שיטות הסתברותיות ואלגוריתמים – תרגולים

מחברים: אלדר פישר, יונתן גולדהירש

1 ביולי 2019

הקדמה וענינים טכנים

הקורס עוסק בשיטות הסתברותיות בקומבינטוריקה ואלגוריתמים. הדגש הוא על לימוד השיטות עצמן ולכן על הסטודנטים לצפות ללמוד גם תוצאות במתמטיקה טהורה וגם תוצאות במדעי המחשב. עיקר החלקים המתמטים בקורס הם לפי הספר הבא:

N. Alon and J. Spencer, The Probabilistic Method (2nd/3rd/4th edition).

הפרק על הילוכים מקריים יסתמך בעיקר על המאמר הבא:

L. Lovász, Random Walks on Graphs: A survey. In: Combinatorics, Paul Erdös is Eighty (Vol. 2), D. Miklós, V.T. Sós and T. Szönyi (editors).

חלקים אלגוריתמים אחרים יהיו בד"כ לפי הספר הבא:

R. Motwani and P. Raghavan, Randomized Algorithms.

ספר נוסף על שיטות הסתברותיות:

M. Mitzenmacher and E. Upfal, Probability and Computing: Randomized Algorithms and Probabilistic Analysis.

הספר הבא מכיל מבוא בסיסי לתורת האנטרופיה שישמש אותנו:

T.M. Cover and J.A. Thomas, Elements of Information Theory.

מומלץ לבצע קריאה מקדימה של פרק השאלות על מרחק בין התפלגויות המופיע בחוברת התרגילים הפתורים של הקורס, אשר יועבר בתרגיל הראשון. נסו לפתור את השאלות בעצמכם לקראת תחילת הקורס.

מתכונת הקורס

הקורס ניתן במתכונת של שעתיים הרצאה, שעה תרגיל ושעה אימון (שעתיים אלו יהיו עוקבות וינתנו ע"י המתרגל). בשעת התרגיל יועברו הוכחות ונושאים הנגזרים מנושאי ההרצאה, ושעת האימון תוקדש למעבר על פתרונות של תרגילים משנים קודמות.

ציון הקורס כולו מבוסס על סמך פתרון דפי תרגילים (בדרך כלל ארבעה), כאשר התרגיל האחרון ניתן לקראת סוף הקורס ויהיה להגשה לאחר הסוף (אין מבחן). יש להגיש את כל דפי התרגיל, הציון יהיה פונקציה של סך כל הנקודות שנצברו בפתרונות השאלות שבדפי תרגילים (לכל שאלה יהיה ניקוד מקסימלי ולא יהיה שקלול לכל דף תרגילים בנפרד). התרגילים ופתרונותיהם יחולקו אלקטרונית. ההגשה תהיה אלקטרונית לכתובת הדוא"ל של המתרגל (במקרים מסויימים ניתן יהיה לקבל אישור להגשה ידנית) והתרגילים הבדוקים יוחזרו (מודפסים עם הערות) בצינורות המקובלים. לנוחותכם יהיה מעקב Webcourse על ציוני השאלות במהלך הסמסטר.

שימושים בלינאריות התוחלת

כאשר יש לנו שני מ"מ X,Y בעלי תוחלת, תמיד מתקיים $\mathrm{E}\left[X+Y\right]=\mathrm{E}\left[X\right]+\mathrm{E}\left[Y\right]$. שימו לב כי שוויון זה מתקיים מבלי להניח דבר נוסף על X ו־Y, ובפרט לא דורש אי תלות ביניהם. מקרה נפוץ בו נשתמש בלינאריות התוחלת הוא הבא. נניח כי ישנה סדרת מאורעות $A_1,A_2,\ldots,A_n$ ולהם בהתאמה מ"מ מציינים (אינדיקטורים) $X_1,X_2,\ldots,X_n$, ואנחנו נתהה כמה מהמאורעות מתרחשים בתוחלת. המ"מ העונה לשאלה זו מתקבל מסכומם $X\triangleq X_1+X_2+\ldots+X_n$, ומכיוון שלכל משתנה אינדיקטור מתקיים $\mathrm{E}\left[X_i\right]=\Pr\left[A_i\right]$ אז נקבל שסכומם מקיים

$$\mathrm{E}\left[X\right]=\mathrm{E}\left[X_1+X_2+\ldots+X_n\right]=\mathrm{E}\left[X_1\right]+\ldots+\mathrm{E}\left[X_n\right]=\Pr\left[A_1\right]+\ldots+\Pr\left[A_n\right]$$

וכך נדע את תוחלת מספר המאורעות שקורים מבלי להניח דבר על תלות או אי תלות ביניהם.

ישום לצביעה מקרית של גרף

בהנתן גרף $G=(V,E)$, צביעה של G היא פונקציה $f:V\to[c]$. נאמר כי קשת $uv\in E$ היא מונוכרומטית אם $f(u)=f(v)$. נראה כי כל גרף ניתן לצבוע ב־c צבעים כך שלכל היותר $\frac{1}{c}$ מהקשתות הן מונוכרומטיות. לכל קשת $e\in E$ נגדיר את המאורע A_e של היות הקשת מונוכרומטית. X_e יהיה משתנה מקרי מציין עבור המאורע A_e. כך, מספר הקשתות המונוכרומטיות בהשמה מקרית הוא המשתנה המקרי $X\triangleq\sum_{e\in E}X_e$. נחשב את תוחלתו: $X_e=1$ אם ורק אם שני צידי הקשת e צבועים באותו הצבע c. ההסתברות לכך היא $\frac{1}{c}$: לכל צבע של הצד הראשון, יש $c-1$ צבעים לצד השני שלא יצרו קשת מונוכרמטית, וצבע אחד שכן. כך $\mathrm{E}[X]=\sum_{e\in E}\mathrm{E}[X_e]=\sum_{e\in E}\frac{1}{c}=\frac{1}{c}|E|$ ולכן קיימת השמה למשתנים המקריים שבה מספר הקשתות המונוכרומטיות קטן או שווה לתוחלת, ובהשמה זו לכל היותר $\frac{1}{c}$ מהקשתות מונוכרומטיות.

דוגמה לישום בתורת המספרים

נראה ישום של שיטת התוחלת בתורת המספרים, הכולל גם קצת אלגברה: נראה שבכל קבוצה A בת n מספרים טבעיים, קיימת תת קבוצה בת $\frac{n}{3}$ מספרים שבה אף מספר אינו סכום של שני מספרים אחרים בקבוצה (קבוצה כזו נקראת בלתי תלויה). ראשית נצטט משפט של Dirichlet: עבור $a,b\in\mathbb{N}$ שעבורם $\gcd(a,b)=1$ ישנם אינסוף מספרים ראשוניים מהצורה $ak+b$. בשביל ההוכחה ניקח p ראשוני מהצורה $3k+2$ שהוא גדול דיו (יותר גדול מ־$\max A$), ונסתכל בתוך $\mathbb{Z}_p$ על הקבוצה $S=\{k+1,\ldots,2k+1\}$. הקבוצה S היא ב"ת מעל $\mathbb{Z}_p$, וכן הקבוצה $iS=\{i(k+1),\ldots,i(2k+1)\}$ (כאן הכפל הוא ב־$\mathbb{Z}_p$) היא ב"ת מעל $\mathbb{Z}_p$ לכל $i\in\mathbb{Z}_p\setminus\{0\}$. בפרט זה נכון עבור קבוצות אלו גם מעל הטבעיים, כאשר מסתכלים כאן על היצוג של איברי iS בתוך $\mathbb{Z}$. לבסוף, עבור i אקראי יוניפורמי מתוך $\mathbb{Z}_p\setminus\{0\}$, הסיכוי של כל איבר $a\in A$ להיות ב־iS (כאשר מסתכלים על היצוג ב־$\mathbb{Z}$) הוא $\frac{k+1}{3k+1}>\frac{1}{3}$ (כי $a\in iS$ אם ורק אם $i^{-1}a\in S$, ולכל $a\neq 0$ הביטוי $i^{-1}a$ מתפלג יוניפורמית ב־$\mathbb{Z}_p\setminus\{0\}$), ולכן התוחלת של גודל $iS\cap A$ היא לפחות $\frac{1}{3}|A|$. לכן קיים i עבורו $iS\cap A$ היא תת הקבוצה המבוקשת של A.

הוכחה נוספת ושיפור ללמת הבידוד

נראה כאן הוכחה קלה לגרסה משופרת של למת הבידוד, אשר נוסחה ע"י Noam Tashma (תלמיד תיכון בעת כתיבת ההוכחה). הגרסה כאן היא עם שינוי של ניצן (סטודנט במחזור קודם בקורס) אשר עושה אותה יותר גמישה להכללות. גם כאן נניח ש־A היא קבוצה בת m איברים, ש־$\mathcal{F}$ היא משפחה של תתי קבוצות של A, ושפונקצית המשקלות $w:A\to\{1,\ldots,n\}$ מוגרלת כך ש־$w(a)$ נבחר באופן יוניפורמי וב"ת לכל $a\in A$. נסו לראות עד היכן ניתן להכליל את ההוכחה לערכים אחרים של הטווח של w. המשפט המשופר קובע כי הסיכוי שתהיה $F\in\mathcal{F}$ יחידה עם משקל מינימלי הוא לפחות $(1-\frac{1}{n})^m$. זה נותן למשל סיכוי חיובי קבוע עבור $m=2n$, דבר שלמת הבידוד המקורית אינה מבטיחה.

באופן מפתיע ההוכחה אינה הסתברותית אלא קומבינטורית טהורה. נסמן ב־W את קבוצת כל פונקציות המשקל האפשריות, נסמן ב־$W' = \{w \in W : \mathrm{Im}(w) \subseteq \{2,\ldots,n\}\}$ את קבוצת פונקציות המשקל שלא מקבלות עבור אף איבר את הערך הנמוך ביותר 1, ונסמן ב־$\hat{W}$ את קבוצת פונקציות המשקל עבורן יש $F \in \mathcal{F}$ יחידה עם משקל מינימלי. אנו נראה שמתקיים $|\hat{W}| \geq |W'|$, ומכאן נובע מייד שהסיכוי לקיום קבוצה מינימלית יחידה הוא לפחות $|\hat{W}|/|W| = (1-\frac{1}{n})^m$ כנדרש.

נבנה אם כן פונקציה $\phi : W' \to \hat{W}$ ונראה שהיא חד־חד ערכית. בהינתן פונקציה $w : A \to \{2,\ldots,n\}$ (כזכור פונקציות מ־W' לא מקבלות ערך 1), נבחר $F \in \mathcal{F}$ שעבורה $w(F)$ היא מינימלית, ומבין הקבוצות הנ"ל נבחר אחת שאינה מוכלת באף קבוצה אחרת $F' \in \mathcal{F}$ עם משקל מינימלי (אם יש מספר קבוצות המקיימות את שני התנאים אז נבחר מהן אחת באופן שרירותי). נגדיר עתה את $\phi(w)$ להיות הפונקציה $\hat{w} : A \to \{1,\ldots,n\}$ כך ש־$\hat{w}(a) = w(a)$ אם $a \in A \setminus F$ ו־$\hat{w}(a) = w(a) - 1$ אם $a \in F$.

נראה ש־F היא הקבוצה היחידה מ־$\mathcal{F}$ עם משקל מינימלי: אם $F' \in \mathcal{F}$ היתה קבוצה אחרת שעבורה $w(F')$ מינימלי, אז מכיוון שהיא אינה מכילה את F, מתקיים $\hat{w}(F') = w(F) - |F \cap F'| > w(F) - |F| = \hat{w}(F)$. אם לעומת זאת $F' \in \mathcal{F}$ אינה בעלת משקל מינימלי לפי w (אבל היא כן יכולה להכיל את F), אז מתקיים $\hat{w}(F') > w(F) - |F \cap F'| \geq w(F) - |F| = \hat{w}(F)$.

הראינו שהתמונה של הפונקציה ϕ מוכלת ב־$\hat{W}$, ולכן נותר רק להראות שהיא חח"ע, וזאת ע"י הגדרת פונקציה הופכית $\phi' : \mathrm{Im}(\phi) \to W'$. בהינתן $\hat{w} = \phi(w) \in \mathrm{Im}(\phi)$, לפי מה שהוכחנו קודם יש עבורה קבוצה יחידה עם משקל מינימלי, שהיא אותה F שנבחרה בהגדרה של $\phi(w)$. נגדיר את $w' = \phi'(\hat{w})$ ע"י כך ש־$w'(a) = \hat{w}(a)$ אם $a \in A \setminus F$, ו־$w'(a) = \hat{w}(a) + 1$ אם $a \in F$. הפונקציה מוגדרת היטב מכיוון שזהות F נגזרת אך ורק מ־$\hat{w}$ בתור הקבוצה היחידה עם משקל מינימלי, וקל לראות שאכן $w = w'$, ז"א ש־ϕ' פונקציה הופכית ל־ϕ.

דה־רנדומיזציה

מוטיבציה

בדרך כלל, בהנתן הוכחה לקיומו של מבנה קומבינטורי מסויים, מצפים לקבל ממנה גם אלגוריתם יעיל למציאתו של מבנה כזה. אולם כאשר ההוכחה היא הסתברותית, לרב גם האלגוריתם המתקבל יהיה הסתברותי, ובמקרים מסויימים לא יהיה בידינו אפילו אלגוריתם הסתברותי (אף בשיטת התוחלת יהיו מצבים שבהם האלגוריתם ההסתברותי אינו מיידי, שכן לא תמיד יש חסם תחתון על הסיכוי שבו ערכו של מ"מ אינו קטן מתוחלת המשתנה). כאן אנו נראה שתי שיטות לבניה של אלגוריתם דטרמיניסטי יעיל מתוך הוכחה הסתברותית.

שיטת התוחלות המותנות

שיטה זו יכולה לעזור כאשר תוצאת הקיום המקורית הוכחה באמצעות התיחסות לתוחלת של משתנה מקרי. נניח שהמבנה הקומבינטורי המוגרל ניתן לאפיון ע"י המשתנים המקריים $X_1,\ldots,X_m$ (למשל הגרף המקרי $G(n,\frac{1}{2})$, כאשר כל X_i הוא משתנה אינדיקטור לקיום קשת מסוימת בגרף). נניח שבנוסף לכל X_i יש תחום ערכים שגודלו חסום ע"י קבוע (בעוד מספר הערכים האפשריים ל־$X_1,\ldots,X_m$ עדיין יכול להיות אקספוננציאלי ב־m).

עבור פונקציה מספרית $f(X)$ של המבנה המקרי שלנו, אם לכל סדרת ערכים $i_1,\ldots,i_k$ שמתקיים עבורם $\Pr\big[X_1 = i_1,\ldots,X_k = i_k\big] > 0$ ניתן לחשב ביעילות את $\mathrm{E}\big[f(X)|X_1 = i_1,\ldots,X_k = i_k\big]$, אז קיים אלגוריתם דטרמיניסטי יעיל למציאת מבנה קומבינטורי C (הנתון ע"י סדרה של ערכים עבור המ"מ) עבורו מתקיים $f(C) \geq \mathrm{E}[f(X)]$. האלגוריתם יפעל בצורה הבאה: בשלב ה־k, האלגוריתם יעבור על כל ערך אפשרי j של המ"מ X_k (בהינתן הערכים $i_1,\ldots,i_{k-1}$ שנבחרו עבור $X_1,\ldots,X_{k-1}$), ויחשב את התוחלת המותנה $\mathrm{E}\big[f(X)|X_1 = i_1,\ldots,X_{k-1} = i_{k-1}, X_k = j\big]$. מכיוון שמתקיים

$$\mathrm{E}\big[f(X)|X_1 = i_1,\ldots,X_{k-1} = i_{k-1}\big] =$$
$$= \sum_{j:\Pr[X_1=i_1,\ldots,X_{k-1}=i_{k-1},X_k=j]>0} \mathrm{E}\big[f(X)|X_1 = i_1,\ldots,X_{k-1} = i_{k-1}, X_k = j\big] \cdot \Pr\big[X_k = j|X_1 = i_1,\ldots,X_{k-1} = i_{k-1}\big]$$

אז קיים j עבורו

$$\mathrm{E}\big[f(X)|X_1=i_1,\ldots,X_{k-1}=i_{k-1},X_k=j\big] \geq \mathrm{E}\big[f(X)|X_1=i_1,\ldots,X_{k-1}=i_{k-1}\big]$$

נבחר את i_k להיות ערך j המקיים זאת. להוכחה שלאחר m השלבים אכן קיבלנו את המבנה המבוקש, נשים לב שאם נסמן $\mathrm{E}_k = \mathrm{E}\big[f(X)|X_1=i_1,\ldots,X_k=i_k\big]$ אז מתקיים

$$\mathrm{E}[f(X)] = \mathrm{E}_0 \leq \mathrm{E}_1 \leq \cdots \leq \mathrm{E}_m = f(i_1,\ldots,i_m)$$

כנדרש. שימו לב שלא דרשנו כאן אי תלות של $X_1,\ldots,X_m$, אולם בהרבה מקרים יהיה בשיטה זו שימוש כאשר המ"מ הם בלתי תלויים, כי אז קל יותר לחשב את התוחלות המותנות.

