נושאים מתקדמים - אביב תשס"ח---2007-08

מבוא: חריגות גדולות משווי משקל (סטיות גדולות) Large Deviations

1 מבוא

עבור מ"א נזכר במשפטי הגבול המוכרים. יהיו $\{X_i\}$ מ"א בת"ס ושווי פילוג, $\mathbb{E}\,X_i = \mu$, $\mathbb{E}\,X_i^2 = \sigma^2$. נגדיר את הסכום המצטבר והממוצע האמפירי

$$S_n = \sum_{i=1}^{n} X_i \ , \quad \mu_n = \frac{1}{n} S_n \ . \tag{1.1}$$

אזי חוק המספרים הגדולים הוא כי

$$\mu_n \to \mu \quad \text{w.p.1 as } n \to \infty \tag{1.2}$$

ומשפט הגבול המרכזי הוא כי

$$\frac{1}{\sigma\sqrt{n}}(S_n - n\mu) \to Z \quad \text{in distribution as } n \to \infty \tag{1.3}$$

כאשר Z הוא מ"א גאוסי סטנדרטי (כלומר ממוצע אפס ווריאנס 1). כלומר, סטיות בסדר גודל של $\sqrt{n}$ מהממוצע הן "נורמליות". לעומת זאת, אנו נדון בסטיות גדולות: אלו מוגדרת כסטיות בסדר גודל של n. למשל, נשאל כיצד מתנהגת ההסתברות של המאורע

$$\{\mu_n \ge \mu + a\} = \{S_n \ge n(\mu + a)\} \tag{1.4}$$

עבור $a > 0$. ברור כי הסתברות זו שואפת לאפס כאשר $n \to \infty$. השאלה היא - באיזה קצב הדבר קורה. לפני שננסח את השאלה במדיוק, הבה נפתח אינטואיציה לגבי קצב זה. נניח לשם פשטות כי $\mu = 0$. מצד אחד, לכל k שלם,

$$\begin{aligned}
\mathbb{P}\{\mu_n \ge a\} &= \mathbb{P}\{X_1 + \ldots + X_n \ge na\} && (1.5)\\
&\ge \mathbb{P}\left\{X_{jk+1} + \ldots + X_{(j+1)k} \ge ak \quad \text{for all } j = 0, 1, \ldots, (n/k) - 1\right\} && (1.6)\\
&= \left[\mathbb{P}\{X_1 + \ldots + X_k \ge ak\}\right]^{n/k} && (1.7)
\end{aligned}$$

בגלל אי התלות של קבוצות המ"א. אולם מחוק המספרים הגדולים, עבור k גדול מספיק הבטוי בסוגריים מרובעים קטן ממש מ- 1. לכן, עבור k כזה, נקבל חסם תחתון השואף לאפס בצורה גאומטרית. מסקנה---קצב הירידה הוא לכל היותר גאומטרי.

מצד שני, נפעיל את אי שוויון צ'בישב. נבחר $\theta > 0$ ונחשב

$$\begin{aligned}
\mathbb{P}\{X_1 + \ldots + X_n \ge na\} &= \mathbb{P}\left\{e^{\theta(X_1 + \ldots + X_n)} \ge e^{\theta na}\right\} && (1.8)\\
&\le e^{-\theta na}\,\mathbb{E}\left[e^{\theta(X_1 + \ldots + X_n)}\right]\\
&= e^{-\theta na}\left[\mathbb{E}\,e^{\theta X_1}\right]^n\\
&= \left[e^{-\theta a}\,\mathbb{E}\,e^{\theta X_1}\right]^n && (1.9)
\end{aligned}$$

בגלל שהמ"א הם iid. אם נפתח כעת את שני האקספוננטים לטור טיילור, נקבל

$$e^{-\theta a}\,\mathbb{E}\,e^{\theta X_1} \approx (1 - \theta a)(1 + \theta^2\sigma^2) \tag{1.10}$$

בגלל שלמשתנים יש תוחלת אפס. כעת עבור הערך הנתון של a נבחר $\theta = (a/\sigma^2)$ ונקבל

$$e^{-\theta a}\,\mathbb{E}\,e^{\theta X_1} \approx 1 - (a/\sigma)^4 < 1. \tag{1.11}$$

נציב זאת ב- (1.8)--(1.9) ונקבל שקצב הירידה הוא לפחות אקספוננציאלי.

הערה 1.1 למעשה אפשר לבחור את θ כך שיתן חסום הדוק יותר: מהחישוב (1.8)--(1.9) נקבל

$$\mathbb{P}\left\{X_1+\ldots+X_n \geq na\right\} \leq \left[\inf_{\theta>0} e^{-\theta a}\,\mathbb{E}\,e^{\theta X_1}\right]^n \tag{1.12}$$

מסתבר שזהו בדיוק קצב הדעיכה הנכון - כפי שנראה בהמשך.

2 סטיות גדולות למ"א

בפרק זה מטרתנו להראות כי (בסימונים הקודמים) עבור כל $a>0$ ו- $\varepsilon>0$,

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left\{a-\varepsilon<\mu_n<a+\varepsilon\right\}\approx-\ell(a) \tag{2.1}$$

עבור פונקציה כלשהיא ℓ, וכן נרצה לקבל ביטוי מפורש ככל האפשר עבור ℓ. נשים לב כי ביטוי כזה הוא קרוב אסימפטוטי-לוגריתמי, אשר ניתן גם לכתוב בצורה

$$\mathbb{P}\left\{a-\varepsilon<\mu_n<a+\varepsilon\right\}=e^{-n\ell(a)+o(n)} \tag{2.2}$$

כאשר $o(n)$ הוא ביטוי המקיים $\frac{o(n)}{n}\to 0$. יהיו $\{X_i\}$ מ"א בת"ס ושווי פילוג. נגדיר את הפונקציה יוצרת מומנטים $M(\theta)$ ואת טרנספורמצית לג'נדר של הלוגריתם שלה $\ell(a)$ (הלוגריתם $\log M(\theta)$ נקרא Comulant generating function)

$$M(\theta)\doteq\mathbb{E}\,e^{\theta X_1} \tag{2.3}$$

$$\ell(a)\doteq-\log\left\{\inf_\theta e^{-\theta a}M(\theta)\right\}=\sup_\theta\left\{\theta a-\log M(\theta)\right\}. \tag{2.4}$$

משפט 2.1 [1] עבור כל $a>\mathbb{E}\,X_1$ וכל שלם n

$$\mathbb{P}\left\{\mu_n\geq a\right\}\leq e^{-n\ell(a)}\ . \tag{2.5}$$

אם $M(\theta)<\infty$ עבור θ בסביבה של 0 אזי לכל ε קיים n_0 כך שלכל $n\geq n_0$

$$\mathbb{P}\left\{\mu_n\geq a\right\}\geq e^{-n(\ell(a)+\varepsilon)} \tag{2.6}$$

הוכחה: על ידי הזזת הממוצע של המשתנים, מספיק להוכיח את המשפט כאשר הממוצע של X_1 הוא 0. את החסם העליון מוכיחים בעזרת אי שוויון צ'בישב, כפי שעשינו בפרק הקודם. כדי להראות שמותר לכלול גם θ שליליים נשתמש באי שוויון ינסן כדי להראות שהמקסימום לא יושג בערכים שליליים.
את החסם התחתון מוכיחים בעזרת שינוי מידה - ובגלל חשיבות הטכניקה נעבור בקצרה על ההוכחה. נשים לב כי אם מתקיים

$$\ell(a)=\sup_\theta\left\{\theta a-\log M(\theta)\right\}=\theta^*a-\log M(\theta^*) \tag{2.7}$$

אזי אפשר לרשום

$$\ell(a)=-\log\mathbb{E}\,e^{\theta^*(X_1-a)} \tag{2.8}$$

נסמן ב- F את הפילוג של X_1 ונגדיר פילוג חדש

$$G(x)\doteq\frac{1}{M(\theta^*)}\int_{-\infty}^{x}e^{\theta^* y}\,dF(y)\ . \tag{2.9}$$

זהו אכן פילוג בגלל ההגדרה של $M(\theta)$. (המעבר מהפילוג המקורי לפילוג החדש נעשה עם פונקצית "צפיפות" אספונ-נציאלית - ולכן נקרא exponential twisting). כעת לכל α בשל ההגדרה של $G(x)$ ואי התלות של המ"א

$$\mathbb{P}\{X_1 \geq \alpha\} = \int_{-\infty}^{\infty} \mathbf{1}_{\{y\geq\alpha\}} dF(y) \tag{2.10}$$

$$= \int_{-\infty}^{\infty} \mathbf{1}_{\{y\geq\alpha\}} e^{-\theta^* y} e^{\theta^* y} dF(y) \tag{2.11}$$

$$= M(\theta^*) \int_{-\infty}^{\infty} \mathbf{1}_{\{y\geq\alpha\}} e^{-\theta^* y} dG(y)\ . \tag{2.12}$$

