
הראה כי N, M הם תהליכי פואסון (רמז: בדוק את ההגדרות).

תהליכים הקשורים לתהליך פואסון

עבור תהליך פואסון N_t נגדיר מעין קרוב לנגזרת

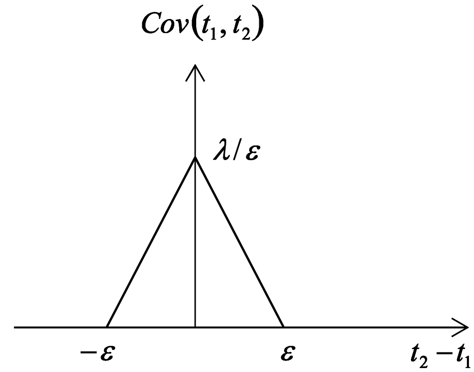
$$Y_\varepsilon(t) \triangleq \frac{N_{t+\varepsilon} - N_t}{\varepsilon}$$

שים לב ש- $Y_\varepsilon(t) \ge 0$, $\mathbb{E}[Y_\varepsilon(t)] = \frac{\lambda(t+\varepsilon) - \lambda t}{\varepsilon} = \lambda$; עבור $\varepsilon \to 0$ נקבל ש $Y_0(t)$ הוא סופרפוזיציה של פונקציות דירק. נחזור ל-$\varepsilon > 0$ ונגדיר

$$Z_\varepsilon(t) \triangleq Y_\varepsilon(t) - \mathbb{E}[Y_\varepsilon(t)] = \frac{N_{t+\varepsilon} - N_t - \varepsilon\lambda}{\varepsilon}$$

$Y_\varepsilon(t)$ ו-$Z_\varepsilon(t)$ אינם תהליכים של הפרשים בלתי תלויים. נעיין ב- $\mathrm{Cov}(Y_\varepsilon(t_1), Y_\varepsilon(t_2)) = \mathbb{E}[Z_\varepsilon(t_1)Z_\varepsilon(t_2)]$ ונבחין בין שני מקרים:

(א) $|t_2 - t_1| \geq \varepsilon$ (ב) $|t_2 - t_1| < \varepsilon$ במקרה (א) ברור ש- $\mathbb{E}[Z_\varepsilon(t_1)Z_\varepsilon(t_2)] = 0$. במקרה (ב) וללא הוכחה (החשבון פשוט) מתקבלת התוצאה המופיעה בציור 5.2



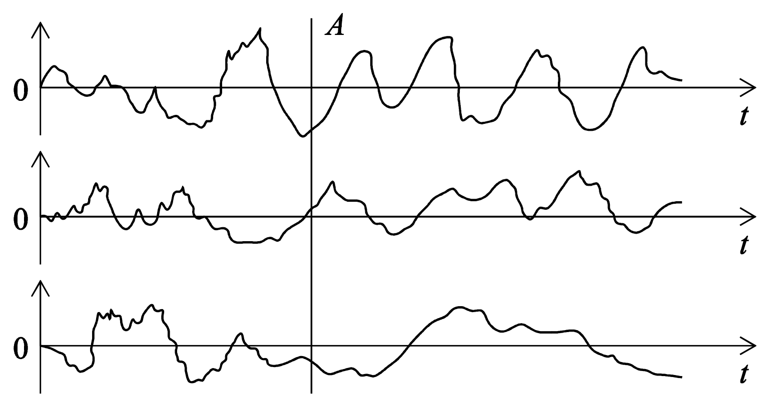
איור 5.2:

<u>שאלות</u>: האם $\mathbb{P}\{Z_\varepsilon(t) \leq \alpha\}$ $(t > 0)$ תלוי ב t?

האם $\mathbb{P}\{Z_\varepsilon(t) \leq \alpha, Z_\varepsilon(t+5) \leq \beta\}$ $(t > 0)$ תלוי ב t?

5.2 סטציונריות:

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 5.3:

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- $\mathbb{P}\{X(t,\omega) \le a\}$ 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 הציור); 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 5.6 התהליך $\{X(t,\omega), \infty < t < \infty\}$ נקרא ת"א סטציונרי אם עבור כל n , כל $t_1,\ldots,t_n$ וכל τ , חוקי ההסתברות של הוקטור האקראי:

$$\Big(X_{t_1}, X_{t_2}, \ldots, X_{t_n}\Big)^T$$

ושל הוקטור האקראי:

$$\Big(X_{t_1+\tau}, X_{t_2+\tau}, \ldots, X_{t_n+\tau}\Big)^T$$

זהים זה לזה, דהיינו:

$$F_{X_{t_1},X_{t_2},\ldots,X_{t_n}}(a_1,a_2,\ldots,a_n) = F_{X_{t_1+\tau},X_{t_2+\tau},\ldots,X_{t_n+\tau}}(a_1,a_2,\ldots,a_n)$$

בצורה מקוצרת נאמר שת"א הוא סטציונרי אם חוק ההסתברות שלו אינוריינטי 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 5.7 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.

<u>דוגמאות:</u>

1. **נעיין בשני התהליכים הבאים שהם ת"א מנוונים:**

$$\{X(t,\omega) = 5\,, \qquad -\infty < t < \infty\}$$

$$\{X(t,\omega) = 5\sin 2\pi t\,, \qquad -\infty < t < \infty\}$$

התהליך הראשון סטציונרי והשני לא. מדוע? אם $A(\omega)$ מ"א כלשהוא אזי התהליך $X(t,\omega) = A(\omega)$ הוא סטציונרי, מדוע?

2. **תהליך פואסון אינו סטציונרי. מדוע?**

3. **(ללא הוכחה): לתהליך פואסון נדביק עוד תהליך פואסון בעל ערך שלילי ובזמנים שליליים כאשר החלק בזמן חיובי והחלק בזמן שלילי בלתי תלויים. את התהליך ב-$(-\infty,\infty)$ נסמן ב-N_t. נגדיר $Y(t) \triangleq N_{t+\varepsilon} - N_t$ או $Y_\varepsilon(t) \triangleq (N_{t+\varepsilon} - N_t)/\varepsilon$ אזי $Y(t)$ ו-$Y_\varepsilon(t)$ (ε קבוע) הם ת"א סטציונריים.**

4. נעיין בתהליך האקראי

$$X(t,\omega) = A\cos\Big(2\pi f_0 t + 2\pi\phi(\omega)\Big), \quad \{-\infty < t < \infty\}$$

כאשר f_0, A קבועים (אינם אקראיים) ו-$\phi(\omega)$ מ"א אקראי המפולג באופן אחיד בתחום $[0,1)$

<u>טענה:</u> התהליך $X(t,\omega)$ הוא סטציונרי.

השווה לטענת הסטציונריות עבור פונקציה מחזורית עם פאזה יוניפורמית, בפרק 3.

<u>הוכחה:</u> ראשית, הוכחה פשוטה מבחינה חישובית. נסמן ב-$\lfloor\alpha\rfloor$ את החלק השלם של המספר α. נקבע τ ונשים לב כי המ"א $\psi \doteq f_0\tau + \phi - \lfloor f_0\tau + \phi\rfloor$ מפולג אחיד ב-$[0,1]$, כלומר מפולג כמו ϕ. בנוסף, בגלל המחזוריות של פונקצית ה-$\cos$

$$A\cos\Big(2\pi f_0(t+\tau) + 2\pi\phi\Big) = A\cos\Big(2\pi[f_0 t + f_0\tau + \phi - \lfloor f_0\tau + \phi\rfloor]\Big) = A\cos\Big(2\pi f_0 t + 2\pi\psi\Big).$$

קיבלנו כי $A\cos\Big(2\pi f_0(t+\tau) + 2\pi\phi\Big)$ מפולג כמו $A\cos\Big(2\pi f_0 t + 2\pi\phi\Big)$-כלומר התהליך סטציונרי.

הוכחה נוספת, מפורטת יותר: נקבע $n \quad a_1,\dots,a_n,t_1,\dots,t_n,$ ונעיין במאורע B המוגדר כדלקמן:

$$B = \Big\{\omega: \ X(t_1,\omega) \le a_1,\dots,X(t_n,\omega) \le a_n\Big\}$$

נגדיר גם את Φ_B כדלקמן:

$$\Phi_B = \Big\{\theta: \ \theta \in [0,1)\Big\} \bigcap \Big\{A\cos(2\pi f_0 t_1 + 2\pi\theta) \le a_1,\dots,A\cos(2\pi f_0 t_n + 2\pi\theta) \le a_n\Big\}$$

דהיינו: Φ_B הוא אוסף כל המספרים θ כך שאם $\phi(\omega) \in \Phi_B$ (כלומר, אם הגרלנו $\phi(\omega)$ כזה השייך ל-Φ) 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 ארך הקטע, ואם Φ_B הוא אחוד של קטעים לא חופפים אזי

$$\mathbb{P}\{B\} = \{\text{סכום ארכי הקטעים המהווים את } \Phi_B\}$$

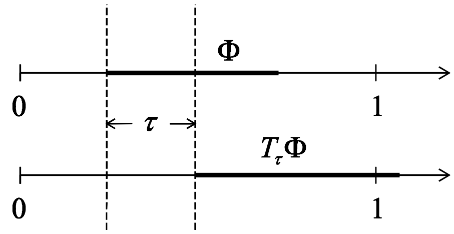
תהייה Φ קבוצת נקודות כלשהיא ב-$[0,1)$ נגדיר את ההזזה של Φ בכמות τ ע"י (ראה איור 5.4):

$$T_\tau\Phi = \{\phi : \phi - \tau \in \Phi\}$$

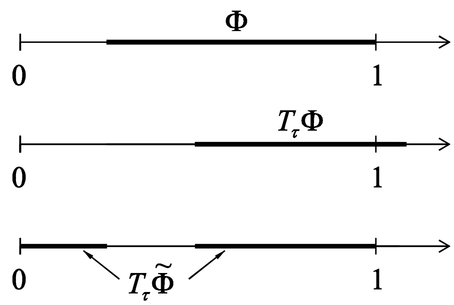
שים לב שאם גם Φ וגם $T_\tau\Phi$ ב-$[0,1)$ אזי $\mathbb{P}\{\phi(\omega) \in \Phi\} = \mathbb{P}\{\phi(\omega) \in T_\tau\Phi\}$ וזאת מאחר ו-$\phi(\omega)$ מפולג אחיד בתחום $[0,1)$. עבור מספר ממשי X נסמן ב-$[X]$ את החלק השלם של X (כגון אם $[X] = 2, X = 2.3$) וב-$\tilde{X}$ נסמן $X - [X]$ (0.3 בדוגמא). 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 5.5)

כעת אם Φ קבוצת נקודות ב-$[0,1)$ אזי

$$\mathbb{P}\Big\{\phi(\omega) \in \Phi\Big\} = \mathbb{P}\Big\{\phi(\omega) \in (T_\tau\tilde{\Phi}))\Big\}$$



איור 5.4:



איור 5.5:

עבור כל τ ושוב, הדבר נכון כי ϕ מפולג אחיד. נחזור עכשיו ל-

$$\Phi_B = \Big\{\phi: \ A\cos(2\pi f_0 t_i + 2\pi\phi) \le a_i, \quad i = 1,\dots,n\Big\}$$

נעיין כעת ב-

$$\Phi_C = \Big\{\phi: \ A\cos(2\pi f_0 (t_i + \tau) + 2\pi\phi) \le a_i, \quad i = 1,\dots,n\Big\}$$

אזי $\Phi_C = T_{-2\pi f_0 \tau}\Phi_B$ ולכן

$$\begin{aligned}\mathbb{P}\{X_{t_i+\tau} \le a_i, \quad i = 1,\dots,n\} &= \mathbb{P}\{\phi \in \tilde{\Phi}_C\} = \mathbb{P}\{\phi \in T_{-2\pi f_0\tau}\tilde{\Phi}_B\} \\ &= \mathbb{P}\{\phi \in \Phi_B\} = \mathbb{P}\{X_{t_i} \le a_i, \quad i = 1,\dots,n\}\end{aligned}$$

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. כך למשל, נתבונן בתהליך $X(t)$ הנתון על ידי (5.2). יהיה U מ"א בת"ס ב-$\{A_n\}$ ומפולג אחיד על $[0,1]$. אזי $X(t+U)$ הוא ת"א סטציונרי.

הגדרה 5.8 זוג תהליכים $\{X_t, Y_t\}$ נקראים סטציונרים במשותף (או, בצורה שקולה, הווקטור $\{(X_t, Y_t)\}$ נקרא סטציונרי) אם

עבור כל n , כל $t_1,s_1,\ldots,t_n,s_n$ וכל τ ,

$$F_{X_{t_1},Y_{s_1},\ldots,X_{t_n},Y_{s_n}}(a_1,b_1,\ldots,a_n,b_n)=F_{X_{t_1+\tau},Y_{s_1+\tau},\ldots,X_{t_n+\tau},Y_{s_n+\tau}}(a_1,b_1,\ldots,a_n,b_n)$$

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.

<u>מומנטים</u>

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).

<u>הגדרות:</u>

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:

$$\mathbf{K}_X(t_1,t_2)\triangleq\mathbb{E}\Big[(X_{t_1}-\mu_X(t_1))(X_{t_2}-\mu_X(t_2))\Big]=\mathbf{R}(t_1,t_2)-\mu_X(t_1)\mu_X(t_2)$$

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 $t_1,\ldots,t_n$ המטריצה:**

$$\begin{pmatrix}\mathbf{R}_X(t_1,t_1) & \mathbf{R}_X(t_1,t_2) & \ldots & \mathbf{R}_X(t_1,t_n)\\ \mathbf{R}_X(t_2,t_1) & & & \\ \vdots & & & \vdots\\ \mathbf{R}_X(t_n,t_1) & \ldots & \ldots & \mathbf{R}_X(t_n,t_n)\end{pmatrix}$$

מטריצה לא שלילית (מדוע? הוכח).

שלושת התכונות נכונות כמובן (כמקרה פרטי) גם עבור מטריצת הקווריאנס $\mathbf{K}$.

<u>סטציונריות במובן הרחב</u>

נחזור לסטציונריות; עבור ת"א סטציונרי ברור שמתקיים $F_{X_t}(a) = F_{X_{t'}}(a) \quad \forall a$ 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:

$$\mathbf{K}_X(t_1,t_2) = K_X(|t_1 - t_2|);\ \mathbf{K}_X(t_1, t_1+\tau) = K_X(|\tau|)$$

דהיינו: עבור ת"א סטציונרי פונקציות האוטוקורילציה והקווריאנס הן פונקציות של משתנה אחד (והוא הפרש הזמנים) בלבד.

<u>הגדרה:</u> 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_2, t_1 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 בדרך כלל; אפשר לראות שישנם ת"א סטציונרים במובן הרחב שאינם סטציונריים.

נזכר בהגדרה (5.1) של גל מרובע ובתהליך האפנון (5.2). לשם פשטות נבחר $T = 1$. ראינו כי $X(t+U)$ הוא ת"א סטציונרי אם U מפולג אחיד על $[0,1]$ ובת"ס ב-$\{A_n\}$. נתבונן כעת באות אחר אשר בבירור אינו סטציונרי. תהי $\{B_n\}$

סדרת מ"א בת"ס בעלי פילוג נורמלי עם ממוצע אפס ווריאנס 1 ובת"ס ב-U. נגדיר

$$Y(t) = \sum_{n=-\infty}^{-1} A_n p(t-n-U) + \sum_{n=0}^{\infty} B_n p(t-n-U) \tag{5.7}$$

$$\doteq \sum_{n=-\infty}^{\infty} C_n p(t-n-U) \tag{5.8}$$

כאשר השוויון האחרון מגדיר את C_n. הפילוג של האות יהיה לכן שונה עבור זמנים שליליים וחיוביים, ולכן האות אינו סטציונרי. נראה כי אות זה הוא סטציונרי במובן הרחב. על ידי התנייה ב-U קל לראות כי $\mathbb{E}\, Y(t) = 0$. כיוון ש-$\{C_n\}$ בת"ס ב-U נקבל

$$\mathbb{E}\, Y(t)Y(s) = \mathbb{E} \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} C_n C_m p(t-n-U)p(s-m-U) \tag{5.9}$$

$$= \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \mathbb{E}\, C_n C_m \int_0^1 p(t-n-u)p(s-m-u)\, du \tag{5.10}$$

$$= \sum_{n=-\infty}^{\infty} \int_0^1 p(t-n-u)p(s-n-u)\, du \tag{5.11}$$

כיוון של-C_n יש ווריאנס 1. כעת נשים לב כי

$$p(t-n-u) = 1 \ \text{ אם ורק אם } \ t-1 < n+u < t \tag{5.12}$$

$$p(s-n-u) = 1 \ \text{ אם ורק אם } \ s-1 < n+u < s \tag{5.13}$$

לכן מכפלתם תהייה שונה מאפס אם ורק אם

$$t-1 < n+u < s\,, \quad s-1 < n+u < t \tag{5.14}$$

כלומר $|t-s| < 1$. הסכום האינסופי כולל לכל היותר שני אברים השונים מאפס. על ידי שרטוט שתי הפונקציות קל לראות כי כי אם תנאי החפיפה מתקיים ואם למשל $t < s$ אזי החפיפה של שני הריבועים היא $(t+1)-s = 1-|s-t|$. חישוב דומה עבור $t > s$ יתן את התוצאה

$$\mathbb{E}\, Y(t)Y(s) = \begin{cases} 1-|t-s| & |t-s| < 1 \\ 0 & |t-s| \geq 1 \end{cases} \tag{5.15}$$

כלומר התהליך סטציונרי במובן הרחב.

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 $[0, 2\pi)$:

$$\begin{aligned}\mathbb{E}[X_t] &= \mathbb{E}\Big[A(\omega)\cos(2\pi f_0 t + \phi(\omega))\Big] = \mathbb{E}\Big[A(\omega)\Big]\,\mathbb{E}\Big[\cos(2\pi f_0 t + \phi(\omega))\Big] \\ &= \mathbb{E}[A(\omega)] \cdot \frac{1}{2\pi}\int_0^{2\pi} \cos(2\pi f_0 t + \theta)d\theta = 0\end{aligned}$$

$$\begin{aligned}\mathbb{E}[X_t X_{t+\tau}] &= \mathbb{E}\Big[A^2(\omega)\cos(2\pi f_0(t+\tau) + \phi) \cdot \cos(2\pi f_0 t + \phi)\Big] \\ &= \frac{1}{2}\,\mathbb{E}\Big[A^2(\omega)\cos 2\pi f_0 \tau\Big] + \frac{1}{2}\,\mathbb{E}\Big[A^2\cos(2\pi f_0(2t+\tau) + 2\phi)\Big] \\ &= \frac{1}{2}\,\mathbb{E}\Big[A^2(\omega)\Big] \cdot \cos 2\pi f_0 \tau + 0\end{aligned}$$

וכיוון ש-$\cos x = \cos(-x)$ התהליך סטציונרי במובן הרחב. כדאי לזכור שעבור תהליך פשוט זה

$$R_X(\tau) = \frac{\mathbb{E}[A^2]}{2}\cos(2\pi f_0 \tau)$$

<u>הערה:</u> חזור על דוגמא זאת כאשר כמקודם ϕ ו- A בלתי תלוים, אולם הפעם $\phi = 0, \pi/2, \pi, 3\pi/2$ בהסתברות 1/4. האם תהליך זה סטציונרי במובן הרחב? האם תהליך זה סטציונרי?