עתה נראה דוגמה קונקרטית. כזכור, הוכחנו בשיטת לינאריות התוחלת שלכל 3CNF נתון בעל n משתנים ו־m פסוקיות קיימת הצבה המספקת לפחות $\frac{7}{8}m$ פסוקיות מתוכן. נראה עתה אלגוריתם דטרמיניסטי למציאת ההצבה הנ"ל עבור 3CNF נתון: עבור הצבה $X=(X_1,\ldots,X_n)$ במשתנים $x_1,\ldots,x_n$, נסמן ב־$f(X)$ את מספר הפסוקיות המסופקות על ידה. כזכור $\mathrm{E}[f(X)]=\frac{7}{8}m$ כאשר ערכי $X_1,\ldots,X_n$ נבחרים מקרית באופן יוניפורמי וב"ת. עתה נריץ את האלגוריתם למעלה, כאשר בשלב ה־k יוחלט האם $X_k=1$ או $X_k=0$. התוחלות המותנות של f בכל שלב אינן קשות לחישוב ע"י שימוש בלינאריות התוחלת. בכך ניתן למצוא באופן דטרמיניסטי הצבה המספקת לפחות $\frac{7}{8}m$ מהפסוקיות.

הערה: שימו לב כי אין זה מובטח שמצאנו את ההצבה המספקת את המספר המירבי של פסוקיות (וזה אינו מפתיע, כי בעית מציאת ההצבה בעלת הסיפוק המקסימלי היא NP־קשה). יתכן למשל שבשלב ה־k בחרנו ערך 0 עבור X_k כי לו היתה התוחלת המותנה הגדולה יותר, בעוד שעבור הבחירה $X_k=1$ היו מעט הצבות המספקות הרבה יותר פסוקיות.

שיטת מרחבי המדגם המוגבלים

שיטה זו ישימה לפעמים כאשר מרחב ההסתברות הוא מכפלה של הרבה מרחבי הסתברות קטנים. בדומה למעלה נניח שהמבנה המוגרל מתאפיין ע"י המ"מ $X_1,\ldots,X_m$, אולם כאן נניח שכל המ"מ האלו הם בלתי תלויים. אם הוכחת קיום המבנה אינה משתמשת בתכונת אי התלות המלאה, אלא רק בתכונה חלשה יותר של המ"מ, אז ניתן לעיתים להחליף את מרחב המכפלה של כל ערכי המ"מ האפשריים במרחב קטן בהרבה, שאותו ניתן לסרוק.

בדוגמה כאן נניח ש־$X_1,\ldots,X_m$ הם משתנים בוליאנים המקבלים את ערכיהם באופן יוניפורמי וב"ת. נניח עתה שהוכחנו שבהסתברות חיובית המבנה המוגרל $X=(X_1,\ldots,X_m)$ מקיים את התכונות הרצויות, ושבהוכחה זו לא השתמשנו באי התלות המוחלטת של המשתנים $X_1,\ldots,X_m$, אלא רק באי תלות בזוגות, ז"א בכך שכל זוג (X_i,X_j) הוא זוג ב"ת. נרצה עתה למצוא באופן דטרמיניסטי מבנה X המקיים את התכונות הרצויות, ואת זאת נעשה ע"י כך שנראה שאותה הוכחת קיום תעבוד גם עבור מרחב הסתברות קטן בהרבה מהמרחב המקורי.

נראה עתה שיטה אלגברית ליצירת מרחב הסתברות שגודלו 2^k, ושעבורו קיימים 2^k-1 משתנים מקרים בוליאנים יוניפורמים וב"ת בזוגות. לכן אם נבחר $k=\lfloor\log_2 m\rfloor+1$ אז נוכל לייצר m מ"מ ב"ת בזוגות, ולעומת זאת לדאוג לכך שהמרחב עצמו יכיל לא יותר מ־$2m$ סדרות ערכים שונות עבור המ"מ, שאותן נוכל פשוט לסרוק. בניית המרחב נעשית כך: להגרלת הערכים ראשית נגריל k מ"מ בוליאנים יוניפורמים וב"ת (באופן מוחלט), ונסמן אותם $Y_1,\ldots,Y_k$. עתה לכל קבוצה $\emptyset\neq I\subseteq\{1,\ldots,k\}$, נגדיר את המ"מ $X_I=\bigoplus_{i\in I}Y_i$. קל להוכיח שכל מ"מ X_I מקבל את ערכו באופן יוניפורמי: נבחר שרירותית $i\in I$, ואז מתקיים

$$\begin{aligned}\Pr\big[X_I=1\big] &= \Pr\big[Y_i=1|\bigoplus_{j\in I\setminus\{i\}}Y_j=0\big]\Pr\big[\bigoplus_{j\in I\setminus\{i\}}Y_j=0\big]\\ &\quad+\Pr\big[Y_i=0|\bigoplus_{j\in I\setminus\{i\}}Y_j=1\big]\Pr\big[\bigoplus_{j\in I\setminus\{i\}}Y_j=1\big]\\ &= \frac{1}{2}\Pr\big[\bigoplus_{j\in I\setminus\{i\}}Y_j=0\big]+\frac{1}{2}\Pr\big[\bigoplus_{j\in I\setminus\{i\}}Y_j=1\big]=\frac{1}{2}\end{aligned}$$

עתה נותר להוכיח שלכל $I \neq J$ מתקיימת אי תלות בין X_I ל־X_J. במקרה זה של שני מ"מ בוליאנים יוניפורמים הדבר שקול לטענה ש־$\Pr\left[X_I \oplus X_J = 1\right] = \frac{1}{2}$, ואותה קל להוכיח מכך שמתקיים $X_I \oplus X_J = X_{(I \setminus J) \cup (J \setminus I)}$.

דוגמה לשימוש: נבחן את ההוכחה הבאה הקובעת שלכל גרף עם m קשתות יש חתך בעל לפחות $\frac{m}{2}$ קשתות. לכל צומת $v \in V$ נגריל באופן אחיד וב"ת משתנה $X_v \in \{0,1\}$, ונבחן את החתך V_0, V_1 המוגדר על ידי $V_k = \{v \in V | X_v = k\}$. לא קשה לוודא שתוחלת מספר הקשתות בחתך היא $\frac{m}{2}$: לכל קשת uv בגרף מגדירים משתנה אינדיקטור המקבל 1 אם היא בחתך (ז"א ש־u ו־v אינם באותו V_i) ו־0 אחרת. התוחלת של משתנה אחד כזה היא $\frac{1}{2}$, ולכן תוחלת סכום המשתנים הנ"ל הוא $\frac{m}{2}$ כנדרש. לכן קיים חתך עם לפחות מספר זה של קשתות (הערה: יש גם הוכחות דטרמיניסטיות פשוטות יותר לטענה זו, אולם שיטת ההוכחה כאן ישימה גם בבעיות אחרות, ומספקת המחשה טובה לשיטת מרחבי המדגם המוגבלים).

נשים לב עתה לכך שתוחלת מספר קשתות החתך היא $\frac{m}{2}$ אפילו אם מניחים רק שהמשתנים X_v נבחרים באופן ב"ת בזוגות. לכן אפשר להשתמש בבניה למעלה כדי להראות ששימוש ב־$k = \lfloor \log_2 |V| \rfloor + 1$ משתנים מקריים $Y_1, \ldots, Y_k$, ובניית ה־X_v־ים מהם, גם תתן מרחב הסתברות שבו תוחלת גודל החתך היא $\frac{m}{2}$, ולכן קיים גם במרחב ההסתברות הקטן יותר חתך בגודל $\frac{m}{2}$ לפחות. עתה נוכל לכתוב אלגוריתם דטרמיניסטי שסורק את כל האפשרויות עבור $Y_1, \ldots, Y_k$ (שמספרן הוא $2^k = O(|V|)$), ולכל אחת מהאפשרויות בודק את גודל החתך. הערה אחרונה: שאלת גודל יחס הקירוב האופטימלי לחתך המקסימלי בגרף (בהנחה כי $\mathrm{P} \neq \mathrm{NP}$) עודנה פתוחה, והאלגוריתם הטוב ביותר הידוע משיג יחס קירוב של בערך $\frac{8}{7}$ (האלגוריתם כאן נותן יחס של 2), ידוע מאידך שזה NP־קשה לקרב את גודל החתך המקסימלי ביחס יותר טוב מ־$\frac{17}{16}$.

מספר הערות לסיכום שיטת מרחבי המדגם המוגבלים: ניתן להקטין את מרחב המדגם גם במקרים יותר כלליים, למשל כאשר יש צורך בכך שכל קבוצה בת k משתנים תהיה ב"ת עבור k קבוע. שימו לב למשל שעבור $k = 3$ הדבר יתן פיתרון אלטרנטיבי לשאלת המציאה של הצבה המספקת $\frac{7}{8}m$ מהפסוקיות בנוסחת 3CNF נתונה. ניתן גם לבנות מרחבים מוגבלים כאשר המ"מ אינם בוליאנים או אינם יוניפורמים – אז משתמשים ב־Reed-Solomon codes להפחתת מרחב המדגם, אם כי אלו יעילים פחות.

הערות אחרונות על דה־רנדומיזציה באופן כללי: מה שראינו כאן הוא רק קצה המזלג. במשך זמן רב אחת השאלות הקשות היתה שאלת הדה־רנדומיזציה (אפילו חלקית) של הוכחות המשתמשות בלמה הלוקלית, שאת התשובה לה תראו בתרגול כאשר נלמד את הלמה. למושג הדה־רנדומיזציה יש גם מוטיבציה בתורת הסיבוכיות, מכיוון ששיטות דה־רנדומיזציה כלליות יכולות לספק הוכחה לכך שמחלקת הסיבוכיות BPP (מחלקת התכונות הניתנות להכרעה באמצעות אלגוריתם הסתברותי פולינומי עם שגיאה חסומה) אינה חזקה כפי שהיא נראית.

הכרסום (nibble) של Rödl

הקדמה ומוטיבציה

תזכורת קלה בנושא היפרגרפים: היפרגרף r־יוניפורמי (פשוט) $H = (V, E)$ מורכב מקבוצת צמתים V וקבוצת קשתות E, כך שכל קשת היא קבוצה (לא סדורה) של r צמתים (ללא חזרות). בפרט, היפרגרף 2־יוניפורמי הוא גרף (פשוט) רגיל. לצומת $v \in V$ נגדיר את דרגתו $d(v)$ כמספר הקשתות המכילות את v. בהיפרגרפים אפשר גם להגדיר דרגות של קבוצות צמתים. למשל, עבור $v \neq w$, נגדיר את הדרגה המשותפת כמספר הקשתות המכילות את שני הצמתים, $d(v, w) = |\{e \in E | \{v, w\} \subseteq e\}|$.

באמצעות הכרסום של Rödl מוכיחים תוצאה כללית למציאת כיסוי כמעט מושלם עבור היפרגרף שמקיים תנאי "אחידות" מתאימים בדרגות צמתיו. ההוכחה מבוססת על פעולה בשלבים ("נגיסות"), כאשר הוכחת ביצוע כל שלב מסתמכת בכבדות על שיטת המומנט השני. ראשית ננסח את הגירסא הכללית, לפי Pippenger שניסח בהסתמך על Frankl ו־Rödl: לכל מספר שלם $r \geq 2$ וממשיים $k \geq 1$ ו־$a > 0$, קיימים $\gamma = \gamma(r, k, a)$ ו־$d_0 = d_0(r, k, a)$ עם התכונה הבאה. נניח שעבור $n, D \geq d_0$ נתון היפרגרף r־יוניפורמי $H = (V, E)$ עם n צמתים וללא צמתים מבודדים, כך שלכל צומת פרט ל־γn מהם דרגתו היא בין $(1 - \gamma)D$ ל־$(1 + \gamma)D$, לא קיים צומת שדרגתו kD או יותר, ולכל זוג צמתים דרגתם המשותפת אינה יותר מ־γD. בהיפרגרף כזה קיימות $(1 + a)\frac{n}{r}$ קשתות שמכסות יחדיו את כל הצמתים.

שימו לב שנובע ממשפט זה גם שעבור $d_0(r,k,\frac{b}{r-1})$ ו־$\gamma(r,k,\frac{b}{r-1})$ והתנאים למעלה קיימות $(1-b)\frac{n}{r}$ קשתות זרות: עבור $\alpha=\frac{b}{r-1}$, בהנתן כסוי עם $(1+\alpha)\frac{n}{r}$ קשתות, נעבור קשת־קשת ובכל שלב נבחר את הקשת רק אם היא אינה חותכת את הקשתות הקודמות שנבחרו. אם נשארנו בסוף עם $(1-\beta)\frac{n}{r}$ קשתות, אז ניתן לראות ש־$n\le(1-\beta)n+(\alpha+\beta)\frac{r-1}{r}n$ ולכן $\beta\le(r-1)\alpha=b$. לא קשה גם לראות (עם בחירת פרמטר מעט קטן יותר מ־$\frac{b}{r-1}$) שעבור קיום קבוצה כזו של קשתות אפשר גם לוותר על התנאי שאין צמתים בודדים.

השימוש המקורי של שיטת הכרסום של Rödl היה בפתרון החיובי של השאלה הבאה: מנסים לכסות את כל תת הקבוצות מגודל l של $\{1,\ldots,m\}$ על ידי תת קבוצות מגודל k. האם אפשר (עבור m גדול דיו) למצוא $(1+o(1))\binom{m}{l}/\binom{k}{l}$ תת קבוצות מגודל k, כך שכל ת"ק מגודל l תהיה מוכלת בלפחות אחת מהן? מהלמה הכללית מוכיחים זאת ע"י בניית ההיפרגרף הבא: בוחרים $r=\binom{k}{l}$. כל צומת v מתאימה לת"ק מגודל l של $\{1,\ldots,m\}$, ולכל ת"ק A מגודל k של $\{1,\ldots,m\}$ לוקחים כקשת את כל הצמתים המתאימים לת"ק מגודל l של A. מתקבל שכל צומת נמצאת ב־$D=\binom{m-l}{k-l}$ קשתות בדיוק, וכל זוג צמתים שונים זה מזה נמצאים יחדיו בלא יותר מ־$\binom{m-l-1}{k-l-1}=o(D)$ קשתות.

למת "הנגיסה הבודדת"

כדי להוכיח את המשפט הכללי מוכיחים את הלמה הבאה, ואחר כך משתמשים ב"הרצות חוזרות" שלה: לכל שלם $r\ge 2$ וממשיים $K\ge 1$, $\epsilon>0$, ו־$\delta'>0$ קיימים $\delta(r,K,\epsilon,\delta')>0$ ו־$D_0(r,K,\epsilon,\delta')>0$, כך שאם היפרגרף r־יוניפורמי H בעל m צמתים כך ש $m,D\ge D_0$ מקיים שלכל צומת פרט ל־δm מהם דרגתו היא בין $(1-\delta)D$ ל־$(1+\delta)D$, לא קיים ב־H צומת שדרגתו היא KD או יותר, ולכל זוג צמתים דרגתם המשותפת אינה יותר מ־δD, אז קיימת ב־H קבוצת קשתות $\tilde{E}$ כך שמתקיים

$$\frac{\epsilon m}{r}(1-\delta')\le|\tilde{E}|\le\frac{\epsilon m}{r}(1+\delta')$$

ומתקיימים עבורה התנאים הבאים: נגדיר את $V'=V\setminus\bigcup_{e\in\tilde{E}}e$, ואת H' להיות ההיפרגרף המושרה על V' (זהו למעשה ההיפרגרף הנותר לאחר הסרת כל הצמתים המוכלים באיברי $\tilde{E}$). אלו נדרשים לקיים

$$me^{-\epsilon}(1-\delta')\le|V'|\le me^{-\epsilon}(1+\delta')$$

(שימו לב שמספר הצמתים קטן לפחות פי פקטור קבוע), ולכל צמתי V' פרט ללא יותר מ־$\delta'|V'|$ מתוכם דרגתם ב־H' מקיימת

$$De^{-\epsilon(r-1)}(1-\delta')\le d_{H'}(v)\le De^{-\epsilon(r-1)}(1+\delta')$$

(הרעיון כאן הוא שתתקיים "אחידות דרגה" מסויימת הדרושה להפעלות נוספות של הלמה).