נשתמש בנוסחה זו לחישוב ההסתברות:

$$\mathbb{P}\{\mu_n \geq a\} = \mathbb{P}\{X_1 + \ldots + X_n \geq na\} \tag{2.13}$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{1}_{\{y_1+\ldots+y_n\geq na\}} dF(y_1)\ldots dF(y_n) \tag{2.14}$$

$$= [M(\theta^*)]^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{1}_{\{y_1+\ldots+y_n\geq na\}} e^{-\theta^* y_1} \cdots e^{-\theta^* y_n} dG(y_1)\ldots dG(y_n)\ . \tag{2.15}$$

כעת לכל $\varepsilon' > 0$,

$$\mathbb{P}\{\mu_n \geq a\} \geq [M(\theta^*)]^n \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{1}_{\{n(a+\varepsilon')\geq y_1+\ldots+y_n\geq na\}} e^{-\theta^* y_1} \cdots e^{-\theta^* y_n} dG(y_1)\ldots dG(y_n) \tag{2.16}$$

$$\geq [M(\theta^*)]^n e^{-n\theta^*(a+\varepsilon')} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathbf{1}_{\{n(a+\varepsilon')\geq y_1+\ldots+y_n\geq na\}} dG(y_1)\ldots dG(y_n) \tag{2.17}$$

$$= [M(\theta^*)]^n e^{-n\theta^*(a+\varepsilon')}\, \mathbb{P}_G\{n(a+\varepsilon') \geq X_1 + \ldots + X_n \geq na\}\ . \tag{2.18}$$

כיוון שהנחנו ש- $M(\theta)$ מוגדר סביב θ^* נובע כי הוא גזיר בכל סדר, ולכן למשתנים עם פילוג G יש מומנטים מכל סדר, ובפרט מומנט שני. בנוסף, כיוון ש- θ^* היא נקודת אקסטרמום של $\mathbb{E}\, e^{\theta(X_1-a)}$ נובע שהנגזרת של ביטוי זה מתאפסת ב- θ^*. לכן

$$\mathbb{E}(X_1 - a)e^{\theta^*(X_1-a)} = 0 \tag{2.19}$$

$$\mathbb{E}\, X_1 e^{\theta^* X_1} = a\, \mathbb{E}\, e^{\theta^* X_1} = aM(\theta^*) \tag{2.20}$$

$$\mathbb{E}_G\, X_1 = a\ . \tag{2.21}$$

כלומר ה- exponential twisting הוא כזה שהופך את המאורע הנדיר למאורע הצפוי ביותר - הממוצע. כיוון שתחת הפילוג G הממוצע הוא a נקבל ממשפט הגבול המרכזי שההסתברות לפי G ב- (2.18) שואפת ל- $1/2$. בפרט, עבור n גדול מספיק היא תהיה לפחות $1/4$. עבור n כזה

$$\mathbb{P}\{\mu_n \geq a\} \geq \frac{1}{4}[M(\theta^*)]^n e^{-n\theta^*(a+\varepsilon')} \tag{2.22}$$

$$= \frac{1}{4} e^{-n\ell(a)} e^{-n\theta^*\varepsilon'}\ . \tag{2.23}$$

כיוון ש- ε' הוא שרירותי ו- $\theta^* > 0$ קיבלנו את החסם (2.6).

הצורה המקובלת של משפטי סטיות גדולות היא כלהלן.

משפט 2.2 יהיו $\{X_i\}$ מ"א בת"ס ושווי פילוג. יהיו O קבוצה פתוחה ו- C קבוצה סגורה. אזי

$$\overline{\lim}_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{\mu_n \in C\} \leq -\inf_{a\in C} \ell(a) \tag{2.24}$$

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{\mu_n \in O\} \geq -\inf_{a\in O} \ell(a)\ . \tag{2.25}$$

שימו לב כי אין שום הנחות על המ"א: תוצאה זו קיימת אך ורק במימד אחד, כלומר עבור מ"א סקלריים. עבור וקטורים יש צורך בתנאים שהם קיום של $M(\theta)$ בסביבה של 0 ותנאי טכני נוסף הדרוש אם $M(\theta)$ אינו סופי עבור כל θ.
לפני שנדון בהוכחת המשפט, הרי כמה תוצאות הנובעות ממנו. ראשית, כיוון שהדרך להגיע לחסם התחתון היא על ידי שינוי זהה של הפילוג עבור כל אחד מהמשתנים, אנו מגיעים למסקנה כי מאורע נדיר במקרה כזה קורה על ידי שיתוף פעולה של כל המשתנים --- כולם משתנים את התנהגותם.
מהעובדה כי החסם נקבע על ידי האינפימום, ברור כי ההסתברות (הלוגריתמית-אסימפטוטית) של המאורע נקבעת על ידי ההסתברות להיות סביב נקודת המינימום של ℓ מתוך הקבוצה. כלומר, לא רק שהמשפט נותן את קצב הירידה של ההסתברות, הוא גם מבהיר כיצד (כלומר סביב איזה ערך של μ_n) קרה המאורע בעל ההסתברות הנמוכה.
כדי להוכיח את המשפט נשים לב תחילה כי את המשפט הקודם נוכל להרחיב לאינטרוולים אין סופיים לשמאל שאינם כוללים את הממוצע של X_1עם הוכחה זהה. מכאן הוכחת החסם התחתון זהה לזו שעשינו: אם $a \in O$ אזי יש סביבה אינטרוול שכולו ב- O, והפעלת חישוב כפי שעשינו מראה ש- $\ell(a)$ הוא חסם לכל a כזה. לגבי החסם העליון ניבדוק את הנקודה הקרובה ביותר לממוצע ונחסום את הסתברות על ידי הסתברות הקטע האין סופי המתחיל בנקודה זו.

דוגמה 2.3 עבור מ"א גאוסיים סטנדרטיים $M(\theta)$ סופי לכל θ ו-

$$M(\theta) = e^{\theta^2/2} \quad \ell(a) = \frac{1}{2}a^2 \ . \tag{2.26}$$

למשתנים פואסוניים עם פרמטר λ $\ell(a)$ סופי עבור $a \geq 0$

$$M(\theta) = e^{\lambda(e^{\theta}-1)} \quad \ell(a) = a\left(\log\frac{a}{\lambda} - 1\right) + \lambda \ . \tag{2.27}$$

למשתנים ברנוליים עם ערכים $\{0,1\}$ בהסתברות $1/2$ נקבל ש- ℓ סופי עבור $1/2 < a < 1$

$$M(\theta) = \frac{1}{2}\left(e^{\theta}+1\right) \quad \ell(a) = a\log a + (1-a)\log(1-a) + \log 2 \tag{2.28}$$

עבור המשתנים בדוגמה $M(\theta)$ סופי לכל θ: אולם למשתנה אקספוננציאלי, למשל, $M(\theta)$ סופי רק אם θ קטן ממש מהפרמטר של המשתנה האקספוננציאלי. יש כמובן משתנים עבורם $M(\theta)$ אינו קיים כלל (אינו סופי לשום $\theta \neq 0$- --וכאלו משתנים נקראים בעלי זנב כבד. חשוב להדגיש שבמקרה זה לא רק שההוכחות אינן תקיפות---ההתנהגות במובן סטיות גדולות שונה לגמרי. לגבי משתנים כאלו, סטיה גדולה קורית על ידי ערך חריג של משתנה אחד - ולא על ידי "שיתוף פעולה" כפי שראינו במקרה הקודם.

תכונות של פונקציית הקצב עבור מ"א.

משפט 2.4 נגדיר את פונקצית הקצב על ידי (2.4), אזי

1. $\ell(a) = 0$ אם ורק אם $a = \mu = \mathbb{E}\, X_1$
2. נסמן $\sigma^2 = \mathrm{Var} X_1$, אזי $\ell'(\mu) = 0, \quad \ell''(\mu) = 1/\sigma^2$
3. ℓ קונווקסי כך שהמינימום ב- μ הוא יחיד. $\ell(a)/|a| \to \infty$ כאשר $|a| \to \infty$. נסמן $D_\ell = \{a : \ell(a) < \infty\}$. אזי $\{a : \ell(a) \leq b\} \cap \bar{D}_\ell$ היא סגורה וחסומה לכל b.

משני המשפטים האחרונים נקבל את ההערכה הבאה. נסמן $a_n \doteq \mu + b\sigma/\sqrt{n}$ כאשר $b > 0$. אזי אם $M(\theta) < \infty$ לכל θ סביב 0

$$\mathbb{P}\left\{\frac{S_n - \mu n}{\sigma\sqrt{n}} \approx b\right\} \approx e^{-n\ell(\mu + b\sigma/\sqrt{n})} \approx e^{-b^2/2} \tag{2.29}$$

כלומר קיבלנו את משפט הגבול המרכזי!

רוב התכונות ניתנות להכללה למשתנים שאינם ממשיים. אולם המשפט הבא הוא ייחודי למ"א ממשיים.

משפט 2.5 *[Bahadur-Rao]* אם $M(\theta) < \infty$ לכל θ סביב 0 אזי

$$\mathbb{P}\{S_n \geq na\} = \frac{1}{\sqrt{2\pi n[\ell'(a)]^2/\ell''(a)}} e^{-n\ell(a)}[1+o(1)] \ . \tag{2.30}$$

משפט זה נותן בצורה מדוייקת יותר את המקדם של האקספוננט.