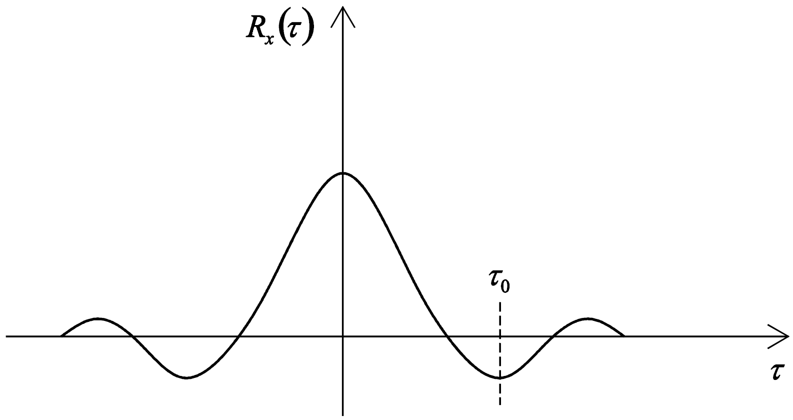
<u>תכונות פונקצית האוטוקורלציה:</u>

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. $R_X(0) \geq |R_X(\tau)|$ (הוכח בעזרת $\mathbb{E}[X_{t+\tau} \pm X_t)^2] \geq 0$).

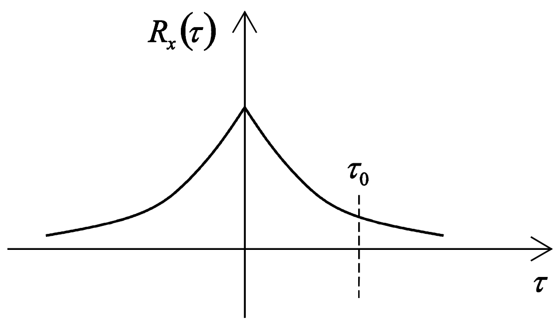
מדוע אנחנו מעונינים בפונקצית האוטוקורלציה? היא מסכמת את המומנטים מסדר שני של התהליך וכפי שנראה אח"כ היא מאפשרת לענות על שאלות מענינות. הסבר (חלקי מאוד) על משמעות פונקצית האוטוקורלציה ניתן לראות מהשיקול הבא: נניח סטציונריות, $\mu_X(t) = 0$, ונניח שעבור τ "גדול מאוד", X_t ו- $X_{t+\tau}$ "כמעט" בלתי תלויים סטטיסטית. במקרה זה

$$R_X(\tau) \underset{|\tau|\to\infty}{\longrightarrow} 0$$

כמו כן ידוע לנו ש-$R_X(0) \geq |R_X(\tau)|$ ולכן מהלך $R_X(\tau)$ יהיה משהו כמתואר באיור 5.6 או כמתואר באיור 5.7. נוכל איפוא לומר שבקרוב גס פונקצית האוטוקורלציה "כמעט" מתאפסת עבור $\tau > \tau_0$ ולכן τ_0 מתאר במקרים אלה את "הזכרון האפקטיבי של התהליך". כדי לראות שפונקצית האוטוקורלציה נותנת מעין תאור גס של "הזכרון הטבעי"



איור 5.6: פונקצית אוטוקורלציה דועכת עם תנודות



איור 5.7: פונקצית אוטוקורלציה דועכת מונוטונית

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 $\mathbf{R}_X(\tau)$ אינו דועך כאשר $|\tau| \to \infty$: למשל האות הסינוסי עם פאזה אקראית.

חשוב לזכור שמומנטים נותנים תאור חלקי בלבד של התהליך. יתכן מצב בו לשני תהליכים יש את אותה פונקצית תוחלת ואוטוקורלציה, אך הפילוגים וכן פונקציות המדגם שונים בתכלית. בפרט, ראינו שיתכן תהליך סטציונרי במובן הרחב, אשר הפילוגים שלו בזמנים שונים הם שונים וכן גם פונקציות המדגם.

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

$$\begin{aligned}\mathbf{R}_Z(t_1,t_2) &= \mathbb{E}\Big[(X_{t_1}+Y_{t_1})(X_{t_2}+Y_{t_2})\Big]\\ &= \mathbf{R}_X(t_1,t_2)+\mathbf{R}_Y(t_1,t_2)+\mathbf{R}_{X,Y}(t_1,t_2)+\mathbf{R}_{Y,X}(t_1,t_2)\end{aligned}$$

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 $\mathbf{R}_Z(t_1,t_2)=\mathbf{R}_X(t_1,t_2)+\mathbf{R}_Y(t_1,t_2)$.

<u>הערה:</u> 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

$$\mathbf{R}_{X,Y}(t_1,t_1+\tau) = R_{X,Y}(\tau) = R_{Y,X}(-\tau)$$

<u>הגדרה:</u> זוג התהליכים האקראיים $\{X_t, Y_t, -\infty < t < \infty\}$ נקראים סטציונרים 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 <u>בלבד</u> .

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

$$X(t) \doteq \sum_{n=-\infty}^{\infty} a_n p(t-nT-U) \tag{5.16}$$

$$Y(t) \doteq \sum_{n=-\infty}^{0} a_n p(t-nT-U) \tag{5.17}$$

$$+\sum_{n=1}^{\infty} b_n p(t-nT-U) \tag{5.18}$$

הוא סטציונרי (במובן הצר). אולם

$$\mathbb{E}[X_1 Y_1] = \mathbb{E}[a_n^2 p(1-U)] = 1 \tag{5.19}$$

$$\mathbb{E}[X_{-1} Y_{-1}] = \mathbb{E}[a_{-1} b_{-1} p(1-U)] = 0. \tag{5.20}$$

כלומר התהליך הווקטורי אינו סטציונרי במובן הרחב, או במילים אחרות זוג התהליכים אינו סטציונרי במשותף במובן הרחב.

5.3 תהליך אקראי גאוסי

<u>הגדרה:</u> ת"א $\{X_t, t \in [a,b]\}$ נקרא גאוסי אם עבור כל n וכל $t_1,\ldots,t_n \in [a,b]$ הוקטור האקראי $(X_{t_1},\ldots,X_{t_n})^T$ הוא ו"א גאוסי.

<u>טענה:</u> אם X_t ת"א גאוסי ב-$[a,b]$ אזי חוק ההסתברות שלו נקבע חד משמעית ע"י פונקציות התוחלת והאוטוקורלציה שלו $\mathbf{R}_X(t_1,t_2), \mu_X(t) \quad t,t_1,t_2 \in [a,b],$

<u>הוכחה:</u> עבור כל n וכל $t_1,\ldots,t_n$ המקיימים $t_i \in [a,b]$ הוקטור האקראי

$$\underline{Z} = \Big(X_{t_1}, X_{t_2}, \ldots, X_{t_n}\Big)^T$$

הוא וקטור אקראי <u>גאוסי</u> (היות והנחנו ש-$\{X_t, t \in [a,b]\}$ הוא ת"א גאוסי). עבור המומנטים מסדר ראשון ושני של $\underline{Z}$ מתקיים:

$$\mathbb{E}[\underline{Z}] = \Big(\mu_X(t_1), \mu_X(t_2), \ldots, \mu_X(t_n)\Big)^T$$

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 $(\nu_1,\ldots,\nu_n)$, הפונקציה האופיינית של הוקטור $\underline{Z}$ נתונה ע"י:

$$\phi_Z(\nu_1,\ldots,\nu_n) = \exp\left\{ i\sum \nu_i \mu_X(t_i) - \frac{1}{2}\sum_i \sum_j \nu_i \nu_j \,\mathbf{K}_X(t_i,t_j)\right\}$$

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 גאוסי סטציונרי במובן הרחב הוא סטציונרי.

טענה: אם $\{X_t, -\infty < t < \infty\}$ ת"א גאוסי, גם התהליכים

$$a(t)X_t$$

$$a(t)X_t + b(t)X_{t+c}$$

תהליכים גאוסיים עבור כל $c,\ b(\cdot),\ a(\cdot)$ דטרמיניסטיים (מדוע? הוכח). לכן נצפה גם שאם הגבול

$$\frac{dX_t}{dt} = \lim_{\varepsilon\to 0}\frac{X_{t+c} - X_t}{\varepsilon}$$

קיים, אזי (בהנחות מתאימות) הגבול גם הוא ת"א גאוסי.

יהיה X_t ת"א גאוסי; אזי

$$\int_a^b X_s ds \approx \sum_i X_{s_i}(s_{i+1} - s_i)$$

הוא מ"א גאוסי 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 גאוסי נותנת תהליך אקראי גאוסי.

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

$$\mathbb{E}\int_0^t |X_s|\,ds < \infty \quad \text{או התנאי השקול} \quad \int_0^t \mathbb{E}\,|X_s|\,ds < \infty \qquad (5.21)$$

5.4 מעבר תהליכים אקראיים דרך מערכות לינאריות

נתונה מערכת כבאיור 5.8. תהליך הכניסה הוא $X(t)$ ותהליך היציאה הוא $Y(t)$.

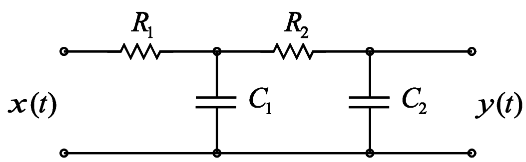
שאלה: ידוע חוק ההסתברות של התהליך $\{X(t), -\infty < t < \infty\}$, נתונה מערכת (לא אקראית, לינארית או לא לינארית) שאפיונה ידוע. מהו חוק ההסתברות של תהליך היציאה $\{Y(t), -\infty < t < \infty\}$? בדרך כלל התשובה אינה

איור 5.8: מעבר אות במערכת דטרמיניסטית

ידועה אפילו אם מדובר במערכת לינארית קבועה בזמן. באופן מעשי נתעניין הרבה פעמים לא בחוק ההסתברות של $Y(t)$ אלא ב-$\mathbb{E}[(Y(t)-\mathbb{E}[Y(t)])^2]$ או $\mathbb{E}[Y^2(t)]$. 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.

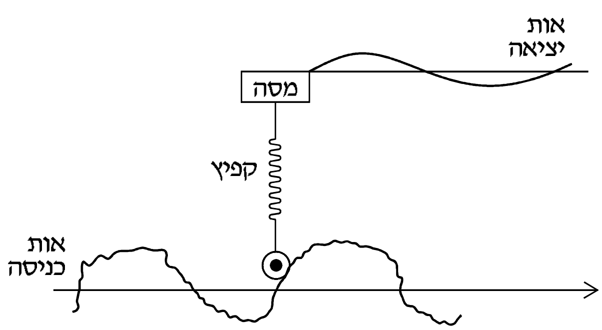
<u>דוגמאות:</u>

(א) מסנן R-C:



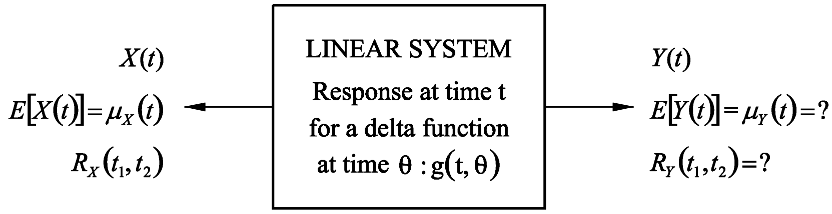
איור 5.9: מסנן

(ב) בולם זעזועים:



איור 5.10: בולם זעזועים

ננסח את השאלה הבאה: נתונה מערכת לינארית לא אקראית שהאיפיון שלה ידוע; מה צריך לדעת על $\{X(t), -\infty < t < \infty\}$, על מנת שנוכל לחשב את ? $\mathbb{E}[Y^2(t)] =$. בהמשך נקבל תשובה <u>מלאה</u> לשאלה זו. ברור שידיעת $\mathbb{E}[X^2(t)]$ אינה מספיקה (מדוע?). את הבעיה נתאר בצורה ציורית כדלקמן:



איור 5.11:

הקשר בין כניסת המערכת ויציאתה נתון ע"י:

$$Y(t) = \int_{-\infty}^{\infty} X(\theta)g(t,\theta)d\theta \cong \sum_i X(\theta_i)g(t,\theta_i)(\theta_{i+1}-\theta_i) \tag{5.22}$$

לא נתיחס כאן לבעיות של דיוק מתמטי ונרשום:

$$\mathbb{E}[Y(t)] \cong \sum_i \mathbb{E}\Big[X(\theta_i)g(t,\theta_i)(\theta_{i+1}-\theta_i)\Big] \to \int_{-\infty}^{\infty} \mu_X(\theta)g(t,\theta)d\theta$$

במילים אחרות, בצענו תוחלת על (5.22) 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 שידיעת $\mathbf{R}_Y(t_1,t_2)$ כוללת ידיעת "הספק היציאה הממוצע"

$$\mathbb{E}[Y^2(t)] = \mathbf{R}_Y(t,t)$$

$$\begin{aligned}\mathbb{E}\Big[Y(t_1)Y(t_2)\Big] &= \mathbb{E}\left[\int_{-\infty}^{\infty} X(\theta)g(t_1,\theta)d\theta \int_{-\infty}^{\infty} X(\eta)g(t_2,\eta)d\eta\right] \\ &= \mathbb{E}\left[\iint_{-\infty}^{\infty} X(\theta)X(\eta)g(t_1,\theta)g(t_2,\eta)n\theta d\eta\right] \\ &= \iint_{-\infty}^{\infty} \mathbf{R}_X(\theta,\eta)g(t_1,\theta)g(t_2,\eta)d\theta d\eta \qquad (5.23)\end{aligned}$$

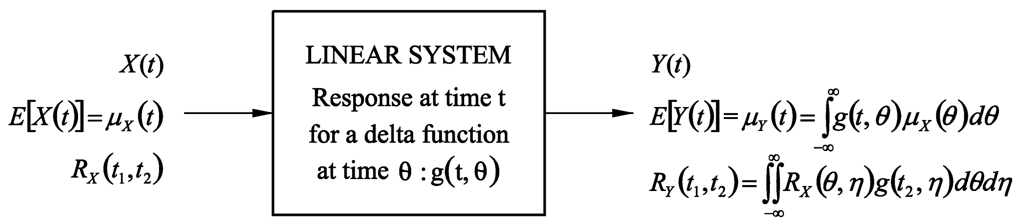
<u>מסקנות:</u>

(א) עבור מערכות לינאריות לא אקראיות (קבועות בזמן או משתנות בזמן, סיבתיות או לא סיבתיות), ידיעת $\mu_X(t), \mathbf{R}_X(t_1,t_2)$ 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 (5.22) ו (5.23) נסכם בציור הבא:



איור 5.12: 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 $\mu_X(t) = \mu_X$ 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 שהאנטגרל האחרון הוא ההגבר DC של המערכת הלנארית. קיבלנו שממוצע התגובה גם הוא אינו תלוי בזמן. באשר לאוטוקורלציה,

$$\begin{aligned}
\mathbf{R}_Y(t,t+\tau) &= \mathbb{E}\Big[Y(t)Y(t+\tau)\Big] \\
&= \mathbb{E}\left[\int_{-\infty}^{\infty} X(t-\theta)h(\theta)\,d\theta \int_{-\infty}^{\infty} X(t+\tau-\eta)h(\eta)\,d\eta\right] \\
&= \iint\limits_{-\infty}^{\infty} \mathbb{E}\Big[X(t-\theta)X(t+\tau-\eta)\Big]h(\theta)h(\eta)\,d\theta\,d\eta \\
&= \iint\limits_{-\infty}^{\infty} R_X(\tau-\eta+\theta)h(\theta)h(\eta)\,d\theta\,d\eta \\
&= [\mathbf{R}_X * h * \check{h}](\tau)
\end{aligned}$$

כאשר הגדרנו $\check{h}(\tau) = h(-\tau)$ והמשוואה האחרונה היא סימון נכון יותר ל-$\mathbf{R}_X(\tau) * h(\tau) * h(-\tau)$. כמו כן

$$\mathbf{R}_{X,Y}(t,t+\tau) = \mathbb{E}\Big[X(t)Y(t+\tau)\Big] \tag{5.24}$$

$$= \mathbb{E}\left[\int_{-\infty}^{\infty} X(t)X(t+\tau-\theta)h(\theta)\,d\theta\right] \tag{5.25}$$

$$= \int_{-\infty}^{\infty} R_X(\tau-\theta)h(\theta)\,d\theta \tag{5.26}$$

$$= [R_X * h](\tau) \tag{5.27}$$

$$= R_{X,Y}(\tau)\ . \tag{5.28}$$

בפיתוחים שעשינו החלפנו סדר בין תוחלת ואינטגרציה. כפי שהזכרנו החלפה זו מותרת אם מתקיים תנאי (5.21). אולם תנאי זה שקול ליציבות BIBO.

<u>מסקנה:</u> 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-

$$\mu_Y = \mu_X \int_{-\infty}^{\infty} h(\theta)d\theta \tag{5.29}$$

$$R_Y(\tau) = \iint\limits_{-\infty}^{\infty} R_X(\tau-\eta+\theta)h(\theta)h(\eta)d\theta d\eta \tag{5.30}$$

בפרט, אם $\{X(t)\}$ ת"א גאוסי סטציונרי במובן הרחב אזי הוא גם סטציונרי, ונובע כי גם $\{Y(t)\}$ גאוסי וסטציונרי וכן $\{X(t), Y(t)\}$ גאוסיים במשותף וסטציונריים במשותף.

אם המערכת סיבתית אזי $h(\theta)=0$ עבור $\theta<0$ ואז

$$R_Y(\tau)=\int_0^\infty\int_0^\infty R_X(\tau-\eta+\theta)h(\theta)h(\eta)d\theta d\eta \tag{5.31}$$

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.

התמרות פוריה

באותות ומערכות למדנו שיש שתי גישות לאפיון מערכות לינאריות קבועות בזמן :

גישה א' : אפיון המערכת בתחום הזמן, ע"י תגובתה להלם $h(t)$ וייצוג כניסה שרירותית X ע"י סופרפוזיציה $X(t)=X*\delta(t)=\int_{-\infty}^{\infty}X(\tau)\delta(t-\tau)\,d\tau$. מתקבלת התגובה $Y(t)=X*h(t)=\int_{-\infty}^{\infty}X(\tau)h(t-\tau)\,d\tau$.