לפני הוכחת הלמה, נראה איך מוכיחים ממנה את המשפט הכללי: נבחר $\epsilon>0$, $0<\delta<\frac{1}{10}$, ומספר שלם t כך ש־$e^{-\epsilon t}<\epsilon$ וכן $(1+4\delta)(\frac{\epsilon}{1-e^{-\epsilon}}+r\epsilon)<1+a$. עתה נבחר סידרה $\delta=\delta_t>\delta_{t-1}>\ldots>\delta_0$ כאשר

$$\delta_i=\min\{\delta_{i+1}e^{-i\epsilon(r-1)},\frac{1}{4}\delta_{i+1},\delta(r,ke^{i\epsilon(r-1)},\epsilon,\delta_{i+1})\}$$

הפונקציה "δ" (עם ארבעת הפרמטרים) שם היא ה־δ של הנגיסה הבודדת, והביטוי $\frac{1}{4}\delta_{i+1}$ במינימום נועד להבטיח שמתקיים $\prod_{i=0}^{t}(1+\delta_i)\le 1+2\delta$. עתה נפעיל את הלמה t פעמים, כאשר בפעם ה־$i+1$ נשתמש בפרמטרים r, $K=ke^{i\epsilon(r-1)}$, ϵ ו־$\delta'=\delta_{i+1}$ (כאשר k הוא הפרמטר המופיע במשפט הכרסום של Rödl), כשבכל פעם הלמה מופעלת על תת ההיפרגרף המושרה על הצמתים שלא כוסו בפעמים הקודמות, וה"מקבילה" של D תהיה $De^{-i\epsilon(r-1)}$. בפרט רואים שהתנאי עבור K מתקיים (שכן $KD'=kD$); קיום התנאי עבור הדרגות המשותפות לשני צמתים מובטח מהביטוי $\delta_{i+1}e^{-i\epsilon(r-1)}$ המופיע בהגדרת δ_i למעלה, והתנאי על דרגות כל הצמתים פרט ל δm מהם נובע מהלמה עצמה.

את d_0 בוחרים כך ש־$De^{-i\epsilon(r-1)}$ יהיה גדול דיו בכל שלב (לפי ה־"D_0" המתאים), ו־γ יהיה שווה ל־δ_0 כפי שנבחר למעלה. את הצמתים שנשארו לאחר t הפעלות נכסה פשוט ע"י לקיחת קשת מכילה מתוך ההיפרגרף

המקורי לכל צומת שנותר. אם נסמן בכל שלב את קבוצת הקשתות שנלקחו ב־E_i ואת הצמתים שנשארו ב־V_i, אז מטענת הלמה ש־$|V_{i-1}|e^{-\epsilon}(1-\delta_i) \leq |V_i| \leq |V_{i-1}|e^{-\epsilon}(1+\delta_i)$ נובע כי $|V_i| \leq |V_0|e^{-i\epsilon}(1+2\delta)$, ולכן מתקבל $|E_i| \leq \frac{\epsilon|V_{i-1}|}{r}(1+\delta_i) \leq \frac{\epsilon n}{r}e^{-(i-1)\epsilon}(1+4\delta)$. מספר הקשתות המלא בכיסוי חסום ע"י

$$\begin{aligned}\sum_{i=1}^{t}|E_i|+|V_t| &\leq (1+4\delta)\frac{\epsilon n}{r}\sum_{i=0}^{t-1}e^{-i\epsilon}+(1+2\delta)ne^{-\epsilon t}\\ &< \frac{n}{r}(1+4\delta)(\frac{\epsilon}{1-e^{-\epsilon}}+r\epsilon)<(1+a)\frac{n}{r}\end{aligned}$$

(המעבר לשורה השניה משתמש בחסימת טור הנדסי ובהנחה $e^{-\epsilon t} < \epsilon$).

הוכחת הנגיסה הבודדת

הפרוצדורה עצמה פשוטה: לוקחים את $\tilde{E}$ להיות תת קבוצה אקראית של E, כאשר כל קשת נבחרת בהסתברות $\frac{\epsilon}{D}$ באופן ב"ת, ומראים שבהסתברות לפחות $\frac{1}{2}$ (ואף גבוהה יותר) הקבוצה הזו מקיימת את הנדרש (אלגוריתם הסתברותי למציאת $\tilde{E}$ כזו יהיה לבחור אותה שוב ושוב עד שתקיים את התנאים הנ"ל).

ראשית נזכור שסכום הדרגות (בכל היפרגרף r־יוניפורמי פשוט) מקיים $\sum_{v\in V}d(x)=r|E|$ ונסיק מכך ש־$\frac{1}{r}(1-2\delta)Dm \leq |E| \leq \frac{1}{r}((1+\delta)+\delta K)Dm$. התוחלת של $|\tilde{E}|$ היא $\frac{\epsilon}{D}|E|$, ומכיוון שהבחירה היא ב"ת נובע כי בהסתברות לפחות $\frac{9}{10}$ (עבור m גדול דיו) הערך הוא קרוב לתוחלת (מיד נוכיח זאת ע"י שיטת המומנט השני). לכן לכל $\delta_1 > 0$ אפשר למצוא פרמטרים δ ו־D_0 כך שמתקיים אי השוויון $\Pr[(1-\delta_1)\frac{\epsilon m}{r} \leq |\tilde{E}| \leq (1+\delta_1)\frac{\epsilon m}{r}] \geq \frac{9}{10}$: ראשית בוחרים δ כך ש־$(1-\frac{\delta_1}{2})\frac{\epsilon m}{r} \leq \mathrm{E}[|\tilde{E}|] \leq (1+\frac{\delta_1}{2})\frac{\epsilon m}{r}$ (אח"ב נקטין את δ עוד). ניתן לראות שהשונות אז מקיימת $\mathrm{V}[|\tilde{E}|] < (1+\frac{\delta_1}{2})\frac{\epsilon m}{r}$ (משתני האינדיקטור עבור הקשתות הם ב"ת) ומכך נובע שעבור m גדול דיו (חשוב לשים לב שזה אינו תלוי ב־D) אכן בהסתברות גבוהה $|\tilde{E}|$ יהיה בתחום המבוקש. בפרט אפשר לדאוג ש־$\delta' \leq \delta_1$ כדי לטפל בדרישה על $|\tilde{E}|$ שבניסוח הלמה. הערה: ה־δ_i שמופיעים כאן אינם קשורים לאלו שהוגדרו בתת־הפרק הקודם (אגב, מכיוון שהמשתנים המקריים כאן הם ב"ת לחלוטין, אפשר היה עד כאן גם להשתמש בחסימת סטיות גדולות, המופיעה בהמשך הקורס).

על מנת לחסום את $|V'|$ לכל $v \in V$ נסמן ב־I_v את משתנה האינדיקטור עבור המאורע ש־v לא מכוסה ע"י $\tilde{E}$. אם $(1-\delta)D \leq d(v) \leq (1+\delta)D$ אז $(1-\frac{\epsilon}{D})^{(1-\delta)D} \leq E\left[I_v\right] \leq (1-\frac{\epsilon}{D})^{(1+\delta)D}$. לכן בסיכום קבוצת הצמתים $(1-\delta)(1-\frac{\epsilon}{D})^{(1-\delta)D}m \leq E\left[|V'|\right] \leq (\delta+(1-\frac{\epsilon}{D})^{(1+\delta)D})m$, ולכל δ_2 בחירה מתאימה של D ושל δ תבטיח שזה בין $(1-\delta_2)me^{-\epsilon}$ ל־$(1+\delta_2)me^{-\epsilon}$. לחשוב השונות נחסום עבור $v \neq w$ את

$$\begin{aligned}\mathrm{Cov}[I_v,I_w] &= \mathrm{E}[I_vI_w]-\mathrm{E}[I_v]\mathrm{E}[I_w]=(1-\frac{\epsilon}{D})^{d(v)+d(w)-d(v,w)}-(1-\frac{\epsilon}{D})^{d(v)+d(w)}\\ &= (1-\frac{\epsilon}{D})^{d(v)+d(w)}\left((1-\frac{\epsilon}{D})^{-d(v,w)}-1\right) \leq (1-\frac{\epsilon}{D})^{-\delta D}-1\end{aligned}$$

עבור בחירה מתאימה של הפרמטרים אפשר לדאוג שזה יהיה קטן מכל δ_3 שנרצה (קודם בוחרים D_0 כך ש־$(1-\frac{\epsilon}{D})^{-D} < 4$ ואז בוחרים δ קטן דיו), ומכאן אפשר להבטיח שבהסתברות $\frac{9}{10}$ לפחות $|V'|$ יהיה בתחום הערכים הנדרש על פי שיטת המומנט השני, כי השונות במקרה זה תהיה חסומה ע"י $(1+\delta_2)me^{-\epsilon}+\delta_3 m^2$.

נותר להוכיח את התנאי על הדרגות. לא ניכנס להוכחה כאן, והנכם מוזמנים לקרוא אותה במהדורה המתאימה של הספר של Alon-Spencer. העיקרון דומה, ומתחיל מכך שלרב הצמתים v מתקיים שרב הקשתות החותכות אותם חותכות קרוב ל־$(r-1)D$ קשתות של H שאינן מכילות את v (עקב התנאים על הדרגות ב־H). לכל צומת "טוב" כזה סופרים את מספר הקשתות המכילות אותו וזרות ל־$\tilde{E}$, ומראים שבסיכוי $1-\delta_4$ (עבור δ_4 מתאים) מספר זה קרוב לתוחלת, על מנת להראות (תוך שימוש באי שוויון מרקוב) שבסיכוי $\frac{9}{10}$ רב הצמתים הטובים ישארו עם דרגות כנדרש.

חסימת סטיות גדולות (large deviation inequalities)

נזכיר כי חסימת סטיות גדולות עוסקת במתן חסמים כמותיים למשפט הגבול המרכזי. בהרצאה ראינו את המקרה הבא: נניח כי $X_1, \ldots X_m$ מקבלים ערכים ב־$\{-1,1\}$ באופן יוניפורמי, ונסמן $X = \sum_{i=1}^{m} X_i$. ברור כי $\mathrm{E}[X] = 0$, אבל היינו רוצים גם לחסום את ההסתברות של סטיה גדולה של X מהתוחלת. בהרצאה ראינו את החסם $\Pr[X > a] < e^{-a^2/2m}$. השימוש הקלאסי בחסם זה הוא כשאנחנו דוגמים משתנים מקריים ומעוניינים לקרב ככל הניתן את ערך התוחלת ה"אמיתי". עבור $a = \omega\left(\sqrt{m}\right)$ נקבל שההסתברות לסטייה של סכום המשתנים מהתוחלת הולכת וקטנה עם מספר הדגימות. הצרה היא כאשר $a = O\left(\sqrt{m}\right)$, ואז הגדלת מספר הדגימות לא תשפר את ההסתברות להצלחה. זו אכן האמת עבור הסיטואציה שהוצגה בהרצאה, אבל במקרים בהם המשתנים מאוד "נדירים", זה נותן הערכה גרועה מדי. בתרגול זה נגזור אי שוויון שיהיה יעיל לסיטואציה כזאת. הטריק יהיה בשימוש חזק יותר בתכונות הקעירות של הפונקציות המעורבות.

נציג סיטואציה קצת יותר כללית: $\Pr[X_i = 1 - p_i] = p_i, \ \ \Pr[X_i = -p_i] = 1 - p_i$ ונסמן $p = \frac{1}{m}\sum_{i=1}^{m} p_i$. ניתן לקבל את הסיטואציה שעסקנו בה בהרצאה על ידי בחירת $p_i = \frac{1}{2}$ והסתכלות על המשתנים המקריים $2X_i$.

נשתמש שוב בפונקציה יוצרת המומנטים

$$\mathrm{E}\left[e^{\lambda X}\right] = \prod_{i=1}^{m} \mathrm{E}\left[e^{\lambda X_i}\right] = \prod_{i=1}^{m}\left(p_i e^{\lambda(1-p_i)} + (1-p_i)\, e^{-\lambda p_i}\right) = e^{-\lambda p m}\prod_{i=1}^{m}\left(p_i e^{\lambda} + (1-p_i)\right)$$

כעת, נחסום את הלוגריתם של המכפלה בצד ימין:

$$\ln\left(\prod_{i=1}^{m}\left(p_i e^{\lambda} + (1-p_i)\right)\right) = \sum_{i=1}^{m}\ln\left(p_i e^{\lambda} + 1 - p_i\right) \le m\ln\left(p e^{\lambda} + 1 - p\right)$$

כאשר אי השוויון האחרון נובע מהקעירות של הפונקציה $\ln\left(x e^{\lambda} + 1 - x\right)$ כאשר $\lambda > 0$ קבוע, ומאי שוויון ינסן (שיופיע שוב בפרק על אנטרופיה). כך, אם ניקח בחזרה חזקה משני צידי אי השוויון נקבל את החסם $\mathrm{E}\left[e^{\lambda X}\right] \le e^{-\lambda p m}\left(p e^{\lambda} + (1-p)\right)^m$, ומאי שוויון מרקוב

$$\Pr\left[X > a\right] = \Pr\left[e^{\lambda X} > e^{\lambda a}\right] < \mathrm{E}\left[e^{\lambda X}\right] e^{-\lambda a} \le e^{-\lambda p m}\left(p e^{\lambda} + (1-p)\right)^m e^{-\lambda a}$$

כעת נקבע $\lambda = \ln(1 + a/pm)$, ובעזרת העובדה ש־$(1 + a/m)^m \le e^a$ נקבל

$$\Pr\left[X > a\right] < e^{a - pm\ln(1+a/pm) - a\ln(1+a/pm)}$$

ואם נפשט עוד, בעזרת אי השוויון $\ln(1 + a/pm) \ge (a/pm) - (a/pm)^2/2$, שנובע מקיצוץ טור טיילור של $\ln\left(1+x\right)$ אחרי שני איבריו הראשונים, נקבל את אי השוויון שרצינו להשיג:

$$\begin{aligned}\Pr\left[X > a\right] &< e^{a - pm\ln(1+a/pm) - a\ln(1+a/pm)} \le \\ &\le e^{a - pm\left((a/pm) - (a/pm)^2/2\right) - a\left((a/pm) - (a/pm)^2/2\right)} = e^{-a^2/2pm + a^3/2(pm)^2}\end{aligned}$$

אם אכן מדובר בסיטואציה בה $p = o\left(1\right)$ אבל $p = \omega\left(\frac{1}{\sqrt{m}}\right)$, אז עבור מקרים בהם $a = \Theta\left(\sqrt{m}\right)$ מקבלים $e^{-a^2/2pm + a^3/2(pm)^2} = o\left(1\right)$, ואכן ההערכה משתפרת עם גידול מספר הדגימות. דוגמה קונקרטית יכולה להיות הערכה של משתנה מקרי בינומי: כזכור, $K \sim B\left(m, p\right)$ מקבל את מספר ניסויי הברנולי המוצלחים מבין m ניסויים בלתי תלויים עם הסתברות הצלחה p. חסימת סטיות גדולות היא כלי קלאסי להערכת המשתנה המקרי הבינומי. במקרה שלנו, $p = o\left(1\right)$ וגם $p = \omega\left(\frac{1}{\sqrt{m}}\right)$, לשם הקונקרטיות נניח $p = \frac{1}{\log m}$.

כזכור, הממוצע של משתנה מקרי בינומי הוא $mp = \frac{m}{\log m}$. נגדיר סדרת משתנים מקריים $X_1, \ldots, X_m$ למטרתנו, כאשר X_i מקבל $1 - \frac{1}{\log m}$ במקרה שהניסוי ה־i הצליח, ו־$-\frac{1}{\log m}$ במקרה שהניסוי ה־i נכשל. כלומר $\sum_{i=1}^{m} X_i = K - \mathrm{E}[K]$. נשתמש באי השוויון שהסקנו כדי להעריך את ההסתברות לסטיה של יותר מ־$\sqrt{m}$ מהתוחלת:

$$\Pr\left[\sum_{i=1}^{m} X_i \geq \sqrt{m}\right] = \Pr\left[K - \frac{m}{\log m} \geq \sqrt{m}\right]$$
$$< e^{-a^2/2pm + a^3/2(pm)^2} = e^{-\log m/2 + \log^2 m/2\sqrt{m}} = o(1)$$

ואכן הסתברות לסטיה כזאת שואפת לאפס כשמספר הניסויים שואף לאינסוף.

