פרק ה'

תהליכים מקריים

5.1 תהליכים מקריים

5.1.1 מבוא

בפרקים הקודמים עסקנו במשתנים מקריים בודדים או בקבוצות קטנות של משתנים מקריים. בפרק הנוכחי נרחיב את הדיון לטיפול בסדרות של משתנים מקריים, סדרה כזאת נקראת **תהליך מקרי** או **תהליך סטוכסטי** (stochastic process).

תהליך מקרי $\mathbf{X} = \{X_t, t \in T\}$ הוא סדרה של משתנים מקריים, דהיינו לכל אינדקס t ב- T, X_t הוא משתנה מקרי. פעמים רבות נתייחס לאינדקס t כמציין זמן ונכנה את הערך ש- X_t מקבל בשם המצב של התהליך בזמן t.
האינדקס t יכול להתאים ליחידות בדידות של זמן (**תהליך מקרי בזמן בדיד** discrete-time process) או לזמן רציף (תהליך מקרי בזמן רציף continuous-time process), וגם כל אחד מהמשתנים המקריים X_t עצמם יכולים להיות בדידים (שאז נאמר כי **מרחב המצבים** הוא **בדיד** (discrete state space) או רציפים. תהליכים מקריים בזמן בדיד מתאימים בין השאר לתיאור של סדרת תוצאות ניסויים ודוגמא לתהליך מקרי כזה היא למשל שערי מדד המניות בסופו של כל יום מסחר (משתנה מקרי בדיד) או הטמפרטורה היומית המקסימלית (משתנה מקרי רציף). תהליכים מקריים בזמן רציף מתאימים לתיאור של משתנים שנמדדים באופן רציף, למשל מתח הממברנה של תא כפונקציה של הזמן (משתנה מקרי רציף), או קיום פוטנציאל פעולה בתא כפונקציה של הזמן (משתנה מקרי בדיד בתהליך מקרי רציף).

דוגמא: הילוך מקרי

המשתנה המקרי X_{n+1} מתפלג על פי

$$X_{n+1} = \begin{cases} X_n + 1 & 0.5 \\ X_n - 1 & 0.5 \end{cases}$$

ההמחשה המקובלת לתהליך כזה היא אדם היוצא מבית המרזח בהיותו מבוסם כיאות ומתנדנד צעד לימין וצעד לשמאל בהסתברויות שוות. לעתים קל יותר לחשוב

על סדרת ההפרשים של התהליך השקולה לסדרה של ניסויי ברנולי במטבע הוגנת $X_{n+1} - X_n \sim U(\{-1,+1\})$.

תהליך סטציונרי

תהליך מקרי סטציונרי הוא תהליך בו ההתפלגות של המשתנה המקרי אינה תלויה מפורשות בזמן. באופן פורמלי תהליך הוא סטציונרי אם ההתפלגות המשותפת של $P(X_{t1+h}, X_{t2+h}, ..., X_{tn+h})$ זהה להתפלגות של $P(X_{t1}, X_{t2}, ..., X_{tn})$, לכל h ולכל בחירה של אינדקסים $t_1, ..., t_n$.

5.1.2 תהליכים פואסוניים

התהליך הפואסוני הוא אחת הדוגמאות הפשוטות ביותר לתהליך מקרי, אך גם שימושי ביותר. באופן בלתי פורמלי נספר כי תהליך מקרי פואסוני הוא תהליך מקרי בזמן רציף בו המשתנה המקרי X_t סופר את מספר הפעמים שאירוע מסויים קרה עד לזמן t, ואירועים כאלו מתרחשים בקצב קבוע ובאופן בלתי תלוי זה בזה. באופן לא מפתיע התפלגות המשתנה המקרי בזמן t היא התפלגות פואסונית.

לצורך הגדרה פורמלית יותר, נאמר כי תהליך X נקרא **תהליך ספירה** (counting process) אם X_t מייצג את מספר האירועים מסוג מסוים שקרו עד לזמן t. סדרת הערכים המתקבלת בתהליך כזה היא סדרה עולה עם הזמן, המתחילה בדרך כלל בערך $X_0=0$ וקופצת בערך של יחידה בכל פעם שארוע התרחש. תהליך ספירה הוא בהכרח א-סטציונרי, היות ונקודת ההתחלה שלו מוגדרת באופן ייחודי. נאמר שתהליך ספירה הוא תהליך בעל independent increments אם מספר האירועים שקורים בשני פרקי זמן לא חופפים הוא בלתי תלוי. נאמר שתהליך ספירה הוא בעל stationary increments אם התפלגות מספר האירועים שקורים בכל פרק זמן אינו תלוי בזמן האבסולוטי אלא רק במשך פרק הזמן. לבסוף, נזכיר כי פונקציה $f(h)$ נקראת $o(h)$ אם מתקיים

$$\lim_{h\to 0} \frac{f(h)}{h} \to 0$$

כעת נוכל להגדיר פורמלית

הגדרה: תהליך פואסוני

תהליך ספירה X_t נקרא פואסוני עם קצב λ אם הוא מקיים (בגבול של h קטן)

1. $X_0 = 0$
2. לתהליך יש independent and stationary increments
3. $P(X_t = 1) = \lambda h + o(h)$
4. $P(X_t = 2) = o(h)$

טענה: משתנה מקרי X_T בתהליך פואסוני מתפלג פואסונית

$$(5.1)\qquad P(X_T = k) = \frac{(\lambda T)^k}{k!} e^{-\lambda T}$$

כדי להוכיח טענה זו, נחלק את הזמן [0..T] ל- n פרקי זמן קצרים באורך T/n. כאשר פרקי זמן אלו קצרים מספיק, אזי לפי תכונה 3 ו- 4 בכל פרק זמן קצר מתרחש לכל היותר אירוע בודד, וההסתברות כי אירוע כזה יתרחש היא $\lambda T/n$. מספר האירועים עד לזמן T הוא לכן סכום של n ניסויי ברנולי (לפי תכונה 2) ולכן הוא מתפלג בינומית. עם זאת, אנחנו מתעניינים בהתפלגות המשתנה המקרי בזמן רציף שהוא הגבול של ההתפלגות הבינומית עבור n שואף לאינסוף והראנו בפרק 1 כי בגבול זה ההתפלגות הבינומית שואפת להתפלגות הפואסונית כנדרש.

זמן בין מאורעות Inter Events Intervals

יהי X תהליך מקרי פואסוני ונסמן ב- Y_1 את הזמן בו התרחש המאורע הראשון, וכן נסמן ב- Y_n את הפרש הזמנים בין התרחשות המאורע ה- n-1 והמאורע ה- n. הסדרה $\{Y_1,...,Y_n\}$ היא סדרת המרווחים בין ארועים Inter-Event-Intervals המכונה גם **זמני המתנה** waiting times (וכאשר האירועים הם פטנציאלי פעולה בתאי עצב אז מרווחים אלו נקראים Inter-Spikes Intervals).