משמעות העובדה כי בשני החסמים מופיע בצד ימין האינפימום של ℓ על הקבוצה עבורה מחשבים הסתברות היא כי

$$\mathbb{P}\{\text{event}\} = \mathbb{P}\{\text{subset of the event contaning the minimizing point}\} \tag{2.31}$$

ס.ר.ס. ורדן S.R.S. Varadhan נתן לתורת הסטיות הגדולות ניסוח מתמטי שהפך לסטנדרטי:

הגדרה 2.6 פונקציה ממשית אי שלילית I על מרחב מטרי, שאינה זהותית ∞ נקראת פונקצית קצב אם היא *lower semi continuous*, כלומר לכל y

$$\liminf_{x\to y} I(x) \geq I(y) \ . \tag{2.32}$$

היא נקראת פונקצית קצב טובה *good rate function* אם הקבוצה

$$\{x : I(x) \leq \alpha\} \tag{2.33}$$

היא קומפקטית לכל α (או בקצרה, ל- I יש *compact level sets*). נאמר שאוסף מידות הסתברות P_n מקיים את עקרות הסטיות הגדולות *Large Deviations Principle—LDP* עם פונקצית קצב I אם לכל קבוצה סגורה C ולכל קבוצה פתוחה O

$$\overline{\lim}_{n\to\infty} \frac{1}{n} \log P_n\{C\} \leq -\inf\{I(x) : x \in C\} \tag{2.34}$$

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log P_n\{O\} \geq -\inf\{I(x) : x \in O\} \tag{2.35}$$

הערה 2.7 המינוח לעיל אינו סטנדרטי: יש הקוראים לפונקצית קצב טובה פונקצית קצב, ואז קוראים לפונקצית קצב פונקצית קצב חלשה.

בצורה שקולה אפשר לחשוב על P_n כעל הפילוג של מ"א X_n ולנסח את הגבולות בצורה

$$\overline{\lim}_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{X_n \in C\} \leq -\inf\{I(x) : x \in C\} \tag{2.36}$$

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{X_n \in O\} \geq -\inf\{I(x) : x \in C\} \tag{2.37}$$

נשים לב שהחסם התחתון לא יכול להיות תקף לקבוצות סגורות, שכן בפרט נקודה בודדת היא קבוצה סגורה, ואם יש צפיפות אזי ההסתברות של נקודה בודדת היא 0. מצד שני, החסם העליון תקף במקרים רבים גם עבור קבוצות פתוחות.

ראינו דוגמה אחת בה עקרון ה- LDP תקף---עבור הממוצע האמפירי של מ"א iid כאשר פונקצית הקצב היא ℓ. דוגמה נוספת היא הפילוג האמפירי. יהיו $\{X_i\}$ מ"א כמו לעיל ונגדיר מידה (אקראית)---המידה האמפירית

$$L_n \doteq \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \ . \tag{2.38}$$

אם X_n מקבלים ערכים בדידים, אזי אפשר לחשוב על L_n כווקטור, אשר רכיביו מתארים את מספר הפעמים (המנורמל) שקיבל כל ערך אפשרי. במקרה הכללי זוהי מידה אקראית (התלויה בראליזציה של התהליך) על מרחב המצב כאשר לכל קבוצה A הגודל $L_n(A)$ הוא מספר הפעמים (המנורמל) שהתהליך היה בקבוצה A. נחשוב על המקרה הבדיד, נסמן ב- ρ את פונקצית הפילוג של X_1 ונגדיר מרחב בין שתי מידות להיות המרחק בנורמת L^1 כלומר $d(\mu,\nu) = \sum_i |\mu_i - \nu_i|$. עבור מידה ν נסמן ב- $B_a(\nu)$ את הכדור הפתוח סביב ν ברדיוס a, כלומר $B_a(\nu) = \{\mu : d(\mu,\nu) < a\}$ ונסמן ב- $B_a^c(\nu)$ את המשלים של $B_a(\nu)$.

משפט 2.8 $\{L_n\}$ מקיימים את עקרון ה- LDP עם פונקצית קצב טובה

$$I_\rho(\nu) \doteq \sum_i \nu_i \log \frac{\nu_i}{\rho_i} \ . \tag{2.39}$$

בנוסף

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\{L_n \in B_a^c(\rho)\} = -\inf\{I_\rho(\nu) : \nu \in B_a^c(\rho)\} \ . \tag{2.40}$$

כמובן שעבור המקרה בו המשתנים בדידים זהו מקרה פרטי של העקרון עבור ווקטורים ממשיים. אך באופן עקרוני הטיפול במידות אמפיריות הוא "ברמה מעל" הטיפול בממוצע האמפירי, כי הממוצע האמפירי הוא פונקציה (לינארית) של הפילוג האמפירי. כפי שנראה מייד, יש דרך לקבל עקרון LDP עבור פונקציה של סדרה מתוך העקרון עבור הסדרה עצמה.

משפט 2.9 *[Contraction Principle]* יהיו X ו- Y מרחבים מטריים שלמים וספרבילים (מרחבים פולניים) ותהי f פונקציה רציפה מ- X ל- Y. תהי $\{P_n\}$ סדרת פילוגים על X, נאמר של מ"א $\{X_n\}$ ויהיו $\{Q_n\}$ הפילוגים (על Y) של $\{f(X_n)\}$. אם $\{P_n\}$ מקיים את ה- LDP עם פונקצית קצב טובה I אזי $\{Q_n\}$ מקיים את ה- LDP עם פונקצית קצב טובה

$$J(y) \doteq \inf\{I(x) : f(x) = y\} \ . \tag{2.41}$$

3 תהליכי קפיצה טהורים מרקוביים

pure jump Markov processes הם תהליכים מרקוביים אשר ערכם משתנה בקפיצות בלבד. במובן מסויים אילו תהליכי פואסון, ובפרק זה נראה כיצד הם מוגדרים.

הגדרה 3.1 תהליך קפיצה טהור מרקובי ניתן לייצוג בצורה הבאה. נסמן ב- S את הערכים האפשריים עבור התהליך. אזי אם בזמן t התהליך נמצא במצב $x \in S$ אזי התהליך ישאר במצב זה למשך זמן המוגדר על ידי מ"א אקספוננציאלי, בת"ס בעבר, עם פרמטר $\lambda(x)$. פילוג המצב הבא נקבע על ידי הסתברות מעבר $p(y|x)$.

תהליך כזה ישמש עבורנו כמודל של רשתות מסוגים שונים. ההנחה הבסיסית היא אם כן כי משך הזמן בין מאורע למאורע הוא אקספוננציאלי. כיוון שנעסוק בדרך כלל בתהליך המתאר אורכי תורים, מספר הגעות וכו', נעסוק בתהליכים שהם וקטורים עם רכיבים שלמים. כדי שנוכל לצטט תוצאות כלליות, נגביל עוד את המודל בצורה הבאה.

הגדרה 3.2 תהליך הקפיצה המרקובי מתואר על ידי הגדלים הבאים:

- תהליך וקטורי x_t אשר רכיביו מקבלים ערכים דיסקרטיים
- אוסף סופי של קפיצות אפשריות של התהליך $\{e_j\}$
- פרמטרים $\lambda_j(x)$ המתארים את הזמן הממוצע לכל קפיצה אפשרית, בתלות במצב

התנהגות התהליך היא כלהלן. אם בזמן t התהליך נמצא במצב x אזי מוגרל עבור כל j מ"א אקספוננציאלי עם פרמטר $\lambda_j(x)$. הקפיצה שתתבצע היא זאת עבורה הזמן שהוגרל הוא הקצר ביותר, כך שהמצב החדש יהיה $x + e_j$.

תאור זה שקול מבחינה הסתברותית לכל אחד משני התאורים הבאים:
במצב x מופעל תהליך פואסון עם פרמטר $\lambda_j(x)$ עבור כל j, כאשר התהליכים בת"ס. הקפיצה הראשונה של אחד מהתהליכים תעביר למצב החדש $x + e_j$, ותוזר חלילה.
במצב x מופעל תהליך פואסון עם פרמטר $\lambda(x) = \sum_j \lambda_j(x)$. עם הקפיצה של תהליך הפואסון מוגרל כיוון הקפיצה של התהליך המרקובי על ידי קוביה בלתי תלויה אשר הסתברות הצד ה-j שלה היא $\lambda_j(x)/\lambda(x)$.

דוגמה 3.3 תהליך פואסון הוא תהליך קפיצה טהור מרקובי (הוא למעשה תהליך מניה, כלומר כל קפיצה היא למעלה בגודל 1). במושגים לעיל התהליך הוא חד ממדי, $e_1 = 1$ ו- $\lambda_1 = \lambda$.

דוגמה 3.4 תור $M/M/1$ הוא תור עם הגעות פואסון ושירות אקספוננציאלי. במושגים לעיל הוא תהליך חד ממדי עם קפיצות למעלה (הגעות) $e_1 = 1$ בקצב $\lambda_1 = \lambda$, וקפיצות למטה (סיום שירות - עזיבות) $e_2 = -1$ בקצב $\lambda_2(x) = \mu \mathbf{1}_{\{x \neq 0\}}$.