גישה ב' : אפיון המערכת בתחום התדר: ע"י תגובתה $H(f)$ לערור הרמוני $(e^{i2\pi ft})$ (כאן f הוא התדר ו-$2\pi f$ הוא התדר הזויתי). יצוג כניסה כללית $X(t)$ ע"י סיכום תנודות הרמוניות (התמרת פוריה הפוכה) $X(t)=\int_{-\infty}^{\infty}X(f)e^{2\pi ft}\,df$ ושוב סופרפוזיציה $Y(t)=\int_{-\infty}^{\infty}H(f)X(f)e^{2\pi ft}\,df$. במקרה הדטרמיניסטי הקשר בין הכניסה והיציאה (גישה א') הוא:

$$Y(t)=\int_{-\infty}^{\infty}X(t-\theta)h(\theta)d\theta$$

ואילו במקרה האקראי הקשר בין פונקציות האוטוקורלציה בכניסה וביציאה נתון ע"י

$$R_Y(\tau)=\iint_{-\infty}^{\infty}R_X(\tau+\theta-\eta)h(\theta)h(\eta)d\theta d\eta$$

שאלה : האם לגישה ב' אפשר לתת מובן במקרה של תהליכים אקראיים? בהמשך נראה שהדבר אכן אפשרי וכשם שבמקרה הדטרמיניסטי יש יתרונות רבים לאנליזה במרחב התדר כן גם במקרה האקראי.

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:

$$\hat{X}(f)=F\{X(t)\}=\int_{-\infty}^{\infty}X(t)e^{-2\pi ift}\,dt$$

והתמרת פוריה ההפוכה נתונה ע"י

$$X(t)=F^{-1}\{\hat{X}(f)\}=\int_{-\infty}^{\infty}\hat{X}(f)e^{2\pi ift}\,df$$

(קיימת נקודה עדינה לגבי קיום האינטגרל האחרון אולם נתעלם מבעיה זו). משפט פרסוול בהתמרות פוריה אומר שאם $\int_{-\infty}^{\infty} |X_i(t)|^2 \, dt < \infty$ עבור $i = 1, 2$ אזי

$$\int_{-\infty}^{\infty} X_1(t) X_2^*(t) \, dt = \int_{-\infty}^{\infty} \hat{X}_1(f) \hat{X}_2^*(f) \, df$$

ל-$\int_{-\infty}^{\infty} X^2(t) \, dt$ נקרא האנרגיה של הפונקציה $X(\cdot)$ (מה האנרגיה של דגם טפוסי של ת"א סטציונרי?)

אם $F\{X(t)\} = \hat{X}(f)$ אזי בתנאים מתאימים קיים

$$F\left\{\frac{dX(t)}{dt}\right\} = 2\pi i f \hat{X}(f)$$

אם $X(\cdot)$ ממשי אזי

$$F\{X(-t)\} = \hat{X}^*(f)$$

$$F\{X(t+\tau)\} = e^{2\pi i f \tau} \hat{X}(f)$$

$$F\{X * h(t)\} = F\left\{\int_{-\infty}^{\infty} X(t-\theta) h(\theta) \, d\theta\right\} = \hat{X}(f) \cdot \hat{h}(f)$$

צפיפות ספקטרלית

כאשר $(X(t), -\infty < t < \infty)$ ת"א סטציונרי, האנרגיה של פונקצית 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 פונקצית האוטוקורלציה. נגדיר, עבור ת"א סטציונריים, $\{Y(t), X(t)\}$ את:

(א) הצפיפות הספקטרלית $S_X(f)$ של $\{X(t), -\infty < t < \infty\}$

$$S_X(f) \triangleq F\{R_X(\tau)\}$$

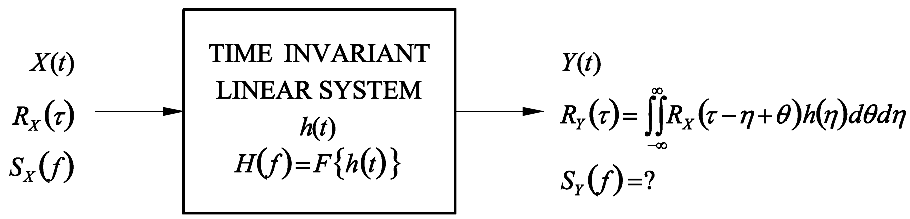
(ב) פונקצית הצפיפות הספקטרלית המצטלבת

$$S_{X,Y}(f) \triangleq F\{R_{X,Y}(\tau)\}$$

וזאת בהנחה שלפונקציות $R_{X,Y}(\cdot),\ R_X(\cdot)$ 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 5.13. אנו נניח כי h ממשית. במקרה הנוכחי:

$$R_{X,Y}(\tau) = \mathbb{E}\Big[X(t)Y(t+\tau)\Big] = \int_{-\infty}^{\infty} R_X(\tau - \theta) h(\theta) \, d\theta = R_X * h(\tau)$$



איור 5.13: מעבר אות סמ"ר במערכת לק"ב יציבה

ולכן (מדוע?)

$$S_{X,Y}(f) = S_X(f) \cdot H(f)$$

הקונוולוציה הפכה למכפלה. לגבי $S_Y(f)$:

$$R_Y(\tau) = \iint\limits_{-\infty}^{\infty} R_X(\tau-\theta)+\eta)h(\theta)h(\eta)\,d\theta\,d\eta = R_X(\tau) * h(\tau) * h(-\tau)$$

על פי ההגדרה $S_Y(f) = \int_{-\infty}^{\infty} R_Y(\tau) e^{-2\pi i f \tau}\,d\tau$ ולכן בשינוי סדר האינטגרציה

$$\begin{aligned}\int_{\eta=-\infty}^{\infty}\int_{\theta=-\infty}^{\infty}\int_{\tau=-\infty}^{\infty} e^{-2\pi i f\tau} R_X(\tau+\theta-\eta)h(\theta)h(\eta)\,d\tau\,d\theta\,d\eta &= \iint\limits_{-\infty}^{\infty} S_X(f)e^{2\pi i f(\theta-\eta)}h(\theta)h(\eta)\,d\theta\,d\eta \\ &= S_X(f)H(f)\cdot H^*(f) \\ &= S_X(f)|H(f)|^2\end{aligned}$$

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:

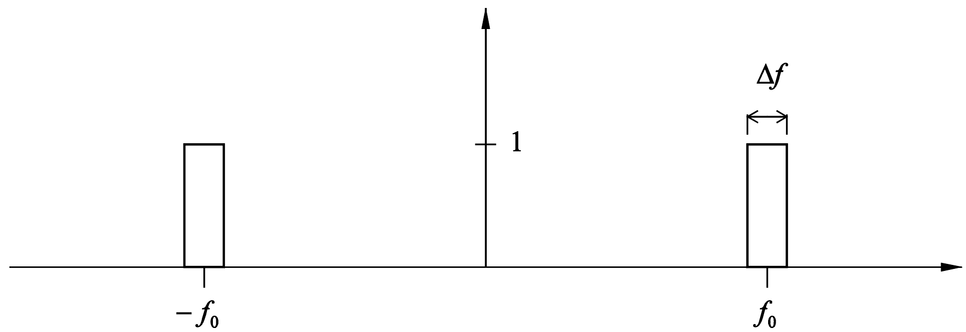
$$R_Y(0) = \int_{-\infty}^{\infty} S_X(f)|H(f)|^2\,df$$

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 $R_X(0)$).



איור 5.14:

(ג) <u>טענה:</u> $S_X(f) \geq 0$.

<u>הוכחה:</u> נבחר ב-$H(f)$ להיות מסנן צר סרט אידיאלי סביב תדר שרירותי f_0:

הספק היציאה במקרה זה יהיה (בערך)

$$0 \leq \mathbb{E}\left[Y^2(0)\right] = \int_{-\infty}^{\infty} S_X(f)|H(f)|^2\, df \approx 2S_X(f_0)\Delta f \tag{5.32}$$

ולכן עבור כל f_0, $S_X(f_0) \geq 0$.

$\mathbb{E}\left[Y^2(t)\right]$ 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 $(\mathrm{Volt})^2/\mathrm{Hz}$ או $(\mathrm{Amp})^2/\mathrm{Hz}$

נעיין בדוגמה: $X(t) = A(\omega)\cos(2\pi f_0 t + \phi(\omega))$ כאשר A, ϕ ב"ת ו-ϕ מפולג אחיד בתחום $[0, 2\pi)$. כזכור

$$R_X(\tau) = \frac{1}{2}\,\mathbb{E}[A^2]\cos 2\pi f_0 \tau \tag{5.33}$$

$$S_X(f) = \frac{1}{4}\,\mathbb{E}[A^2][\delta(f - f_0) + \delta(f + f_0)] \tag{5.34}$$

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 מתכונות אלו נובע שלא כל פונקציה זוגית יכולה להיות אוטוקורלציה: כדי להיות כזאת הפונקציה צריכה, למשל, להיות בעלת התמרת פוריה לא-שלילית.

<u>לסיכום:</u> מושג הצפיפות הספקטרלית נותן תאור של תכולת ההספק בתדרים השונים, דהיינו, גם אם לא ניתן לדבר על התמרת פוריה של פונקצית מדגם של התהליך $\{X(t), -\infty < t < \infty\}$, פונקצית 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 <u>הממוצע</u> ביציאה ממסננת bandpass אידאלית בעלת רחב סרט Δ בתחום $(f_0, f_0+\Delta)$.

<u>שים לב:</u> כאשר $\phi(t)$ פונקציה דטרמיניסטית בעלת אנרגיה סופית, אנו יכולים לדעת את תכולת <u>האנרגיה</u> בתחום תדרים מסוים ע"י עיון באינטגרל של $|\Phi(f)|^2$ על תחום התדרים המעניין. כאן אנו עוסקים בתכולת <u>ההספק</u>!

כדי להווכח כי $S_X(f)$ אכן מתאר את תכולת ההספק של <u>פונקציות המדגם</u>, 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

$$\mathbb{E}\,\varepsilon^2 = \int S_X(f)\,|1-H(f)|^2\ df = 0\,.$$

<u>הערות:</u>

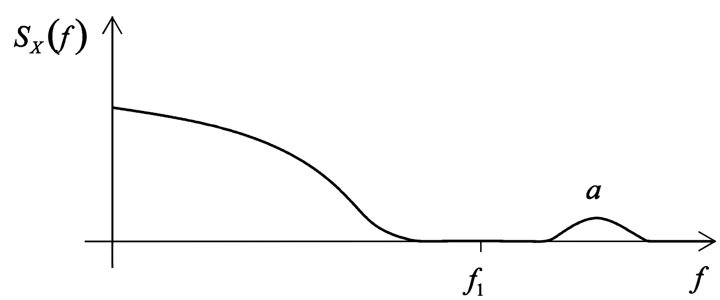
(א) גוזר $Y(t)=dX(t)/dt$ מבליט תדרים גבוהים ומתקיים עבורו $F\{Y(t)\} = 2\pi i f F\{X(t)\}$.

(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 הציר האופקי על מנת לברר אם הוא מבטא זמן או תדר.

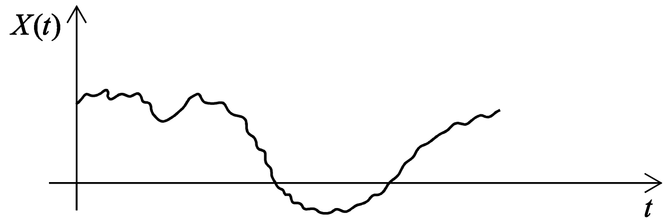
(ד) כזכור הגדרנו את הצפיפות הספקטרלית עבור תהליך אקראי בעל תוחלת אפס. אפשר ליחס צפיפות ספקטרלית גם לתהליכים (סטציונריים במובן הרחב) בעלי תוחלת שונה מאפס. במקרה זה תהיה לצפיפות הספקטרלית פונקצית דירק בתדר אפס. דהיינו, אם לת"א $X(t)$ נוסיף מ"א C בלתי תלוי ב $X(t)$, לקבלת $Y(t)=X(t)+C$, אזי: $S_Y(f) = S_X(f) + \mathbb{E}[C^2]\delta(f)$.

<u>דוגמה:</u> נעסוק באות אקראי $X(t)$ סטציונרי במובן הרחב. תהיה $S_X(f)$ הצפיפות הספקטרלית---ראה איור 5.15--- של אות זה (כולל החלק α).



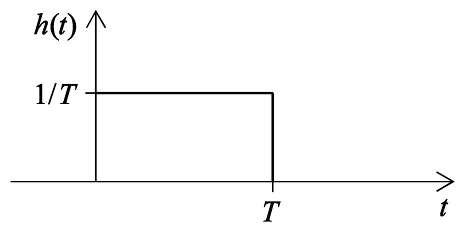
איור 5.15: צפיפות ספקטרלית של אות הכולל רעש

אנו מניחים שה"תוספת" המסומנת ב-α נובעת מרעש שאיננו מעוניינים בו. בהנחה ש-$S_X(f)$ (כולל החלק α) מתיחס לתהליך אקראי גאוסי, דגם טיפוסי יראה כמתואר באיור 5.16.



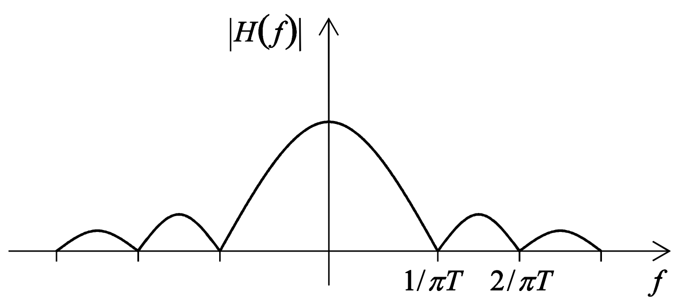
איור 5.16: פונקצית מדגם של האות הכולל רעש

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-$|H(f)|^2 = (2\pi f)^2$ 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 5.17 ואז $H(f) = F\{h(t)\}$ נראה כמתואר



איור 5.17: תגובת הלם של המסנן

באיור 5.18. אם נבחר ב-T גדול מאוד, נפגע בסיגנל עצמו, אם נבחר ב-T קטן מאוד לא נחליק את הרעש. מתוך



איור 5.18: תגובת תדר של המסנן

שיקולים ענייניים נראה שכדאי לבחור את T כך ש- $f_1 = 1/\pi T$ כאשר f_1 מופיע בציור לעיל.

<u>שאלה:</u> 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

$$X_T(t) = \begin{cases} X(t) & \text{if} \quad |t| < T \\ 0 & \text{if} \quad |t| \geq T \end{cases}$$

$$\hat{X}_T(f) = F\{X_T(t)\}$$

אזי

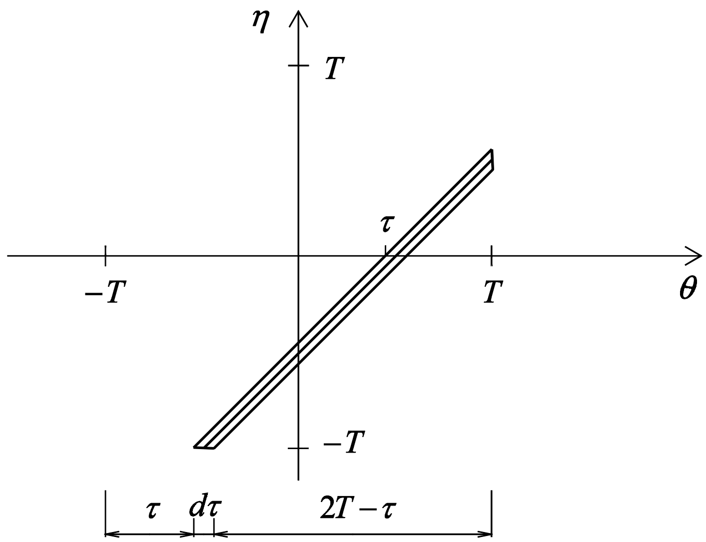
$$\mathbb{E}\left[\frac{1}{2T}\,|\hat{X}_T(f)|^2\right] \underset{T\to\infty}{\longrightarrow} S_X(f)$$

כאן רואים את הקשר בין התמרת פוריה של האות עצמו לצפיפות הספקטרלית. החלוקה באורך אינטרוול הזמן מעבירה אותנו מאנרגיה להספק.

הוכחה:

$$\mathbb{E}\left[\frac{1}{2T}|\hat{X}_T(f)|^2\right] = \frac{1}{2T}\,\mathbb{E}\left[\iint\limits_{-T}^{T} X_\theta X_\eta e^{-2\pi i f(\theta-\eta)}\,d\theta\,d\eta\right] = \frac{1}{2T}\iint\limits_{-T}^{T} R_X(\theta-\eta)e^{-2\pi i f(\theta-\eta)}\,d\theta\,d\eta$$

נסמן $\tau = \theta - \eta$ ונעיין באיור 5.19: נבצע קודם אינטגרציה על הפס הצר שעבורו $\tau \leq \theta - \eta \leq \tau + d\tau$ לכן



איור 5.19: שינוי סדר אינטגרציה

$$\mathbb{E}\left[\frac{1}{2T}|\hat{X}_T(f)|^2\right] = \frac{1}{2T}\int_{-2T}^{2T} R_X(\tau)e^{-2\pi i f\tau}(2T-|\tau|)\cdot\sqrt{2}\,\frac{d\tau}{\sqrt{2}}$$

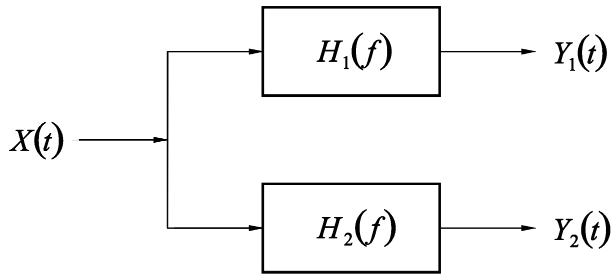
$$= \int_{-2T}^{2T} R_X(\tau)e^{-2\pi i f\tau}\left(1-\frac{|\tau|}{2T}\right)d\tau \underset{T\to\infty}{\longrightarrow} S_X(f) = F\{R_X(\tau)\}(f)\ .$$

מ.ש.ל.

נחזור למושג הצפיפות הספקטרלית המצטלבת

$$S_{X,Y}(f)=F\{R_{X,Y}(\tau)\}=F\{\mathbb{E}[X(t)Y(t+\tau)]\}$$

נאמר כי W,Z חסרי קורלציה אם $\mathbb{E}\,W(t_1)Z(t_2)\equiv 0$ (כלומר שווה לאפס לכל t_1,t_2) או, באופן שקול: $S_{WZ}(f)\equiv 0$.
נעיין במיוחד במקרה המתואר באיור 5.20.