טענה: זמני המתנה בתהליך פואסוני

זמני ההמתנה בתהליך פואסוני עם קצב λ הם בעלי התפלגות אקספוננציאלית עם ממוצע λ.

הוכחה

עבור המשתנה הראשון, מתקיים $Y_1>t$ רק אם לא התרחש שום אירוע בפרק הזמן [0,t] ומתוך משוואה (5.1) נובע כי $P(Y1 > t) = P(N(t) = 0) = e^{-\lambda t}$ (כאשר N(t) הוא מספר האירועים עד לזמן t). מתוך דרישת הנרמול לסכום הסתברויות אחת נובע כי Y_1 מתפלג אקספוננציאלית עם פרמטר λ. כעת עבור המשתנה השני

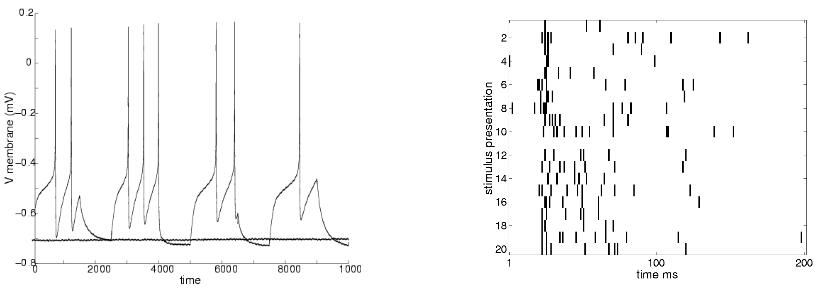
$$(5.2)\qquad \begin{aligned} &P(Y2 > t \mid Y_1 = s) = \\ &= P\{\text{no event occured in the interval } (s,s+t\,] | Y_1 = s\} \\ &= P\{\text{no event occured in the interval } (s,s+t\,]\} \\ &= e^{-\lambda t} \end{aligned}$$

כאשר המעבר הראשון נובע מתכונת האי תלות (independent increments) והמעבר השני מהסטציונריות (stationary increments). ובאופן דומה ניתן להמשיך לכל Y_n, וקיבלנו כי Y_n מתפלג אקספוננציאלית עם פרמטר λ כנדרש. נעיר כי ניתן להוכיח את הפילוג האקספוננציאלי של זמני ההמתנה באופן דומה להוכחת הטענה הקודמת על ידי חלוקת הקטע לקטעים אינפיטיסימליים.

5.2 שרשרות ירי עצביות כתהליכים מקריים

תאי עצב מגיבים באופן סטוכסטי לקלטים שהם מקבלים. לא נוכל להכנס כאן למגבלות שהתנהגות מקרית כזאת כופה על ביצוע חישובים על ידי תאי עצב, אך נתאר בקצרה מודל נפוץ, אם כי בהחלט לא מדוייק של שרשרות ירי עצביות כתהליכים מקריים.

ראשית נשים לב כי שרשרות ירי של תאי עצב הם תהליכים מקריים בזמן רציף, שניתן למדל אותם כ**תהליכים נקודתיים**, דהיינו סדרה של מאורעות נקודתיים בזמן (זאת אם נבחר עבור כל פוטנציאל פעולה רגע אחד שבו הוא מתרחש, למשל הרגע בו מתח הממברנה הוא מקסימלי). האיור הבא (שמאל) מתאר את מתח הממברנה במשך זמן בו הוסרק לתא זרם שגרם לפוטנציאלי פעולה. קל לראות כיצד ניתן לתרגם תהליך זה לתהליך נקודתי. האיור שמימין מתאר תגובות של תא עצב במערכת השמיעה לגירוי שהוצג עשרים פעמים. התגובה לכל אחת מהצגות הגירוי מוצגת בשורה נפרדת כתהליך נקודתי (raster plot).



רישום תוך תאי באפליזיה. נתונים של Chechik, Globerson and Jacobson 1999

הקצב שבו פוטנציאלי הפעולה בתא עצב מתרחשים אינו קבוע אלא תלוי בקלטים שהתא מקבל. נכליל אם כן את התהליך הפואסוני שהגדרנו לעיל לתהליך בו אנו מרשים לפרמטר הקצב λ להשתנות כפונקציה של הזמן $\lambda(t)$, ונגדיר תהליך פואסוני לא הומוגני בזמן.

הגדרה: תהליך פואסוני לא הומוגני (Inhomogeneous Poisson)

בדומה להגדרת תהליך פואסוני נאמר כי תהליך ספירה X_t הוא פואסוני לא הומוגני עם קצב $\lambda(t)$ אם הוא מקיים

1. $X_0=0$
2. לתהליך יש independent increments
3. $P(X_t=1)=\lambda(t)h+o(h)$
4. $P(X_t=2)=o(h)$

תהליך פואסוני לא הומוגני מתאים אם כן לתיאור התפלגות מספר פוטנציאלי פעולה שהתרחשו כפונקציה של הזמן, אם פוטנציאלי הפעולה מתרחשים באופן שאנו תלוי זה בזה, אך בקצב שיכול להשתנות בזמן. כמו במקרה ההומוגני, ניתן להראות כי התפלגות זמני ההמתנה של משתנה מקרי פואסוני לא הומוגני אף היא אקספוננציאלית, אך עם תוחלת התלויה באינטגרל על הקצב $\lambda(t)$.

תהליך מתחדש (renewal process)

ראינו כי זמני ההמתנה בתהליך פואסוני הם בלתי תלויים ומתפלגים התפלגות אקספוננציאלית. הרחבה מתבקשת של מודל זה הם תהליכים שבהם זמני ההמתנה בלתי תלויים אבל מתפלגים לפי פילוג כלשהו. תהליכים כאלה נקראים renewal processes. למשל, בקירוב לא רע, תעלות היונים בתא בעקבות פוטנציאל פעולה עוברות reset כך שהתא "שוכח" את המצב בו הוא היה לפני פוטנציאל הפעולה, ולכן זמני ההמתנה יהיו בלתי תלויים. מצד שני תאי עצב אינם נוטים לירות מספר פוטנציאלי פעולה סמוכים מאוד (תופעה הנובעת מהתקופה הרפרקטורית היחסית), ולכן תהליך שבו התפלגות זמני ההמתנה היא אקספוננציאלית אינה מתאימה לתיאור התנהגות של תאים בעלי תקופה רפקטורית משמעותית. תיאור טוב יותר ניתן להשיג למשל אם זמני ההמתנה מתפלגים התפלגות גאמא.