חסמי צ'רנוף Chernoff כפליים

המושג "חסם צ'רנוף" משמש כיום כ"מותג" עבור משפחה די גדולה של חסמי סטיות גדולות, כולל כמה שכבר למדנו. נראה כאן מספר חסמים מאוד פופולארים ושימושיים שנהוג להתייחס אליהם בשם זה.

נתחיל מהמשתנים $X_1, \ldots, X_m$ שהוגדרו למעלה עם $p_1, \ldots, p_m$ המתאימים, ושאר הסימונים. כזכור, לקראת סוף הפיתוח הגענו לאי השוויון $\Pr[X > a] \leq e^{a - pm\ln(1+a/pm) - a\ln(1+a/pm)}$. כאן נמשיך לפתח את זה בכיוון קצת שונה: $e^{a - pm\ln(1+a/pm) - a\ln(1+a/pm)} = \left(\frac{e^{a/pm}}{(1+a/pm)^{1+a/pm}}\right)^{pm}$. נהוג לכתוב $\mu = \mathrm{E}[X] = pm$ ו־$\delta = a/\mu$, ואז מקבלים צורה מוכרת של אי השוויון:

$$\Pr[X > \delta\mu] < \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}$$

עם פיתוח דומה למדי (מחליפים את X_i ב־$-X_i$), מקבלים גם כיוון שני:

$$\Pr[X < -\delta\mu] < \left(\frac{e^{\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}$$

עבור שימוש נוח, בד"כ כותבים עבור $\delta \geq 1$ את המסקנה $\Pr[X \geq \delta\mu] < e^{-\delta\mu/3}$, ועבור $0 < \delta \leq 1$ את המסקנות $\Pr[X \geq \delta\mu] < e^{-\delta^2\mu/3}$ ו־$\Pr[X \leq -\delta\mu] < e^{-\delta^2\mu/2}$.

מרטינגלים

כזכור, סדרה $\underline{X}$ של משתנים מקריים היא מרטינגל אם $\mathrm{E}[X_{i+1}|X_0, \ldots, X_i] = X_i$ לכל i. שימו לב כי מדובר בשוויון בין משתנים מקריים. כלומר, בהנתן ערכי המשתנים עד כה, תוחלת הצעד הבא שווה לערך הצעד הנוכחי. דוגמה קלאסית למרטינגל היא הכד של פוליה (Polya). נניח כי יש לנו כד ובו w כדורים לבנים ו־b כדורים שחורים. את התסריט בו מוציאים כדור מהכד, בוחנים את צבעו ומחזירים אותו לכד הכרנו בקורס בסיסי בהסתברות, וגם את התסריט בו מוציאים כדור מהכד, בוחנים את צבעו ולא מחזירים אותו לכד. בכד של פוליה אנחנו מוציאים כדור מהכד, בוחנים את צבעו, ומחזירים אותו לכד יחד עם כדור נוסף באותו הצבע. נסמן ב־$\delta_{n,w}$ את השינוי במספר הכדורים הלבנים לאחר n צעדים, ונגדיר את המשתנה המקרי המנורמל $X_n = \frac{w+\delta_{n,w}}{w+b+n}$. סדרת משתנים זו היא מרטינגל: נשים לב כי אם אנו יודעים את ערכו של X_i, אז מהגדרתו נקבל $\delta_{i,w} = X_i(w+b+i) - w$ כלומר אנחנו יודעים את מספר הכדורים הלבנים (ומכך גם את השחורים) שבכד כעת, שכן w ו־b נקבעו בהתחלה ו־i ידוע. כעת, נחשב את תוחלת X_{i+1} בהנתן ערכו של X_i.

ההסתברות לבחור כדור לבן בזמן זה היא $\frac{w+\delta_{i,w}}{w+b+i}$ ובמקרה זה ערך המשתנה יהיה $\frac{w+\delta_{i,w}+1}{w+b+i+1}$. ההסתברות לבחור כדור שחור בזמן זה היא $\frac{b+i-\delta_{i,w}}{w+b+i} = 1 - \frac{w+\delta_{i,w}}{w+b+i}$ ובמקרה זה ערך המשתנה יהיה $\frac{w+\delta_{i,w}}{w+b+i+1}$. כך

$$\begin{aligned}\mathrm{E}\left[X_{i+1}|X_0,\ldots,X_i\right] &= \left(\frac{w+\delta_{i,w}}{w+b+i}\right)\left(\frac{w+\delta_{i,w}+1}{w+b+i+1}\right)+\left(\frac{b+i-\delta_{i,w}}{w+b+i}\right)\left(\frac{w+\delta_{i,w}}{w+b+i+1}\right)\\ &= \frac{\left(w+\delta_{i,w}\right)\left(w+\delta_{i,w}+1+b+i-\delta_{i,w}\right)}{\left(w+b+i\right)\left(w+b+i+1\right)} = \frac{w+\delta_{i,w}}{w+b+i}\end{aligned}$$

שזה בדיוק ערכו של X_i.

אי שוויון מקדיארמיד ויישום

נביט במקרה פרטי של מרטינגל החשיפה שנותן באופן מיידי מספר תוצאות חזקות. נניח כי המבנה הקומבינטורי שלנו הוא סדרה של n משתנים $Z_1,\ldots Z_n$ המקבלים בהתאמה ערכים $z_1,\ldots,z_n$ באיזה תחום סופי D. המבנה $C = (z_1,\ldots,z_n)$ ניתן לזיהוי עם וקטור מתוך D^n, והחשיפה מתבצעת משתנה־משתנה, כלומר $\mathcal{D}_i = \{Z_1,\ldots,Z_i\}$. אנחנו רוצים לחסום את הסטיה של פונקציה $f : D^n \to \mathbb{R}$ מתוחלתה.

נגדיר מרטינגל חשיפה $\underline{X}$ כך: $X_0 \triangleq \mathrm{E}\left[f\left(Z_1,\ldots,Z_n\right)\right]$, ובכל שלב נחשוף משתנה אחד, כלומר באופן כללי $X_i \triangleq \mathrm{E}\left[f\left(Z_1,\ldots,Z_n\right)|Z_1,\ldots,Z_i\right]$ כאשר התוחלת נלקחת על בחירת $Z_{i+1},\ldots,Z_n$. כך X_i הוא משתנה מקרי שנקבע לפי ערכי המשתנים המקריים $Z_1,\ldots,Z_i$. בהיותו מקרה פרטי של מרטינגל החשיפה שהוצג בהרצאה, $\underline{X}$ מקיים את תנאי חוסר הזכרון והוא מרטינגל.

בדומה למה שראינו בהרצאה, אם אפשר להוסיף את ההנחה כי כל אחד מערכי המשתנים נבחר באופן בלתי תלוי בשני, וכי שינוי בקואורדינטה ה־i של הפונקציה לא יוביל לשינוי של יותר מ־c_i בערכה, אז ניתן להפעיל את אי שוויון אזומה ולקבל לכל $\lambda > 0$ שמתקיים $\Pr\left[f\left(Z_1,\ldots,Z_n\right)-\mu > \lambda\sqrt{\sum_{i=1}^n c_i^2}\right] < e^{-\lambda^2/2}$. כלומר, עבור כל פונקציה שניתן לחסום את השינוי בערכה שעלול להגרם משינוי בקואורדינטה אחת, ניתן גם לקבל חסם הדועך אקספוננציאלית להסתברות שהשמה מקרית לה תסטה מהתוחלת. נעיר (בלי הוכחה) שהתוצאה תקפה גם למקרה בו הערכים $(z_1,\ldots,z_n)$ נלקחים מתחום אינסופי ועם תומך לאו דווקא סופי. מקרה פרטי זה של אי שוויון אזומה נקרא לעיתים אי שוויון מקדיארמיד (McDiarmid) וניתן לגזור ממנה תוצאות רבות בקלות.

בעיית אופטימיזציה קלאסית היא בעיית האריזה בתאים (bin packing). נתונים n משתנים (אצלנו נתייחס למצב שאלו משתנים מקריים) $Z_1,\ldots,Z_n \in [0,1]$, ואנחנו רוצים לארוז אותם בכמה שפחות תאים, כאשר סכום המשתנים בתא נתון הוא לכל היותר 1. במאמר של Rhee, Talagrand מ־1987 הם השתמשו במרטינגל החשיפה על מנת לנתח את הבעיה. נסמן ב־$f\left(z_1,\ldots,z_n\right)$ את הפונקציה המתאימה לערכים $z_1,\ldots,z_n$ את מספר התאים המינימלי בו ניתן לארוז אותם. נגדיר מרטינגל חשיפה על המשתנים כפי שעשינו בפסקה הקודמת.

נניח כי נקבעו כל הערכים, ואנחנו מעוניינים לשנות את ערך המשתנה Z_i. ברור ש־f מונוטונית לא יורדת בכל משתניה, ולכן הערך המקסימלי יתקבל מהשמה $Z_i = 1$. שינוי הערך מ־Z_i ל־1 יוסיף לכל היותר 1 לערך הפונקציה f, שכן נוכל להשתמש בסידור הקיים של יתר המשתנים בתאים, ולארוז את המשתנה Z_i בתא נפרד. הערך המינימלי יתקבל מהשמה $Z_i = 0$, ושינוי הערך כך יחסר לכל היותר 1 מערך הפונקציה f, שכן הדבר שקול לאריזת המשתנים מבלי לארוז את Z_i, ואריזה כזאת אפשר להפוך לאריזה הכוללת גם את Z_i על ידי הוספת תא מיוחד לארוז בו אותו, במקרה הגרוע ביותר. לכן שינוי בקואורדינטה אחת של הפונקציה מביא לשינוי בערך של f ב־1 לכל היותר. כך מאי שוויון מקדיארמיד לכל $\lambda > 0$ מתקיים $\Pr\left[f\left(Z_1,\ldots,Z_n\right)-\mu > \lambda n\right] < e^{-\lambda^2/2}$. כלומר, פתרון אופטימלי לקלט מקרי של בעיית האריזה בתאים קרוב בערכו, בהסתברות גבוהה, לתוחלת הפתרון האופטימלי על פני ההשמות המקריות האפשריות.

חסימת מרטינגל חשיפה של פרמוטציה

נניח שאנחנו דנים בפונקציה מספרית f של קבוצת כל הפרמוטציות מעל $\{1,\ldots,n\}$, ונניח שאנחנו בונים עבורה מרטינגל חשיפה של פרמוטציה σ המוגרלת יוניפורמית (מבין $n!$ האפשרויות), כאשר $\mathcal{D}_i = \{1,\ldots,i\}$.

היינו רוצים שיטה כללית לחסום את ההפרש $|X_i - X_{i-1}|$ כפי שהדבר נעשה למרטינגל חשיפה בכיתה, אולם כאן יש לנו בעיה עם זה שההתפלגות של הפרמוטציה σ אינה מקיימת אי תלות בין ערכי $\sigma(i)$, מכיוון שעליהם להיות שונים זה מזה. נראה שאם f מקיימת שלכל זוג פרמוטציות σ ו־σ' המתקבלות זו מזו ע"י החלפת שני ערכים מתקיים $|f(\sigma) - f(\sigma')| \leq c$, אז מתקיים גם התנאי שאנחנו רוצים, $|X_i - X_{i-1}| \leq c$ לכל $1 \leq i \leq n$. דוגמה לפונקציה כזו עם $c = 1$ היא הפונקציה הסופרת את מספר העגילים הזרים בפירוק של σ.

לשם כך ראשית נראה עבור כל $j_1, \ldots j_{i-1}$ שונים זה מזה, ו־j, k שונים זה מזה ומ־$j_1, \ldots, j_{i-1}$, שההפרש $|\mathrm{E}[f(\sigma)|\sigma(1) = j_1, \ldots, \sigma(i-1) = j_{i-1}, \sigma(i) = j] - \mathrm{E}[f(\sigma)|\sigma(1) = j_1, \ldots, \sigma(i-1) = j_{i-1}, \sigma(i) = k]|$ חסום ע"י c. בשביל זה נראה התאמה חח"ע ועל בין כל הפרמוטציות ש־i הערכים הראשונים שלהן הם $j_1, \ldots, j_{i-1}, j$ לבין כל הפרמוטציות ש־i הערכים הראשונים שלהן הם $j_1, \ldots, j_{i-1}, k$. בהינתן פרמוטציה σ השייכת לקבוצה הראשונה, נגדיר את σ' באופן הבא: נבחר את ה־l עבורו $\sigma(l) = k$, ונשים לב שמתקיים $l \in \{i+1, \ldots n\}$ (הנחנו שגם k אינו בין $j_1, \ldots j_{i-1}$). נגדיר את $\sigma'(i) = k$, את $\sigma'(l) = j$, ושאר ערכי σ' יהיו זהים לאלו של σ. לא קשה לראות ש־σ' היא פרמוטציה המתקבלת מ־σ ע"י החלפת שני ערכים, ושההתאמה הזו היא חח"ע ועל בקבוצת כל הפרמוטציות הנ"ל.

נשים עתה לב שהתוחלת $\mathrm{E}[f(\sigma)|\sigma(1) = j_1, \ldots, \sigma(i-1) = j_{i-1}, \sigma(i) = k]$ זהה לחלוטין לתוחלת $\mathrm{E}[f(\sigma')|\sigma(1) = j_1, \ldots, \sigma(i-1) = j_{i-1}, \sigma(i) = j]$, בגלל שהמדובר בהעתקה חח"ע ועל. כמו כן, לכל σ ש־i הערכים הראשונים שלה הם $j_1, \ldots, j_{i-1}, j$, מתקיים $|f(\sigma) - f(\sigma')| \leq c$ לפי מה שהנחנו על f. משני הנתונים האלו נובע החסם המבוקש על ההפרש.

לסיום ניזכר בהגדרה של המשתנים המקריים של המרטינגל כפונקציות של מבנה נתון. במקרה זה עבור פרמוטציה $\tilde{\sigma}$ (שאח"כ מגרילים אותה) מגדירים: $X_i = a_i(\tilde{\sigma}) = \mathrm{E}[f(\sigma)|\sigma(1) = \tilde{\sigma}(1), \ldots, \sigma(i) = \tilde{\sigma}(i)]$. לפי זה נכתוב את $a_{i-1}(\tilde{\sigma})$:

$$\begin{aligned} a_{i-1}(\tilde{\sigma}) &= \mathrm{E}[f(\sigma)|\sigma(1) = \tilde{\sigma}(1), \ldots, \sigma(i-1) = \tilde{\sigma}(i-1)] \\ &= \frac{1}{n+1-i} \sum_{k \in \{1,\ldots,n\} \setminus \{\tilde{\sigma}(1),\ldots,\tilde{\sigma}(i-1)\}} \mathrm{E}[f(\sigma)|\sigma(1) = \tilde{\sigma}(1), \ldots, \sigma(i-1) = \tilde{\sigma}(i-1), \sigma(i) = k] \end{aligned}$$

ראינו כאן ש־$a_{i-1}(\tilde{\sigma})$ הוא ממוצע של ערכים שכל אחד מהם נמצא במרחק של לא יותר מ־c מהערך של $a_i(\tilde{\sigma})$, ולכן גם $a_{i-1}(\tilde{\sigma})$ עצמו נמצא במרחק של לא יותר מ־c מ־$a_i(\tilde{\sigma})$, כנדרש.

המרטינגל האדפטיבי

נביט במרחב וקטורי V ונניח כי נתונים לנו סדרת וקטורים $v_1, \ldots, v_n \in V$. אנחנו בוחרים תת קבוצה $I \subseteq [n]$ באקראי (כלומר בוחרים כל אינדקס בהסתברות $\frac{1}{2}$ באופן ב"ת באחרים, וצמתחים את מימד המרחב הנפרש על ידי הוקטורים בקבוצה $v_I = \{v_k | k \in I\}$. נסמן את $\mathrm{E}[\dim(v_I)]$ ב־ρ ואת $\dim(v_{[n]})$ ב־d. ברור כי $\rho < d$, שכן יש הסתברות חיובית שהמימד לא יהיה מלא (למשל אם לא יבחר אף וקטור). כמו כן $d/2 \leq \rho$: נקבע בסיס למרחב הנפרש B, ונביט במשתנה המקרי שהוא מספר הוקטורים מ־B המופיעים ב־v_I. זה בבירור חסם תחתון ל־$\dim(v_I)$, ולכן תוחלתו, שהיא $d/2$, היא חסם תחתון ל־ρ.