נגדיר סדרת תהליכים התלויה בפרמטר, כדי לחקור את ההתנהגות האסימפטוטית. בניגוד לחקירה של הממוצע האמפירי, שם עסקנו בסכום גדל ומנורמל של משתנים iid, כאן נעסוק בתהליכים אשר עברו שינויי סקלה. השינוי זהה לזה שהופיע בקרובי נוזלים.

הגדרה 3.5 התהליך המנורמל $x^n(t)$ מוגדר מבחינה הסתברותית בעזרת אותם גדלים המתארים את התהליך $x(t)$ אולם

- תנאי ההתחלה הוא $x^n(0)$
- הקפיצות הן $\{e_j/n\}$
- קצבי הקפיצות הם $n\lambda_j(x)$

כלומר התהליך החדש מוגדר על ידי קצבים המואצים פי n וגדלי קפיצה המוקטנים פי n ביחס לתהליך המקורי.

אזהרה: באופן כללי הפילוג של התהליך $x^n(t)$ שונה מהפילוג של התהליך $x(nt)/n$!

דוגמה 3.6 בהמשך לדוגמאות *3.3*--*3.4*, סדרות התהליכים תוגדרנה כך. לתהליך פואסון x^n הקפיצות הן בגודל $1/n$ והקצב הוא $n\lambda$. לתהליך $M/M/1$ המנורמל, גודל הקפיצות גם הוא $1/n$ והקצבים הם $\lambda_1 = n\lambda$, $\lambda_2(x) = n\mu \mathbf{1}_{\{x \neq 0\}}$. עבור תהליכים אלו אכן מתקיים $x^n(t) = x(nt)/n$ כאשר השוויון הוא במובן שלשני הצדדים אותו הפילוג.

4 סטיות גדולות למסלולים

כיוון שמדובר בתהליכי קפיצה, הרי שהמסלולים מקבלים ערכים במרחב D. כיוון שזהו מרחב שקשה לעסוק בו, נבצע את הקירוב הבא. נסמן ב- $\tilde{x}^n$ את התהליך שמתקבל מ- x^n על ידי אינטרפולציה לינארית של הערכים בין כל שתי נקודות קפיצה. זהו כמובן תהליך רציף, ולכן פונקציות המדגם הן במרחב C. נקרא לשתי קבוצות תהליכים $\{y^n\}$ ו- $\{z^n\}$ exponentially equivalent עם קיום עקרון LDP עבור $\{y^n\}$ עם פונקצית קצב I גורר קיום אותו עקרון עבור $\{z^n\}$ עם אותה פונקצית קצב (זאת לא ההגדרה המדויקת של exponentially equivalent אך היא מספיקה לצרכינו).

משפט 4.1 אם $\lambda_j(x)$ חסום לכל j אזי $\{x^n\}$ ו- $\tilde{x}^n$ הם *exponentially equivalent*.

כיוון שכך, מרגע זה לא נעשה הבחנה בין שני התהליכים, ומחמת נוחיות נעסוק בתהליך האינטרפולציה, שהוא רציף. לפני שנפתח באופן אינטואיטיבי עקרון LDP עבור סדרת התהליכים $\{x^n\}$ נזדקק לתוצאה טכנית.

טענה 4.2 נניח שסדרות המספרים החיוביים $\{a_n\},\ \{b_n\}$ מקיימות

$$\lim_{n\to\infty} \frac{1}{n} \log a_n = a \quad \lim_{n\to\infty} \frac{1}{n} \log b_n = b. \tag{4.1}$$

אם $a > b$ אזי

$$\lim_{n\to\infty} \frac{1}{n} \log(a_n + b_n) = a. \tag{4.2}$$

הוכחה: כיוון שהלוג היא פונקציה מונוטונית עולה, התנאי מבטיח כי עבור n גדול, $a_n \geq b_n$ וכן כי $b_n/a_n \to 0$. לכן

$$\frac{1}{n}\log a_n \leq \frac{1}{n}\log(a_n+b_n) = \frac{1}{n}\log a_n(1+b_n/a_n) \leq \frac{1}{n}\log a_n + \frac{1}{n}\frac{b_n}{a_n} \to \frac{1}{n}\log a_n. \tag{4.3}$$

טענה זו מוצאת שימוש בהקשר הבא.

הגדרה 4.3 סדרה $\{z^n\}$ נקראת *exponentially tight* אם לכל M קיימת קבוצה קומפקטית K_M כך ש-

$$\overline{\lim}_{n\to\infty}\frac{1}{n}\log\mathbb{P}\{z^n \notin K_M\} \leq -M\ . \tag{4.4}$$

השמוש של הגדרה זו הוא הבא. נניח שאנו רוצים לחשב את הגבול של $\frac{1}{n}\log\mathbb{P}\{x^n \in S\}$. אזי נוכל לחסום אותו כך:

$$\frac{1}{n}\log\mathbb{P}\{x^n \in S\cap K_M\} \leq \frac{1}{n}\log\mathbb{P}\{x^n \in S\} = \frac{1}{n}\log\mathbb{P}\{\{x^n \in S\cap K_M\}\cup\{x^n \in S\cap K_M^c\}\} \tag{4.5}$$

$$\leq \frac{1}{n}\log\left(\mathbb{P}\{x^n \in S\cap K_M\} + \mathbb{P}\{x^n \in K_M^c\}\right) \tag{4.6}$$

אם $\{x^n\}$ מקיים את ה- LDP והוא exponentially tight אזי בזכות הטענה הקודמת ההסתברות האחרונה זניחה, ואפשר להתייחס לכל קבוצה כאילו היא פרה-קומפקטית.

נזכר שקבוצה ב- C היא פרה קומפקטית אם ורק אם הפונקציות חסומות ובעלות מקדם רציפות אחיד.

משפט 4.4 אם $\lambda_j(x)$ כולם חסומים אזי $\{x^n\}$ הוא *exponentially tight*.

לפני שנשרטט את השיטה להוכחת משפט LDP לתהליכי קפיצה מרקוביים, נתאר גרסה של חוק המספרים הגדולים לתהליכים כאלו. נשים לב שבאופן ממוצע, אם התהליך במצב x אזי הוא נע לכיוון e_j בקצב $\lambda_j(x)$. כיוון שהקצבים הם רציפים, עבור n גדול תהיינה קפיצות רבות מבלי שנתרחב מ- x ולכן אפשר לדבר על המסלול הממוצע: הוא ינוע לכיוון הממוצע, שהוא

$$\sum_j \lambda_j(x)e_j \tag{4.7}$$

ואכן, משפט קורץ Kurtz טוען שסדרת התהליכים מתכנסת לתהליך הממוצע. נגדיר משוואה דיפרנציאלית, ונסמן את הפתרון שלה ב- x^∞:

$$\frac{dx^\infty(t)}{dt} = \sum_j \lambda_j(x^\infty(t))e_j, \quad x^\infty(0) = \lim_n x^n(0). \tag{4.8}$$

אם הקצבים הם רציפים במובן ליפשיץ אזי למשוואה זו יש תמיד פתרון יחיד.

משפט 4.5 *[Kurtz]* נניח שהקצבים הם רציפים במובן ליפשיץ. נניח שהמצב ההתחלתי $x^n(0)$ מתכנס לקבוע $x^\infty(0)$. אזי לכל T סופי קיים קבוע חיובי C_1 ופונקציה חיובית C_2 כך ש-

$$\mathbb{P}\left\{\sup_{0\leq t\leq T}|x^n(t)-x^\infty(t)| \geq \varepsilon \mid |x^n(0) = x^\infty(0)\right\} \leq C_1 e^{-nC_2(\varepsilon)} \tag{4.9}$$

$$\lim_{\varepsilon\downarrow 0}\frac{C_2(\varepsilon)}{\varepsilon^2} \in (0,\infty) \qquad \lim_{\varepsilon\uparrow\infty}\frac{C_2(\varepsilon)}{\varepsilon} = \infty \tag{4.10}$$

כלומר הסיכוי שהתהליך המנורמל רחוק מהתהליך הדטרמיניסטי (הפתרון של המשוואה הדיפרנציאלית) על פני אינטרוול סופי שואף לאפס במהירות אקספוננציאלית.

הרעיון בהוכחת עקרון ה- LDP הוא כלהלן. ראשית, את החסם התחתון מספיק להוכיח עבור כדורים במקום עבור קבוצה פתוחה כללית. עבור החסם העליון, נטען כך. ראשית, בגלל exponentially tightness מספיק להוכיח את החסם עבור כל קבוצה קומפקטית K. כיוון ש- K קומפקטית, נכסה אותה על ידי כדורים פתוחים, ונמצא תת כיסוי סופי. נוכיח את החסם עבור כל כדור בנפרד, והתוצאה תתקבל בזכות טענה 4.2.