5.3 תהליכים מרקוביים

5.3.1 מבוא

נטפל כעת בתהליך מקרי X שיכול לקבל מספר סופי m של ערכים בכל צעד זמן, שאותם נסמן בפשטות ב- $\{1,2,\ldots,m\}$, ואם $X_n=i$ אז נאמר שהתהליך נמצא במצב i בזמן n. נניח כי בכל פעם שהמערכת נמצאת במצב i אז ישנה הסתברות קבועה לכך שהמערכת תמצא במצב j בצעד הזמן הבא

$$a_{ji} \equiv P\left(\text{transition from state } i \text{ to state } j\right) = \\ = P\left(X_n = j \mid X_{n-1} = i\right) \tag{5.3}$$

נשים לב כי הסתברות זו קבועה ואינה תלויה במצבים בהם היתה המערכת לפני זמן n-1. כלומר באופן פורמלי נוכל לרשום

$$P\left(X_n = j \mid X_{n-1} = i\right) = \\ = P\left(X_n = j \mid X_{n-1} = i,\, X_{n-2} = i_2,\, \ldots,\, X_0 = i_n\right) \tag{5.4}$$

נדגיש כי תכונה זו היא **אי-תלות מותנית**: אמנם קיימת תלות בין מצב המערכת בזמן n לשרשרת המצבים בזמנים הקודמים $1,\ldots n-1$, אבל אם ידוע לנו מצב המערכת בזמן $n-1$ אז המצב בזמן n כבר אינו תלוי במצב בזמנים $1,\ldots,n-2$. תכונה זאת של אי תלות מותנית נקראת התכונה המרקובית, ותהליך המקיים תכונה זאת נקרא **שרשרת מרקוב** (Markov Chain).

במקרה הכללי שרשרות מרקוב מוגדרות גם למערכת בעלת מספר אינסופי (אך בן מניה) של מצבים אפשריים, כך שמטריצת הסתברויות המעבר היא מטריצה ממימד אינסופי. בנוסף לכך מגדירים תהליך מרקובי מסדר K, אם מצב המערכת אינו תלוי בהסטוריה בהנתן K המצבים הקודמים.

תכונות מטריצת הסתברויות המעבר

נשים לב כי היות וישנה הסתברות של 1 לצאת ממצב i בכל צעד הרי שמתקיים

$$\sum_{j=1}^{n} a_{ji} = 1 \tag{5.5}$$

ואם נרשום את הסתברויות המעבר כמטריצה $\boldsymbol{A} = \{a_{ji}\}$ אזי הסכום של כל עמודה יסתכם לאחת. מתכונה זו נובע כי למטריצה A יש וקטור עצמי שמאלי שהערך העצמי שלו הוא 1, זאת היות ואם נסמן את וקטור האחדות $\overline{\mathbf{1}} = (1,1,\ldots,1)$ אז מתקיים

$$\overline{\mathbf{1}} \cdot \mathbf{A} = (1,1,\ldots,1) = \overline{\mathbf{1}} \cdot 1$$

היות והערכים העצמיים השמאליים שווים לערכים העצמיים הימניים, אז קיים גם וקטור עצמי ימני (שאותו נסמן $\mathbf{v}_1$) שלו מתאים ערך עצמי 1, וניתן להראות כי שאר הערכים העצמיים של A קטנים מ- 1.

5.3.2 סוגי מצבים

נאמר שניתן להגיע ממצב i אל מצב j אם קיימת הסתברות גדולה ממש מאפס לרצף מעברים בין מצבים שמתחיל במצב i ומסתיים במצב j. לדוגמא, במטריצת המעברים הבאה ניתן להגיע ממצב 1 למצב 4 דרך מעבר במצב 2 ו- 3 אך לא ניתן להגיע למצב 5. לעומת זאת ניתן להגיע בצעד אחד בהסתברות 0.5 למצב אחת מתוך מצב 5.

$$A_1 = \begin{matrix} 0.5 & 0.4 & 0 & 0.9 & 0.5 \\ 0.5 & 0.3 & 0 & 0 & 0 \\ 0 & 0.3 & 0.2 & 0.1 & 0 \\ 0 & 0 & 0.8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 \end{matrix}$$

נשים לב כי למעשה לא ניתן להגיע למצב 5 בתהליך זה משום מצב אחר, כך שאם נתחיל ממצב 5 לא נוכל לחזור אליו. תהליך נקרא recurrent אם לא קיימים בו מצבים מסוג זה, כלומר אם לכל מצב ההסתברות כי נחזור אליו בזמן סופי היא אחת.

שני מצבים i ו- j נקראים מתקשרים, אם ניתן להגיע ממצב אחד למשנהו וגם להפך (למשל מצבים 1 ו- 2 במטריצה לעיל). וניתן להראות בקלות כי יחס ההתקשרות הינו יחס שקילות. תהליך מרקובי נקרא **אי-פריק** (irreducible) אם קיימת בו מחלקת שקילות אחת בלבד, דהיינו אם ניתן להגיע מכל מצב לכל מצב במספר סופי של צעדים בעלי הסתברות חיובית. כדי להבהיר רעיון זה, נביט לדוגמה במטריצת המעברים הבאה המתארת תהליך מרקובי שבו שתי מחלקות שקילות

$$A_2 = \begin{matrix} 0.5 & 0.25 & 0.3 & 0 & 0 \\ 0.4 & 0.25 & 0.3 & 0 & 0 \\ 0.1 & 0.5 & 0.4 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{matrix}$$

בתהליך זה ישנן בברור שתי קבוצות של מצבים, שלא ניתן לעבור ביניהן: אם במקרה התחלנו במצב מבין שלושת המצבים הראשונים, הרי שבהכרח נשאר בקבוצה זאת, וכך גם אם התחלנו בשני המצבים האחרונים. בדוגמא זאת, סדרת המצבים של המערכת תלויה באופן חזק במצב ההתחלתי של המערכת. דוגמא

נוספת לתהליך בו סדרת המצבים תלויה במצב ההתחלתי, היא של **תהליך מחזורי**, כפי שמוצג במטריצה הבאה:

$$A_3 = \begin{matrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{matrix}$$

סדרת המצבים של תהליך זה היא בהכרח מהצורה $1,2,3,1,2,3,\ldots$ אך הפאזה של התהליך תלויה במצב ההתחלתי. ההגדרה הפורמלית של תהליך שאיננו מחזורי (a-periodic) מייגעת מעט, ולכן נסתפק כאן באינטואיציה הקובעת כי לא קיימים מחזורים בסדרת המצבים של התהליך.

בניגוד לדוגמאות לעיל בהן המצב ההתחלתי קובע באופן קריטי את סדרת מצבי המערכת, נראה כעת כי בתהליכים אי-פריקים ולא מחזוריים המערכת שוכחת את תנאי ההתחלה שלה לאחר זמן מספיק.