אנחנו מעוניים להראות כי בהסתברות גבוהה מימד המרחב הנפרש על ידי הוקטורים שבחרנו יהיה קרוב ל־ρ. אם נשתמש באי שוויון אזומה עבור מרטינגל חשיפה רגיל, נקבל שרק בהסתברות $o(1)$ מימד זה יכול יחרוג מ־ρ כדי יותר מ־$O(\sqrt{n})$. עם זאת, נדמה כי d הוא המאפיין הנכון יותר לבעיה, ובמקרה שבו d ו־ρ קטנים בהרבה מ־n סטיה מסדר גודל של $\sqrt{n}$ גם היא לא צריכה להיות סבירה. היינו רוצים לקבל חסם על הסטיה במונחי d. אינטואטיבית, אם נביט במרטינגל חשיפה על תוחלת המימד החושף את הוקטורים שנבחרו, אז ברור שנשנה את תוחלת המימד רק אם נחשוף וקטור שאינו תלוי באלו שנחשפו עד כה. יש מעט וקטורים כאלה, ואחרי שנחשוף את כולם לא ישתנה עוד הערך שחושפים. אם כך, נרצה להגדיר מרטינגל שראשית יחשוף את הוקטורים שאינם תלויים בקודמים, ואז את כל היתר. נגדיר במפורש את הרעיונות הללו — בהגדרות להלן,

מדובר במרחב הסתברות μ מעל קבוצת כל הפונקציות $f:\mathcal{D}\to\mathcal{R}$, שנסמן ב־$\mathcal{C}$, ואנו מנסים לחסום את הסטיה מהתוחלת של פונקציונל ממשי $\mathcal{F}:\mathcal{C}\to\mathbb{R}$ המוגדר עבור פונקציות אלו.

הגדרה (סכמת חשיפה): נסמן ב־G את קבוצת כל הפונקציות מתת קבוצה כל שהיא של $\mathcal{D}$ ל־$\mathcal{R}$. סכמת חשיפה (אדפטיבית) היא משפחה (עם חזרות) של תתי קבוצות של $\mathcal{D}$ עם אינדכסים ב־G, כך שלכל $g:\mathcal{D}'\to\mathcal{R}$ (עבור $\mathcal{D}'\subseteq\mathcal{D}$ כל שהוא) מתקיים $\mathcal{D}'\subseteq\mathcal{D}_g$, וההכלה היא ממש אלא אם כן תחום זה הוא $\mathcal{D}$.

באופן כללי לא נדרוש איבר $\mathcal{D}_g$ לכל פונקציה g, אבל נדרוש $\mathcal{D}_\epsilon$ המוגדר עבור הפונקציה הריקה $\epsilon:\emptyset\to\mathcal{R}$, ואם $\mathcal{D}_h$ מוגדר אז $\mathcal{D}_g$ צריך להיות מוגדר לכל פונקציה $g:\mathcal{D}_h\to\mathcal{R}$ המרחיבה את h. נניח כי $\mathcal{D}_g$ שלא הוגדר שווה ל־$\mathcal{D}$.

הגדרה (סידרת חשיפה): בהנתן $\mathcal{D},\mathcal{R},\mu,\mathcal{F}$, סכמת חשיפה $\mathrm{D}=\{\mathcal{D}_f:\mathcal{D}'\subseteq\mathcal{D},f:\mathcal{D}'\to\mathcal{R}\}$ ופונקציה $h:\mathcal{D}\to\mathcal{R}$, נגדיר את $\mathcal{D}_0(h),\mathcal{D}_1(h),\ldots$ ואת $a_0(h),a_1(h),\ldots$ באופן הבא.

1. לכל $\mathcal{D}'\subseteq\mathcal{D}$ נגדיר את פונקצית התוחלת המותנה $a_{\mathcal{D}'}(h)=\mathrm{E}_{f\sim\mu}\left[\mathcal{F}(f)\,|f|_{\mathcal{D}'}=h|_{\mathcal{D}'}\right]$.
2. מגדירים $\mathcal{D}_0(h)=\emptyset$ ובהתאמה $a_0(h)=a_{\mathcal{D}_0(h)}(h)=\mathrm{E}_{f\sim\mu}\left[\mathcal{F}(f)\right]$.
3. בהנתן $\mathcal{D}_{i-1}(h)$ מגדירים אינדוקטיבית $\mathcal{D}_i(h)=\mathcal{D}_{h|_{\mathcal{D}_{i-1}(h)}}$ ו־$a_i(h)=a_{\mathcal{D}_i(h)}(h)$.

נשים לב כי $\mathcal{D}_{i-1}(h)\subseteq\mathcal{D}_i(h)$ עם שוויון אם ורק אם $\mathcal{D}_{i-1}(h)=\mathcal{D}$. כמו כן נשים לב שתמיד מתקיים $\mathcal{D}_{|\mathcal{D}|}(h)=\mathcal{D}$ ולכן $a_{|\mathcal{D}|}(h)=\mathcal{F}(h)$ לכל $h:\mathcal{D}\to\mathcal{R}$ (יש סכימות חשיפה עבורן ניתן להבטיח זאת לאינדכסים קטנים יותר, כגון אלו הקשורות בחשיפת צמתים של גרף).

הגדרה (מרטינגל חשיפה אדפטיבי): בהנתן $\mathcal{D},\mathcal{R},\mu,\mathcal{F}$ וסכמת חשיפה $\mathrm{D}=\{\mathcal{D}_f:\mathcal{D}'\subseteq\mathcal{D},f:\mathcal{D}'\to\mathcal{R}\}$, מרטינגל החשיפה האדפטיבי של $\mathcal{F}$ מעל μ לפי D הוא סדרה של משתנים מקריים $\underline{X}=(X_0,X_1,\ldots)$ שערכיהם נקבעים לפי התהליך הבא:

1. בוחרים את $\hat{f}:\mathcal{D}\to\mathcal{R}$ לפי ההתפלגות μ.
2. לכל $i\geq 0$ קובעים עתה את הערך $X_i=a_i(\hat{f})$.

ההבדל בין הגדרה זו להגדרה הרגילה של מרטינגל החשיפה של Doob היא שהתחומים $\mathcal{D}_i$ עשויים להיות תלויים גם הם בפונקציה $\hat{f}$ שבחרנו לפי μ. ההגדרה האחרונה שנצטרך:

הגדרה (תנאי ליפשיץ ביחס לסכמת חשיפה): נאמר כי $\mathcal{F}$ היא ליפשיץ ביחס ל־D אם לכל $f,g:\mathcal{D}\to\mathcal{R}$ המקיימות כי הן מזדהות על $(\mathcal{D}\setminus\mathcal{D}_{i+1}(f))\cup\mathcal{D}_i(f)$ עבור i כלשהו מתקיים כי $|\mathcal{F}(f)-\mathcal{F}(g)|\leq 1$.

נוח יותר להשתמש בתנאי החזק יותר: לכל $h:\mathcal{D}'\to\mathcal{R}$ ו־$f,g:\mathcal{D}\to\mathcal{R}$ (כאשר $\mathcal{D}'\subseteq\mathcal{D}$) המסכימות עם h על $\mathcal{D}'$ ומסכימות זו עם זו על $\mathcal{D}\setminus\mathcal{D}_h$ מתקיים $|\mathcal{F}(f)-\mathcal{F}(g)|\leq 1$.

בדומה למקרה שכבר הכרנו, ניתן להראות שסדרת משתנים מקריים כזאת היא אכן מרטינגל, ושאם μ הוא כזה שכל ערך של הפונקציה נבחר מבלי תלות באחרים ו־$\mathcal{F}$ היא ליפשיץ ביחס ל־D, אז גם המרטינגל מקיים את תנאי ההפרשים החסומים. נחזור אם כך למקרה איתו התחלנו. אנו מעוניינים להראות:

משפט: לכל $\alpha>0$ קיים $\beta>0$ כך ש־$\mathrm{Pr}_I\left[|\dim(v_I)-\rho|>\alpha d\right]<e^{-\beta d}$.

נניח בלי הגבלת הכלליות כי $n>10d$, כי אחרת ניתן להשתמש באי שוויון אזומה ומרטינגל חשיפה רגיל. נקבע $\mathcal{D}=[n]\,,\mathcal{R}=\{0,1\}$ ו־μ תהיה ההתפלגות על תתי הקבוצות כפי שהוגדרה קודם. על מנת להגדיר את סכמת החשיפה D, תהא $h:\mathcal{D}'\to\mathcal{R}$ פונקציה עבורה אנחנו רוצים להגדיר את $\mathcal{D}_h$, ונסמן ב־J_h את האינדקסים החברים בקבוצה המתאימה לה. נבחין בין שני מקרים:

- אם $[n]\setminus\mathcal{D}'$ מכיל אינדקס j_h עבורו v_{j_h} בלתי תלוי ב־v_{J_h}, נבחר j_h כזה ונקבע $\mathcal{D}_h=\mathcal{D}'\cup\{j_h\}$.
- אם אין j_h כזה, אז בהכרח $\dim\left(v_{J_h}\right)=\dim\left(v_{J_h\cup(\mathcal{D}\setminus\mathcal{D}')}\right)$. במקרה הזה נקבע $\mathcal{D}_h=\mathcal{D}$.

זו בבירור סכמת חשיפה כפי שהגדרנו. כעת נסמן ב־$\underline{X} = (X_1, \ldots, X_n)$ את מרטינגל החשיפה ל $\dim(V_I)$ מעל סכמה זו. זה מרטינגל כפי שכבר ציינו. נראה כי הוא ליפשיץ ביחס ל־D:

במקרה הראשון ליצירת $\mathcal{D}_h$ מתקיים $|\mathcal{D}_h \setminus \mathcal{D}'| = 1$, וכך אם f, g נבדלים לכל היותר בקואורדינטה אחת אז מימדי קבוצות הוקטורים המתאימות יבדל גם כן לכל היותר ב־1. במקרה השני נשים לב כי אם f, g מזדהות עם h על $\mathcal{D}'$ אז המימד של שתי קבוצות הוקטורים המתאימות שווה ל־$\dim(v_{J_h})$ בכל מקרה.

לסיכום נראה כיצד ניתן "לקצר" את המרטינגל כדי לקבל ריכוז במונחי d ולא במונחי n, כלומר נראה כי בהסתברות לפחות $1 - e^{-d}$ מתקיים כי $X_n = X_{10d}$:

תהא $\hat{f}$ פונקציה הנבחרת באופן אקראי כבהגדרת מרטינגל החשיפה. לכל אינדקס $0 \le i \le 10d$ נסמן $J_i = \left\{k : \hat{f}(k) = 1\right\} \cap \mathcal{D}_i(\hat{f})$ ו־$d_i = \dim(v_{J_i})$. כך, אם $\mathcal{D}_i(\hat{f}) = \mathcal{D}$ אז $X_i(\hat{f}) = X_n(\hat{f})$. אם $X_i(\hat{f}) \neq X_n(\hat{f})$ אז מהאופן שבחרנו את D אנו יודעים כי $\mathcal{D}_i(\hat{f}) \setminus \mathcal{D}_{i-1}(\hat{f})$ מכילה איבר בודד, שנסמן ב־j_i. כיוון ש־$\hat{f}(j_i)$ נבחר באופן בלתי תלוי ב־$\hat{f}|_{\mathcal{D}_{i-1}}$, הוא שווה ל־1 בהסתברות $\frac{1}{2}$, ללא תלות בערכי $J_1, \ldots, J_{i-1}$. מהתנאי על אי תלות i ו־j עולה כי בהסתברות $\frac{1}{2}$ יש לנו $d_i = d_{i-1} + 1$, ללא תלות בערכים הקודמים. כך, ההסתברות ל־$X_{10d} \neq X_n$ חסומה על ידי ההסתברות ש־$10d$ הטלות מטבע יוניפורמיות יסתכמו בפחות מ־d, ומחסימת סטיות גדולות היא נמוכה מ־e^{-d}. כעת, מאי שוויון אזומה $\Pr\left[|X_0 - X_{10d}| > \alpha d\right] < 2e^{-\alpha^2 d/20}$, ולכן מחסם האיחוד מעל שני מאורעות "רעים" אלה אנחנו מקבלים את הדרוש, שכן $X_0 = \rho$ ו־X_n מתפלג כמו $\dim(v_I)$.

למעוניינים נציין כי הצגה שונה של טכניקה זו מופיעה בספר של Alon, Spencer כמשפט 7.4.3.

הפרדיגמה של פואסון

כאשר אנחנו דנים בסדרת משתנים מקריים שהם "בלתי תלויים למדי" ו"נדירים", היינו רוצים לאמר שהתפלגותם דומה לזו של משתנה מקרי פואסוני. זה בניגוד למקרה הרגיל שבו אנחנו מתבססים על כך שהתפלגותם דומה למשתנה מקרי נורמלי. נפרמל את האינטואיציה הזאת בעזרת אי שוויון ינסון (Janson), אבל ראשית נגדיר את הסיטואציה במדויק:

נסמן ב־Ω את העולם הסופי שלנו, ונגדיר $R \subset \Omega$ שנבחר באופן הבא: $\Pr\left[r \in R\right] = p_r$ כאשר כל איבר ב־Ω נבחר להיות ב־R בהגרלה בלתי תלויה באיברים האחרים. נסמן ב־$\{A_i\}_{i \in I}$ אוסף של תת קבוצות של Ω, וב־B_i את המאורעות המתאימים להם, כלומר B_i הוא המאורע ש־$A_i \subseteq R$. נגדיר בהתאמה X_i כמשתנה האינדיקטור של B_i ואת $X = \sum_{i \in I} X_i$, מספר הקבוצות שמקיימות $A_i \subseteq R$. לכל $i, j \in I$ נסמן $i \sim j$ אם $i \neq j$ וכן $A_i \cap A_j \neq \emptyset$. כלומר אם $i \nsim j$ ו־$i \neq j$ אז B_i, B_j הם מאורעות בלתי תלויים. נוסף על כך, אם $i \notin J \subset I$ וגם לכל $j \in J$ מתקיים $i \nsim j$, אז B_i בלתי תלוי בכל צירוף של $\{B_j\}_{j \in J}$. זאת פשוט מכיוון שהם נקבעים על ידי הגרלות שונות ובלתי תלויות.

נגדיר את "הערך שהיה ל־$\Pr\left[\bigwedge_{i \in I} \neg B_i\right]$ לו היו ה־B_i בלתי תלויים", $M \triangleq \prod_{i \in I} \Pr\left[\neg B_i\right]$, ומדד לתלות המאורעות $\Delta \triangleq \sum_{i \sim j} \Pr\left[B_i \wedge B_j\right]$, כאשר כאן נספרים הזוגות כסדורים, כלומר יש ספירה כפולה. נסמן גם כרגיל $\mu = \mathrm{E}\left[X\right] = \sum_{i \in I} \Pr\left[B_i\right]$. אי שוויון ינסון מפרמל את האינטואיציה שאם המאורעות "מאוד לא סבירים" אז ההתנהגות של איחודם דומה לזו של משתנה מקרי פואסוני.

אבחנה פשוטה היא ש־$M \le e^{-\mu}$: נשים לב כי $\Pr\left[\neg B_i\right] = 1 - \Pr\left[B_i\right] \le e^{-\Pr[B_i]}$ וכך נקבל את החסם $M = \prod_{i \in I} \Pr\left[\neg B_i\right] \le \prod_{i \in I} e^{-\Pr[B_i]} = \exp\left(-\sum_{i \in I} \Pr\left[B_i\right]\right) = \exp\left(-\mu\right)$.

אי שוויון ינסון: בסימונים לעיל, אם לכל $i \in I$ מתקיים החסם $\Pr\left[B_i\right] \le \epsilon$, אז מתקיימים אי השוויונות $M \le \Pr\left[\bigwedge_{i \in I} \neg B_i\right] \le M e^{(\Delta/2(1-\epsilon))}$ וכן $\Pr\left[\bigwedge_{i \in I} \neg B_i\right] \le e^{-\mu + \Delta/2}$.