נזכר כי במרחב C כדור $B_\varepsilon(r)$ הוא אוסף כל המסלולים שמרחקם מ- r במובן נורמת הסופרמום קטן מ- ε. בנוסף, אנו נתייחס לתת הקבוצה של המסלולים אשר נקודת ההתחלה שלהם קבועה---נסמן אותה ב- x_0. גם את התהליכים המנורמלים נתחיל באותה נקודה. לצורך חישוב הגבולות, נשאל---מהו המאורע המבטיח שהתהליך ישהה בכדור נתון לאורך אינטרוול הזמן, נניח $[0,T]$, וכיצד נראה התהליך אם אכן הוא בכדור זה? נניח כעת את שתי ההנחות המהותיות לצורך החישוב:
הקצבים $\lambda_j(x)$ חסומים מלמעלה, חסומים מלמטה הרחק מ- 0 ו-
הקצבים הם פונקציות רציפות של x.
נחלק את אינטרוול הזמן לקטעים קצרים, ונשים לב כי כדי שהתהליך ישאר בכדור הנתון, עליו לקיים שני תנאים: להתחיל ב- x_0, ובכל קטע לשמור על שיפוע קרוב לשיפוע של r באותו קטע. כיוון ש- r כמובן פונקציה רציפה, משמעות הדבר היא שבקטע זמן קצר מספיק, נניח $[t,t+\Delta]$ הקצבים הם בקרוב קבועים. כלומר מספר הקפיצות (לכל כיוון) הוא מ"א פואסוני, עם קצב "קבוע" שנסמן ב- $n\Delta\lambda_j$. כיוון שמשתנה כזה אפשר לייצג כסכום של $n\Delta$ משתנים פואסוניים עם קצב λ_j, הרי שהאנקרמנט של התהליך בכל קטע קצר כזה הוא בקרוב סכום מהצורה $\frac{1}{n}\sum_{i=1}^{n\Delta} y_i$, ועלינו לבדוק אם סכום זה קרוב לאינקרמנט של r שהוא שווה בקירוב $r'\cdot\Delta$. אם כך

$$\mathbb{P}\left\{x^n\in B_\varepsilon(r)\right\}\approx\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n\Delta}y_i^k\approx r'\Delta,\ \text{all } k\right\}\approx\mathbb{P}\left\{\Delta\frac{1}{n\Delta}\sum_{i=1}^{n\Delta}y_i^k\approx r'\Delta,\ \text{all } k\right\}\tag{4.11}$$

כאשר y_i^k הם משתנים iid שהם פואסוניים עם פרמטרים $\lambda_j(r(k\Delta))$. אולם אם כך הסכומים (עבור k-ים שונים) הם בת"ס! קיבלנו

$$\mathbb{P}\left\{x^n\in B_\varepsilon(r)\right\}\approx\prod_k\mathbb{P}\left\{\frac{1}{n\Delta}\sum_{i=1}^{n\Delta}y_i^k\approx r'\right\}\tag{4.12}$$

לכן

$$\frac{1}{n}\log\mathbb{P}\left\{x^n\in B_\varepsilon(r)\right\}\approx\sum_{k=1}^{T/\Delta}\Delta\frac{1}{\Delta}\frac{1}{n}\log\mathbb{P}\left\{\frac{1}{n\Delta}\sum_{i=1}^{n\Delta}y_i^k\approx r'\right\}\tag{4.13}$$

נסמן כעת ב $\ell(r(t),r'(t))$ את פונקצית הקצב המוכרת (למ"א) עבור משתנים פואסוניים עם קצבים $\lambda_j(r(t))$ וכאשר אנו מחשבים את פונקצית הקצב בנקודה $r'(t)$ (כלומר $r'(t)$ הוא הנקודה a בה חישבנו את הערך, או הגודל והכיוון של הסטיה הגדולה עבור המ"א). נקבל אם כן כאשר תחילה $n\Delta\to\infty$ ואח"כ $\Delta\to 0$

$$\frac{1}{n}\log\mathbb{P}\left\{x^n\in B_\varepsilon(r)\right\}\approx\sum_{k=1}^{T/\Delta}-\Delta\ell(r(k\Delta),r'(k\Delta))\to-\int_0^T\ell(r(t),r'(t))\,dt\doteq-I(r)\tag{4.14}$$

הערה: בתחום סטיות גדולות, שני גבולות (או יותר) הוא ארוע נפוץ: וברוב המקרים (לשמחת המתמטיקאים) אין אפשרות להחליף את סדר הגבולות: לא רק מסיבות טכניות, אלא מהותית שינוי סדר הגבולות יתן תוצאה שונה.

משפט 4.6 עבור תהליך קפיצה מרקובי נניח שהקצבים חסומים מלמעלה, רציפים ליפשיץ ואינם מתקרבים לאפס. נסמן ב- $\ell(x,\cdot)$ את פונקצית הקצב של וקטור אקראי המתקבל מסכום של וקטורים e_j כאשר כל אחד מוכפל במשתנה אקספוננציאלי עם פרמטר $\lambda_j(x)$. אזי אוסף תהליכי הקפיצה x^n מקיים את עקרון ה- LDP עם פונקצית קצב טובה

$$I(r)\doteq\begin{cases}\int_0^T\ell(r(t),r'(t))\,dt & r\ \textit{absolutely continuous}\\ \infty & \textit{otherwise.}\end{cases}\tag{4.15}$$

בהקשר של תהליכים, I נקראת פונצית הקצב ו- ℓ נקראת פונקצית הקצב המקומית.

הערה: כאמור אנו מקפיאים את המצב ההתחלתי, הן של התהליכים והן של המסלולים בקבוצה עבורה מחשבים הסתברות, וכל המושגים הם ביחס לתת הקבוצה הזו של C. שים לב שתת הקבוצה עם תנאי התחלה קבוע אינה מרחב לינארי!

בעית חישוב קצב הדעיכה של הסתברות היא מהסוג

$$\min\{I(r) : r \in S\} \tag{4.16}$$

כאשר S היא קבוצת המסלולים המעניינת אותנו. עבור תהליך קפיצה, I הוא אינטגרל: הבעייה של מציאת מינימום של אינטגרל נקראת בעיית ווריאציה Variational problem: בעיות כאלו נחקרו רבות במתמטיקה, אם כי רק לעיתים נדירות אפשר למצוא פתרונות מפורשים בצורה אנליטית.

דוגמה 4.7 בהמשך לדוגמאות *3.4*--*3.3*, נשים לב שתהליך פואסון מקיים את ההנחות של המשפט. את פונקצית הקצב למשתנה פואסון חישבנו כבר ב- *2.27* . לעומת זאת תור $M/M/1$ אינו מקיים את ההנחות, מכיוון ש- $\lambda_2(x)$ אינו רציף. כפי שנראה בהמשך, חוסר רציפות מסוג זה הוא המקור העיקרי לקשיים בהפעלת התאוריה עבור תורים. מצד שני, במקרה זה נוכל להשתמש ב- *contraction principle* שכן ה- *reflection map* היא פונקציה רציפה, והיא תתן תהליך זה כפונקציה של תהליך הנקרא "התהליך החפשי".
אלו קבוצות S מעניינות בהקשר של תור $M/M/1$? הקבוצה

$$S_e = \{r : r(0) = x_0, r(t) \geq M \ for\ some\ 0 \leq t \leq T\} \tag{4.17}$$

היא קבוצת המסלולים אשר עברו את הגודל M בשלב כלשהו. לכן חישוב אסימפטוטי כזה יתן תשובה לקצב הדעיכה של המאורע שהתהליך $x(t)$ (ללא נרמול!) חרג מגודל של nM בפרק זמן של nT.

הגדרה 4.8 התהליך החופשי של תור $M/M/1$ הוא תהליך קפיצה מרקובי טהור עבורו $e_1 = 1,\ e_2 = -1$ ו- $\lambda_1 = \lambda,\ \lambda_2 = \mu$.

כלומר התהליך החופשי הוא הפרש של שני תהליכי פואסון.

5 תכונות פונקצית הקצב והתהליך החופשי

ראינו שעבור תהליך פואסון ועבור התהליך החפשי הקצבים אינם תלויים ב- x. לכך יש משמעות רבה מבחינה תאורטית וחישובית.

משפט 5.1 אם $\ell(x, y) = \ell(y)$ אינו תלוי ב- x אזי המסלול r^* המביא למינימום בבעיית הווריאציה

$$\min\{I(r) : r(0) = a,\ r(T) = b\} \tag{5.1}$$

הוא לינארי.

הוכחה: יהיה U מ"א המפולג יוניפרמית על $[0, T]$. אזי כיוון ש- ℓ קונווקסית, נובע מאי שיוויון ינסן כי לכל r

$$\int_0^T \ell(r'(t))\,dt = T\,\mathbb{E}\left[\ell(r'(U))\right] \geq T\ell\left[\mathbb{E}(r'(U))\right] = T\ell\left(\frac{1}{T}\int_0^T r'(t)\,dt\right) = T\ell\left(\frac{r(T) - r(0)}{T}\right) = T\ell\left(\frac{b-a}{T}\right)\ . \tag{5.2}$$

כלומר במקרה שהקצבים אינם תלויים במקום (כמו בתהליך החופשי!), הדרך הסבירה ביותר להגיע ממקום למקום היא בקו ישר. נוסף לכך, כדי למצוא את הערך של פונצית הקצב אין צורך לפתור בעיית ווריאציה - אלא רק לחשב ערך של הפונקציה ℓ.