5.3.3 התפלגות במצב שווי משקל

נטפל בתהליך מרקובי בלתי פריק, שאינו מחזורי והוא recurrent ובו נסמן את התפלגות המשתנה המקרי X_n על ידי הוקטור $P^n = \left[P(X_n=1), P(X_n=2), \ldots, P(X_n=m)\right]$. נביט על התפלגות המצבים בצעד הזמן הבא. הסיכוי להיות במצב j בצעד הזמן הבא תלוי רק בצעד הזמן הקודם, ולכן

$$P(X_n=j) = \sum_{i=1}^{m} A_{ji} P(X_{n-1}=i) = \sum_{i=1}^{m} A_{ji} P_i^{n-1}$$

ובכתיב מטריציוני

$$P^n = \mathbf{A} P^{n-1}$$

ובאינדוקציה נקבל

$$P^n = \mathbf{A} P^{n-1} = \mathbf{A}\mathbf{A} P^{n-2} = \ldots = \mathbf{A}^n P^0$$

כדי לתאר את תוצאת ההפעלה החוזרת של המטריצה $\mathbf{A}$ על וקטור התפלגות התחלתי, קל יותר לעבור לבסיס הוקטורים העצמיים של המטריצה $\mathbf{A}$. כאשר מטריצת המעברים $\mathbf{A}$ ניתנת ללכסון, אזי ניתן לרשום אותה כמכפלת מטריצות מהצורה $\mathbf{A} = \mathbf{T}^{-1}\mathbf{\Delta}\mathbf{T}$, כאשר Δ היא מטריצה אלכסונית ובה הערכים העצמיים של A שאותם נסמן $\lambda_1, \lambda_2, \ldots, \lambda_m$. כשנעלה את המטריצה בחזקת n נוכל לרשום $\mathbf{A}^n = \mathbf{T}^{-1}\mathbf{\Delta}^n\mathbf{T}$ ומכאן שהערכים העצמיים של $\mathbf{A}^n$ הם $\lambda_1{}^n, \lambda_2{}^n, \ldots, \lambda_m{}^n$, וביניהם הערך העצמי הגדול ביותר הוא 1, ושאר הערכים העצמיים שואפים לאפס בקצב אקספוננציאלי כפונקציה של n. נרשום שוב את ההתפלגות לאחר n צעדים

$$P^n = \mathbf{A}^n P^0 = \mathbf{T}^{-1}\mathbf{\Delta}^n\mathbf{T} P^0$$

ומאלגברה בסיסית ידוע כי משמעות ביטוי זה היא כי הוקטור P^0, מוטל לבסיס הוקטורים העצמיים של $\mathbf{A}$, וכל רכיב בו מוכפל בערך העצמי המתאים של $\mathbf{A}^n$. היות וקיים רק ערך עצמי יחיד של $\mathbf{A}^n$ שאינו שואף לאפס, אזי

$$\lim_{n\to\infty} P^n = \mathbf{v}_1 \tag{5.6}$$

כאשר v_1 הוא הוקטור העצמי בעל הערך העצמי 1 של המטריצה $\mathbf{A}$. ההתפלגות P^n בגבול נקראת ההתפלגות הסטציונרית של התהליך המרקובי.

5.4 מודלים מרקובים חבויים

5.4.1 הקדמה

מודלים מרקובים חבויים Hidden Markov Models מהווים טכניקה סטטיסטית ללימוד של סדרות זמניות. המודלים הללו מאפשרים לעבוד עם רוב ההסתברויות הקלאסיות, מחיר המימוש שלהם הוא נמוך יחסית (לינארי באורך הנתונים) תכונות אלו ואחרות הפכו את המודלים מן הסוג הזה למושכים במיוחד עבור בעיות של זיהוי שפה ודיבור, מודלים של רצפי DNA בתחום הביואינפורמטיקה ועיבוד אותות באופן כללי. למרות שהמודלים הללו הוצגו כבר בסוף שנות השישים ובתחילת שנות השבעים, הם נעשו פופולרים במיוחד בתחילת שנות השמונים ובעזרתם הושגו תוצאות מן הטובות בתחום.

בדומה למודלים המרקוביים, גם במודלים המרקוביים החבויים המערכת יכולה להיות באחד מבין כמה מצבים, אך בניגוד למודלים המרקוביים המצב בו נמצאת המערכת בכל צעד זמן אינו ידוע לנו, ועלינו לנחש אותו מתוך תצפיות המקושרות באופן לא דטרמיניסטי למצב המערכת.

5.4.2 הגדרות

התיאור הבסיסי של המודלים המרקובים החבויים הוא כלהלן: נתונה מערכת אשר ניתנת לתיאור כאוסף של N מצבים $s_1, s_2, \ldots, s_N$. בכל רגע נתון המערכת נמצאת באחד (בדיוק) ממצבים אלו - נסמן מצב מסוים זה ב- q_t. ביחידות זמן קבועות מראש המערכת עוברת שינוי מצב (כאשר קיימת גם האפשרות להשאר במצב הקודם), כאשר שינוי זה הוא על פי הסתברויות קבועות מראש הקשורות לכל מצב. ההסתברויות מסומנות בדרך כלל בצורה $a(i,j)$ ומציינות את ההסתברות למעבר ממצב s_i למצב s_j. ההסתברויות מקיימות את ההנחה המרקובית, דהיינו שהן תלויות אך ורק במצב האחרון i (ולא במצבים קודמים אשר המערכת עברה דרכם), והן אינן תלויות בזמן. המערכת מופעלת על ציר זמן המתחיל מ- $t=1$ ומסתיים ב- t=T. נוסף להסתברויות המעבר, קיימות גם הסתברויות התחלתיות, דהיינו ההסתברות להיות בזמן $t=1$ במצב i היא $\pi(i)$. המצבים המתוארים לעיל אינם נצפים בצורה ישירה במודל זה (הם חבויים), אלא דרך התפלגות נוספת. בכל מצב ניתן לראות תצפיות (אשר יכולות להיות סקלריות או ווקטוריות, בדידות או רציפות) המאפיינות את המצב. עבור המקרה הסקלרי והבדיד נהוג לסמן את מספר התצפיות האפשרי ב- M , ולסמן את התצפיות השונות ב- $v_1, v_2, \ldots, v_M$. נסמן ב- $b(i,j)$ את ההסתברות לצפות בתצפית בדידה v_j כאשר נמצאים במצב s_i. בכל רגע נתון המערכת מוציאה תצפית מסוימת, שאותה נסמן ב- o_t.

מודל מרקובי חבוי Λ מוגדר באופן פורמלי כמורכב מחמישה מרכיבים $\pi(S,V,A,B,\Lambda)$

1. S היא קבוצה של N מצבים $s_1, s_2, \ldots, s_N$.

2. V היא קבוצה של M ערכים אפשריים לתצפיות $v_1, v_2, ..., v_M$.
3. A היא מטריצה בגודל NxN הכוללת את הסתברויות המעבר בין מצבים. $a(i,j)$ היא ההסתברות לעבור ממצב s_i למצב s_j
4. B היא מטריצה בגודל NxM המכילה את ההסתברויות ליצירה של תצפית מסוימת במצב מסוים. $b(i,j)$ היא ההסתברות לצפות בתצפית בדידה v_j כאשר נמצאים במצב s_i .
5. ווקטור הסתברויות מעבר התחלתיות: π בעל אורך N.