איור 5.20: כניסה משותפת לשתי מערכות במקביל

$$Y_i(t)=\int_{-\infty}^{\infty}h_i(\theta)X(t-\theta)\,d\theta$$

$$R_{Y_1,Y_2}(\tau)=\mathbb{E}\Big[Y_1(t)Y_2(t+\tau)\Big]=\mathbb{E}\left[\iint_{-\infty}^{\infty}X(t-\theta)X(t+\tau-\eta)h_1(\theta)h_2(\eta)\,d\theta\,d\eta\right]$$

$$=\iint_{-\infty}^{\infty}R_X(\tau+\theta-\eta)h_1(\theta)h_2(\eta)\,d\theta\,d\eta=[R_X*\check{h}_1*h_2](\tau)$$

כאשר השתמשנו בסימון $\check{h}_1(\tau)=h_1(-\tau)$. לכן עבור h ממשית

$$S_{Y_1,Y_2}(f)=\iiint R_X(\tau+\theta-\eta)e^{-2\pi i f\tau}h_1(\theta)h_2(\eta)\,d\theta\,d\eta\,d\tau$$

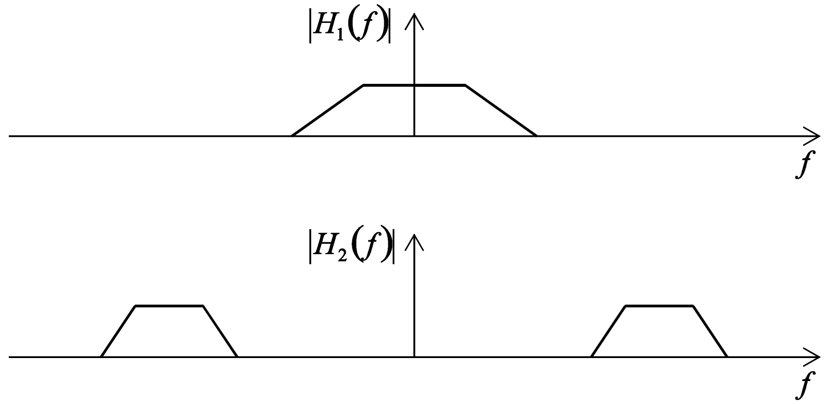
$$=\iint S_X(f)e^{2\pi i f(\theta-\eta)}h_1(\theta)h_2(\eta)\,d\theta\,d\eta=S_X(f)H_1^*(f)H_2(f)$$

במיוחד אם $H_2(f)H_1(f)\equiv 0$, כגון כנתון באיור 5.21 אזי התהליכים האקראיים $Y_1(t)$ ו-$Y_2(t)$ חסרי קורלציה (ולכן במקרה של תהליכים גאוסיים במשותף הת"א $\{Y_1(t),-\infty<t<\infty\}$ ו-$\{Y_2(t),-\infty<t<\infty\}$ בלתי תלויים).

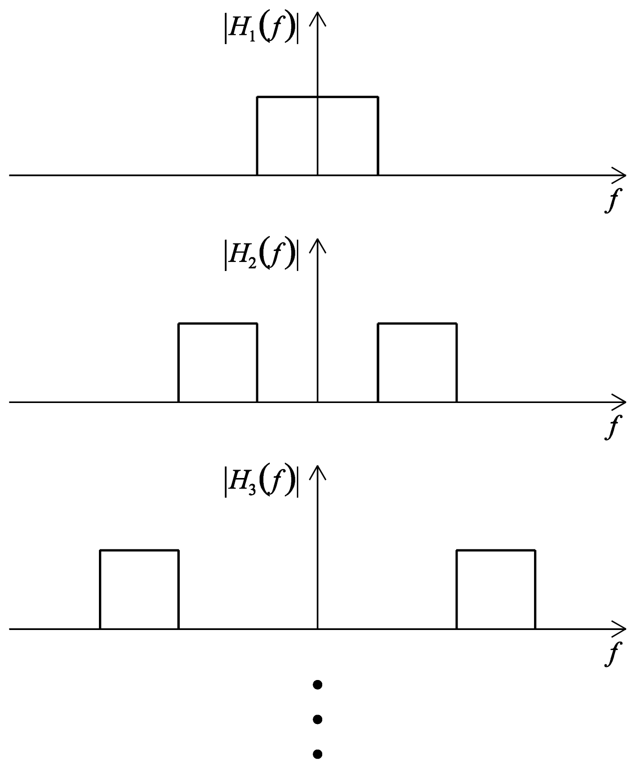
<u>הערה:</u> הוכח שהמשתנים האקראיים $X(t)$ ו-$dX(t)/dt$ חסרי קורלציה עבור אותו t, אולם כתהליכים אקראיים הם לא חסרי קורלציה.

נגדיר אוסף של מסננים $H_1(f),H_2(f),\ldots$ ע"י איור 5.22. מסננים אלו מקיימים:

$$\forall i\neq j,\quad H_i(f)\cdot H_j(f)\equiv 0\,;\quad \sum_i H_i(f)=1$$



איור 5.21: שני מסננים ללא חפיפה בתגובת התדר



איור 5.22: אוסף מסננים ללא חפיפה במישור התדר

כלומר המסננים מחלקים את מישור התדר כך בכל 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 5.21 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.

בעזרת מושג הצפיפות הספקטראלית 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 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 הגאוסי, בת"ס), ומטפלים בכל אות בנפרד. זוהי שיטה שמושית בבעיות מעשיות רבות, למשל בדחיסת אינפורמציה.

5.5 הזזת 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

$$F\Big\{g(t)\cos 2\pi f_0 t\Big\} = \frac{1}{2}\Big[G(f+f_0) + G(f-f_0)\Big]$$

ולכן

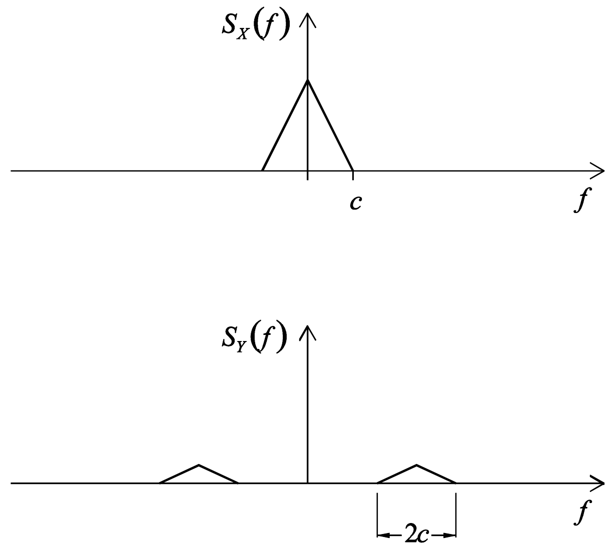
$$S_Y(f) = \frac{1}{4}\Big[S_X(f+f_0) + S_X(f-f_0)\Big]$$

מקרה א': אות צר סרט מאופנן על ידי אות בתדר גבוה כמתואר באיור 5.23.

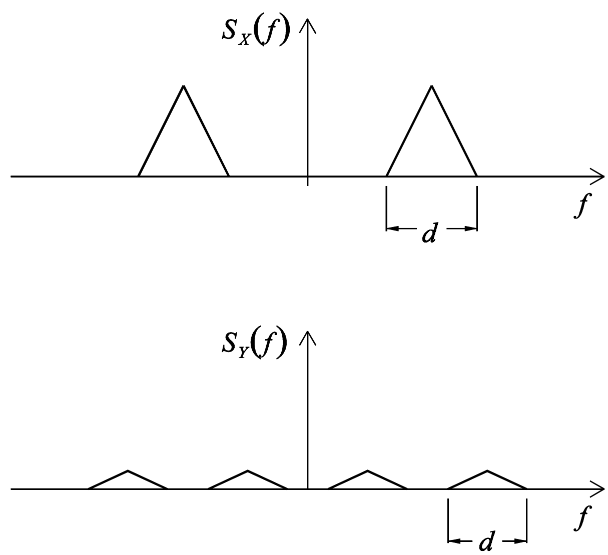
מקרה ב': אות המרוכז סביב תדר מסויים, מאופנן על ידי אות בתדר גבוה כמתואר באיור 5.24.

5.6 סינון לינארי אופטימלי

נעבור כעת לבעיה כללית של סינון לינארי. זוהי הרחבה (לאותות) של בעיית השערוך הלינארי, כאשר כעת המדידה היא תהליך. נניח שנתונים שני ת"א $\{X(t), -\infty < t < \infty\}$ ו- $\{n(t), -\infty < t < \infty\}$, לכל אחד מהם תוחלת אפס וכל אחד מהם סטציונרי במובן הרחב. כן נניח שהתהליכים חסרי קורלציה ($\mathbb{E}[X(t_1)n(t_2)] = 0$ לכל t_1 ו- t_2) ולכן הם סטציונריים במשותף במובן הרחב. נקלט האות $Y(t)$ שהוא סכום האות הרצוי $X(t)$ והרעש $n(t)$, כלומר, $Y(t) = X(t) + n(t)$. 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 5.26. 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 5.26 בצורה השקולה, המופיעה באיור 5.27. בגלל



איור 5.23: ספקטרום של אות צר סרט ושל האות המאופנן



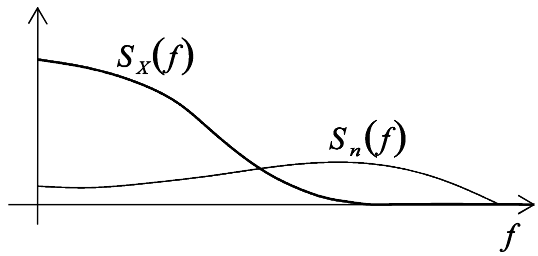
איור 5.24: ספקטרום של אות ושל האות המאופנן

אי התלות הלינארית בין $X(\cdot)$ לבין $n(\cdot)$ נוכל לרשום

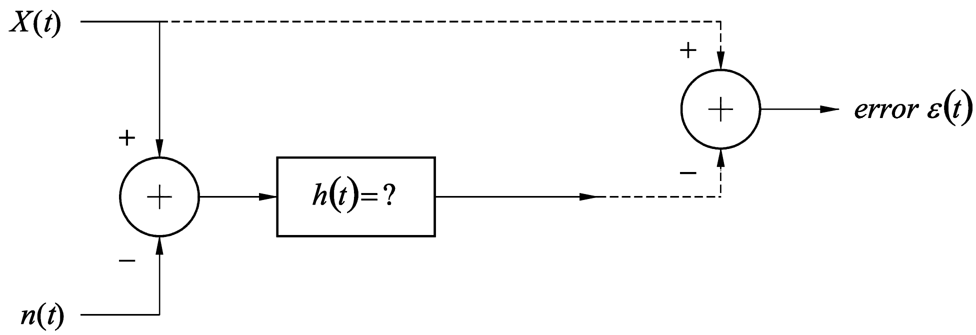
$$\mathbb{E}\left[\varepsilon^2(t)\right] = \mathbb{E}\left[z_1^2(t)\right] + \mathbb{E}\left[z_2^2(t)\right]$$

ולכן מאיור 5.27

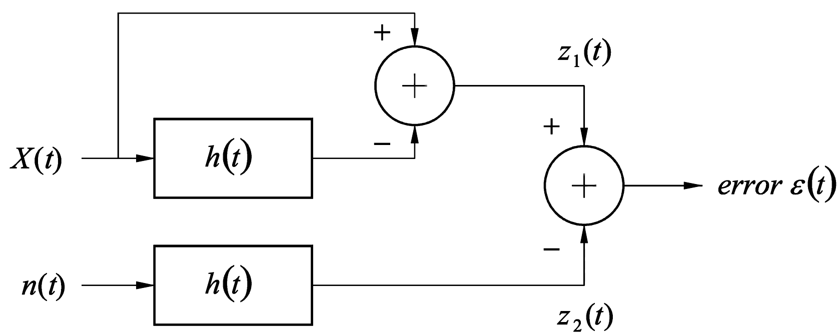
$$\mathbb{E}\left[\varepsilon^2(t)\right] = \int_{-\infty}^{\infty} S_n(f)|H(f)|^2\, df + \int_{-\infty}^{\infty} S_X(f)|1-H(f)|^2\, df$$



איור 5.25: צפיפות ספקטרלית טיפוסית של אות ושל רעש



איור 5.26: אות בתוספת רעש עוברים במערכת לינארית



איור 5.27: הצגה שקולה של אות בתוספת רעש העוברים במערכת לינארית

.

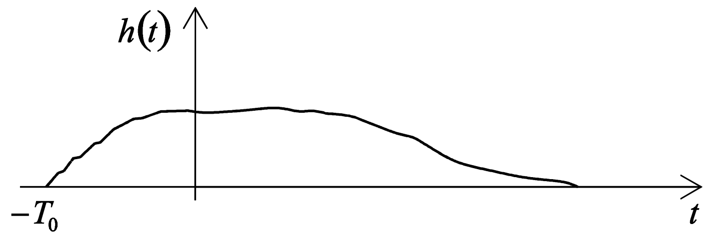
קבלנו ביטוי מפורש עבור השגיאה הריבועית הממוצעת הכוללת. עכשיו נוכל להשוות הצעות שונות ל-$H(f)$. למשל, עבור $H(f) = (1 + if/f_0)^{-1}$ (דהיינו, מסננת R-C) נוכל לחשב את השגיאה כפונקציה של הפרמטר f_0, לבצע אופטימיזציה ולמצוא את הפרמטר f_0 הטוב ביותר שיגרום לשגיאה ריבועית ממוצעת מינימלית עבור מסנן כזה.

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.

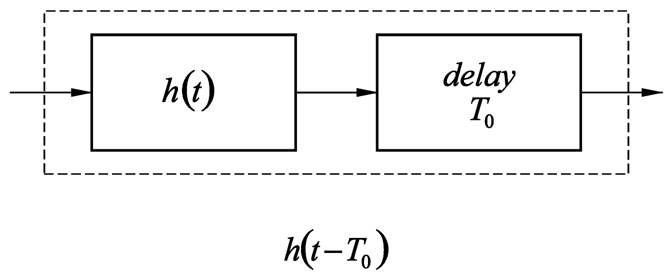
ב. קיימת מגבלה ש-$h(t)$ יהיה סיבתי.

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 Wiener 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 5.28 כאשר עבור $h(t) = 0,\ t < -T_0$ (או זניח) אזי $h(t)$ אינה



איור 5.28: תגובת הלם לא סבתית עם חלק שלילי סופי

סיבתית אולם $h(t-T_0)$ סיבתית. כלומר את המערכת 5.29 נוכל לממש. לכן, בבעיות בהן אפשר לסבול השהייה (כגון



איור 5.29: קרוב מעשי למערכת לא סבתית על ידי השהיה

בעיות תקשורת בניגוד לבעיות בקרה) יש לפתרון ללא סיבתיות משמעות פיזיקלית. לפתרון זה קוראים "מסננת וינר עם השהייה אינסופית".

נחזור לבעית מציאת המסננת האופטימלית המביאה את השגיאה הריבועית הממוצעת למינימום. השגיאה הריבועית הממוצעת ניתנת כאמור ע"י:

$$\mathbb{E}\Big[\varepsilon^2(t)\Big] = \int_{-\infty}^{\infty} S_n(f)|H(f)|^2\,df + \int_{-\infty}^{\infty} S_x(f)|1-H(f)|^2\,df$$

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:

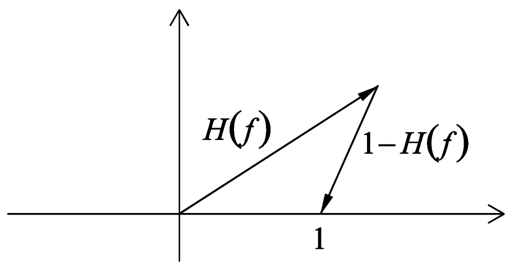
$$\mathbb{E}\Big[\varepsilon^2(t)\Big] = \int_{-\infty}^{\infty} \big[S_n(f)|H(f)|^2 + S_x(f)|1-H(f)|^2\big]\ df \tag{5.35}$$

ונשים לב כי כל הביטויים תחת האינטגרל חיוביים. אם כך, אנו מחפשים $H(f)$ כך שהשטח מתחת לפונקציה

$$S_n(f)|H(f)|^2 + S_x(f)|1-H(f)|^2$$

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 ממציאת מינימום של אינטגרל למציאת מינימום של פונקציה.

שלב שני: עבור f מסוים נעיין ב-$|1-H(f)|$. נשרטט זאת באיור 5.30. שים לב שבין כל הנקודות על המעגל (סביב



איור 5.30: בחירת הפאזה של H

הראשית) $|H(f)| = \text{const}$, הקרובה ביותר לנקודה 1 היא הנקודה (הממשית) $|H(f)|$. לכן אם עבור אותו f נחליף את $H(f)$ ב-$|H(f)|$ נגרום להקטנת $|1-H(f)|$. באופן אלגברי:

$$|1-H(f)| \geq 1 - |H(f)|$$

$$|1-H(f)| \geq |H(f)| - 1$$

ולכן

$$|1-H(f)| \geq |1-|H(f)||\,.$$

ע"י החלפה זו לא נשנה את הביטוי $S_n(f)|H(f)|^2$ אבל נקטין את האיבר השני $S_X(f)|1-H(f)|^2$, 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

$$S_n(f)\Big[H(f)\Big]^2 + S_X(f)\Big[1-H(f)\Big]^2$$

הוא מינימלי. נסמן (עבור f נתון) ב-$a = H(f)$ את אותו $H(f)$ עבורו האינטגרנד מינימלי. נמצא את המינימום על ידי

$$\frac{\partial}{\partial a}\Big\{S_n(f)a^2 + S_X(f)(1-a)^2\Big\} = 0$$

ונקבל

$$a=\frac{S_X(f)}{S_X(f)+S_n(f)}$$

ולכן

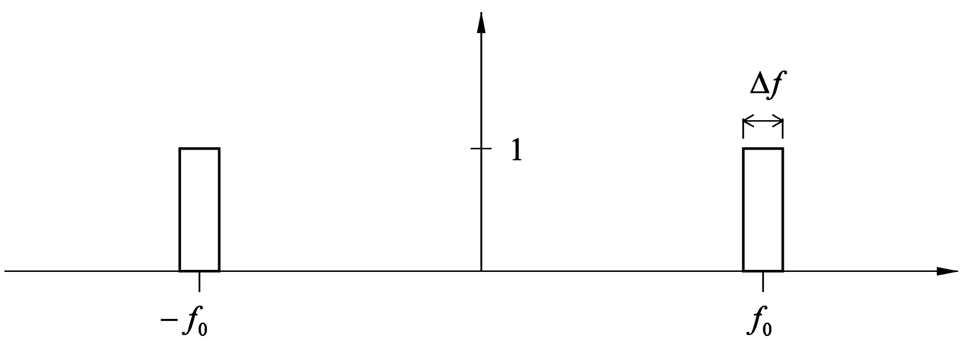
$$H_{\text{opt}}(f)=\frac{S_X(f)}{S_X(f)+S_n(f)} \tag{5.36}$$

והשגיאה המינימלית מתקבלת ע"י הצגת $H_{\text{opt}}(f)$ לתוך נוסחת השגיאה הרבועית הממוצעת:

$$\begin{aligned}\mathbb{E}\left[\varepsilon^2_{\min}(t)\right] &= \int_{-\infty}^{\infty}\left(\frac{S_n(f)S_X^2(f)}{(S_X(f)+S_n(f))^2}+\frac{S_X(f)S_n^2(f)}{(S_X(f)+S_n(f))^2}\right)df\\ &= \int_{-\infty}^{\infty}\frac{S_X(f)S_n(f)}{S_X(f)+S_n(f)}\,df\end{aligned}$$

נשים לב כי מאיורים 5.26--5.27. 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 שערוך לינארי (גאוסי אופטימלי שהוא לינארי) של משתנה אקראי כל סמך מדידה רועשת שלו---ראה (2.7) והמקדם האופטימלי (2.10). 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 שערוך של מ"א.