פונקצית הקצב עבור המודל שלנו היא אכן פונקצית קצב טובה. אך בניגוד למקרה של משתנים אקראיים, פונקצית הקצב אינה קונווקסית. הדבר מקשה על בעיית האופטימיזציה.

5.1 התהליך החפשי

התהליך החופשי הוא הפרש של שני תהליכי פואסון, עם פרמטרים

$$e_1 = 1,\ e_2 = -1, \quad \lambda_1 = \lambda,\ \lambda_2 = \mu \tag{5.3}$$

ממשפט קורץ נקבל מייד את ההתנהגות ה"צפוייה" של התהליך:

$$\frac{d}{dt}x^\infty(t) = \lambda - \mu, \quad x^\infty(0) = x^\infty(0) + (\lambda-\mu)\cdot t \tag{5.4}$$

כלומר סדרת התהליכים מתכנסת לפונקציה לינארית. שיפוע הפונקציה הוא חיובי אם $\lambda > \mu$ ושלילי אחרת.

עבור תהליך זה נתן לחשב במפורש

$$\ell(x,y) = \ell(y) = y\log\left(\frac{y+\sqrt{y^2+4\lambda\mu}}{2\lambda}\right) + \lambda + \mu - \sqrt{y^2+4\lambda\mu} \tag{5.5}$$

כיוון ש-

$$(\lambda-\mu)^2 + 4\lambda\mu = (\lambda+\mu)^2, \tag{5.6}$$

קיבלנו (כצפוי) ש- $\ell(\lambda-\mu) = 0$.

דוגמה 5.2 נחשב את (הגבול של) ההסתברות $\mathbb{P}\{x^n(T) \geq a \mid x^n(0) = 0\}$ עבור $a > (\lambda-\mu)T$ (אם תנאי זה אינו מתקיים אזי לפי משפט קורץ, ההסתברות שואפת ל- 1). כלומר עלינו לחשב

$$\inf\{I(r) : r(0) = 0,\ r(T) \geq a\} \tag{5.7}$$

כיוון ש ℓ תלוי רק במשתנה המהירות ולא במשתנה המקום, אנו יודעים כי עבור כל ערך אפשרי של $r(T)$ המסלול המיטבי הוא קו ישר. לכן בעיית האופטימיזציה הופכת להיות

$$\inf\{T\ell(b/T) : b \geq a\}. \tag{5.8}$$

אולם ℓ היא קונווקסית, נקודת המינימום שלה היא ב- $\lambda-\mu$ וכיוון ש- $a/T > \lambda-\mu$ הרי ש- ℓ פונקציה עולה מימין ל- a/T. לכן המינימום הוא $T\ell(a/T)$, והמסלול הסביר להגיע ל- a בזמן T הוא קו ישר, עם השיפוע המתאים.

דוגמה 5.3 בעית הזמן החפשי. נניח ש- $\mu > \lambda$. האם ניתן באותם כלים לחשב, למשל, לחשב את ההסתברות של המאורע (הנדיר) שהתהליך מתחיל ב- $x(0) = 0$ ויעבור רמה $a > 0$ בזמן כלשהו? כדי לעבור מבעיה בזמן נתון לבעיה עם "זמן חפשי" יש לוודא שתי נקודות. האחת---שההסתברות מרוכזת סביב זמן סופי, כלומר שאיננו במצב בו, אם ניתן למערכת לרוץ ללא מגבלת זמן, אזי היא תעבור את a. כמובן שאם זה המצב - אזי זה לא מאורע נדיר.
בנוסף עלינו לבדוק שהכלים שפיתחנו עדיין תקפים. אפשר להראות שעבור התהליך החפשי והבעיה של מעבר סף חיובי, שני התנאים מתקיימים. חישוב ההסתברות (האסימפטוטית) הופך לכן לחישוב הבא. אם המינימום מושג ב- T כלשהו, אזי בהכרח הוא מושג על ידי קו ישר, ובהתאם לדוגמה הקודמת קו זה מסתיים בדיוק ב- a. לכן הבעיה היא שוב בעית אופטימיזציה רגילה (ולא של פונקציות), וניתן לרשום אותה כך:

$$\inf\{T\ell(a/T) : T > 0\} \tag{5.9}$$

בצורה שקולה, נסמן ב- $c = a/T$ את השיפוע ונקבל בעיה שקולה

$$\inf\left\{a\frac{1}{c}\ell(c) : c > 0\right\} = \inf_{c>0}\left\{\frac{a}{c}\left(c\log\left(\frac{c+\sqrt{c^2+4\lambda\mu}}{2\lambda}\right) + \lambda + \mu - \sqrt{c^2+4\lambda\mu}\right)\right\} \tag{5.10}$$

$$= a\inf_{c>0}\left\{\left(\log\left(\frac{c+\sqrt{c^2+4\lambda\mu}}{2\lambda}\right) + \frac{1}{c}\left[\lambda+\mu-\sqrt{c^2+4\lambda\mu}\right]\right)\right\} \tag{5.11}$$

למציאת המינימום נגזור את הביטוי ללא המקדם a לפי c ונשווה לאפס:

(5.12)
$$\frac{2\lambda}{c+\sqrt{c^2+4\lambda\mu}}\cdot\frac{1}{2\lambda}\left(1+\frac{1}{2}\left[\sqrt{c^2+4\lambda\mu}\right]^{-1}\cdot 2c\right)-\frac{1}{c^2}(\lambda+\mu-\sqrt{c^2+4\lambda\mu})-\frac{1}{c}\frac{1}{2}\left[\sqrt{c^2+4\lambda\mu}\right]^{-1}\cdot 2c$$

(5.13)
$$=\frac{1}{c+\sqrt{c^2+4\lambda\mu}}\cdot\left(1+[\sqrt{c^2+4\lambda\mu}]^{-1}\cdot c\right)-\frac{1}{c^2}(\lambda+\mu-\sqrt{c^2+4\lambda\mu})-[\sqrt{c^2+4\lambda\mu}]^{-1}=0$$

כעת נבדוק את הפתרון $c=\mu-\lambda>0$. עבור ערך זה $\sqrt{c^2+4\lambda\mu}=\lambda+\mu$, ולכן נקבל

$$=\frac{1}{2\mu}\left(1+\frac{\mu-\lambda}{\lambda+\mu}\right)-0-\frac{1}{\lambda+\mu} \tag{5.14}$$

$$=\frac{1}{2\mu}\frac{2\mu}{\lambda+\mu}-\frac{1}{\lambda+\mu}=0 \tag{5.15}$$

ואכן זהו השיפוע האופטימלי. הצבה בנוסחה (5.11) תתן שהערך המינימלי המתקבל הוא

$$\inf\left\{a\frac{1}{c}\ell(c):c>0\right\}=a\log\frac{\mu}{\lambda} \tag{5.16}$$

נזכר בפרשנות כי המסלול האופטימלי (במקרה זה, קו ישר) נותן את הדרך הסבירה ביותר למאורע הנדיר. אם כך, הצורה שהתהליך החפשי עם ממוצע שלילי מגיע לגובה חיובי בזמן חופשי הוא בשיפוע הפוך לשיפוע הממוצע. אפשר להראות כי, במובן מסויים, לאורך מסלול זה הקצבים מתהפכים. כצפוי, כאשר μ ו- λ מתקרבים, המינימום מתקרב לאפס, ומצד שני משך הזמן הדרוש שואף לאינסוף.

6 תור עם הגעות פואסון ושרות אקספוננציאלי

תור $M/M/1$ הוא תהליך כמו התהליך החפשי, אלא שקפיצות כלפי מטה כאשר אורך התור הוא אפס מבוטלות. בגלל חוסר הזכרון של הפילוג האקספוננציאלי, תאור זה זהה לתאור הקודם. ראינו גם שקצבי הקפיצה אינם רציפים (בגלל הדרישה שהתור לא יהיה שלילי) ולכן המשפט הכללי אינו תקף. למזלנו ניתן להתגבר על כך בעזרת ה- reflection map. אולם נתחיל ממשפט קורץ. ברור שכל עוד התור חיובי, התהליך זהה לתהליך החפשי. כמו כן ברור שאם התור יציב, אזי מרגע שהתרוקן, אם יהפוך לחיובי שוב אזי הוא יתלכד שוב עם תהליך חופשי. לכן אפשר לקבל ממשפט קורץ

$$\frac{d}{dt}x^\infty(t)=\begin{cases}\lambda-\mu & x^\infty(t)>0\\ 0 & x^\infty(t)=0\end{cases} \tag{6.1}$$

$$x^\infty(t)=\begin{cases}x^\infty(0)+(\lambda-\mu)\cdot t & x^\infty(t)>0\\ 0 & t\geq x^\infty(0)/(\mu-\lambda)\end{cases} \tag{6.2}$$

$$=\max\{x^\infty(0)+(\lambda-\mu)\cdot t,0\} \tag{6.3}$$

אם נשאל מהו הסיכוי שתור כזה יעקוב אחרי אוסף מסלולים אשר אינם מגיעים ל- 0, אזי חזרנו לתהליך החפשי: כל עוד התור עוקב אחרי מסלולים כאלו הוא לא יתרוקן ולכן הוא זהה לתהליך החפשי. לכן, בפרט החישוב עבור התהליך החופשי עונה על השאלה: מהיא ההסתברות ומהו המסלול הסביר ביותר שהתהליך יגיע מנקודה אחת לשניה, מבלי לעבור דרך 0.