לרוב, שני המרכיבים הראשונים הם קבועים וידועים, ולכן נסמן לרוב את המודל על ידי $\pi(A, B, \Lambda)$.

שימוש במודל ליצור תצפיות:

בהנתן מודל מרקובי חבוי ניתן להפעיל אותו בצורה הבאה על מנת לקבל סדרת תצפיות $O = o_1, o_2, ..., o_T$ וסדרת מצבים $Q = q_1, q_2, ..., q_T$:

1. בחר מצב התחלתי $q_1 = s_i$ על פי הסתברות המעבר ההתחלתית $\pi(i)$. וקבע את הזמן לנקודת ההתחלה t=1.
2. בחר תצפית לזמן t, $o_t = v_k$ לפי הסתברות $b(i,k)$ לקבלת תצפית v_k במצב הנכחי s_i .
3. עבור למצב חדש $q_{t+1} = s_j$ על פי הסתברות המעבר $a(i,j)$ ממצב s_i למצב s_j. הגדל את האינדקס של צעד הזמן הנוכחי $t \to t+1$.
4. אם עבר כל הזמן (t=T) אשר בו פועל המודל הפסק, אחרת חזור לשלב 2.

התיאור דלעיל מתאר הן כיצד המודל פולט תצפיות והן כיצד סדרת תצפיות מסוימת יוצרה על ידי המודל .שימוש חישובי במודל המרקובי החבוי מצריך את היכולת לפתור מספר בעיות הקשורות בו

בעיית הנראות:

בהנתן סדרת תצפיות $O = o_1, o_2, ..., o_T$ ומודל $\Lambda = (A, B, \pi)$, מה ההסתברות שהסדרה נוצרה על ידי המודל $P(O \mid \Lambda)$ (מה הנראות של התצפיות לפי המודל)?

בעיית הפענוח:

בהנתן סדרת תצפיות $O = o_1, o_2, ..., o_T$ ומודל $\Lambda = (A, B, \pi)$ מהי סדרת המצבים (המסלול) $q_1, q_2, ..., q_T$ הסבירה ביותר שיצרה את סדרת התצפיות?

בעיית הלמידה:

בהנתן סדרת תצפיות $O = o_1, o_2, ..., o_T$ מהו המודל $\Lambda = (A, B, \pi)$ הסביר ביותר אשר יצר אותן?

5.4.3 פתרון בעיית הנראות

פתרון נאיבי לבעיה זו הוא

$$\begin{aligned} P(O \mid \Lambda) &= \sum_{allQ} P(O \mid Q, \Lambda) P(Q \mid \Lambda) \\ &= \sum_{q_1, q_2, ..., q_T} \pi(q_1) b(q_1, o_1) a(q_1, q_2) b(q_2, o_2) \cdots a(q_{T-1}, q_T) b(q_T, o_T) \end{aligned} \tag{5.7}$$

כאשר הסכום על Q הוא כל הפרמוטציות של סדרת המצבים החבויים (כל המסלולים האפשריים). במצב זה קל לחשב את ההסתברות של כל מסלול (כפי שכתוב בביטוי לעיל). פתרון זה דורש מספר פעולות בסדר גודל של TN^T עובדה אשר הופכת את כל התהליך לבלתי ניתן למימוש אפילו עבור מספר קטן של מצבים ואורך סדרה T קצר יחסית.

נתאר כעת פתרון בעל סיבוכיות זמן ריצה שהיא לינארי באורך הסדרה T . הפתרון מתבסס על עקרונות האופטימליות של מכונת מצבים מרקובית מסדר ראשון. על פי עקרון זה אין חשיבות לסדרת המצבים אשר הובילה לפתרון אופטימלי אלא רק למצב האחרון בשרשרת המצבים האופטימלים.

נגדיר לכן תחת השם "משתנים קדמיים" (Forward Variables)

$$\alpha_t(i) \equiv P(o_1 o_2 \cdot\cdot o_t, q_t = s_i \mid \Lambda) \tag{5.8}$$

כלומר, ההסתברות לקבל את סדרת התצפיות ולהגיע למצב s_i בזמן t בתלות במודל. אם נוכל לחשב את $a_t(i)$ בצורה יעילה הרי שהפתרון לבעיית הנראות (על פי ההגדרה) ניתן על ידי סכום המשתנים הקדמיים בזמן t=T . $P(O \mid \Lambda) = \sum_{i=1}^{N} \alpha_T(i)$

האלגוריתם לחישוב $a_t(i)$ הוא כדלקמן

1. אתחול

$$\alpha_1(i)=\pi_i b(i,o_1) \qquad 1\le i\le N$$

2. אינדוקציה

$$\alpha_{t+1}(j)=\left[\sum_{i=1}^{N}\alpha_t(i)a(i,j)\right]b(j,o_{t+1}) \qquad 1\le t\le T-1, \qquad 1\le j\le N$$

3. סיום

$$P(O\,|\,\Lambda)=\sum_{i=1}^{N}\alpha_T(i)$$

שלב האינדוקציה משתמש בצורה ישירה בעקרון המרקוביות. מספר הפעולות הדרושות על מנת לממש אלגוריתם זה הוא N^2T אשר מהווה מספר סביר של פעולות. אלגוריתם זה הוא מקרה פרטי של אלגוריתם תכנות דינמי Dynamic Programing.

בצורה דומה ניתן גם להגדיר "משתנים אחוריים" (Backward Variables)

$$\beta_t(i)\equiv P(o_{t+1}o_{t+2}\cdots o_T\,|\,q_t=S_i,\Lambda) \qquad (5.9)$$

ולבצע חישוב דומה "מן הסוף להתחלה". באופן כללי ניתן לכתוב

$$P(O\,|\,\Lambda)=\sum_{i=1}^{N}\alpha_t(i)\beta_t(i) \quad \forall t \qquad (5.10)$$

5.4.4 פתרון בעיית הפענוח

על מנת לפתור את בעיית הפענוח יש להגדיר בצורה טובה את המשמעות של סדרת המצבים הסבירה ביותר (מסלול). הגדרה אפשרית היא המסלול אשר מכיל בכל רגע נתון את המצב הסביר ביותר. מסלול זה הוא בעייתי משום שהוא יכול להיות מסלול בלתי אפשרי לפי המודל (לדוגמא, כאשר $\{q_t=s_i,q_{t+1}=s_j\}$ אבל $a(i,j)=0$).