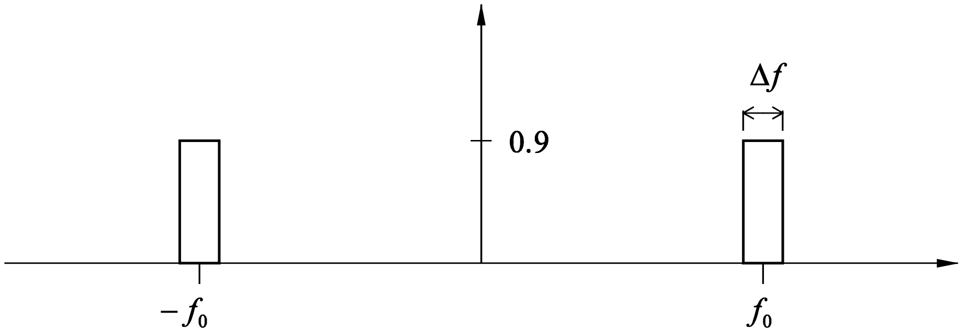
<u>דוגמה:</u> יהיה $X(t)$ תהליך אקראי בעל צפיפות הספק כנתון באיור 5.31. אליו מתווסף רעש $n(t)$ בעל צפיפות הספק



איור 5.31: תהליך צר סרט סביב תדר f_0

$S_n(f)=1/9$, בלתי תלוי ב $X(t)$. המסננת האופטימלית $H_{\text{opt}}(f)$ לשחזור $X(t)$ מתוך $Y(t)=X(t)+n(t)$, נתונה ע"י (5.36), ונראית כנתון באיור 5.32.

שים לב ש-$H_{\text{opt}}(f)$ 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).



איור 5.32: מסנן אופטימלי

דוגמה 5.9 הצפיפות הספקטרלית של האות ושל הרעש נתונים על ידי

$$S_X(f) \doteq \frac{S_0}{1+(f/f_0)^2}$$
$$S_n(f) \doteq N_0 \ .$$

אזי המסנן האופטימלי נתון על ידי (5.36):

$$H_{\rm opt}(f) = \frac{S_X(f)}{S_X(f)+S_n(f)} = \frac{\frac{S_0}{1+(f/f_0)^2}}{\frac{S_0}{1+(f/f_0)^2}+N_0} = \frac{S_0}{S_0+N_0\left(1+\left(\frac{f}{f_0}\right)^2\right)} \tag{5.37}$$

$$= \frac{S_0}{S_0+N_0}\ \frac{S_0+N_0}{S_0+N_0\left(1+\left(\frac{f}{f_0}\right)^2\right)} = \frac{S_0}{S_0+N_0}\ \frac{1}{1+\left(\frac{N_0}{S_0+N_0}\left(\frac{f}{f_0}\right)^2\right)} \tag{5.38}$$

$$= \frac{S_0}{S_0+N_0}\ \frac{1}{1+\left(\frac{f}{f_0\gamma}\right)^2}\ , \qquad \gamma = \sqrt{1+\frac{S_0}{N_0}} \tag{5.39}$$

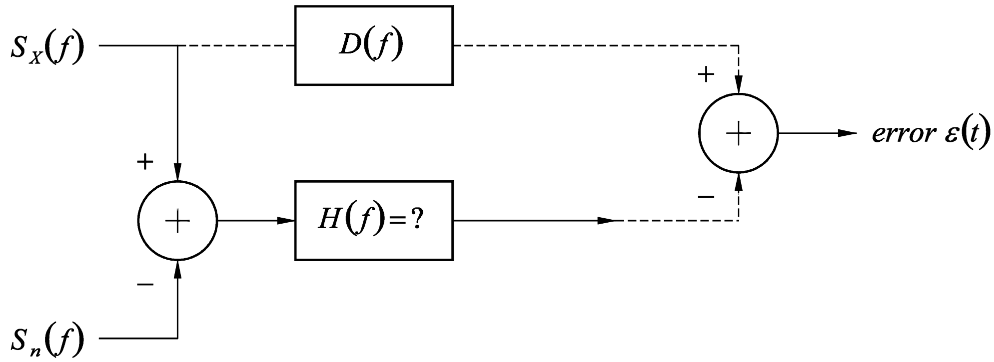
כאשר העצמה היחסית של האות גדולה, כלומר $S_0/N_0 \to \infty$ נקבל כי $\gamma \to \infty$ ובנוסף $\frac{S_0}{S_0+N_0} \to 1$ ולסיכום $H_{\rm opt} \to 1$. לעומת זאת כאשר העצמה היחסית של האות קטנה, כלומר $S_0/N_0 \to 0$ נקבל כי $\gamma \to 1$ ובנוסף $\frac{S_0}{S_0+N_0} \to 0$ ולסיכום $H_{\rm opt} \to 0$. חישוב דומה נותן כי השגיאה בדוגמה זו היא

$$\mathbb{E}[\epsilon^2(t)] = \int_{-\infty}^{\infty} \frac{S_X(f)S_n(f)}{S_X(f)+S_n(f)}\ df \tag{5.40}$$

$$= \frac{S_0N_0}{S_0+N_0}\int_{-\infty}^{\infty} \frac{1}{1+(f/f_0\gamma)^2}\ df \tag{5.41}$$

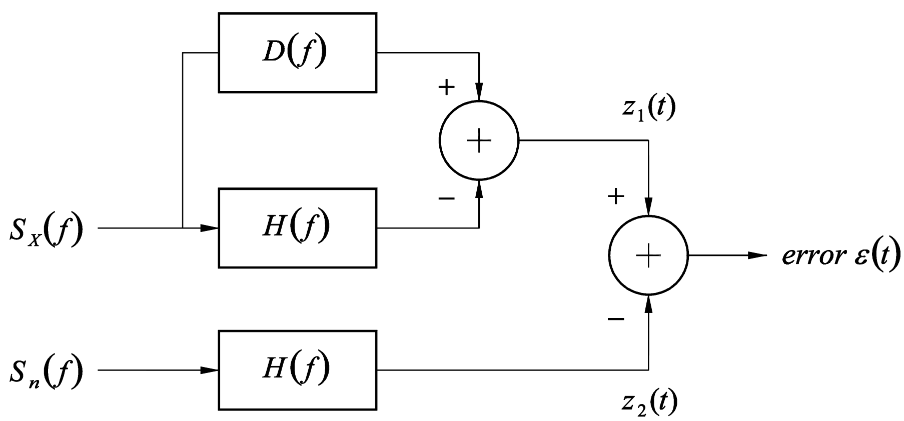
$$= \frac{\pi S_0N_0f_0\gamma}{S_0+N_0} = \frac{\pi S_0 f_0}{\gamma} \tag{5.42}$$

ניתן להרחיב במקצת את בעית מציאת המסננת האופטימלית. נמשיך להניח $\mathbb{E}[n(t)] = \mathbb{E}[X(t)] = 0$ וכן ש-$X(t_1), n(t_2)$ 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 גוזר. המערכת מתוארת (במרחב התדר) באיור 5.33. כאשר $D(f) = F\{d(t)\} = 1$ נקבל את המקרה



איור 5.33: אות תגובה מעובד ומסונן

הקודם. ע"י סופרפוזיציה נוכל להחליף את איור 5.33 באיור 5.34 ולכן



איור 5.34: אות תגובה מעובד ומסונן

$$\mathbb{E}\Big[\varepsilon^2(t)\Big] = \int_{-\infty}^{\infty} S_n(f)|H(f)|^2\,df + \int_{-\infty}^{\infty} S_X(f)|D(f) - H(f)|^2\,df$$

את המשוואה לעיל נרשום בצורה הבאה:

$$\mathbb{E}\Big[\varepsilon^2(t)\Big] = \int_{-\infty}^{\infty} \left|\frac{H(f)}{D(f)}\right|^2 S_n(f)|D(f)|^2\,df + \int_{-\infty}^{\infty} S_X(f)|D(f)|^2\;\left|1 - \frac{H(f)}{D(f)}\right|^2\,df$$

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:

$$\left(\frac{H(f)}{D(f)}\right)_{\text{opt}} = \frac{S_X(f)|D(f)|^2}{|D(f)|^2[S_X(f)+S_n(f)]} = \frac{S_X(f)}{S_X(f)+S_n(f)}$$

ולכן

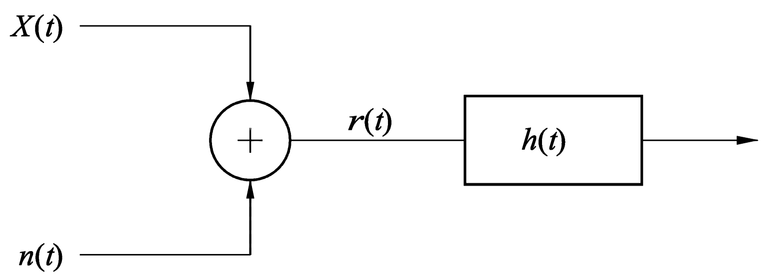
$$H_{\text{opt}}(f) = \frac{D(f)S_X(f)}{S_X(f)+S_n(f)} \tag{5.43}$$

והשגיאה הריבועית הממוצעת המינימלית המתקבלת היא:

$$\mathbb{E}\Big[\varepsilon^2_{\min}(t)\Big] = \int_{-\infty}^{\infty} \frac{|D(f)|^2 S_X(f) S_n(f)}{S_X(f)+S_n(f)}\, df$$

עקרון ההשלכה

לסיכום סעיף זה על סינון לינארי אופטימלי, נחשב בחשבון ישיר וקצר את המסנן האופטימלי (כולל המקרה של קורלציה בין האות לרעש) בעזרת עקרון ההשלכה. המערכת מתוארת במישור הזמן באיור 5.35. כזכור,



איור 5.35: סינון אות רועש---מערכת במישור הזמן

.

$r(t)=n(t)+X(t)$. נסמן ב-$d(t)$ את התמרת פוריה ההפוכה של $D(f)$ אזי

$$\mathbb{E}\Big[\varepsilon^2(t)\Big] = \mathbb{E}\left[\int_{-\infty}^{\infty} X(t-\theta)\, d(\theta)\, d\theta - \int_{\infty}^{\infty} r(t-\theta) h(\theta)\, d\theta\right]^2 .$$

ע"י עקרון ההשלכה נקבל שהמסננת האופטימלית חייבת לקיים:

$$\mathbb{E}\left[\left(\int_{-\infty}^{\infty} X(t-\theta)\, d(\theta)\, d\theta - \int_{-\infty}^{\infty} r(t-\theta) h_{\text{opt}}(\theta)\, d\theta\right) r(\eta)\right] = 0$$

עבור כל η, $-\infty<\eta<\infty$ ולכן:

$$\int_{-\infty}^{\infty} R_{r,X}(t-\theta-\eta)\, d(\theta)\, d\theta = \int_{-\infty}^{\infty} R_r(t-\theta-\eta) h_{\text{opt}}(\theta)\, d\theta$$

נציג $\tau = t - \eta$ ונבצע התמרת פוריה על שני אגפי המשוואה, נקבל:

$$S_{r,X}(f)D(f) = S_r(f)H_{\text{opt}}(f)$$

ולכן:

$$H_{\text{opt}}(f) = \frac{D(f)S_{r,X}(f)}{S_r(f)}$$

<u>הערה 1:</u> שים לב כי לא השתמשנו בקשר $r = X + n$, ולמעשה פתרנו את המקרה הכללי בו r, X ת"א סמ"ר במשותף.

<u>הערה 2:</u> במקרה הפרטי בו $r = X + n$ והאותות $n(t)$ ו- $X(t)$ חסרי קורלציה מתקיים $S_{r,X}(f) = S_X(f)$ וכן $S_r(f) = S_X(f) + S_n(f)$ ומקבלים את התוצאה (5.43) שהשגנו על ידי ניתוח במישור התדר.

לכאורה הניתוח על ידי עקרון ההשלכה הוא כללי יותר, ובנוסף הוא מאפשר לקחת בחשבון אילוצים בתחום הזמן. אולם לעיתים בבעיות הנדסיות קיימים אילוצים בתחום התדר (למשל---לא ניתן ליישם מגבר עם הגבר גדול מסף נתון, כאשר ההגבר מוגדר דרך תגובת התדר). בבעיות כאלו נוח יותר הניתוח במישור התדר.

לסיום, חשוב להדגיש כי הניתוח והשיטות שפיתחנו בפרק זה, עבור תהליכים בזמן רציף, ישימים גם בזמן בדיד (עם מספר שינויים קלים), אך הפרטים אינם חלק מהקורס הנוכחי.

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

$$\frac{1}{N}\sum_{i=1}^{N} f\Big(X_i(\omega)\Big) \underset{N\to\infty}{\longrightarrow} \mathbb{E}\Big[f(X_1)\Big] \tag{5.44}$$

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.

<u>טענה:</u>

$$\mathbb{E}\left(\frac{1}{N}\sum_{1}^{N} X_i(\omega) - \mathbb{E}[X_1]\right)^2 \underset{N\to\infty}{\longrightarrow} 0$$

<u>הוכחה:</u> צ"ל

$$\mathbb{E}\left(\frac{1}{N}\sum_{1}^{N} Y_i\right)^2 \underset{N\to\infty}{\longrightarrow} 0$$

אולם

$$\mathbb{E}\left(\frac{1}{N}\sum_{1}^{N} Y_i\right)^2 = \frac{1}{N^2}\sum_{1}^{N}\mathbb{E}[Y_i^2] = \frac{1}{N}\,\mathbb{E}[Y_1^2] \to 0$$

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 $N\to\infty$.

שים לב שאם במקום לטפל ב-X_i נטפל ב-$f(X_i)$, ונדרוש $\mathbb{E}[|f(X_1)|^2]<\infty$, נקבל:

$$\frac{1}{N}\sum_{i=1}^{N}\Big(\mathbb{E}[f(X_i)]-\mathbb{E}[f(X_1)]\Big)\to 0$$

תוצאה זאת מהווה "פרוש" אפשרי ל- (5.44).

מתבקשת השאלה אם (5.44) נשאר נכון כאשר יש תלות בין האיברים השונים בסדרה האקראית.

<u>הגדרה:</u> 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 $t_1,t_2,\ldots,t_k$ מתקיים (בהסתברות 1):

$$\frac{1}{2T}\int_{-T}^{T}f\Big(X(t_1+s,\omega),\ldots,X(t_k+s,\omega)\Big)\,ds\underset{T\to\infty}{\longrightarrow}\mathbb{E}\,f\Big(X(t_1),\ldots,X(t_k)\Big)$$

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 שוה לתוחלת). את התכונה הארגודית נבטא ע"י הסיסמה: עבור תהליכים ארגודיים, ממוצע הזמן שוה לממוצע האנסמבל (=התוחלת).

<u>דוגמאות לתהליכים לא ארגודיים:</u>

1. משתנה אקראי $X_t(\omega)=C(\omega)$
2. $X_t(\omega)=A(\omega)\cos(2\pi f_0t+\phi(\omega))$ כאשר ϕ בלתי תלוי ב-A ומפולג אחיד בתחום $[0,2\pi]$. אם $A(\omega)$ מ"א לא מנוון אזי

$$\frac{1}{2T}\int_{-T}^{T}X_t^2dt\to\frac{A^2(\omega)}{2}\neq\mathbb{E}\left[\frac{A^2(\omega)}{2}\right]$$

אפשר להראות שעבור $A(\omega)\equiv\mathrm{const}$ התהליך אכן ארגודי.

3. יהיו

$$X_t(\omega)=\cos(2\pi f_0t+\phi)$$

$$Y_t(\omega)=\cos(2\pi f_0t+\psi)$$

כאשר ψ,ϕ בת"ס ומפולגים אחיד בתחום $[0,2\pi]$.

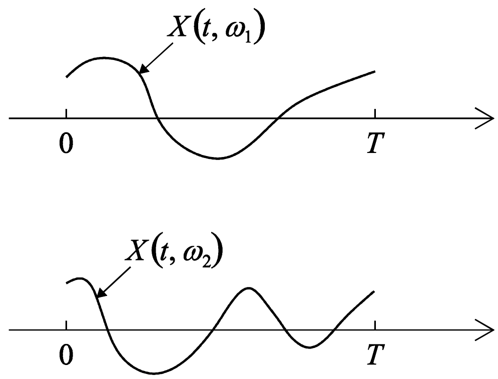
נגדיר

$$Z_t(\omega) \triangleq X_t(\omega)Y_t(\omega)$$

נקל לראות ש-$Z_t(\omega)$ איננו ת"א ארגודי.

4. תערובת של שני תהליכים שונים שכל אחד ארגודי לא חייבת להיות ארגודית.

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 5.36, אזי אם נסתכל על



איור 5.36: שתי פונקציות מדגם

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 הארגו-דיות:

<u>הגדרה:</u> ת"א $\{X(t), -\infty < t < \infty\}$ סטציונרי (או אפילו סטציונרי במובן הרחב) ייקרא ארגודי לגבי הממוצע אם

$$\mathbb{E}\left(\frac{1}{2T}\int_{-T}^{T} X(t)\,dt - \mu\right)^2 \underset{T\to\infty}{\longrightarrow} 0$$

כאשר $\mu = \mathbb{E}[X(t)]$, דהיינו, אם

$$\mathbb{E}\left(\frac{1}{2T}\int_{-T}^{T}\Big(X(t)-\mu\Big)\,dt\right)^2 \underset{T\to\infty}{\longrightarrow} 0\ .$$

<u>משפט:</u> אם הקווריאנס

$$\mathbf{K}_X(t_1,t_2) = \mathbb{E}\Big[(X(t_1)-\mu)(X(t_2)-\mu)\Big]$$

מקיים:

$$\frac{1}{4T^2}\int_{-T}^{T}\int_{-T}^{T}\mathbf{K}_X(t_1,t_2)dt_1dt_2 \underset{T\to\infty}{\longrightarrow} 0$$

אזי התהליך ארגודי לגבי הממוצע.

<u>הוכחה:</u> 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

$$\frac{1}{2T}\int_{-2T}^{2T}\psi(\tau)\left(1-\frac{|\tau|}{2T}\right)d\tau \underset{T\to\infty}{\longrightarrow} 0 \tag{5.45}$$

אזי התהליך ארגודי לגבי הממוצע. ההוכחה מתבססת על הקשר

$$\int_{-T}^{T}\int_{-T}^{T}Q(t_1-t_2)dt_1dt_2 = \int_{-2T}^{2T}Q(\theta)(2T-|\theta|)d\theta$$

<u>הערה:</u> ההוכחה לקשר זה מצויה סביב איור 5.19.