כדי לקבל עקרון LDP עבור תור M/M/1 עלינו להשתמש ב- reflection map. נזכר בהגדרה: עבור תהליך $x(t)$

$$R[x](t)=x(t)-\inf_{0\leq s\leq t}x(s). \tag{6.4}$$

הפונקציונל R הוא רציף כמיפוי ממרחב הפונקציות הרציפות (עם נורמת הסופרמום) לאותו מרחב, כי

$$|R[x](t) - R[y](t)| = |x(t) - y(t) - [\inf_{0\le s\le t} x(s) - \inf_{0\le s\le t} y(s)]| \tag{6.5}$$

$$\le |x(t) - y(t)| + \sup_{0\le s\le t} |x(s) - y(s)| \le 2 \sup_{0\le t\le T} |x(t) - y(t)| \tag{6.6}$$

$$\sup_{0\le t\le T} |R[x](t) - R[y](t)| \le 2 \sup_{0\le t\le T} |x(t) - y(t)|. \tag{6.7}$$

לכן נוכל להשתמש ב- contraction mapping theorem (זוהי אחת מהסיבות להשתמש בתהליך הרציף: R אינו רציף כפונקציונל על D כיוון שבמרחב זה חיבור אינו רציף!). נסמן ב- ℓ_f את פונקציית הקצב המקומית של התהליך החפשי---נזכר שפונקציה זו תלויה רק במשתנה המהירות, אך לא במשתנה המקום.

משפט 6.1 תור $M/M/1$ יציב, כלומר $\lambda < \mu$ מקיים את עקרון ה- LDP עם פונקציית קצב טובה

$$I(r) \doteq \begin{cases} \int_0^T \ell(r(t), r'(t))\, dt & r \textit{ absolutely continuous} \\ \infty & \textit{otherwise} \end{cases} \tag{6.8}$$

$$\ell(x,y) \doteq \begin{cases} \ell_f(y) & \textit{if } x > 0 \textit{ or } y > 0 \\ 0 & \textit{if } x = 0 \textit{ and } y = 0 \\ \infty & \textit{if } x < 0 \textit{ or } x = 0 \textit{ and } y < 0\ . \end{cases} \tag{6.9}$$

הוכחה: מכיוון שאת התור נקבל על ידי הפעלת המיפוי R על התהליך ומכיוון שהמיפוי רציף, לפי משפט ה- contraction mapping theorem אכן מתקיים עקרון ה-LDP. פונקציית הקצב מוגדרת על ידי

$$I(r) = \inf\{I_f(u) : R[u] = r\} \tag{6.10}$$

כאשר I_f היא פונקציית הקצב של התהליך החפשי, אשר יש לה את הצורה האנטגרלית, עם פונקציית קצב מקומית ℓ_f. כעת נשים לב כי אם $r(t) < 0$ עבור t כלשהו אזי למשוואה $R[u] = r$ אין פתרון u. בחירת הערך האינסופי עבור פונקציית הקצב המקומית אכן תתן תכונה זו. בנוסף, אם מתקיים $R[u] = r$ וכן $r(t) > 0$ או לחילופין $r'(t) > 0$ אזי בהכרח גם $r'(t) = u'(t)$ שכן האינפימום אינו משתנה סביב כזה t. בנוסף, לפי משפט קורץ אם $r(t) = 0$ אזי המצב הסביר ביותר עבור התור הוא להשאר שם - ולכן ערך פונקציית הקצב המקומית הוא 0. אם $u(t) = 0$ כלומר התור ריק, ומתחיל להתמלא, אזי כאמור $r'(t) = u'(t)$. אולם תופעה זו לא יכולה להתרחש על אינטרוול - אם התהליך עולה אזי הוא לא ישאר שווה ל- 0. לכן אפשר להתעלם מנקודות הזמן בהן התור ריק אך מתחיל להתמלא. אם כך, עבור r חיוביים,

$$\inf\{I_f(u) : R[u] = r\} = \inf\{\int_0^T \ell_f(u'(t))\mathbf{1}_{\{r(t)>0\}}\, dt : R[u] = r\} \tag{6.11}$$

$$= \inf\{\int_0^T \ell_f(r'(t))\mathbf{1}_{\{r(t)>0\}}\, dt : R[u] = r\} \tag{6.12}$$

$$= \inf\{\int_0^T \ell(r(t), r'(t))\, dt\} \tag{6.13}$$

וכאמור עבור r המקבלים ערכים שליליים ההסתברות היא 0 ובהתאם פונקציית הקצב היא אינסופית.

בשלב זה נעבור לחישובים הקשורים לשאלות מעשיות, ונראה שיטות לקיצורי דרך-כדי להמנע מפתרון בעיות וריאציה. השאלה היא: מהו ערך פונקציית הקצב, ומהו המסלול הסביר ביותר, עבורו התהליך מתחיל בנקודה x בזמן 0 ומסיים בנקודה y בזמן T? אם נתרגם זאת לשפת הסטיות הגדולות, עלינו למצוא את הערך וכן את המסלול המביא למינימום בבעיה הבאה:

$$\inf\{I(r) : r(0 = x,\ r(T) = y\}. \tag{6.14}$$

כדי להמנע מטענות טריוויאליות נניח $x > 0, y > 0$.

טענה 6.2 יהי r מסלול כלשהו המקיים $r(0)=x,\ r(T)=y$. אזי קיים מסלול $\hat{r}$ המקיים תכונות אלו, $I(\hat{r})\le I(r)$ ובנוסף $\hat{r}(t)=0$ על קטע אחד לכל היותר. אם קיימים זמנים $t_1<t_p<t_2$ כך ש- $r(t_1)=r(t_2)=0$ אולם $r(t_p)>0$ כלומר r אינו מקיים תכונה זו, אזי הוא אינו אופטימלי.

הוכחה: נסמן $t_1=\inf\{t>0: r(t)=0\}$ ו- $t_2=\sup\{t<T: r(t)=0\}$. יש מה להוכיח אם שניהם סופיים. נגדיר את $\hat{r}$ להיות שווה ל- r מחוץ לאינטרוול $[t_1,t_2]$ ושווה ל- 0 באינטרוול זה. אזי

$$I(r)=\int_0^T \ell(r(t),r'(t))\,dt \tag{6.15}$$

$$=\int_0^{t_1} \ell(r(t),r'(t))\,dt+\int_{t_1}^{t_2} \ell(r(t),r'(t))\,dt+\int_{t_2}^{T} \ell(r(t),r'(t))\,dt \tag{6.16}$$

$$\ge\int_0^{t_1} \ell(r(t),r'(t))\,dt+\int_{t_2}^{T} \ell(r(t),r'(t))\,dt \tag{6.17}$$

$$=\int_0^T \ell(\hat{r}(t),\hat{r}'(t))\,dt \tag{6.18}$$

כי $\ell(a,b)\ge 0$ ושווה לאפס כאשר $a=b=0$. אם $r(t_p)>0$ אזי קיים קטע באינטרוול $(t_1,t_p]$ עליו r חיובי וכן r' חיובי. אולם אז $\ell(r,r')>0$ ונקבל אי שוויון ממש ב- (6.16)--(6.17).

טענה 6.3 המסלול האופטימלי מ- x ל- y הוא או קו ישר, או מסלול המורכב משלושה קטעים: קטע מ- x ל- 0 אשר שיפועו $\lambda-\mu$, קטע בו המסלול שווה 0 וקטע מ- 0 ל- y ששיפועו $\lambda-\mu$.

הוכחה: המסלולים שאינם יורדים ל- 0 הם אלו עבורם התהליך החופשי והתור הם תהליכים זהים, לפחות עבור n גדול. כיוון שעבור התהליך החופשי הפתרון האופטימלי הוא תמיד קו ישר, נובע כי מבין כל המסלולים האלו, האופטימלי יהיה קו ישר.

עבור מסלולים העוברים דרך 0 ראינו כי הם בנויים משלושה קטעים. נתבונן בקטע הראשון: לאורכו המסלול אינו שווה 0, ולכן אותו טיעון תופס-כדי שהמסלול יהיה אופטימלי, על הקטע הראשון להיות קו ישר. טיעון זהה תקף לגבי הקטע האחרון.