הוכחה: ראשית עלינו לנצל את אי השוויון הבא: לכל תת קבוצה $J \subset I$ כך ש־$i \notin J$ מתקיים כי $\Pr\left[B_i \middle| \bigwedge_{j \in J} \neg B_j\right] \le \Pr\left[B_i\right]$. נותיר את אי השוויון הזה לעת עתה ללא הוכחה, שכן הוא נובע ממשפט

FKG שיוכח בהמשך הקורס. כעת, נניח בלי הגבלת הכלליות כי $I=[m]$, ונביט בקבוצת האינדקסים הקטנים ממש מ־i. קבוצה זו בוודאי לא מכילה את i, וכך מאי השוויון לעיל נקבל $\Pr\left[B_i|\bigwedge_{1\leq j<i}\neg B_j\right]\leq\Pr\left[B_i\right]$, ועל ידי מעבר למאורעות המשלימים $\Pr\left[\neg B_i|\bigwedge_{1\leq j<i}\neg B_j\right]\geq\Pr\left[\neg B_i\right]$. לפי חוק בייס

$$\Pr\left[\bigwedge_{i\in I}\neg B_i\right]=\prod_{i=1}^{m}\Pr\left[\neg B_i|\bigwedge_{1\leq j<i}\neg B_j\right]\geq\prod_{i=1}^{m}\Pr\left[\neg B_i\right]=M$$

וכך מתקבל החסם התחתון. נשים לב שאי השוויון שהשתמשנו בו קודם תקף גם את נתנה את שני צדדיו במאורע B_k עבור $k\notin J$ כך שלכל $l\in J$ מתקיים $k\nsim l$, שכן מכיוון שהמאורע B_k בלתי תלוי במאורעות $\{B_j\}_{j\in J}$, התניה כזאת שקולה לקביעת $p_r=1$ לאיברים ב A_k וזו אינה משפיעה על ההסתברויות הרלוונטיות. בזאת מקבלים לכל $J\subset I$ כך ש־$i\in J$, $k\notin J$ ולכל $l\in J$ מתקיים $k\nsim l$, את אי השוויון $\Pr\left[B_i|B_k\wedge\bigwedge_{j\in J}\neg B_j\right]\leq\Pr\left[B_i|B_k\right]$.

כעת נעבור לחסם העליון. עבור i נתון, נסמן ב־D_i את קבוצת ה j־ים הקטנים מ־i המקיימים $j\sim i$. מחוק בייס, לכל שלושה מאורעות A,B,C מתקיים $\Pr\left[A|B\wedge C\right]\geq\Pr\left[A\wedge B|C\right]$. נציב את המאורעות הרלוונטיים לנו כאן $A=B_i,B=\bigwedge_{j\in D_i}\neg B_j,C=\bigwedge_{j\notin D_i,j<i}\neg B_j$, ונשים לב כי A לא תלוי ב C ואז:

$$\begin{aligned}\Pr\left[B_i|\bigwedge_{1\leq j<i}\neg B_j\right]&=\Pr\left[B_i|\bigwedge_{j\in D_i}\neg B_j\wedge\bigwedge_{k\notin D_i,k<i}\neg B_k\right]\\&\geq\Pr\left[B_i\wedge\bigwedge_{j\in D_i}\neg B_j|\bigwedge_{k\notin D_i,k<i}\neg B_k\right]\\&=\Pr\left[B_i|\bigwedge_{j\notin D_i,j<i}\neg B_j\right]\Pr\left[\bigwedge_{j\in D_i}\neg B_j|B_i\wedge\bigwedge_{k\notin D_i,k<i}\neg B_k\right]\\&=\Pr\left[B_i\right]\Pr\left[\bigwedge_{j\in D_i}\neg B_j|B_i\wedge\bigwedge_{k\notin D_i,k<i}\neg B_k\right]\end{aligned}$$

נמשיך ונחסום:

$$\Pr\left[\bigwedge_{j\in D_i}\neg B_j|B_i\wedge\bigwedge_{k\notin D_i,k<i}\neg B_k\right]\geq1-\sum_{j\in D_i}\Pr\left[B_j|B_i\wedge\bigwedge_{k\notin D_i,k<i}\neg B_k\right]\geq1-\sum_{j\in D_i}\Pr\left[B_j|B_i\right]$$

כאשר המעבר האחרון הוא מאי השוויון השני שהצגנו בתחילת ההוכחה. נחבר הכל יחדיו ונקבל

$$\Pr\left[B_i|\bigwedge_{1\leq j<i}\neg B_j\right]\geq\Pr\left[B_i\right]\left(1-\sum_{j\in D_i}\Pr\left[B_j|B_i\right]\right)=\Pr\left[B_i\right]-\sum_{j\in D_i}\Pr\left[B_j\wedge B_i\right]$$

נעבור למאורעות המשלימים:

$$\Pr\left[\neg B_i|\bigwedge_{1\leq j<i}\neg B_j\right]\leq\Pr\left[\neg B_i\right]+\sum_{j\in D_i}\Pr\left[B_j\wedge B_i\right]\leq\Pr\left[\neg B_i\right]\left(1+\frac{1}{1-\epsilon}\sum_{j\in D_i}\Pr\left[B_j\wedge B_i\right]\right)$$

כאשר המעבר האחרון מוצדק מכוון ש $\Pr\left[\neg B_i\right]\geq1-\epsilon$. כעת, נשתמש ב $1+x\leq e^x$ כדי להסיק $\Pr\left[\neg B_i|\bigwedge_{1\leq j<i}\neg B_j\right]\leq\Pr\left[\neg B_i\right]\exp\left(\frac{1}{1-\epsilon}\sum_{j\in D_i}\Pr\left[B_j\wedge B_i\right]\right)$. לבסוף נציב זאת לכל $1\leq i\leq m$

לתוך צד ימין של:

$$\begin{aligned}\Pr\left[\bigwedge_{i\in I}\neg B_i\right] &= \prod_{i=1}^{m}\Pr\left[\neg B_i|\bigwedge_{1\le j<i}\neg B_j\right]\\ &\le \prod_{i=1}^{m}\left(\Pr\left[\neg B_i\right]\exp\left(\frac{1}{1-\epsilon}\sum_{j\in D_i}\Pr\left[B_j\wedge B_i\right]\right)\right)\\ &= \prod_{i=1}^{m}\Pr\left[\neg B_i\right]\prod_{i=1}^{m}\exp\left(\frac{1}{1-\epsilon}\sum_{j\in D_i}\Pr\left[B_j\wedge B_i\right]\right)\\ &= M\exp\left(\frac{1}{1-\epsilon}\sum_{i=1}^{m}\sum_{j\in D_i}\Pr\left[B_j\wedge B_i\right]\right)\end{aligned}$$

לפי בחירת הקבוצות D_i האיברים באקספוננט מסתכמים ל-$\Delta/2$ ומתקבל החסם העליון הראשון. על מנת לקבל את החסם העליון השני נחסום כל איבר במכפלה באופן הבא:

$$\begin{aligned}\Pr\left[\neg B_i|\bigwedge_{1\le j<i}\neg B_j\right] &\le 1-\Pr\left[B_i\right]+\sum_{j\in D_i}\Pr\left[B_j\wedge B_i\right]\\ &\le \exp\left(-\Pr\left[B_i\right]+\sum_{j\in D_i}\Pr\left[B_j\wedge B_i\right]\right)\end{aligned}$$

וכאשר נחזור למכפלה, חזקות האקספוננט יסתכמו, כמו קודם. האיבר השני בחזקה יסתכם ל $\Delta/2$, והאיבר הראשון יסתכם פשוט ל $-\mu$.

ישום לגרפים מקריים

תכונה בסיסית בחקר גרפים היא היותם חסרי משולשים. נניח כי אנחנו בוחרים גרף לפי ההתפלגות $G(n,p)$ ורוצים לחשב את ההסתברות שהגרף חסר משולשים. עבור שלושה צמתים נתונים u,v,w, ההסתברות שלא יהיה ביניהם משולש היא $1-p^3$. היינו רוצים להסיק מכך שההסתברות שלא יהיו כלל משולשים בגרף היא $\left(1-p^3\right)^{\binom{n}{3}}$, אבל טענה זו אינה נכונה, שכן המאורעות אינם בלתי תלויים (ואכן בהסתברות $(1-p)^{\binom{n}{2}}$ הגרף יהיה ריק ובפרט חסר משולשים). אם נביט בצומת נוסף, z, אז בבירור יש תלות גבוהה בין המאורעות "u,v,w הם משולש" ו"u,v,z הם משולש". נשתמש באי שוויון ינסון על מנת לכמת את התלות הזאת:

הקבוצה R אצלנו היא קבוצת הקשתות האפשריות בגרף המקרי, ולכל קשת אפשרית הסתברות שווה של p להיות בקבוצה. תתי הקבוצות A_i הן כל שלשות הקשתות בגרף המתאימות למשולשים האפשריים. נחשב את $\Delta=\sum_{i\sim j}\Pr\left[B_i\wedge B_j\right]$: נקבע את i, כלומר שלשת צמתים a,b,c. המאורעות שעבורם $i\sim j$ הם אלה החולקים קשתות עם B_i. ישנם שלושה צמתים הקובעים את B_i, ומאורע יחלוק איתו קשתות אם ורק אם הוא יחלוק איתו שניים מהצמתים. לכן יש $3(n-3)$ מאורעות כאלה. נביט במאורע כזה, B_j, ונניח כי הוא נקבע על ידי הצמתים a,b,d. על מנת ששני המאורעות יקרו צריכות להתקיים חמש קשתות $(a,b),(b,c),(b,d),(c,a),(d,a)$, כלומר ההסתברות לכך היא p^5. עבור i נתון סכום ההסתברויות של המאורע $B_i\wedge B_j$ לכל $i\sim j$ הוא $p^5 3(n-3)$, ובסה"כ $\Delta=\sum_{i\sim j}\Pr\left[B_i\wedge B_j\right]=3\binom{n}{3}p^5(n-3)$. כמו כן, לכל i מתקיים $\Pr\left[B_i\right]=p^3$, וכאמור למעלה $M=\left(1-p^3\right)^{\binom{n}{3}}$. אי שוויון ינסון נותן לנו

$$\left(1-p^3\right)^{\binom{n}{3}}\le\Pr\left[\bigwedge\neg B_i\right]\le\left(1-p^3\right)^{\binom{n}{3}}e^{3\binom{n}{3}p^5(n-3)/2\left(1-p^3\right)}$$

.

עבור $p=1/n$ נקבל לדוגמה

$$\Pr\left[\bigwedge\neg B_i\right]\le\left(1-\frac{1}{n^3}\right)^{\binom{n}{3}}\cdot e^{\frac{n^4\cdot n^{-5}}{4\left(1-n^{-3}\right)}}=\left(1-\frac{1}{n^3}\right)^{\binom{n}{3}}e^{\Theta\left(n^{-1}\right)}$$

כלומר, עבור n גדול נקבל שהחישוב חסר התלות קרוב לערך הנכון. לעומת זאת, עבור $p=1/2$ נקבל

$$\Pr\left[\bigwedge\neg B_i\right]\le\left(1-\frac{1}{2^3}\right)^{\binom{n}{3}}e^{\frac{8\cdot 3\cdot n^4}{6\cdot 7\cdot 2\cdot 2^5}}=\left(\frac{7}{8}\right)^{\binom{n}{3}}e^{\Theta\left(n^4\right)}=\omega\left(1\right)$$

הסטיה מהערך של החישוב חסר התלות הולכת לאינסוף ולמעשה לא קיבלנו כל מידע על המצב (עם זאת ברור שהחסם התחתון אינו הדוק). אם נרצה לדאוג לסטיה קבועה לכל היותר מהערך של החישוב חסר התלות, נצטרך לדאוג שהחזקה באקספוננט תהיה קבועה, כלומר $3\binom{n}{3}p^5\left(n-3\right)/2\left(1-p^3\right)\le c$, ואם נפריד בין p ל n נקבל $p^5/\left(1-p^3\right)\le\frac{c}{n^4}$ (עם שינוי קטן בקבוע). אם נניח $p=n^{-\delta}$ אז $n^{-5\delta}/\left(1-n^{-3\delta}\right)\le\frac{c}{n^4}$, ועבור n גדול דיו ניתן להפטר מהמכנה בצד שמאל, ולפשט את הביטוי במחיר הגדלה של הקבוע, ולקבל $n^{-5\delta}\le\frac{c}{n^4}$, כלומר $n^{4-5\delta}$ צריך להיות חסום על ידי קבוע. מכך אנחנו רואים שבשיטה זו נוכל לקבל חסם עבור כל $\delta\ge\frac{4}{5}$.

מקרה נוח לשימוש של הגרסה הלא־סימטרית של הלמה הלוקלית הכללית

עבור הלמה הלוקלית הלא סימטרית יש ניסוח "קל לשימוש" שמזכיר את המקרה הסימטרי, ומכסה מקרה פרטי מאוד נפוץ של שימוש לא סימטרי בלמה: אם נתונים מאורעות $B_1,\dots,B_m$ ורשימת תלויות עבורם $D_1,\dots,D_m$, כך שלכל i מתקיים $\Pr[B_i]<\frac{1}{2}$ וכן $\sum_{j\in D_i}\Pr[B_j]\le\frac{1}{4}$, אז מתקיים $\Pr[\bigwedge_{i=1}^m\neg B_i]>0$.

הוכחה: לכל $1\le i\le m$ נגדיר $x_i=2\Pr[B_i]$, ונוודא ישירות את קיום תנאי הלמה הלוקלית הלא־סימטרית עבור $x_1,\dots,x_m$.

$$x_i\prod_{j\in D_i}(1-x_j)=2\Pr[B_i]\prod_{j\in D_i}(1-2\Pr[B_j])\ge 2\Pr[B_i](1-2\sum_{j\in D_i}\Pr[B_j])\ge 2\Pr[B_i](1-\frac{1}{2})=\Pr[B_i]$$

אי השוויון השמאלי הוא המקרה הפשוט ביותר של הכלה והפרדה (הוא גם מוכר מהכלל על איחוד מאורעות). עתה, מכיוון שנתון $\Pr[B_i]<\frac{1}{2}$ לכל i, נקבל לבסוף $\Pr[\bigwedge_{i=1}^m\neg B_i]\ge\prod_{i=1}^m(1-2\Pr[B_i])>0$ כנדרש.

בחוברת התרגילים יש דוגמה לשימוש ב"ממשק" זה של הלמה הלוקלית.

גרסת בניה של הלמה הלוקלית

שימוש אפשרי של השיטה ההסתברותית, הוא בתסריט מהסוג הבא: נאמר ויש לנו סדרה של מאורעות "רעים" $A_1,\dots,A_n$ וידוע כי לכל אחד מהם $\Pr\left[A_i\right]\le p_i<1$, וכן כי כל המאורעות בלתי תלויים זה בזה. במקרה זה קיימת הסתברות חיובית כלשהי כי אף אחד מהמאורעות לא יתרחש. אם כל אחד מהמאורעות מתאים לאיזו תכונה "רעה" של איזה מבנה קומבינטורי, אז המסקנה היא שקיים מבנה קומבינטורי בלי אף תכונה "רעה". הלמה הלוקלית מאפשרת לנו להחליש את דרישת האי־תלות — אם יש רק "קצת" תלות בין המאורעות, גם אז נוכל לקבל כי ניתן להתחמק מכולם בהסתברות חיובית. הצרה כאשר מדובר במבנה קומבינטורי היא שאמנם הוכחנו את עצם קיומו, אבל איננו יודעים כיצד לבנותו באופן קונסטרוקטיבי דטרמיניסטי, ואף לא באופן מקרי בעל הסתברות גבוהה.

נצטמצם למקרה של נוסחת $k-\mathrm{CNF}$ בת n משתנים ו־m פסוקיות. המדובר בחיתוך ("וגם") של פסוקיות שכל אחת מהן היא איחוד ("או") של k ליטרלים (משתנים או שלילתם). נניח כי כל פסוקית חולקת משתנים

עם $2^k e^{-1} - 1$ פסוקיות לכל היותר. נגריל השמה מקרית ונגדיר לכל פסוקית i את המאורע ה"רע" A_i שהפסוקית לא הסתפקה. המאורע A_i תלוי לכל היותר במאורעות A_j המתאימים לפסוקיות איתם הוא חולק משתנים, ויש לכל היותר $2^k e^{-1} - 1$ כאלה. כמו כן, $\Pr\left[A_i\right] \leq 2^{-k}$. על כן תנאי הלמה הלוקלית הסימטרית יוצא $e^{-1} \geq \left(2^k e^{-1}\right) 2^{-k}$ והוא אכן מתקיים, ולכן קיימת השמה המספקת את הפסוק. ב־1991 הראה Beck אלגוריתם אקראי שמוצא השמה כזאת, אבל רק אם נחליש את גדלי החיתוכים של הפסוקיות — הוא הרשה לכל פסוקית להחתך עם $O\left(2^{k/48}\right)$ פסוקיות אחרות לכל היותר. מאוחר יותר באותה השנה הראה Alon אלגוריתם אקראי שמסתפק בחסם של $O\left(2^{k/8}\right)$ פסוקיות נחתכות, וב־2008 הראה Srinivasan אלגוריתם שדי לו ב־$O\left(2^{k/4}\right)$. ב־2008 הגיעה פריצת דרך של Moser שהציג אלגוריתם אקראי למציאת השמה מספקת כמעט בלי להחליש את ההנחות — האלגוריתם שלו דורש כי כל פסוקית תחתך עם $2^{k-5} - 1$ פסוקיות לכל היותר, וכן הוכחתו אינה משתמשת בהוכחה הלא־קונסטרוקטיבית של הלמה הלוקלית, וכך מספקת הוכחה חדשה וקונסטרוקטיבית ללמה. ב־2009 הושלמה הסאגה עם אלגוריתם של Moser, Tardos שמתאים לכל מקרה של הלמה הלוקלית שניתן לתאר באופן קונסטרוקטיבי. נתאר את האלגוריתם ההסתברותי הפשוט למדי מ־2008, שתוחלת זמן הריצה שלו פולינומית. לשם פשטות, נראה אלגוריתם שבסיכוי גבוה עוצר בזמן פולינומי, וממנו המעבר לאלגוריתם עם תוחלת זמן ריצה פולינומית הוא פשוט (אם עבר זמן רב מדי ללא עצירה, מפסיקים את ריצת התוכנית ומתחילים מהתחלה).