ברוח ההגדרה של ארגודיות לגבי הממוצע נגדיר גם:

<u>הגדרה:</u> יהיה $X(t)$ ת"א סטציונרי. נסמן

$$z_\lambda(t) = X(t+\lambda)X(t)$$

אם עבור כל $z_\lambda(t), \lambda$ ארגודי לגבי הממוצע אזי נגיד ש-$X(t)$ ארגודי לגבי הקורלציה. לכן, על מנת ש-$X(t)$ יהיה ארגודי לגבי הקורלציה צריך להתקיים

$$\mathbb{E}\left(\frac{1}{2T}\int_{-T}^{T}X(t+\tau)X(t)dt - R_X(\tau)\right)^2 \underset{T\to\infty}{\longrightarrow} 0$$

נסמן ב-$R_z(\tau)$ את פונקצית האוטוקורילציה של $z_\lambda(t)$:

$$R_z(\tau) = \mathbb{E}\Big[X(t+\tau+\lambda)X(t+\tau)X(t+\lambda)X(t)\Big]$$

אזי

$$\psi(\tau) = R_z(\tau) - R_X^2(\lambda)$$

צריך לקיים את התנאי (5.45) לעיל.

<u>הערה:</u> על מנת לדעת את $\psi(\tau)$ עלינו לדעת לא רק את המומנטים מסדר שני של $X(t)$ אלא גם את המומנטים מסדר רביעי. במקרה הגאוסי יש קשר בין השניים: מתקיים (כאשר $X(t)$ גאוסי 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 הקורילציה.

<u>משמעות הארגודיות לגבי מדידות</u>

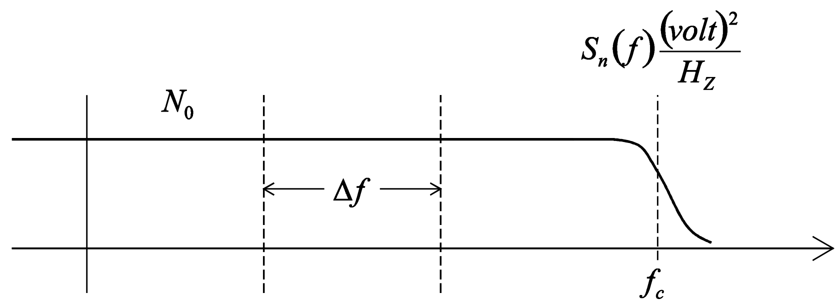
נתונה מערכת כגון מקלט המיוצר בכמויות. על מנת לבדוק את תגובת המקלטים לאותות המשודרים אליהם בתנאי רעש חיצוני, אנו בונים מערכת מדידה המכילה מקור המייצר את האותות המשודרים. המקור מורכב מגנרטור אות

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

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

$$\sqrt{R_{\text{opt}}(0)} \cong \sqrt{K^2 N_0 2\Delta F}$$

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 $R_{\text{opt}}(0)$. מקור רעש שהצפיפות הספקטרלית שלו היא N_0 בכל התדרים נקרא "רעש לבן".

שים לב שאם

$$S_n(f) = N_0$$

אזי

$$R_n(\tau) = N_0\delta(\tau)$$

ואז $R_n(0) \simeq \infty$. 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:

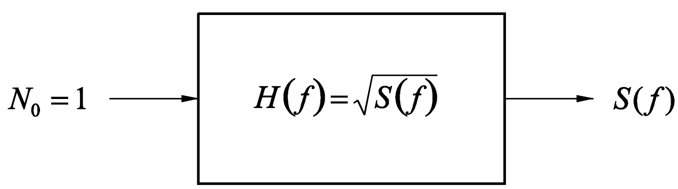
$$S(f)_{\text{חד-צדדי}} = \begin{cases} 0, & f < 0 \\ 2S(f), & f > 0 \end{cases}$$

ואז

$$\text{הספק היציאה} = N_0 \cdot \Delta f \cdot K^2_{\text{חד-צדדי}}$$

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

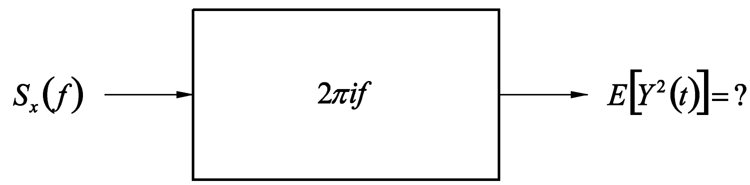
$$H(f) = \begin{cases} \dfrac{1}{1+i2\pi f} \\ \\ \dfrac{1}{1-i2\pi f} \\ \\ \dfrac{1}{\sqrt{1+(2\pi f)^2}} \end{cases}$$

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.

נעיין בגוזר



איור 6.4: גזירה של רעש

נעיין במקרים הבאים עבור $S(f)$

$$S(f) = \begin{cases} N_0 \\ \\ \dfrac{1}{1+f^2} \\ \\ \dfrac{1}{1+f^4} \end{cases}$$

אילו ממקורות אלה אפשר לגזור?

מוסר ההשכל הוא שאם מדובר בגוזר הרי שהקרוב של רעש לבן איננו תופס.

<u>אינטגרלים של רעש לבן</u>

יהיה $\{n(t), -\infty < t < \infty\}$ רעש עם צפיפות ספקטרלית N_0 ותהינה $h_1(t), h_2(t), \ldots$ פונקציות דטרמיניסטיות בעלות אנרגיה סופית $\int_{-\infty}^{\infty} h_i^2(t)dt < \infty, i = 1, 2, \ldots$.

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 מאד")

$$
\begin{aligned}
E[X_i] &= 0 \\
E[X_i, X_j] &= E\left[\int_{-\infty}^{\infty} n(t)h_i(t)dt \cdot \int_{-\infty}^{\infty} n(s)h_j(s)ds\right] \\
&= \iint_{-\infty}^{\infty} h_i(t)(t)h_j(s)E\Big[n(t)n(s)\Big]dt\,ds \\
&= N_0 \iint_{-\infty} h_i(t)h_j(s)\delta(t-s)dt\,ds \\
&= N_0 \int_{-\infty}^{\infty} h_i(t)h_j(t)dt
\end{aligned}
$$

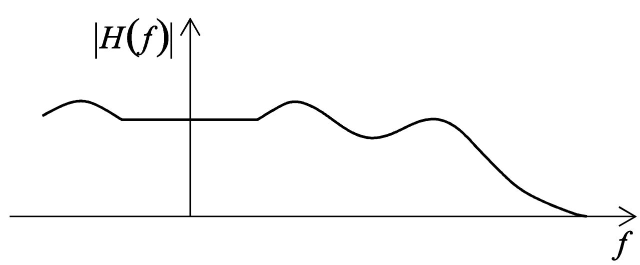
ובמיוחד, אם $n(\cdot)$ רעש לבן גאוסי (X_1, X_2) וקטור אקראי <u>גאוסי</u> עם תוחלת אפס ומטריצת קורלציה

$$
\begin{aligned}
E[X_1^2] &= N_0 \int_{-\infty}^{\infty} h_1^2(s)ds \\
E[X_2^2] &= N_0 \int_{-\infty}^{\infty} h_2^2(s)ds \\
E[X_1 X_2] &= N_0 \int_{-\infty}^{\infty} h_1(s)h_2(s)ds
\end{aligned}
$$

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?

6.2 רוחב סרט אפקטיבי לרעש

נתונה פונקצית תמסורת $H(f)$ מסוג Low-Pass דהיינו $H(f)$ כמעט אפס בתדרים גבוהים ושונה מאפס בתדרים נמוכים כמתואר באיור 6.20.



איור 6.5: פונקצית תמסורת מעבירה נמוכים מעשית

רוצים להגדיר "רוחב סרט אפקטיבי" לפונקצית תמסורת כזו. הגדרה מקובלת היא ההגדרה הבאה:

$$\Delta F \triangleq \frac{\displaystyle\int_{-\infty}^{\infty} |H(f)|^2 df}{2|H(0)|^2}$$

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.

<u>דוגמא:</u>

$$H(f) = \frac{1}{1 + i\frac{f}{f_0}}$$

אזי f_0 הוא רוחב הסרט של $3\,\mathrm{db}$. $H(0) = 1$ ולכן

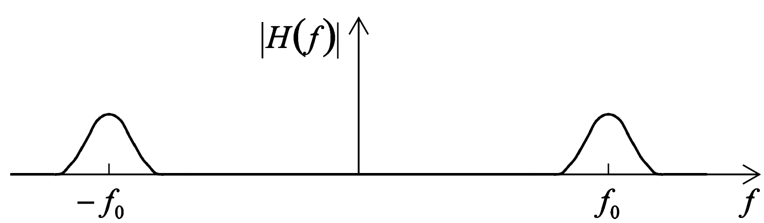
$$\Delta F = \frac{1}{2}\int_{-\infty}^{\infty} |H(f)|^2 df = \frac{f_0}{2}\int_{-\infty}^{\infty} \frac{df/f_0}{1 + (f/f_0)^2} = \frac{\pi}{2} f_0$$

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:

$$\Delta F = \frac{\displaystyle\int_{-\infty}^{\infty} |H(f)|^2 df}{2|H(f_0)|^2}$$

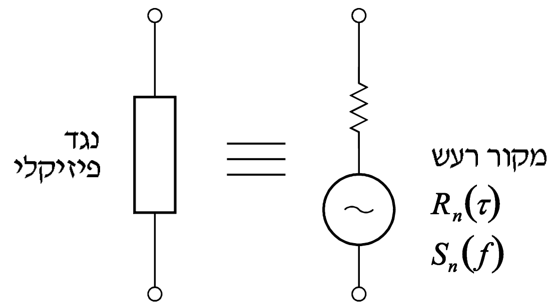
הגדרה זו נוחה במיוחד לפונקציות תמסורת (כולל כמובן מגברים) מטפוס Band Pass, דהיינו לפונקציות תמסורת מהטיפוס:



איור 6.6: מסנן מעביר סרט

6.3 רעש טרמי (רעש הנגד, רעש Nyquist)

בכל נגד פיזיקלי ישנם אלקטרונים הנמצאים בתנועה אקראית. אלקטרונים אלה חייבים להיות בנגד על מנת שהנגד יוכל אכן להעביר זרם. תנועת האלקטרונים האקראית בנגד יוצרת מתחים על פני הנגד שהממוצע שלהם אפס אולם המתח הרגעי הוא מתח רעש שאיננו אפס. לכן עלינו לראות כל נגד פיזיקלי כמורכב מנגד אידיאלי בצרוף מקור רעש:

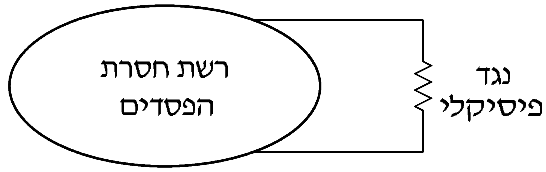


איור 6.7: מודל לנגד רועש

בתנאי "שיווי משקל" עם הסביבה נוכל להניח שרעש הנגד הוא תהליך אקראי סטציונרי והבעיה היא למצוא את $S_n(f)$. באופן כללי $S_n(f)$ 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:

<u>משפט א'</u>: רשת חסרת הפסדים איננה יוצרת רעש.

<u>הוכחה</u>: נעיין ברשת חסרת הפסדים המחוברת לנגד:



איור 6.8: 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 התרמודינימיקה. לכן מצב כזה בלתי אפשרי ולכן רשת חסרת הפסדים איננה יכולה לייצר רעש.

<u>משפט ב'</u>: עבור מקור הרעש הטורי עם הנגד:

$$S_n(f) = 2kTR$$

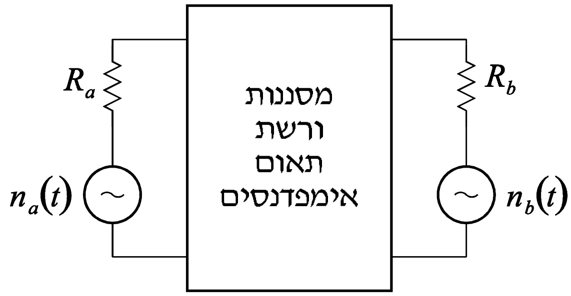
עד לתדירויות "גבוהות מאוד" כאשר k היא הקונסטנטה של בולצמן $1.38\cdot 10^{23}\ \mathrm{Joul}/^oK$. (הערה: "הנוסחה המעשית" היא $S_n(f) = 4kTR$ חד-צדדי ; מכל מקום, תמיד נכון ש-$4kTR\Delta f$ = {ממוצע המתח בריבוע בתחום התדרים Δf).

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.

<u>הוכחה חלקית</u>: 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 6.9: מעגל לחישוב רעש נגד

נשרטט מערכת שקולה פשוטה יותר, כדלקמן:



איור 6.10: מעגל מפושט לחישוב רעש נגד

כאשר $n_b(t), n_a(t)$ הם הרעשים הרגעיים של הנגד אחרי מעבר דרך מסננת צרת סרט, דהיינו:

$$E\left[n_a^2(t)\right] \approx S_n^a(f_0) \cdot 2\Delta f$$

$$E\left[n_b^2(t)\right] \approx S_n^b(f_0) \cdot 2\Delta f$$

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):

$$\frac{1}{2} n_a(t) \frac{n_a(t)}{2R_a}$$

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

$$\frac{S_n^a(f_0)}{R_a} = \frac{S_n^b(f_0)}{R_b}$$

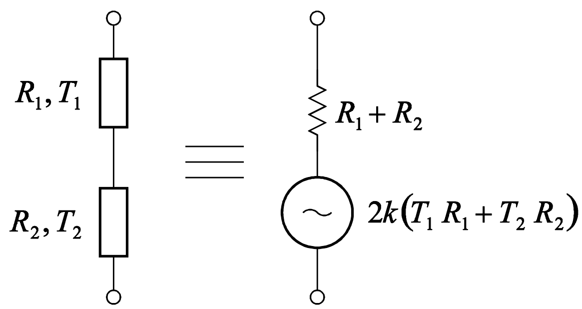
<u>מסקנה:</u> היחס $S_n(f)/R$ הוא יחס אוניברסלי שיכול להיות תלוי בתדר f או בטמפרטורה T אבל איננו יכול להיות תלוי בסוג הנגד (ז.א. לא תלוי בחומר ובגיאומטריה). נקבל ללא הוכחה שע"י אנליזה של נגד עם גיאומטריה פשוטה אפשר להראות

$$S_n(f) = 2kTR$$

ובגלל השיקול שהבאנו, התוצאה מתקיימת עבור כל הנגדים. עד איזו תדירות התוצאה הזו נכונה? בתדרים מספיק גבוהים כאשר מימדי הנגד הם בסדר גודל של אורך הגל ממילא ה"נגד מפסיק להיות נגד". באופן בסיסי התוצאה נכונה עד קרוב לתדירויות אופטיות.

<u>הערות:</u>

א. שים לב שמתחי הרעש מסתכמים "על בסיס של הספקים" דהיינו:

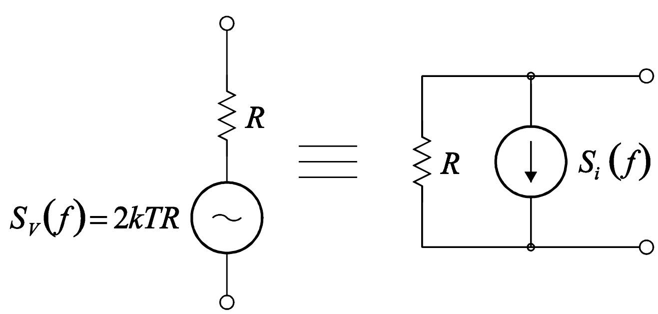


איור 6.11: חיבור נגדים בטור-מעגל שקול

ולכן, נגדיר טמפרטורה אפקטיבית של שני הנגדים כך שהן ההתנגדות השקולה והן הספק הרעש יקבלו את הערכים הנכונים, כלומר $2k(T_1R_1 + T_2R_2) = 2kT_{\text{eff}}(R_1 + R_2)$ או

$$T_{\text{eff}} = \frac{T_1R_1 + T_2R_2}{R_1 + R_2}$$

ב. מעגל תמורה מקבילי במקום מעגל תמורה טורי



איור 6.12: 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

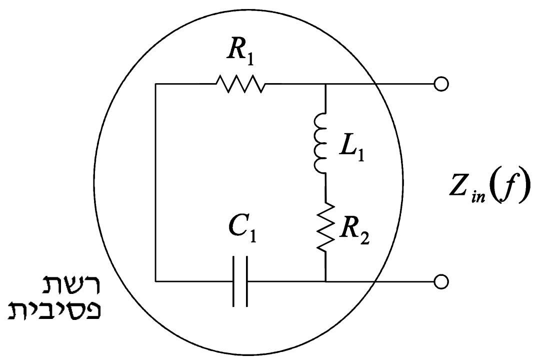
$$R^2 E\Big[i(t+\tau)i(t)\Big] = E\Big[v(t+\tau)v(t)\Big]$$

ומכאן

$$S_i(f) = \frac{S_v(f)}{R^2} = 2ktTG$$

(ועבור $S_v(f) = 4kTR$ נקבל $S_i(f) = 4kTG$ כאשר $G = R^{-1}$).

נעבור למשפט ג': נעיין ברשת פסיבית



איור 6.13: רעש של מעגל

ציור 6.12

הרשת נמצאת בשיווי משקל תרמודינמי עם סביבתה. מה 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:

<u>משפט ג'</u>: אם אימפדנס הכניסה של רשת (פסיבית הנמצאת בשיווי משקל תרמודינמי) הוא

$$Z_{\text{in}}(f) = R_{\text{in}}(f) + iX_{\text{in}}(f)$$

אזי הצפיפות הספקטרלית של מתח הרעש על פני הדקי המעגל יהיה:

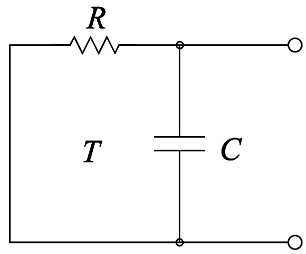
$$S(f) = 2kT\,R_{\text{in}}(f)$$

וסה"כ הספק הרעש יהיה $\overline{v}^2 = \int_{\infty}^{\infty} 2kTR_{\text{in}}(f)df$

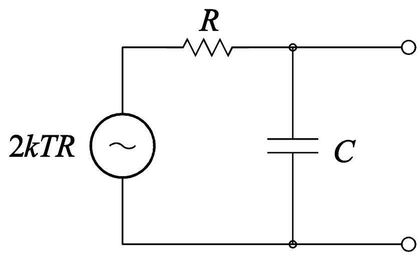
<u>הוכחה</u>: נחבר את המעגל הנתון עם נגד R באמצעות מסננת צרת סרט מסביב ל-f_0 ועם תאום אימפדנסים. השיקול של שיווי משקל תרמודינמי (ההספק הנמסר שווה להספק המתקבל) נותן לנו מיד את התוצאה המבוקשת.