נזכר כי עבור התהליך החופשי המסלול האופטימלי לעליה הוא בשיפוע $\lambda-\mu$. אם השיפוע המקורי תלול פחות, אזי ניתן להקטין אותו על ידי הארכת הקטע בו המסלול שווה 0, וזאת ללא תוספת עלות. אם השיפוע המקורי תלול יותר וניתן למתן את השיפוע על ידי הקטנת הקטע בו התהליך מתאפס, אזי שינוי זה יקטין את העלות. אם לא ניתן, ניצור מסלול חדש, אשר יבנה מ- T ואחורה: עד T שיפועו יהיה $\lambda-\mu$ כלומר השיפוע האופטימלי לעליה. תחילת העליה תהיה בנקודה בה קו זה חוצה את הקטע היורד. המסלול החדש שנוצר הוא מצד אחד עם עלות נמוכה יותר מהמקורי, ומצד שני אינו מגיע ל 0 ולכן ניתן למצוא מסלול עם עלות נמוכה עוד יותר - הקו הישר! לסיכום, שיפוע החלק האחרון חייב להיות השיפוע האופטימלי.
נזכר כי אם השפוע הוא $\mu-\lambda$ (שהוא שיפוע שלילי) אזי $\ell=0$ ולכן אין שום תרומה לאינטגרל. נניח שהשיפוע חיובי יותר-אזי על ידי הקטנת השיפוע ניתן להקטין את האינטגרל כי הפונקציה ℓ חיובית מימין ל $\mu-\lambda$. בצורה דומה, אם השיפוע שלילי יותר, על ידי הפיכתו לשיפוע שעליו פונקצית הקצב מתאפסת, נוכל להקטין את האינטגרל - זאת כל עוד קיים אינטרוול עליו המסלול שווה 0. אם שינוי השיפוע יגרום לחציה מימין ל- t_2 אזי המסלול היורד בשיפוע האופטימלי עד לחציית הקטע העולה וממשיך לעלות משם הוא עדיף; אולם אז משיקולי התהליך החופשי, מסלול זה אינו יכול להיות אופטימלי ועדיף המסלול הישר!

המסקנה: יש שני מסלולים המועמדים להיות אופטימליים: קו ישר או מסלול בן שלושה קטעים. נשים לב כי ערך האינטגרל עבור המסלול בן שלושת הקטעים אינו תלוי ב- T: עבור T גדול מספיק - שכן מחיר שני הקטעים הראשונים הוא 0 ומחיר השלישי תלוי רק ב- y. עבור T קטן הקטע האמצעי יעלם - ואז מסלול זה אינו אופטימלי. עבור זמן ארוך רק מסלול זה אופטימלי, שכן מחיר הקו הישר גדל עם עלית T. יתכן תחום ביניים של T עבורו ישנם שני פתרונות אופטימליים. בכל מקרה, חישוב העלות הוא טריוויאלי ואינו כרוך באופטימיזציה כלשהי.

התנהגות ההסתברות והמסלול האופטימלי כפונקציה של T היא כלהלן. עבור זמנים קצרים המסלול האופטימלי הוא קו ישר, והעלות יורדת כפונקציה של T, עד למינימום (שהוא כמובן לא שלילי). המינימום יהיה 0 אם ורק אם $x>y$ שכן אז יש זמן עבורו ניתן לרדת בשיפוע $\mu-\lambda$. לאחר מכן מתחילה עליה מונוטונית, אך היא חסומה מלמעלה על ידי העלות של המסלול עם שלושה קטעים.

את השאלה: כמה זמן יקח לתהליך לעבור מ- x ל- y ובפרט ל- $y=0$ אפשר לנסח שוב כשאלה לגבי מסלולים.

7 חזרה להילוך שיכור

נחזור להילוך שיכור ונראה כיצד אפשר לקבל תוצאה פשוטה עבור מסלולים. יהי $\{S_n = x_0 + \ldots + x_n\}$ הילוך שיכור כלומר תהליך עם הפרשים בת"ס ושווי פילוג $\{x_n\}$. מהתוצאות עבור הממוצע האמפירי אנו יודעים כי

$$\mathbb{P}\{S_n \geq na\} \approx e^{-n\ell(a)} \tag{7.1}$$

עבור $a > \mathbb{E}\, x_1$ כאשר ℓ היא פונקצית הקצב.

משפט 7.1 $\mathbb{P}\{S_j \geq ja \; for \; all \; j \leq n\} \approx e^{-n\ell(a)}$.

הוכחה: נזקק לטענת עזר: יהיו $\{r_i\}$ מספרים המקיימים $r_1 + \ldots + r_n \geq na$. נגדיר

$$i^* \doteq \arg\inf\{r_1 + \ldots + r_i - ia : 1 \leq i \leq n\}. \tag{7.2}$$

ונסמן $k_n = k \mod n$. אזי

$$r_{(i^*+1)_n} + \ldots + r_{(i^*+j)_n} \geq ja \tag{7.3}$$

לכל $1 \leq j \leq n$ (להוכחה צייר את הערכים המצטברים). נפעיל את הלמה על הסדרה ונקבל שעל המאורע ש- $S_n \geq na$ קיים i^* אקראי כך ש- $S_{(i^*+j)_n} \geq ja$ לכל $1 \leq j \leq n$. אם כך

$$\begin{aligned}
&\mathbb{P}\{S_j \geq ja \text{ for all } j \leq n\} && (7.4)\\
\leq &\mathbb{P}\{S_n \geq na\} && (7.5)\\
= &\mathbb{P}\left\{S_{(i^*+j)_n} \geq ja \text{ for all } j \leq n\right\} && (7.6)\\
= &\sum_{k=1}^{n} \mathbb{P}\left\{S_{(i^*+j)_n} \geq ja \text{ for all } j \leq n \text{ and } i^* = k\right\} && (7.7)\\
= &\sum_{k=1}^{n} \mathbb{P}\left\{S_{(k+j)_n} \geq ja \text{ for all } j \leq n\right\} && (7.8)\\
= &n\, \mathbb{P}\{S_j \geq ja \text{ for all } j \leq n\}. && (7.9)
\end{aligned}$$

8 חשבון וריאציות

כיצד לפתור בצורה אנליטית את בעיית האופטימיזציה? התאוריה הרלוונטית נקראת חשבון וריאציות calculus of variations ויש לה הסטוריה ארוכה, תאוריה מפותחת אך מעט תוצאות חישוביות. כלומר, בדרך כלל לא ניתן למצוא פתרון מפורש. חשבון וריאציות כולל בתוכו (ובמידה רבה הוא מקביל) לתורת הבקרה האופטימלית. "הבעיה הפשוטה" the simplest problem בתחום היא להביא למינימום את האינטגרל

$$\int_0^T \ell(r(t), r'(t))\, dt \tag{8.1}$$

כאשר ℓ היא פונקציה כלשהיא של שני משתנים (לאו דוקא עם התכונות שיש אצלנו), כאשר המינימזציה היא על כל המסלולים (הפונקציות) r המקיימות $r(0) = a, \; r(T) = b$ ו- a, b נתונים. השאלה הראשונה המתבקשת היא: מהם המסלולים המותרים r. חלק ניכר מהספרות עוסק במסלולים שהם גזירים ברציפות למקוטעין. מבחינתינו זוהי דרישה לא קבילה שכן אנו יכולים להראות כי עבור המודלים שלנו כל הגבולות הם (בהסתברות 1) רציפים בהחלט, אך לא בהכרח גזירים ברציפות.

באופן אינטואיטיבי, אפשר לפתח תנאי מספיק לאופטימליות בשיטה הבאה. נניח ש r אופטימלי. יהי x מסלול המקיים $x(0) = x(T) = 0$. אזי לכל δ המסלול $r + \delta x$ מקיים את תנאי השפה כי $(r + \delta x)(0) = r(0) + \delta x(0) = a$ ובצורה דומה עבור T. נגדיר פונקציה ממשית

$$F(\delta) = \int_0^T \ell(r(t) + \delta x(t), r'(t) + \delta x'(t))\, dt \tag{8.2}$$

שים לב ש- r ו- x הם קבועים. הפונקציה F היא ערך הפונקציונל שלנו בנקודה $r + \delta x$. כיוון שזו פונקציה ממשית, וכיוון שהיא מגיעה למינימום ב- $\delta = 0$, הרי שנגזרתה שם צריכה להתאפסץ נחשב

$$0 = \frac{d}{d\delta} \int_0^T \ell(r(t) + \delta x(t), r'(t) + \delta x'(t))\, dt \bigg|_{\delta=0} \tag{8.3}$$

$$= \int_0^T \left[x(t) \frac{d}{dr} \ell(r(t), r'(t)) + x'(t) \frac{d}{dr'} \ell(r(t), r'(t)\, dt \right] \tag{8.4}$$

$$= \int_0^T x(t) \left[\frac{d}{dr} \ell(r(t), r'(t)) - \frac{d}{dt} \frac{d}{dr'} \ell(r(t), r'(t)\, dt \right] \tag{8.5}$$

כאשר בשוויון האחרון השתמשנו באינטגרציה בחלקים ובעובדה ש- $x(0) = x(T) = 0$. כיוון שתנאי זה הוא הכרחי לכל x נסיק מכך שהאינטגרנד חייב להתאפס, וקיבלו את תנאי אוילר Euler necessary condition

$$\frac{d}{dr} \ell(r(t), r'(t)) - \frac{d}{dt} \frac{d}{dr'} \ell(r(t), r'(t) = 0. \tag{8.6}$$

זוהי משוואה דיפרנציאלית (סתומה) עבור המסלול האופטימלי r. כמובן שאם למשוואה זו פתרון יחיד, אזי הוא בהכרת המסלול האופטימלי.

רשימת מקורות

[1] A. Shwartz, A. Weiss, Large deviations for performance evaluation, CRC 1995, second edition.

[2] F. den Hollander, Large deviations, Fields Institute Monographs, AMS, 2000