האלגוריתם של מוזר

נתחיל בתיאור פונקציית עזר רקורסיבית, המקבלת את הפסוק F, פסוקית C ואת ההשמה הנוכחית α:

$\mathrm{LocalFix}\left(\mathrm{F}, \alpha, \mathrm{C}\right)$

1. החלף את כל ערכי המשתנים המופיעים ב־C בהשמה מקרית.
2. כל עוד קיימות פסוקיות מופרות ב־F הנחתכות עם C:
   (א) סמן ב־D את הפסוקית הראשונה לקסיקוגרפית מבין אלו.
   (ב) בצע $\mathrm{LocalFix}(\mathrm{F}, \alpha, \mathrm{D})$.
3. החזר את α.

וכעת האלגוריתם הכללי:

$\mathrm{SolveLovasz}\left(\mathrm{F}\right)$

1. בחר באקראי השמה α.
2. כל עוד קיימות פסוקיות מופרות ב־F:
   (א) סמן ב־D את הפסוקית המופרת הראשונה לקסיקוגרפית.
   (ב) בצע $\mathrm{LocalFix}(\mathrm{F}, \alpha, \mathrm{D})$.
3. החזר את α.

אין לנו כל סיבה להאמין שהאלגוריתם לא ימשיך לרוץ לעד, אבל אנחנו נראה כי בהסתברות גבוהה זמן הריצה פולינומי ב־n. האינטואיציה היא שתהליך התיקון המקומי "מכווץ" את האקראיות, ומכיוון שיש לנו "אקראיות אמיתית" לא ניתן לכווץ אותה מתחת לגודלה. אנחנו נתמודד עם המקרה בו כל פסוקית נחתכת עם 2^{k-b} פסוקיות אחרותף כאשר b קבוע שנבחר בקרוב.

ניתוח האלגוריתם של מוזר

הניתוח שנציג לעיל שונה מזה שמוצג במאמרים המקוריים, אבל הפך למקובל. נסמן את מספר הקריאות (כולל הרקורסיביות) לפונקציית התיקון המקומי ב־s. האלגוריתם משתמש במחרוזת אקראית x שאורכה $n+sk$: n ביטים לדגימת ההשמה המקרית הראשונה, ועוד k ביטים לכל קריאה לפונקציית התיקון המקומי. אם נדע מה היא הפסוקית שמתוקנת, נדע שהיא הייתה מופרת ולכן נדע בדיוק אילו ביטים מ־x היו בה לפני שתוקנה, שהרי לכל פסוקית יש השמה לא־מספקת יחידה. כך תיאורו של x ניתן על ידי סדרת הפסוקיות שהתיקון המקומי מתקן ולאחריהן n הביטים של ההשמה האחרונה.

עכשיו נשתמש בהנחה על החיתוך עם מעט פסוקיות על מנת למצוא תיאור יעיל לסדרת הפסוקיות המתוקנות. כדי לתאר את הפסוקית C שעבורה נקרא התיקון המקומי מהאלגוריתם הכללי נזדקק ל־$\log m$ ביטים, ואת יתר הפסוקיות ברקורסיה המתחילה כאן נוכל לתאר בפחות ביטים, שכן אלו פסוקיות הנחתכות עם C. נסדר אותן לקסיקוגרפית. יש לכל היותר $2^{k-b}-1 \geq r$ כאלה, ולכן נדרש ל־$\log r+c$ (עבור c קבוע גדול דיו) ביטים על מנת לתאר כל אחת מהן, יחד עם סימן מיוחד לסוף הרקורסיה. נשים לב שמרגע שקריאת LocalFix על פסוקית C מסתיימת, אז הפסוקית מסופקת ונשארת כזו. זאת מכוון שקריאת LocalFix בהכרח תחזיר השמה המספקת את C, שכן אנחנו ממשיכים בביצוע תיקונים עד שכל הפסוקיות שנחתכות עם C מסופקות, ובפרט C עצמה. אם תיקון מאוחר יותר של פסוקית אחרת יקלקל את סיפוק C, אז הוא בהכרח תיקון לפסוקית שנחתכת עם C, ולכן נשוב ונתקן אותה רקורסיבית לפני החזרה ממנו.

על כן, תיאור כזה של המחרוזת האקראית ידרש ל־$m\log m$ ביטים לרשימת הפסוקיות עליהן נקרא התיקון המקומי מתוך הלולאה הכללית, $s\left(\log r+c\right)$ ביטים לרשימת הפסוקיות עליהן נקרא התיאור המקומי רקורסיבית, ו־n ביטים לתיאור ההשמה הסופית. מכוון שהמחרוזת x היא אקראית, לא ניתן לתאר אותה בפחות ביטים מאורכה (ראו הסבר בהמשך), ולכן:

$$\begin{aligned} m\log m+s\left(\log r+c\right)+n &\geq n+sk \\ \frac{m\log m}{k-\left(\log r+c\right)} &\geq s \\ \frac{m\log m}{b-c} &\geq s \end{aligned}$$

נבחר $b=c+1$, וכך נקבל חסם $s\leq m\log m$, על מספר הקריאות הכולל לתיקון המקומי במונחי גודל הנוסחה. לסיום ההוכחה נותר להסביר את מושג האקראיות המדויק לו אנחנו נדרשים. במקרה זה אנחנו צריכים שהמחרוזת תהיה אקראית־קולמוגורוב (Kolmogorov), כלומר שהיא תהיה קצרה יותר מכל תוכנית מחשב שיכולה לייצר אותה. בפרט, היא תהיה קצרה יותר מתוכנית מחשב שתייצר אותה בעזרת היסטוריית התיקונים הנ"ל. נציין ללא הוכחה כי ההסתברות שמחרוזת שנבחרת באקראי ובאופן יוניפורמי תהיה אקראית־קולמוגורוב היא גבוהה. כלומר, בהנתן שהתרחש המאורע שהמחרוזת האקראית היא אכן אקראית קולמוגורוב, אז זמן הריצה חסום. למרבה הצער, אין זמן בקורס זה למבוא לסיבוכיות קולמוגורוב. ספר מומלץ למתעניינים:

Ming Li and Paul Vitanyi, An Introduction to Kolmogorov Complexity and Its Applications.

משפט FKG

משפט FKG בהפוך

משפט FKG נותן לנו קורלציה בין ערכיהן הממוצעים של שתי פונקציות עולות. נרצה להסיק תוצאה הפוכה עבור שתי פונקציות שהאחת מהן עולה והשניה יורדת. נניח כי $f:\mathcal{P}\left(S\right)\to\mathbb{R}$ אי־שלילית ומונוטונית לא־יורדת, $h:\mathcal{P}\left(S\right)\to\mathbb{R}$ אי־שלילית ומונוטונית לא־עולה, ו־$\mu:\mathcal{P}\left(S\right)\to\mathbb{R}$ אי־שלילית ולוג־סופר־מודולרית.

נגדיר $\alpha = \max_{A \subseteq S} h\left(A\right)$ ואז $g\left(A\right) = \alpha - h\left(A\right)$ היא אי־שלילית ומונוטונית לא־יורדת. נשתמש במשפט FKG על g, f ונקבל

$$\left(\sum_{A\subseteq S} \mu(A)f(A)\right)\left(\sum_{A\subseteq S} \mu(A)g(A)\right) \leq \left(\sum_{A\subseteq S} \mu(A)f(A)g(A)\right)\left(\sum_{A\subseteq S} \mu(A)\right)$$

ואם נפתח את הביטוי עבור g נקבל

$$\left(\sum_{A\subseteq S} \mu(A)f(A)\right)\left(\sum_{A\subseteq S} \mu(A)\alpha\right) - \left(\sum_{A\subseteq S} \mu(A)f(A)\right)\left(\sum_{A\subseteq S} \mu(A)h(A)\right) \leq$$
$$\left(\sum_{A\subseteq S} \mu(A)f(A)\alpha\right)\left(\sum_{A\subseteq S} \mu(A)\right) - \left(\sum_{A\subseteq S} \mu(A)f(A)h(A)\right)\left(\sum_{A\subseteq S} \mu(A)\right)$$

עכשיו נחסר את הביטוי $\alpha\left(\sum_{A\subseteq S} \mu(A)f(A)\right)\left(\sum_{A\subseteq S} \mu(A)\right)$ המופיע בשני צידי האי שוויון (אבל עם מיקום שונה ל α) ונכפיל במינוס אחת כדי לקבל אי שוויון הפוך ל FKG:

$$\left(\sum_{A\subseteq S} \mu(A)f(A)\right)\left(\sum_{A\subseteq S} \mu(A)h(A)\right) \geq \left(\sum_{A\subseteq S} \mu(A)f(A)h(A)\right)\left(\sum_{A\subseteq S} \mu(A)\right)$$

חסם תחתון באי שוויון ינסון

נזכר כי בהוכחת החסם התחתון באי שוויון ינסון הסתמכנו על כך שאם יש לנו קבוצת מאורעות $\{B_i\}_{i\in I}$ הנקבעים על ידי הכללות קבוצה A_i בקבוצה אקראית R, אז לכל קבוצת אינדקסים $J \subset I$ ולכל $i \notin J$ מתקיים $\Pr\left[B_i | \bigwedge_{j\in J} \neg B_j\right] \leq \Pr\left[B_i\right]$. נוכיח טענה כללית יותר על בסיס משפט קלייטמן ומשפט FKG וממנה נגזור את הנדרש. כדי להוכיח את ההכללה של משפט קלייטמן, כדאי לפרש אותו מחדש באופן הבא: נניח כי $\mathcal{A}$ היא משפחה של תת קבוצות של $[n]$, ונגדיר את ההסתברות שלה $\Pr\left[\mathcal{A}\right] = \frac{|\mathcal{A}|}{2^n}$. כלומר, זאת ההסתברות שאם נבחר יוניפורמית תת קבוצה כלשהי של $[n]$ אז היא תהיה ב $\mathcal{A}$. כך משפט קלייטמן בעצם נותן חסמים על הסתברויות של חיתוכי משפחות במונחי הסתברויות המשפחות הנחתכות.

נרצה לתרגם אותו לתסריט של אי שוויון ינסון: ההתפלגות אינה אחידה על פני תתי הקבוצות, אלא כל איבר נבחר לתת קבוצה באופן בלתי תלוי ובהסתברות ייחודית לו. עבור וקטור ממשי $p = \left(p_1, \ldots, p_n\right)$ כאשר $0 \leq p_i \leq 1$ לכל i, נגדיר מרחב הסתברות בדומה לאי שוויון ינסון, בו האיברים הם כל תתי הקבוצות של $[n]$, ומגדירים את ההסתברויות שלהן $\Pr_p\left[A\right] = \prod_{i\notin A}\left(1-p_i\right)\prod_{j\in A} p_j$. כלומר, ההסתברות שבהגרלה שבה לכל i האיבר i נבחר בהסתברות p_i באופן בלתי תלוי באחרים, קיבלנו את הקבוצה A בדיוק. נסמן את ההסתברות למשפחה $\mathcal{A}$ במרחב הסתברות זה ב־$\Pr_p\left[\mathcal{A}\right] = \sum_{A\in\mathcal{A}} \Pr_p\left[A\right]$. כלומר זוהי ההסתברות שקבוצה שנבחרה באקראי באופן זה היא ב־$\mathcal{A}$.

נגדיר $\mu_p : P\left([n]\right) \to \mathbb{R}^+$ על ידי $\mu_p\left(A\right) = \Pr_p\left[A\right]$. זאת פונקציה לוג־סופר־מודולרית, שכן מתקיים $\mu_p\left(A\right)\mu_p\left(B\right) = \mu_p\left(A\cup B\right)\mu_p\left(A\cap B\right)$, שכן התרומה הכפלית של כל $i \in [n]$ לשני הצדדים היא זהה: במקרה בו $i \in A \setminus B$ אז i תורם p_i ל־$\mu_p\left(A\right), \mu_p\left(A\cup B\right)$ ו־$1-p_i$ ל־$\mu_p\left(B\right), \mu_p\left(A\cap B\right)$, במקרה $i \in A\cap B$ אז i תורם p_i לכל האיברים, ובאופן דומה מוכיחים את המקרים ההפוכים. אם $\mathcal{A}$ משפחה מונוטונית עולה ו־$\mathcal{B}$ משפחה מונוטונית יורדת, אז בהפעלה של משפט FKG בגרסה שהוכחנו זה עתה על הפונקציות המציינות שלהם נקבל את אי השוויון $\Pr_p\left[\mathcal{A}\cap\mathcal{B}\right] \leq \Pr_p\left[\mathcal{A}\right]\Pr_p\left[\mathcal{B}\right]$.

נגדיר את המשפחה העולה $\mathcal{A}$ להיות משפחת כל הקבוצות המכילות את A_i, ואת המשפחה היורדת $\mathcal{B}$ להיות משפחת כל הקבוצות שלכל $j \in J$ הקבוצה A_j אינה מוכלת בהן. במקרה זה מתקיים השוויון

$\Pr_p[\mathcal{A}\cap\mathcal{B}] = \Pr\left[B_i \wedge \bigwedge_{j\in J}\neg B_j\right]$ וכן מתקיים $\Pr_p[\mathcal{A}] = \Pr[B_i], \Pr_p[\mathcal{B}] = \Pr\left[\bigwedge_{j\in J}\neg B_j\right]$, ואז מההכללה שהוכחנו למשפט קלייטמן מתקיים כי

$$\Pr\left[B_i \wedge \bigwedge_{j\in J}\neg B_j\right] = \Pr_p[\mathcal{A}\cap\mathcal{B}] \le \Pr_p[\mathcal{A}]\Pr_p[\mathcal{B}] = \Pr[B_i]\Pr\left[\bigwedge_{j\in J}\neg B_j\right]$$

כעת,

$$\Pr\left[B_i|\bigwedge_{j\in J}\neg B_j\right] = \frac{\Pr\left[B_i \wedge \bigwedge_{j\in J}\neg B_j\right]}{\Pr\left[\bigwedge_{j\in J}\neg B_j\right]} \le \frac{\Pr[B_i]\Pr\left[\bigwedge_{j\in J}\neg B_j\right]}{\Pr\left[\bigwedge_{j\in J}\neg B_j\right]} = \Pr[B_i]$$

וסיימנו את הוכחת הטענה.

אנטרופיה

בהרצאה ראינו כי עבור שני משתנים מקריים X, Y המקבלים ערכים ב־S, T בהתאמה מתקיימת תת־אדיטיבות של האנטרופיה, קרי $\mathrm{H}[X,Y] \le \mathrm{H}[X] + \mathrm{H}[Y]$, ובאינדוקציה נוכל לקבל כי אם $X = (X_1,\ldots,X_n)$ משתנה מקרי המקבל ערכים ב־$S = S_1\times\ldots\times S_n$ אז מתקיים $H[X] \le \sum_{i=1}^n \mathrm{H}[X_i]$. אי שוויון המכליל זאת הוכח ב־1986 על ידי Shearer.

משפט: תחת הסימונים לעיל, אם $\mathcal{A}$ משפחה של תתי קבוצות של $[n]$ וכל $i\in[n]$ שייך ללפחות k איברים של $\mathcal{A}$, אז

$$k\mathrm{H}[X] \le \sum_{A\in\mathcal{A}} \mathrm{H}[X(A)]$$

כאשר $X(A)$ הוא הוקטור המקרי המתאים לאינדקסים של A מתוך הוקטור X, כלומר הוקטור $\langle X_i | i\in A\rangle$.

הוכחה: נוכיח זאת באינדוקציה על k. עבור $k=0$ אי השוויון מיידי מאי השליליות של האנטרופיה.