<u>דוגמא</u>: נעיין במעגל

נוכל לחשב את הרעש על פני ההדקים ע"י מעגל התמורה:



איור 6.14: רעש במעגל פשוט



איור 6.15: מעגל תמורה - רעש במעגל פשוט

וכאשר $H(f)$ פונקצית 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

$$Z_{\text{in}}(f) = \frac{R \cdot \frac{1}{i2\pi f C}}{R + \frac{1}{i2\pi f C}} = \frac{R(1 - i2\pi f RC)}{1 + (2\pi f RC)^2}$$

ולכן

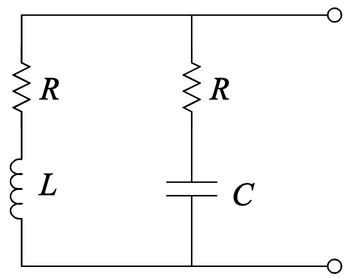
$$R_{\text{in}}(f) = \frac{R}{1 + (2\pi f RC)^2}$$

ומתקבלת, כמובן אותה תוצאה.

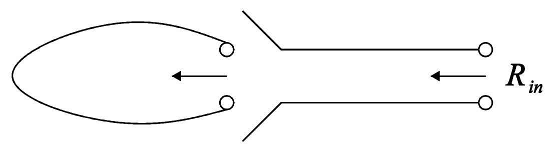
<u>תרגיל:</u>

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.

<u>הערת סיום לסעיף זה:</u> 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 6.16: מעגל טורי-מקבילי



איור 6.17: אנטנה כיוונית

נניח שאימפדנס הכניסה לאנטנה הוא אוהמי טהור, $R_{\rm in}$. נניח שהאנטנה צופה אל אחד הכוונים הבאים:

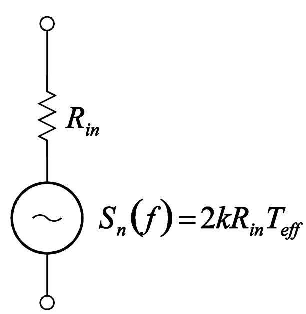
(א) החלל (בניצב לשביל החלב) - (**בסדר גודל של** $1000^o\,\mathrm{K}$ **ב**-$100\,\mathrm{MHz}$, **יורד לכ**-$5^o\,\mathrm{K}$ **ב**-$1000\,\mathrm{MHz}$).

(ב) **שביל החלב** - (**בסדר גודל של** $4000^o\,\mathrm{K}$ **ב**-$100\,\mathrm{MHz}$, **יורד לכ**-$15^o\,\mathrm{K}$ **ב**-$1000\,\mathrm{MHz}$).

(ג) **כדור הארץ** - (כ-$300^o\,\mathrm{K}$)

(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 6.18: מודל של אנטנה כיוונית

כאשר בכל אחד מהמקרים יש להציג את הטמפטורות המתאימות.

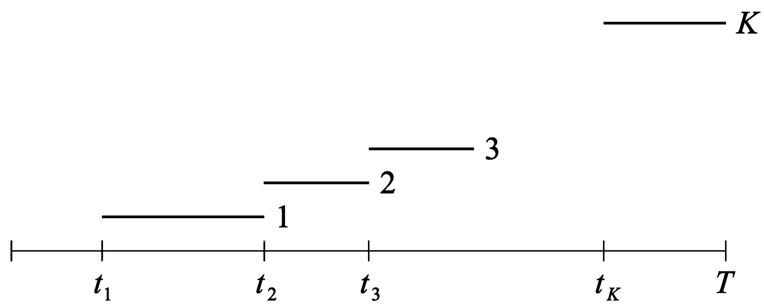
6.4 רעש הדיודה (Shot noise)

לפני שנתיחס 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.

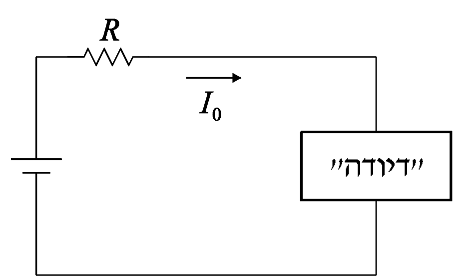
נבצע סימולציה של תהליך אקראי (על מחשב) כדלקמן: נקבע T מסוים, נגריל מספר אקראי K לפי פילוג פואסון עם תוחלת λT. אחרי שהגרלנו את K נבחר באקראי K זמנים ב-$[0, T]$, בלתי תלויים ומפולגים בצורה אחידה בתחום הזה. נסמן את הזמנים האלה $(\theta_1, \dots, \theta_K)$, נסדר את הזמנים בסדר עולה ונסמנם מחדש $(t_1, \dots, t_K)$ (כעת $t_1 \leq t_2 \leq \dots \leq t_k$) ונבנה מהם פונקצית מדרגות עולה כדלקמן:



איור 6.19: תהליך פואסון

<u>טענה</u> (ללא הוכחה): 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-$(0, T)$).

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 עוברים מהקטודה לאנודה.



איור 6.20: מקור רחב סרט

ציור 6.17

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

$$S(f) = I_0^2 \delta(f)$$

מאחר והזרם נגרם ע"י אלקטרונים בעלי מטען סופי נצפה לצפיפות ספקטרלית מהצורה

$$S(f) = I_0^2 \delta(f) + S_n(f)$$

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 נפלטים 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:

$$E\Big[I(t)\Big] = E\left[E\left[\sum_{k=1}^{K} i_e(t - t_k)\middle| K\right]\right] = E\left[K \frac{1}{T}\int_0^T i_e(t-\theta)d\theta\right] = \frac{q}{T}E[K] = \frac{q}{T}\frac{I_0}{q}T = I_0$$

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:

$$R(\tau) = E\Big[I(t+\tau)I(t)\Big] = E\left[\sum_{k=1}^{K} i_e(t+\tau-t_k)\sum_{j=1}^{K} i_e(t-t_j)\right]$$
$$= E\left[\sum_{k=1}^{K} i_e(t+\tau-t_k) i_e(t-t_k)\right] + E\left[\sum\sum_{k\neq j} i_e(t+\tau-t_k) i_e(t-t_j)\right]$$

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)

$$R(\tau) = E\left[\sum_1^K \frac{1}{T}\int_{-T/2}^{T/2} i_e(t+\tau-\theta) i_e(t-\theta)d\theta\right] + E\left[\sum\sum_{k\neq j}\frac{1}{T^2}\int_{-T/2}^{T/2} i_e(t+\tau-\theta)d\theta \int_{-T/2}^{T/2} i_e(t-\eta)d\eta\right]$$
$$\cong \frac{E[K]}{T}\int_{-\infty}^{\infty} i_e(\tau+\theta) i_e(\theta)d(\theta) + \frac{E[K^2-K]}{T^2}q^2$$

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 T זמן המעבר של אלקטרון מהקטודה לאנודה, עבור תדרים f המקיימים $\tau f \ll 1$ יתקיים $G_e(f)\cong G_e(0)$ 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.

<u>הערות:</u>

(א) שים לב שבעזרת התוצאה שקבלנו וע"י מדידת רעש הדיודה אפשר למצוא באופן נסיוני את מטען האלקטרון q.

(ב) בספרי עזר מוצאים הנוסחא $2I_0q$ 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,

$$I=I_0(e^{q\,V/KT}-1)$$

ולמעשה, ניתן למדל את הזרמים בדיודה כך:

זרם סחיפה $\xrightarrow{I_0\,e^{q\,V/kT}}$

זרם דיפוזיה $\xleftarrow{-I_0}$

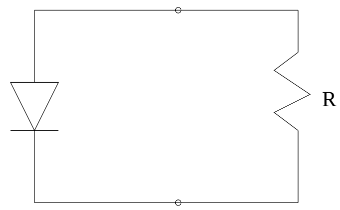
זרם הסחיפה וזרם הדיפוזיה הם זרמים פואסניים, בת"ס. לכן,

$$S_n(I)=qI_0(1+e^{q\,V/kT})$$

ועבור $V=0$,

$$\boxed{S_n(I) = 2qI_0} \tag{6.1}$$

נחבר כעת את המעגל ראשית, נשים לב כי עבור הדיודה,



$$\frac{\partial I}{\partial V} = \frac{qI_0}{kT} \cdot e^{q\,V/kT}$$

ולכן, סביב מתח 0, מתנהגת הדיודה כמו נגד:

$$\boxed{\left.\frac{\partial I}{\partial V}\right|_{V=0} = \frac{qI_0}{kT} = \frac{1}{R}} \tag{6.2}$$

לכן, עם בחירת R כזה, מתקיים תאום אמפדנסים. לכן,

$$2qI_0 = S_n^{R,I}$$

ומכאן

$$S_n^{R,I} = \frac{2\,kT}{R}$$

ומכאן, במעבר לרעש מתח,

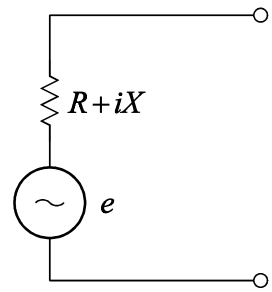
$$\boxed{S_n^{R,V} = 2kTR} \tag{6.3}$$

6.6 אפיון רעש מגבר

נניח שנתונים שני מגברים, האחד עם הגברה גבוהה ורעש גבוה ביציאתו והשני עם הגברה נמוכה ורעש נמוך ביציאתו. איזה מהם אחד לאחר השני איזה משניהם עדיף לשים לפני השני? בבעיות מסוג זה נעסוק בסעיף הנוכחי.

<u>הספק מצוי, הגבר הספק</u>

נעיין בתדר מסוים ונגדיר: הספק מצוי של המקור = הספק מכסימלי שאפשר לקבל מהמקור. הספק זה יתקבל בתנאי התאמת אימפדנסים ולכן עבור איור 6.21 יתקבל ההספק המצוי:



איור 6.21: מעגל והרעש 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.

<u>הגדרות:</u>

א. הטמפרטורה האפקטיבית של מקור כלשהוא (פסיבי או אקטיבי)

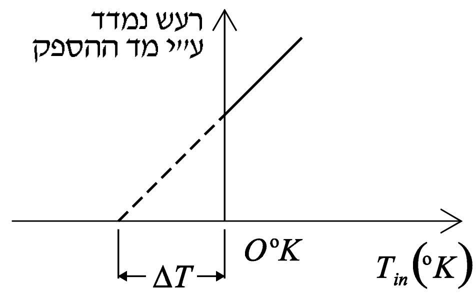
$$T_{\text{eff}} = \frac{1}{k \cdot \Delta f} \cdot [\text{הספק הרעש המצוי ברתב סרט } \Delta f]$$

ב. הגבר הספק: $G_a = \dfrac{\text{הספק מצוי ביציאה}}{\text{הספק מצוי בכניסה}}$

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 6.22: רעש המגבר

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.



איור 6.23: רעש המגבר - מדידה

לאור ההגדרה, הספק הרעש המצוי ביציאת המגבר ברוחב סרט Δf והגבר G יהיה:

$$G \cdot k \cdot (T_{\text{in}} + \Delta T) \cdot \Delta f$$

לכן: אם המגבר מחובר דרך קו תמסורת ואנטנה חסרת הפסדים והאנטנה צופה ל- $T_{(\text{מקור})}$, אזי,

אם $T_{(\text{מקור})} \ll \Delta T$ אזי שיפור ב-ΔT ישפר את הקליטה.

אם $\Delta T \ll T_{(\text{מקור})}$ אזי שיפור ΔT לא ישפר למעשה.

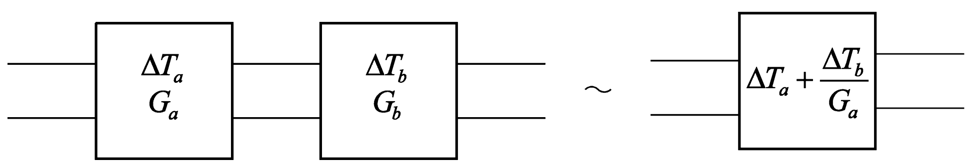
מספרי ΔT טיפוסיים:

מקלט מיקרוגלים לא מעולה $\Delta T \cong 10 \times \mathrm{T}_0 \quad T_0 = 300^o\,\mathrm{K}$

מקלט מיקרוגלים מעולע $-4 \times T_0$

מגבר פרמטרי $\Delta T \sim 80^o\mathrm{K}$

<u>טענה</u>: הטמפרטורה האפקטיבית הכוללת של שני מגברים בזה אחר זה (קסקדה):



איור 6.24: חיבור מגברים בטור

<u>הוכחה</u>: ראשית כפי שכבר הערנו, ההספק המצוי של הרעש ביציאה של מגבר (בודד), הוא $Gk\Delta f(T_{\text{מקור}} + \Delta T)$, (כאשר G - ההגבר המצוי ו- $T_{\text{מקור}}$ טמפרטורת הרעש של המקור). עבור קסקדה, הספק הרעש המצוי ביציאת מגבר a נתון ע"י $G_a k\Delta f(T_{\text{מקור}} + \Delta T_a)$, ולכן ביציאה של מגבר b יהיה הספק רעש מצוי:

$$G_b G_a k\Delta f(T_{\text{מקור}} + \Delta T_a) + G_b k\Delta f \Delta f \Delta T_b = G_a G_b k\Delta f\left(T_{\text{מקור}} + \Delta T_a + \frac{\Delta T_b}{G_a}\right)$$

ולכן:

$$\Delta T_{ab} = \Delta T_a + \frac{\Delta T_b}{G_a}$$

שים לב שמהנוסחא נוכל להחליט אם עדיף לשים את a לפני b או את b לפני a.

מאפין נוסף לרעש המגבר הוא ספרת הרעש F המוגדרת כדלקמן: נחבר לכניסת המגבר נגד בטמפרטורת הסביבה (300^oK). ההספק המצוי של המקור יהיה כמובן $300kG\Delta f$. נגדיר כעת:

$$F = \frac{\text{הספק הרעש המצוי ביציאת המגבר (כולל רעש המגבר)}}{300k\Delta f G}$$

ולכן F קשור ל ΔT ע"י

$$F = 1 + \frac{\Delta T}{300}$$

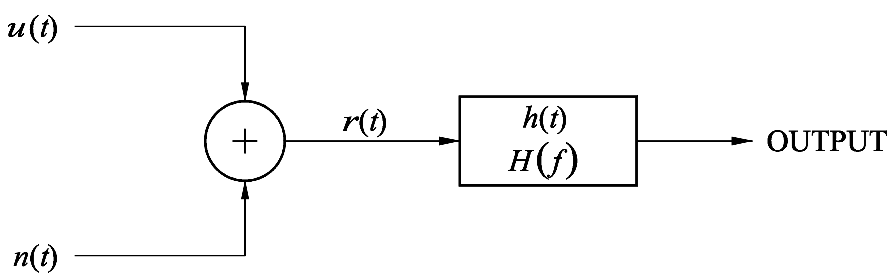
עבור מגבר אידיאלי שאינו יוצר רעש $\Delta T = 0$ ו-$F = 1$. שים לב ש-ΔT מאפיין את רעש המגבר ע"י <u>תוספת</u> הרעש שגורם המגבר, ואילו F מאפיין את רעש המגבר ע"י <u>מקדם כפל</u> לרעש המקור, כאשר המקור בטמפרטורה של 300^oK. מכיון ש-F מוגדר ביחס לטמפרטורה שרירותית ואילו הגדרת ΔT אינה דורשת נקודת יחוס שרירותית, ΔT הוא מושג בסיסי יותר).

<u>תרגיל:</u> הוכח שעבור חיבור שני מגברים בקסקדה

$$F_{ab} = F_a + \frac{F_b - 1}{G_a}$$

6.7 מסננת מתואמת Matched Filter

<u>אות הכניסה:</u> 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 6.25



איור 6.25: אות ורעש

<u>רעש הכניסה:</u> 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:

$$\left(\frac{S}{N}\right)_{\substack{\text{out}\\t=0}} = \frac{(Y(0))^2}{\text{הספק רעש ממוצע ביציאה}}$$

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

$$\left(\frac{S}{N}\right)_{\substack{\text{out}\\t=0}} = \frac{\left[\int_{0}^{\infty} u(-\theta)h(\theta)d\theta\right]^2}{N_0 \int_{0}^{\infty} h^2(\theta)d\theta} = \frac{\left[\int_{0}^{\infty} U^*(f)H(f)df\right]^2}{N_0 \int_{0}^{\infty} |H(f)|^2 df}$$

הבעיה: מצא מערכת לינארית (מאופיינת ע"י $h(\theta)$ או $H(f)$) כך שהיחס $\left(\frac{S}{N}\right)_{\substack{\text{out}\\t=0}}$ יהיה מכסימלי.

תזכורת: אי השיוויון של שוורץ (לפונקציות קומפלקסיות)

$$\left(\int_{T_1}^{T_2} f_1(t)f_2^*(t)dt\right)^2 \le \int |f_1(t)|^2 dt \cdot \int |f_2(t)|^2 dt$$

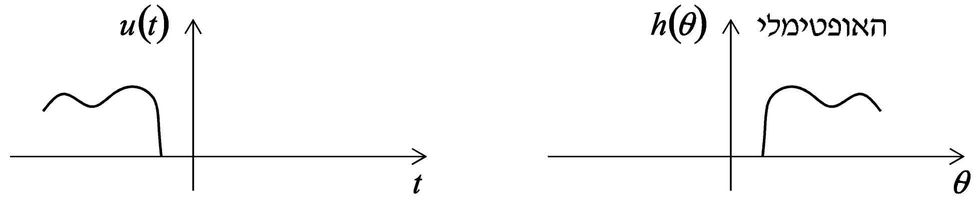
ואי השיוויון הופך לשיוויון כאשר $f_2(t) = \alpha f_1(t)$ (α מקדם פרופורציה שרירותי). לכן, נקבל מאי השיוויון של שוורץ ומהביטויים לעיל

$$\left(\frac{S}{N}\right)_{\substack{\text{out}\\t=0}} \le \frac{\int_{0}^{\infty} (u(-\theta))^2 d\theta \int_{0}^{\infty} h^2(\theta)d\theta}{N_0 \int_{0}^{\infty} h^2(\theta)d\theta} = \frac{1}{N_0}\int_{0}^{\infty} \Big(u(-\theta)\Big)^2 d\theta$$

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.



איור 6.26: אות ומסננת מתואמת

$h(t)$ האופטימלי נקרא: המסננת המתואמת.

הערה: בעמוד הקודם מופיע ביטוי במרחב הזמן וביטוי במרחב התדר. במקום להשתמש בביטוי במרחב הזמן, השתמש בבטוי 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 $\alpha u(-t)$.

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 <u>מרחב המדגם</u> 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 <u>אוסף המאורעות</u> $\mathbb{F}$ הוא אוסף של תת קבוצות של Ω, המקיימות את התנאים הבאים.

1. $\Omega \in \mathbb{F}$, כלומר המקרה שתמיד קורה הוא מאורע, או-תת 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 $i = 1, 2, \ldots$ אזי גם $\cup_{i=1}^{\infty} A_i \in \mathbb{F}$.

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 <u>הקבוצה הריקה</u>, 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 פונקצית הסתברות, או הסתברות $\mathbb{P}\{A\}$ היא פונקציה המוגדרת עבור כל מאורע (ב- $\mathbb{F}$), ומקיימת את התנאים הבאים:

1. $0 \le \mathbb{P}\{A\} \le 1$ לכל $A \in \mathbb{F}$.
2. $\mathbb{P}\{\Omega\} = 1$.
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 $\int_0^1 n^2(t)\,dt \le 3$ 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 $\{\omega : X(\omega) \le a\}$ היא מאורע.