הגדרה מקובלת למסלול אופטימלי היא המסלול אשר הסתברותו הכוללת היא הגבוהה ביותר, כלומר המסלול המביא למקסימום את $P(Q\,|\,O,\Lambda)$. פתרון הבעיה תחת הגדרה זו דומה מאד לפתרון בעיית הנראות אך שונה ממנו בשני אלמנטים. ראשית, בכל רגע נתון בחישוב אנו מעוניינים במסלול הסביר ביותר ולא בכל המסלולים. שנית, יש צורך לשמור מידע על המסלול הסביר ביותר במהלך החישוב.

האלגוריתם הבא (הידוע בשם אלגוריתם Viterbi) נותן פתרון לבעיית הפענוח.

נגדיר את המשתנה הבא (בדומה למשתנה הקדמי)

$$\delta_t(i) \equiv \max_{q_1,q_2,\ldots,q_{t-1}} P\left[q_1 q_2 \cdots q_t = i, o_1 o_2 \cdots o_t \mid \Lambda\right] \tag{5.11}$$

ואת משתנה העזר $\psi_t(j)$ אשר מכיל מצביע למצב s_k אשר הוא המצב הסביר ביותר (בזמן t-1) שממנו ניתן להגיע בזמן t למצב s_j .

1. אתחול

$$\delta_1(i) = \pi_i b_i(o_1), \psi_1(i) = 0 \qquad 1 \le i \le N$$

2. רקורסיה

$$\delta_t(j) = \max_{1\le i\le N}\left[\delta_{t-1}(i)a(i,j)\right]b(j,o_t), \qquad 2 \le t \le T \qquad 1 \le j \le N$$

$$\psi_t(j) = \arg\max_{1\le i\le N}\left[\delta_{t-1}(i)a(i,j)\right], \qquad 2 \le t \le T \qquad 1 \le j \le N$$

3. סיום

$$P^* = \max_{1\le i\le N}\left[\delta_T(i)\right]$$

4. שחזור סדרת המצבים הסבירה ביותר

$$q_t^* = \psi_{t+1}\left(q^*_{t+1}\right), \qquad t = T-1, T-2, \ldots, 1$$

5.4.5 פתרון בעיית הלימוד

לא ידוע פתרון אופטימלי (בזמן סביר) לבעיית הלימוד. ניתן למצוא פתרון אשר מהווה אופטימום מקומי במרחב הפרמטרים של המודל. על מנת לחשב אופטימום זה יש לקבוע מודל התחלתי ולשפר אותו. נתאר כאן את טכניקת Baum-Welch אשר מתבססת על רעיונות של EM – (Expectation Maximization) אשר כבר הוזכרו בכיתה. קיימת גם שיטה תת-אופטימלית אשר זמני הריצה שלה הם קצרים בהרבה הקרויה Segmental K-means, אך לא נתאר אותה כאן.

הפתרון האופטימלי על פי אלגוריתם EM הוא לנחש אוסף פרמטרים לבעיה (נסמן אותם ב- Λ_0) וכעת להפעיל אלגוריתם איטרטיבי על האוסף. על פי אלגוריתם זה בכל שלב יבוצע שיפור של הנראות של התצפיות עד הגעה להתכנסות. בסעיף זה נראה תחילה פתרון אפשרי לבעיית הלימוד ונראה את האינטואיציה שמאחוריו, לאחר מכן נראה כיצד פתרון זה נגזר מתוך הוכחת אלגוריתם ה- EM.

בהנתן אוסף פרמטרים $\Lambda=\left(A,B,\pi\right)$ (המגדירים מודל HMM), נרצה לחשב אוסף חדש $\hat{\Lambda}=\left(\hat{A},\hat{B},\hat{\pi}\right)$. אומד מקסימום נראות לחישוב $\hat{\pi}$ יתקבל על ידי ספירת מספר הפעמים בהם התקבל המצב s_i בזמן t=1, שהוא נקודת הזמן הראשונה. באופן דומה נחשב את $\hat{A}$ על ידי ספירת מספר המעברים ממצב s_i ל- s_j וחלוקת תוצאה זו במספר הפעמים בו שהתה המערכת במצב s_i. לבסוף נחשב את $\hat{B}$ על ידי ספירת מספר הפעמים בהם שהתה המערכת במצב s_i והתצפית היתה v_k וחלוקה במספר הפעמים בהם היה המצב s_i. כל הספירות המוזכרות לעיל מתבצעות על פני כל המסלולים האפשריים ,כאשר כל מסלול מוכפל בהסתברותו. הסתברות זו תלויה בסדרת התצפיות O, ובמודל של השלב הקודם Λ.

על מנת לבצע את החישוב האמור נגדיר תחילה את המשתנה הבא

$$\xi_t\left(i,j\right)\equiv P\left(q_t=s_i,q_{t+1}=s_j\mid O,\Lambda\right)=\frac{P\left(O,q_t=s_i,q_{t+1}=s_j\mid\Lambda\right)}{P\left(O\mid\Lambda\right)}\qquad(5.12)$$

משתנה זה מגדיר את ההסתברות להיות במצב s_i בזמן t ובמצב s_j בזמן t+1 בהנתן כל סדרת התצפיות והפרמטרים מן השלב הקודם - Λ.

ניתן לחשב משתנה זה באמצעות המשתנים הקדמיים והאחוריים בצורה הבאה

$$\xi_t\left(i,j\right)=\frac{\alpha_t\left(i\right)a\left(i,j\right)b\left(j,o_{t+1}\right)\beta_{t+1}\left(j\right)}{P\left(O\mid\Lambda\right)}=$$

$$=\frac{\alpha_t\left(i\right)a\left(i,j\right)b\left(j,o_{t+1}\right)\beta_{t+1}\left(j\right)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t\left(i\right)a\left(i,j\right)b\left(j,o_{t+1}\right)\beta_{t+1}\left(j\right)}=$$

$$=\frac{\alpha_t\left(i\right)a\left(i,j\right)b\left(j,o_{t+1}\right)\beta_{t+1}\left(j\right)}{\sum_{i=1}^{N}\alpha_t\left(i\right)\beta_t\left(i\right)}$$

משתנה נוסף שבו נעזר בו יהיה

$$\begin{aligned}\gamma_t\left(i\right)&\equiv P\left(q_t=s_i\mid O,\Lambda\right)=\frac{P\left(O,q_t=s_i\mid\Lambda\right)}{P\left(O\mid\Lambda\right)}=\frac{\alpha_t\left(i\right)\beta_t\left(i\right)}{P\left(O\mid\Lambda\right)}=\\&=\frac{\alpha_t\left(i\right)\beta_t\left(i\right)}{\sum_{i=1}^{N}\alpha_t\left(i\right)\beta_t\left(i\right)}\end{aligned}\qquad(5.13)$$

משמעות משתנה זה היא ההסתברות להיות במצב s_i בזמן t בהנתן כל סדרת התצפיות והפרמטרים מן השלב הקודם - Λ. המשתנה $\gamma_t(i)$ מקושר למשתנה $\xi_t(i,j)$ בצורה הבאה

$$\gamma_t(i)=\sum_{j=1}^{N}\xi_t(i,j)$$

סכימה של המשתנה $\gamma_t(i)$ על פני כל הזמנים t נותנת את מספר הפעמים שמבקרים במצב s_i (כמו גם את מספר הפעמים בהו יוצאים מאותו מצב). סכימה של המשתנה $\xi_t(i,j)$ נותנת את מספר המעברים הצפוי ממצב s_i למצב s_j.

בהתחשב בעובדות אלו מוגדר השערוך של $\hat{\Lambda}=(\hat{A},\hat{B},\hat{\pi})$ בהנתן $\Lambda=(A,B,\pi)$ על ידי

$$\hat{\pi}_i=\gamma_1(i)$$

$$\hat{a}(i,j)=\frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)}$$

$$\hat{b}(j,k)=\frac{\sum_{t=1,o_t=v_k}^{T}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)}$$

היינו שמחים לקבל כי הנראות של הפרמטרים $\hat{\Lambda}$ על פי כלל העדכון שהצענו כאן גבוהה או שווה לנראות של הפרמטרים בצעד הקודם Λ. כדי להוכיח עובדה זו נשתמש בפונקצית העזר שבה השתמשנו להוכחת ההתכנסות של אלגוריתמי EM, ונסמן אותה כאן ב- J

$$J(\Lambda,\hat{\Lambda})\overset{def}{=}\sum_{Q}P(Q\mid O,\Lambda)\log\left[P(O,Q\mid\hat{\Lambda})\right]$$

נראה כי אותן משוואות עדכון מתקבלות מתוך ביצוע מקסימיזציה של פונקציית העזר (J) על $\hat{\Lambda}$. המקסימיזציה מסתמכת על הטענה שהוכחנו עבור אלגוריתם ה-EM, כי מתקיים

$$\max_{\Lambda}\left[J(\Lambda,\hat{\Lambda})\right]\rightarrow P(O\mid\hat{\Lambda})\geq P(O\mid\Lambda)$$

על מנת למצוא את נקודת המקסימום של הפונקציה J נגזור אותה על פי $\hat{\Lambda}$. הגזירה מתבצעת על פי השלבים הבאים:

שלב א': כתיבה מחדש של פונקציה העזר

נכתוב תחילה את הפונקציה J בצורה נוחה יותר (על פי נוסחת הסתברות מותנה)

$$\sum_Q P(Q|O,\Lambda)\log\left[P\left(O,Q|\hat{\Lambda}\right)\right]=$$

$$=\sum_Q \frac{P(O,Q|\Lambda)}{P(O|\Lambda)}\log\left[P\left(O,Q|\hat{\Lambda}\right)\right]=$$

$$=\frac{1}{P(O|\Lambda)}\sum_Q P(O,Q|\Lambda)\log\left[P\left(O,Q|\hat{\Lambda}\right)\right]$$

ונשים לב שלצורך הגזירה ניתן להתעלם מן הגורם הראשון - $\frac{1}{P(O|\Lambda)}$ משום שהוא קבוע עבור גזירה ב- $\hat{\Lambda}$.

נרשום כעת בצורה מפורטת יותר את הביטוי התלוי ב- $\hat{\Lambda}$

$$\log\left[P\left(O,Q|\hat{\Lambda}\right)\right]=$$

$$=\log\left[\hat{\pi}(q_0)\prod_{t=1}^{T-1}\hat{a}(q_t,q_{t+1})\prod_{t=1}^{T-1}\hat{b}(q_t,o_t)\right]=$$

$$=\log\hat{\pi}(q_0)+\sum_{t=1}^{T-1}\log\,\hat{a}(q_t,q_{t+1})+\sum_{t=1}^{T-1}\log\,\hat{b}(q_t,o_t)$$

שלב ב': שימוש בתכנות דינמי

נשתמש בעובדה כי אנו יודעים למצע על כל המסלולים האפשריים בצורה יעילה (על ידי אלגוריתם של תכנות דינמי). ונכתוב את שלושת הסכומים בצורה שונה

$$J\left(\Lambda,\hat{\Lambda}\right)=J_n\left(\Lambda,\hat{\pi}\right)+\sum_{i=1}^{N}J_{a_i}\left(\Lambda,\hat{a}_i\right)+\sum_{i=1}^{N}J_{b_i}\left(\Lambda,\hat{b}_i\right)$$

כאשר

$$\hat{\pi}=\left[\hat{\pi}(s_1),...,\hat{\pi}(s_N)\right]$$

$$\hat{a}_i=\left[\hat{a}(s_i,s_1),...,\hat{a}(s_i,s_N)\right]$$

$$\hat{b}_i=\left[\hat{b}(s_i,v_1),...,\hat{a}(s_i,v_M)\right]$$

ו-

$$J_{\pi}\left(\Lambda,\hat{\pi}\right)=\sum_{i=1}^{N}P\left(O,q_0=s_i\mid\Lambda\right)\ \log\hat{\pi}\left(s_i\right)$$

$$J_{a_i}\left(\Lambda,\hat{a}_i\right)=\sum_{j=1}^{N}\sum_{t=1}^{T-1}P\left(O,q_t=s_i,q_{t+1}=s_j\mid\Lambda\right)\ \log\hat{a}\left(s_i,s_j\right)$$

$$J_{b_i}\left(\Lambda,\hat{b}_i\right)=\sum_{t=1}^{T}p\left(O,q_t=s_i\mid\Lambda\right)\ \log\hat{b}\left(s_i,o_t\right)=$$

$$=\sum_{k=1}^{M}\sum_{t=1}^{T}P\left(O,q_t=s_i\mid\Lambda\right)\ \delta\left(o_t,v_k\right)\ \log\hat{b}\left(s_i,v_k\right)$$

יש לשים לב שבביטוי האחרון עבור $J_{b_i}\left(\Lambda,\hat{b}_i\right)$ הכנסנו את פונקצית δ אשר מאפשרת לנו להתייחס לווקטור התצפיות בצורה כללית v_k ולא לתצפית הספציפית o_t אשר הופיעה בזמן t.