שים לב שמקובל להשמיט את המשתנה ω, כלומר מקובל לרשום X כאשר מתכוונים ל-$X(\omega)$. בצורה דומה, לפעמים נרשום בקיצור (למשל) $\{X(\omega) \le a\}$, כאשר אנו מתכוונים לתאר את המאורע $\{\omega : X(\omega) \le a\}$.

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 $Y = \max_{0 \le t \le 1} n(t, \omega)$.

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 פונקצית 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):

- פונקצית 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 (צייר את פונקציית הפילוג של קוביה).

הגדרה 8.13 אם קיימת פונקציה (אינטגרבילית רימן), שנסמן ב-$f_X(\theta)$, כך שלכל a

$$F_X(a) = \int_{-\infty}^{a} f_X(\theta)\, d\theta \tag{8.1}$$

אזי f_X תקרא פונקצית 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:

$$\underline{X}(\omega) = \{X_1(\omega), X_2(\omega), \ldots, X_N(\omega)\}'$$

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 פונקצית הפילוג של ווקטור אקראי $\underline{X}(\omega) = \{X_1(\omega), X_2(\omega), \ldots, X_N(\omega)\}'$ היא פונקציה של ווקטור ממשי $\underline{a} = \{a_1, a_2, \ldots, a_N\}'$, המקבלת ערכים בן 0 ו-1, והגדרתה

$$F_{\underline{X}}(\underline{a}) = \mathbb{P}\{X_1(\omega) \le a_1, X_2(\omega) \le a_2, \ldots, X_N(\omega) \le a_N\} \tag{8.2}$$

לווקטור 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 $\{\omega : X \le a\}$ ו-$\{\omega : Y \le b\}$ הם בת"ס. במילים אחרות, המשתנים האקראיים X, Y הם בת"ס אם לכל a, b מתקיים

$$\mathbb{P}\{X \le a, Y \le b\} = \mathbb{P}\{X \le a\} \cdot \mathbb{P}\{Y \le b\}$$

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 <u>התוחלת</u> של משתנה אקראי X נסמן באחת מהצורות $\mathbb{E}\, X = \overline{X} = m_X$. 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

$$-n = \alpha_1^{(n)} < \alpha_2^{(n)} < \cdots < \alpha_i^{(n)} < \alpha_{i+1}^{(n)} < \cdots < \alpha_{k(n)-1}^{(n)} < \alpha_{k(n)}^{(n)} = n$$

$$\max_i \alpha_{i+1}^{(n)} - \alpha_i^{(n)} \to 0$$

כאשר $n \to \infty$ (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

$$\mathbb{E}[X] = \lim_{n\to\infty} \sum_{i=1}^{k(n)} \alpha_i^{(n)} \left[F_X(\alpha_{i+1}^{(n)}) - F_X(\alpha_i^{(n)}) \right] \tag{8.3}$$

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** Stiltjes-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 פונקצית הפילוג של X.

לא לכל מ"א יש תוחלת סופית, ולא תמיד ניתן להגדיר תוחלת. לדוגמה, עבור מ"א עם פונקציות פילוג סגולי

$$f_X(\alpha) = \begin{cases} 0 & \text{אם } \alpha \geq 0 \\ \dfrac{2/\pi}{1+\alpha^2} & \text{אם } \alpha < 0 \end{cases}$$

$$f_Y(\alpha) = \frac{1/\pi}{1+\alpha^2}$$

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 אין-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:

$$I_A(\omega) = \begin{cases} 1 & \text{אם } \omega \in A \\ 0 & \text{אם } \omega \in A^c. \end{cases}$$

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: $\mathbb{E}\left[\left(X-\overline{X}\right)^n\right]$
- **שונות** (ווריאנס): $\mathrm{Var}(X)=\mathbb{E}\left[\left(X-\overline{X}\right)^2\right]$. **סטיית התקן** $\sigma_X=\sqrt{\mathrm{Var}(X)}$
- מ"א X נקרא **מנורמל** אם $\mathbb{E}[X]=0$ ו-$\mathbb{E}\left[X^2\right]=1$.

לפי ההגדרה, אם X הוא מ"א כל שהוא בעל מומנט שני סופי, אזי המ"א

$$Z=\frac{X-\mathbb{E}[X]}{\sigma_X} \tag{8.4}$$

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-$(\mathbb{E}[|X|^n])^{1/n}$ אינו עולה מהר מדי, אזי המומנטים מגדירים את חוק הפילוג באופן חד משמעי.

יהיו X, Y זוג משתנים אקראיים.

הגדרה 8.23 <u>הקווריאנס</u> של זוג מ"א מוגדר ע"י

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \overline{X})(Y - \overline{Y})]$$

אם $\mathrm{Cov}(X, Y) = 0$ נאמר שהמ"א הם <u>חסרי קורלציה</u>, או <u>בלתי מתואמים</u>, או <u>בלתי תלויים לינארית</u> (להבדיל מבלתי תלויים סטטיסטית!). <u>מקדם הקורלציה</u>, או <u>מקדם המיתאם</u> ρ בין המ"א מוגדר ע"י

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

אם X, Y בלתי תלויים לינארית אזי $\rho = 0$. הבה נראה כי $|\rho| \leq 1$. כיוון שמקדם המתאם אינו תלוי בממוצע, 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

$$\lambda^* = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(Y)}$$

אם נציב ערך זה, נקבל

$$\begin{aligned}
0 &\leq \mathbb{E}[(X - \lambda^* Y)^2] \\
&= \mathbb{E}[X^2] - 2\lambda^* \,\mathrm{Cov}(X, Y) + (\lambda^*)^2\, \mathbb{E}[Y^2] \\
&= \mathbb{E}[X^2] - \frac{(\mathrm{Cov}(X, Y))^2}{\mathrm{Var}(Y)}
\end{aligned}$$

ומכאן ש-

$$\mathrm{Var}(X)\,\mathrm{Var}(Y) \geq (\mathrm{Cov}(X, Y))^2$$

או $|\rho| \leq 1$ כנדרש.

כסיכום ביניים, נתאר קשר מפורסם בין מומנטים לבין הסתברויות. קשר זה מראה כיצד מומנטים יכולים לסייע לשם קירוב הסתברויות מסויימות.

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,

$$\mathbb{P}\{X \geq \alpha\} \leq \frac{\mathbb{E}[g(X)]}{g(\alpha)} \tag{8.5}$$

<u>הוכחה:</u> נשתמש בהוכחה זו בסימון הכללי (ראה 8.3) 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 אין-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,

$$\mathbb{P}\{|X| \geq \alpha\} = \mathbb{P}\left\{|X|^2 \geq \alpha^2\right\}$$

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

$$\begin{aligned}\mathbb{P}\{X \geq \alpha\} &\leq \frac{\mathbb{E}\left[e^{(\theta \cdot X)}\right]}{e^{\theta \cdot \alpha}} \\ &= \mathbb{E}\left[e^{\theta(X-\alpha)}\right]\end{aligned}$$

8.7 פונקציה אפיינית

הפונקציה האפיינית ϕ_X של מ"א X היא פונקציה של המשתנה הממשי ν. היא מוגדרת ע"י

$$\phi_X(\nu) = \mathbb{E}\left[e^{i\nu X}\right] = \int_{-\infty}^{\infty} e^{i\nu\alpha}\, dF_X(x).$$

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

$$\left|e^{i\alpha}\right| = 1$$

לכל α. אם קיימים מומטים מכל סדר, אזי ניתן לייצג את הפונקציה האפיינית על ידי טור חזקות

$$\phi_X(\nu) = \sum_{k=0}^{\infty} \frac{(i\nu)^k}{k!}\, \mathbb{E}\left[X^k\right]$$

כאשר (תחת תנאים טכניים)

$$\begin{aligned}
\phi_X(0) &= 1 \\
\left.\frac{\partial \phi_X(\nu)}{\partial \nu}\right|_{\nu=0} &= i\, \mathbb{E}[X] \\
\left.\frac{\partial^k \phi_X(\nu)}{\partial \nu^k}\right|_{\nu=0} &= (i)^k\, \mathbb{E}\left[X^k\right]
\end{aligned}$$

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

$$\mathbb{P}\{A \mid B\} = \frac{\mathbb{P}\{A \cap B\}}{\mathbb{P}\{B\}}$$

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 $\{A_k,\ k = 1, 2, \ldots, K\}$ המקיימים $\mathbb{P}\{A_k\} \neq 0$ מתקיים

$$\begin{aligned}
\mathbb{P}\{A_1 \cap A_2\} &= \mathbb{P}\{A_1 \mid A_2\} \cdot \mathbb{P}\{A_2\} \\
\mathbb{P}\{A_1 \cap A_2 \mid A_3\} &= \mathbb{P}\{A_1 \mid A_2 \cap A_3\} \cdot \mathbb{P}\{A_2 \mid A_3\} \\
\mathbb{P}\left\{\cap_{k=1}^{K} A_k\right\} &= \mathbb{P}\left\{A_1 \mid \cap_{m=2}^{K} A_m\right\} \cdot \mathbb{P}\left\{A_2 \mid \cap_{m=3}^{K} A_m\right\} \times \cdots \\
&\quad \times \mathbb{P}\{A_{K-2} \mid A_{K-1} \cap A_K\} \cdot \mathbb{P}\{A_{K-1} \mid A_K\} \cdot \mathbb{P}\{A_K\} \\
\mathbb{P}\left\{\cap_{k=1}^{K-1} A_k \mid A_K\right\} &= \mathbb{P}\left\{A_1 \mid \cap_{m=2}^{K} A_m\right\} \cdot \mathbb{P}\left\{A_2 \mid \cap_{m=3}^{K} A_m\right\} \times \cdots \\
&\quad \times \mathbb{P}\{A_{K-2} \mid A_{K-1} \cap A_K\} \cdot \mathbb{P}\{A_{K-1} \mid A_K\}
\end{aligned}$$

נניח כעת כי X,Y הם משתנים אקראיים, ונגדיר עבור β קבוע את המאורע $B=\{\omega : X(\omega)=\beta\}$. אזי ניתן לרשום

$$\mathbb{P}\{A \mid B\} = \mathbb{P}\{A \mid X(\omega)=\beta\} = \frac{\mathbb{P}\{A\cap B\}}{\mathbb{P}\{B\}}$$

אולם במיקרים רבים (למשל אם למשתנה X יש צפיפות), ההיסתברות 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

$$\mathbb{P}\{A \mid X(\omega)=\beta\} = \lim_{\epsilon\to 0}\mathbb{P}\{A \mid \beta\le X(\omega)\le \beta+\epsilon\}$$

אנו נניח שגבול זה תמיד קיים, ומגדיר הסתברות מותנית כך שכל התכונות מתקיימות.

הגדרה 8.31 <u>הפילוג המותנה</u> של משתנה Y בהנתן המשתנה X מוגדר כ-

$$F_{Y|X}(\alpha \mid \beta) = \mathbb{P}\{Y\le \alpha \mid X=\beta\}$$

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 אים המאורעות A,B הם בלתי תלויים אזי

$$\mathbb{P}\{A \mid B\} = \mathbb{P}\{A\} \tag{8.6}$$

לכן, אם המשתנים X,Y הם בלתי תלויים סטטיסטית, אזי

$$F_{Y|X}(\alpha \mid \beta) = F_Y(\alpha)$$

שתי הטענות נובעות מהגדרת הסתברות מותנית: אם A,B בת"ס אזי

$$\mathbb{P}\{A \mid B\} = \frac{\mathbb{P}\{A\cap B\}}{\mathbb{P}\{B\}} = \frac{\mathbb{P}\{A\}\mathbb{P}\{B\}}{\mathbb{P}\{B\}} = \mathbb{P}\{A\}$$

הטענה השניה נובעת מכך ומהגדרת הפילוג המותנה.

הגדרה 8.33 אם קיימת פונקציה f כך ש-

$$F_{Y|X}(\alpha \mid \beta) = \int_{-\infty}^{\alpha} f_{Y|X}(\theta \mid \beta)\,d\theta$$

אזי f תקרא <u>פילוג סגולי מותנה</u> או <u>צפיפות מותנית</u> 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:

$$\begin{aligned}
f_{Y|X}(\alpha \mid \beta) &= \lim_{\delta\to 0} \frac{F_{Y|X}(\alpha+\delta \mid \beta) - F_{Y|X}(\alpha \mid \beta)}{\delta} \\
&= \lim_{\delta\to 0} \frac{\mathbb{P}\{\alpha \le Y \le \alpha+\delta \mid X=\beta\}}{\delta} \\
&= \lim_{\delta\to 0} \left[\lim_{\epsilon\to 0} \frac{\mathbb{P}\{\alpha \le Y \le \alpha+\delta, \beta \le X \le \beta+\epsilon\}}{\delta \cdot \mathbb{P}\{\beta \le X \le \beta+\epsilon\}} \right] \\
&= \frac{\delta\cdot\epsilon\cdot f_{X,Y}(\alpha,\beta)}{\delta\cdot\epsilon\cdot f_X(\beta)} \\
&= \frac{f_{X,Y}(\alpha,\beta)}{f_X(\beta)}
\end{aligned}$$

לסיכום, אם יש צפיפות משותפת אזי

$$\boxed{f_{Y|X}(\alpha \mid \beta) = \frac{f_{X,Y}(\alpha,\beta)}{f_X(\beta)}} \tag{8.7}$$

ההיסתברות המותנית $\mathbb{P}\{Y \le \alpha \mid X=\beta\}$ וכן גם הפילוג המותנה $F_{Y|X}(\alpha \mid \beta)$ והצפיפות המותנית $f_{Y|X}(\alpha \mid \beta)$ 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

$$\mathbb{P}\{Y \le \alpha \mid X\} = \mathbb{P}\{Y \le \alpha \mid X=\beta\}|_{\beta=X}$$

מהגדרה זו נובע שהשוויון

$$F_{Y|X}(\alpha \mid X) = \mathbb{P}\{Y \le \alpha \mid X=\beta\}$$

יתקיים אם ורק אם ω מקיים $X(\omega)=\beta$. בצורה דומה יש להבין את המשמעות של $f_{Y|X}(\alpha \mid X)$.

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

$$\mathbb{E}[Y \mid X=\beta] = \int_{-\infty}^{\infty} \alpha\, dF_{Y|X}(\alpha \mid \beta) = \int_{-\infty}^{\infty} \alpha F_{Y|X}(d\alpha \mid \beta)$$

אים יש פילוג סגולי מותנה, אזי

$$\mathbb{E}[Y \mid X=\beta] = \int_{-\infty}^{\infty} \alpha f_{Y|X}(\alpha \mid \beta)\, d\alpha$$

עבור פונקציה כלשהיא g מתקיים

$$\mathbb{E}\left[g(Y) \mid X=\beta\right] = \int_{-\infty}^{\infty} g(\alpha) F_{Y|X}(d\alpha \mid \beta) = \int_{-\infty}^{\infty} g(\alpha) f_{Y|X}(\alpha \mid \beta)\, d\alpha$$

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.

$$\mathbb{E}\left[Y \mid X\right] = \mathbb{E}\left[Y \mid X=\beta\right]\big|_{\beta=X}$$

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 כדי שהתוחלת המותנית תהיה קיימת. בדומה לתוחלת הרגילה,

1. לכל מאורע A ולכל משתנה אקראי X,

$$\mathbb{E}[I_A \mid X] = \mathbb{P}\left\{A \mid X\right\}$$

2. אם $Y=C$ קבוע שאינו אקראי, אזי $\mathbb{E}[Y \mid X]=C$.

3. לינאריות: יהיו a,b זוג קבועים. אזי

$$\mathbb{E}\left[aZ+bY \mid X=\beta\right] = a\cdot\mathbb{E}[Z \mid X=\beta] + b\cdot\mathbb{E}[Y \mid X=\beta]$$

$$\mathbb{E}\left[aZ+bY \mid X\right] = a\cdot\mathbb{E}[Z \mid X] + b\cdot\mathbb{E}[Y \mid X]$$

4. אם $Z \ge Y$ (או $\mathbb{P}\left\{Z \ge Y \mid X=\beta\right\}=1$) אזי $\mathbb{E}[Z \mid X=\beta] \ge \mathbb{E}[Y \mid X=\beta]$.

התכונות הבאות הן מיוחדות לתוחלת המותנית.

5. $\mathbb{E}[X \mid X=\beta]=\beta$ ולכן גם $\mathbb{E}[X \mid X]=X$.

6. עבור פונקציות g,h כלשהן,

$$\mathbb{E}\left[g(X)\cdot h(Y) \mid X=\beta\right] = g(\beta)\,\mathbb{E}\left[h(Y) \mid X=\beta\right]$$

ובאופן כללי יותר, אם h תלוייה בשני המשתנים X,Y

$$\mathbb{E}\left[g(X)\cdot h(X,Y)\mid X=\beta\right]=g(\beta)\,\mathbb{E}\left[h(\beta,Y)\mid X=\beta\right]$$

לכן, לפי ההגדרות, מתקיים בשני המקרים בהתאמה

$$\mathbb{E}\left[g(X)\cdot h(Y)\mid X\right]=g(X)\,\mathbb{E}\left[h(Y)\mid X\right]$$

$$\mathbb{E}\left[g(X)\cdot h(X,Y)\mid X\right]=g(X)\,\mathbb{E}\left[h(X,Y)\mid X\right]$$

.7

$$\mathbb{E}\left[\mathbb{E}(Y\mid X)\right]=\mathbb{E}[Y]$$

ואם המשתנים X,Y בלתי תלויים סטטיסטית אזי

$$\mathbb{E}[Y\mid X]=\mathbb{E}[Y]$$

.8

$$\mathbb{E}\left[g(X)\cdot h(X,Y)\right]=\mathbb{E}\left[g(X)\cdot\mathbb{E}(h(X,Y)\mid X)\right]$$

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

$$\mathbb{E}\left[g(Z)\cdot h(Z,Y)\mid X=\beta\right]=g(\beta)\,\mathbb{E}\left[h(\beta,Y)\mid X=\beta\right]$$

לפי ההגדרה,

$$\begin{aligned}\mathbb{E}\left[g(Z)\cdot h(Z,Y)\mid X=\beta\right]&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(z)h(z,y)F_{Y,Z|X}(dz\,dy\mid\beta)\\&=\int_{-\infty}^{\infty}g(\beta)h(\beta,y)F_{Y,Z|X}(dy\mid\beta)\end{aligned}$$

מכיוון מהגדרת הפילוג המותנה המשותף (ראה בפרט את הנוסחה לצפיפות המותנית), הפילוג מתרכז ב-$X=Z$ כך שהצפיפות המותנית היא פונקצית 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

$$\begin{aligned}\mathbb{E}\left[g(Z)\cdot h(Z,Y)\mid X=\beta\right] &= g(\beta)\cdot\int_{-\infty}^{\infty} h(\beta,y)F_{Y\mid X}(dy\mid\beta)\\ &= g(\beta)\,\mathbb{E}\left[h(\beta,Y)\mid X=\beta\right]\end{aligned}$$

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.