שלב ג': ביצוע אופטימיזציה תחת אילוצים

קיבלנו אם כן סכום של שלושה ביטויים אשר יש לבצע עבורו אופטימיזציה, וזאת תחת האילוצים הנובעים מדרישות ההסתברות

$$\sum_{j=1}^{N}\hat{\pi}_j=1$$

$$\sum_{j=1}^{N}\hat{a}\left(i,j\right)=1\qquad\forall i$$

$$\sum_{j=1}^{M}\hat{b}\left(i,j\right)=1\qquad\forall i$$

נשים לב כי אין תלות בין הסכומים ולכן ניתן לבצע אופטימיזציה לכל סכום בנפרד. נשתמש בכופלי לגרנג' ונראה באופן כללי איך נראית האופטימיזציה של ביטוי מהצורה $\sum_{j=1}^{N}\omega_j\log y_j$ תחת האילוץ $\sum_{j=1}^{N}y_j=1$. נרשום את הלגרנג'יאן

$$\sum_{j=1}^{N}\omega_j\log y_j-\lambda\left(\sum_{k=1}^{N}y_k-1\right)$$

וכאשר נשווה את נגזרת הלגרנג'יאן ביחס לכל אחד מן המשתנים y_j לאפס נקבל

$$\frac{\omega_j}{y_j}-\lambda=0$$

או

$$y_j=\frac{\omega_j}{\lambda}$$

נחשב את λ בצורה הבאה

$$\sum_{j=1}^{N} y_j = 1 \Rightarrow \sum_{j=1}^{N} \frac{\omega_j}{\lambda} = 1 \Rightarrow \lambda = \sum_{j=1}^{N} \omega_j$$

ולסיכום נקבל כי הפתרון הוא

$$y_j = \frac{\omega_j}{\sum_{j=1}^{N} \omega_j} \quad \forall \mathrm{j} \ .$$

שלב ד': פתרון

עבור הבעיה של הפרמטרים למודל החדש נקבל כי

$$\hat{\pi}(s_i) = \frac{P(O, q_0 = s_i \mid \Lambda)}{\sum_{i=1}^{N} P(O, q_0 = s_i \mid \Lambda)} = \frac{P(O, q_0 = s_i \mid \Lambda)}{P(O \mid \Lambda)} = \gamma_1(i)$$

$$\hat{a}(s_i, s_j) = \frac{\sum_{t=1}^{T-1} P(O, q_t = s_i, q_{t+1} = s_j \mid \Lambda)}{\sum_{j=1}^{N} \sum_{t=1}^{T-1} P(O, q_t = s_i, q_{t+1} = s_j \mid \Lambda)}$$

$$= \frac{\sum_{t=1}^{T-1} P(O, q_t = s_i, q_{t+1} = s_j \mid \Lambda)}{\sum_{t=1}^{T-1} P(O, q_t = s_i \mid \Lambda)} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma(i)}$$

$$\hat{b}(s_i, v_k) = \frac{\sum_{t=1}^{T-1} P(O, q_t = s_i \mid \Lambda) \delta(o_t, v_k)}{\sum_{k=1}^{M} \sum_{t=1}^{T-1} P(O, q_t = s_i \mid \Lambda) \delta(o_t, v_k)}$$

$$= \frac{\sum_{t=1}^{T-1} P(O, q_t = s_i \mid \Lambda) \delta(o_t, v_k)}{\sum_{t=1}^{T-1} P(O, q_t = s_i \mid \Lambda)} = \frac{\sum_{t=1, o_t = v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

מסקנה

קיבלנו בדיוק את אותן נוסחאות עדכון שהצענו האינטואיטיבית. נשים לב כי בפתרונות עבור $\hat{a}(s_i, s_j)$ ו- $\hat{b}(s_i, s_j)$ צמצמנו גורם משותף מן המונה והמכנה $P(O \mid \Lambda)$.

תרגילים

1. בניסוי אלקטרופיזיולוגי רשמו את הפעילות העצבית של תא עצב במשך תקופה ארוכה. נניח כי ידוע לנו ששרשרת הירי נוצרת על ידי תהליך מקרי פואסוני, שהקצב שלו הוא λ_1 או λ_2, ואנחנו רוצים להכריע בין שני מצבי עולם אלו. אנו מעוניינים להשוות שתי אסטרטגיות החלטה לבעית ההכרעה הזו. על פי האסטרטגיה הראשונה, נעריך את ההתפלגות של המרווחים בין פוטנציאלי פעולה (ISI), ונשווה אותה להתפלגות הצפויה על פי כל אחד מקצבי הירי. על פי האסטרטגיה השניה נעריך את התפלגות מספר הספייקים בחלונות באורך שניה אחת.

   א. חשב את המרחק הסטטיסטי בין התפלגויות מספרי פוטנציאלי הפעולה עבור שני קצבי הירי.
   ב. חשב את המרחק הסטטיסטי בין התפלגויות המרווחים בין פה"פ עבור שני קצבי הירי.
   ג. בפרק 2 הראנו כי המרחק הסטטיסטי מאפשר להעריך את כמות הדגימה הדרושה כדי להכריע בין שני מצבי העולם. איזו מבין שתי האסטרטגיות תדרוש זמן רישום קצר יותר?
   ד. האם מספר פה"פ מספק את כל האינפורמציה על שרשרות הירי? האם התפלגות המרווחים מספקת את כל האינפורמציה?
   ה. האם תשתנה התשובה אם השרשרות נוצרו על ידי תהליך שאיננו תהליך פואסוני?

2. בהמשך לשאלה הקודמת, נתונות לנו שרשות ירי שנלקחו מתוך תהליך התחדשות שבו התפלגות המרווחים בין פוטנציאלי הפעולה היא התפלגות גאמא עם פרמטרים α ו- β.

   א. מצא סטטיסטים מספיקים במשותף עבור הפרמטרים.
   ב. האם מספר פה"פ מספק את כל האינפורמציה על שרשרות הירי? האם התפלגות המרווחים מספקת את כל האינפורמציה?

3. תיארנו אלגוריתם מסוג של תכנות דינמי לבעית מציאת השרשרת הטובה ביותר. הרחב את האלגוריתם לבעית מציאת 2 השרשרות הטובות ביותר. מהי הסיבוכיות כעת?

4. בניסוי אלקטרופיזיולוגי רשמו את הפעילות של מספר קטן של תאי עצב בקליפת המוח הקדמית של חולדה, בזמן ניסוי שבו החולדה מחפשת את דרכה במבוך. מהסתכלות על התנהגות החיה, נראה כי היא נמצאת במספר מצבים התנהגותיים נפרדים ומובחנים היטב: ריצה במסדרון, התלבטות בנקודות התפצלות, ואכילה נמרצת של הגבינה שבקצהו של המבוך. תאר כיצד ניתן לבנות מודל מרקובי חבוי שיתאר את הפעילות העצבית של התאים, המבוסס על ההנחה כי גם הפעילות העצבית נובעת ממספר מצבים קוגניטיביים מובחנים היטב (שים לב שאלו אינם בהכרח מקבילים בדיוק למצבים ההתנהגותיים). תאר במפורט מה יהיו התצפיות במודל, וכיצד תאמן את המודל. אילו מסקנות מדעיות ניתן להסיק ממודל כזה?