כעת נניח את אי השוויון עבור $k-1\ge 0$ ונוכיח אותו עבור k. אם $[n]\in\mathcal{A}$ אז נוכל להסיר אותו משני אגפי אי השוויון ולהשתמש בהנחת האינדוקציה. אחרת, נסמן ב־A, B שני איברים ב־$\mathcal{A}$. מאי השוויון $\mathrm{H}[X|Y,Z] \le \mathrm{H}[X|Y]$ אנו מקבלים כי

$$\mathrm{H}[X(A\setminus B)|X(A\cap B), X(B\setminus A)] \le \mathrm{H}[X(A\setminus B)|X(A\cap B)]$$

כעת נשתמש בכלל השרשרת ונקבל:

$$\begin{aligned}
\mathrm{H}[X(A)] - \mathrm{H}[X(A\cap B)] &= \mathrm{H}[X(A\setminus B, A\cap B)] - \mathrm{H}[X(A\cap B)] \\
&= \mathrm{H}[X(A\setminus B)|X(A\cap B)] \\
&\ge \mathrm{H}[X(A\setminus B)|X(A\cap B), X(B\setminus A)] \\
&= \mathrm{H}[X(A\setminus B), X(A\cap B), X(B\setminus A)] - \mathrm{H}[X(A\cap B), X(B\setminus A)] \\
&= \mathrm{H}[X(A\cup B)] - \mathrm{H}[X(B)]
\end{aligned}$$

ובהעברת אגפים

$$\mathrm{H}[X(A\cup B)] + \mathrm{H}[X(A\cap B)] \le \mathrm{H}[X(A)] + \mathrm{H}[X(B)]$$

כלומר, אם נחליף את A, B באיחודם וחיתוכם הסכום $\sum_{A\in\mathcal{A}}\mathrm{H}[X(G)]$ לא יגדל. נוכל להמשיך בכך כל עוד $[n]\notin\mathcal{A}$, ובהכרח גם נגיע לשלב זה שכן איחוד כל האיברים ב $\mathcal{A}$ שווה ל $[n]$. את המקרה בו $[n]\in\mathcal{A}$ כבר הוכחנו ובכך מסתיימת הוכחת הטענה.

חסמים תחתונים בעזרת אנטרופיה לקודים הניתנים לפענוח מקומי

כעת נראה משפט מתוך מאמר של Katz ו־Trevisan המשתמש בשיטת האנטרופיה כדי לחסום את הקצב של סוג מסוים של קודים. משפט זה מדגים את האינטואיציה לפיה שיטת האנטרופיה מתאימה לספירת "מימד" של אובייקטים קומבינטוריים.

אנחנו מעוניינים בקודים שניתנים לפענוח מקומי. נרצה קודים, ז"א פונקציות $C:\{0,1\}^n\to R$ שעבורן קיים אלגוריתם פענוח אקראי $A:R\times[n]\to\{0,1\}$, שעבור קלט מהצורה $(C(x),i)$ לאיזה $x\in\{0,1\}^n$ יתן לנו את הקואורדינטה ה־i של x בהסתברות גבוהה. למעשה לא נדרוש הרבה מהאלגוריתם: נניח ש־x נבחר יוניפורמית ושהאלגוריתם מקבל את הקידוד הנכון שלו, ובסה"כ נדרוש שלכל i הסיכוי ה"ממוצע" לנכונות הפיענוח (ביחס לאקאיות הקלט והאלגוריתם כאחד) יקיים $\Pr_{A,x}[A(\mathrm{C}(x),i)=x_i]\geq 1/2+\epsilon$. שימו לב שערך של $\frac{1}{2}$ בדיוק יכול להיות מושג ע"י "אלגוריתם" שעונה תשובה הנבחרת באופן מקרי ויוניפורמי ללא תלות כל שהיא בקלט.

צעד חשוב במאמר הנזכר למעלה הוא הוכחת המשפט הבא, אשר מגביל באופן מהותי את מספר הביטים שאפשר "לחסוך" גם כאשר מסתפקים בדרישת קידוד חלשה כזו. בהקשר של קודים לתיקון שגיאות זה בעייתי, שכן הדבר עלול להקשות על תיקון שגיאות בקוד.

משפט: תהא $\mathrm{C}:\{0,1\}^n\to R$ פונקציה, ונניח כי קיים אלגוריתם כך שלכל אינדקס $i\in[n]$ מתקיים כי $\Pr_{A,x}[A(\mathrm{C}(x),i)=x_i]\geq 1/2+\epsilon$, כאשר ההסתברות נלקחת גם על האקראיות של A וגם על בחירה אקראית של מחרוזת x. אז מתקיים $\log|R|\geq(1-H(1/2+\epsilon))\,n$.

הוכחה: מהגדרת המידע המשותף $\mathrm{I}[x,\mathrm{C}(x)]\leq\mathrm{H}[\mathrm{C}(x)]$, וכפי שראינו בהרצאה $\mathrm{H}[\mathrm{C}(x)]\leq\log|R|$. בכיוון השני מהגדרת האינפורמציה המשותפת ותת אדיטיביות נקבל

$$\mathrm{I}[x,\mathrm{C}(x)]=\mathrm{H}[x]-\mathrm{H}[x|\mathrm{C}(x)]\geq\mathrm{H}[x]-\sum_{i=1}^{n}\mathrm{H}[x_i|\mathrm{C}(x)]$$

בנוגע לאיבר האחרון נשים לב לכך שבהנתן קידוד $\mathrm{C}(x)$, אם הגרלנו את האקראיות של האלגוריתם אז x_i יהיה שווה לערך $A(\mathrm{C}(x),i)$ בהסתברות לפחות $1/2+\epsilon$, ולכן אפשר לראות בו משתנה מקרי המקבל בהסתברות מסויימת את הערך $A(\mathrm{C}(x),i)$ ובהסתברות המשלימה את הערך ההפוך. מכוון שהאנטרופיה גדלה ככל שההתפלגות קרובה יותר ליוניפורמיות, מתקיים $\mathrm{H}[x_i|\mathrm{C}(x)]\leq H(1/2+\epsilon)$ וכך (יחד עם $\mathrm{H}[x]=n$) מקבלים את אי השוויון הדרוש.

קודים חסרי רישות

אחד השימושים החשובים של האנטרופיה הוא הבנת מושג הכיווץ. בפרט, נראה בתרגול זה כיצד ניתן להשתמש במושג האנטרופיה כדי להשיג חסמים על טיבם של קודי חסרי רישות.

הגדרה: פונקציה $C:\mathcal{D}\to\{0,1\}^*$ תקרא קוד בינארי. אם מתקיים בנוסף שלכל $x,y\in\{0,1\}^*$ אם $C(x)$ הוא רישא של $C(y)$ אז $x=y$, אז נאמר כי זה הוא קוד חסר רישות.

אנחנו נחשוב על $\mathcal{D}$ כעל קבוצה סופית של אותיות, והקוד ימפה את האותיות לקידוד בינארי. היתרון של קוד חסר רישות הוא שניתן תמיד לפענח סדרת אותיות שקודדו ברצף. בהקשר זה נניח כי נתונה התפלגות p על פני האותיות $\mathcal{D}$, ומטרתנו היא למזער את תוחלת אורך הקידוד של האותיות, קרי את $\ell=\mathbb{E}_{x\sim p}[|C(x)|]$. נזכור כי בקורס אלגוריתמים 1 נתקלנו בקוד האפמן (Huffman), שהוא קוד חסר רישות אופטימלי, כלומר, הוא קוד חסר רישות שממזער את ℓ. היום נבין את הקשר בין ℓ לבין $\mathrm{H}(p)$.

אבחנה: ניתן לייצג כל קוד חסר רישות בינארי באמצעות עץ בינארי מסודר T, כאשר כל אות מזוהה עם עלה, והקידוד של אות הוא המסלול מהשורש אל העלה המתאים לה.

אי השוויון המרכזי שנשתמש בו הוא אי שוויון קראפט (Kraft): לכל קוד חסר רישות בינארי, אורכי מילות הקוד $l_1,l_2,\ldots,l_m$ (עם כפילויות) חייבים לקיים את אי השוויון $\sum_{i=1}^{m}2^{-l_i}\leq 1$. בכוון ההפוך, לכל סדרת אורכי מילות קוד המקיימת אי שוויון זה, קיים קוד חסר רישות שאלה אורכי המילים בו.

הוכחה: נניח כי $l_1 \le l_2 \le \ldots \le l_m$. נביט בעץ בינארי מלא מעומק l_m, שבו יש גם לכל צומת פנימי סימונים "0" ו־"1" על שני הבנים בהתאמה. ניתן לזהות את מילות הקוד עם צמתים בעץ זה: עבור צומת v בעץ, נסתכל על המילה הנוצרת ממעבר על סימוני הבנים במסלול מהשורש ל־v, ונסמן אותה ב־s_v. עבור מחרוזת x מגודל חסום ע"י l_m, נזהה את x עם הצומת v עבורו $s_v = x$.

יהא v_i צומת המתאים למילת הקוד ה־i. בפרט, זה צומת בעומק l_i. מכיוון שמדובר בקוד חסר רישות, עבור כל צומת v_j המתאים למילת קוד אחרת, לא יתכן ש־v_i הוא צאצא שלו או אב קדמון שלו. על כן, מתקיים שתת העץ המושרש ב־v_i זר לזה המושרש ב־v_j. מספר העלים בתת העץ המושרש ב־v_i הוא $2^{(l_m - l_i)}$. מספר העלים הכולל בעץ המלא מגובה l_m הוא 2^{l_m}, ולכן $\sum_{i=1}^m 2^{(l_m-l_i)} \le 2^{l_m}$. נחלק את האגפים ב־$2^{l_m}$ וסיימנו.

בכוון השני נבנה את הקוד כך – נבחר צומת מעומק l_1, נקבע אותו כקידוד של האות 1 ונמחק את תת העץ המושרש בו. נחזור על התהליך עם האורכים לפי הסדר. ההנחה $\sum_{i=1}^m 2^{-l_i} \le 1$ מבטיחה לנו שכל עוד לא סיימנו, ישנם עלים מעומק l_m בעץ, ולכן גם צמתים מכל העומקים הקטנים יותר. כמו כן, הם לא יהיו אבות קדמונים של צמתים קודמים כי בחרנו אותם בסדר עומקים לא־יורד, והם לא יהיו צאצאים כי כל פעם מחקנו את כל תת העץ המתאים.

כעת, עלינו לקשור אי שוויון זה למושג האנטרופיה. הכוון הראשון הוא באי השוויון הבא: $\ell \ge \mathrm{H}(p)$.

הוכחה: נכתוב במפורש, תחת הסימונים הקודמים:

$$
\begin{aligned}
\ell - H(p) &= \sum_{i=1}^m p_i l_i - \sum_{i=1}^m p_i \log \frac{1}{p_i} \\
&= -\sum_{i=1}^m p_i \log \left(2^{-l_i}\right) - \sum_{i=1}^m p_i \log \frac{1}{p_i} \\
&= -\sum_{i=1}^m p_i \log \left(\frac{2^{-l_i}}{p_i}\right) \\
&\ge -\frac{1}{\ln 2} \sum_{i=1}^m p_i \left(\frac{2^{-l_i}}{p_i} - 1\right) \\
&= -\frac{1}{\ln 2} \left(\sum_{i=1}^m 2^{-l_i} - \sum_{i=1}^m p_i\right) \\
&\ge -\frac{1}{\ln 2}(1-1) = 0
\end{aligned}
$$

כאשר השתמשנו באי השוויון $\ln x \le x - 1$ ובאי שוויון קראפט.

כעת נראה שכמעט ואפשר להשיג חסם תחתון זה. נבחר $l_i = \lceil \log \frac{1}{p_i} \rceil$. עם בחירה זו מתקיים אי שוויון קראפט $\sum_{i=1}^m 2^{-\lceil \log \frac{1}{p_i} \rceil} \le \sum_{i=1}^m 2^{-\log \frac{1}{p_i}} = \sum_{i=1}^m p_i = 1$ ולכן קיים קוד חסר רישות שזה אורך המילים בו. נבחן את אורך המילה הממוצע בעיני האנטרופיה: $\ell = \sum_{i=1}^m p_i \lceil \log \frac{1}{p_i} \rceil \le \sum_{i=1}^m p_i \left(\log \frac{1}{p_i} + 1\right) = H(p) + 1$.
שימו לב שאם כל ה־p_i הם חזקות שלמות של 2 אז אנחנו נשיג במדויק את האנטרופיה.

הילוכים מקריים

על ההתכנסות להתפלגות סטציונרית

בתרגול זה נראה שההתפלגות של כל הילוך מקרי על גרף קשיר שאינו דו־צדדי מתכנסת להתפלגות הסטציונרית. נסמן את מטריצת ההילוך המקרי של הגרף G ב־P. ראשית נציג את הוקטור הסטציונרי, כלומר הוקטור המקיים $\pi = P^T \pi$. נקבע $\pi(v) = \frac{d(v)}{2m}$ ונשים לב כי $P^T = AD^{-1}$ כאשר A מטריצת הסמיכויות של G

ו־D המטריצה האלכסונית שאלכסונה הוא דרגות צמתי הגרף. זה נכון שכן אם נכפיל את P^T מימין ב־D נקבל את מטריצת ההילוך מוכפלת בדרגות הצמתים — זו מטריצת הסמיכויות. כעת, אם d וקטור הדרגות, $P^Td = AD^{-1}d = A\left(1,1,\dots,1\right)^T = d$, ולכן π (ככפולה של d) הוא וקטור עצמי עם ערך עצמי 1. מכיוון שסכום הדרגות הוא $2m$, אנחנו מקבלים שהוקטור π הוא אכן התפלגות סטציונרית. חשובה להמשך גם העובדה שכל הקורדינטות של π הן חיוביות ממש.

כעת נראה תנאי לכך שכל התפלגות אחרת מתכנסת ל־π. נראה כי אם G קשיר אז π הוא הוקטור העצמי היחידי עם ע"ע 1 עד כדי כפל בסקלר, ולאחר מכן נראה כי כל הערכים העצמיים חסומים בערכם המוחלט על ידי 1. על מנת להשלים את ההוכחה נראה כי אם G קשיר ולא דו־צדדי, אז -1 אינו וקטור עצמי. נזכר שכפי שראינו בהרצאה, כל הערכים העצמיים של P הם ממשיים, ונסמן אותם לפי סדר לא־עולה $\lambda_1,\dots,\lambda_n$.

אם כך, נניח כי קיים וקטור עצמי נוסף עם ע"ע 1, ונסמנו v. עבור $\alpha \in \mathbb{R}$ נביט בוקטור $v + \alpha\pi$. זה גם וקטור עצמי של 1, כצירוף לינארי של כאלה. אם ניקח את α להיות שלילי מאוד אז כל ערכי הוקטור $v + \alpha\pi$ יהיו שליליים (בגלל שאין ערכי אפס ב־π), ואם ניקח אותו להיות חיובי מאוד אז כל הערכים יהיו חיוביים. לכן לכל קורדינטה קיים α כך שערכה ב־$v + \alpha\pi$ מתאפס, ומכיוון שהמדובר בפונקציה לינארית ב־α, סדרת הערכים $\alpha_1,\dots,\alpha_k$ עבורה זה קורה היא סופית (וחסומה ע"י n). נסמן ב־β את ה־α_i המקסימלי בסדרה. כל הערכים השונות מאפס ב־$v + \beta\pi$ יהיו חיוביים, כי אחרת נוכל להגדיל את β עד שערך של עוד קורדינטה יתאפס, בסתירה להיותו מקסימלי. כעת ננרמל את $v + \beta\pi$ לקבלת וקטור התפלגות w, שגם לו ערך עצמי 1, ונביט ב־$P^Tw = w$. נניח כי $w_i = 0, w_j > 0$. נביט ב־$(wP)_i = \sum_{k=1}^m w_kP_{k,i} = w_i = 0$, ומכיוון שכל ערכי w אי שליליים, משמעות הדבר היא שלכל $k \in [m]$ מתקיים לפחות אחד מהשניים: או ש־$P_{k,i} = 0$ או ש־$w_k = 0$. ידוע כי $w_j > 0$ ולכן $P_{j,i} = 0$. כלומר, הסתברות המעבר מ־j ל־i היא אפס. מכאן שאין קשתות העוברות בין צמתים עם הסתברות 0 ב־w לצמתים עם הסתברות חיובית, וזוהי סתירה לקשירות G. נשים לב (וזה יהיה חשוב להמשך) שאותם טיעונים היו עובדים גם אם היינו מרשים קשתות מקבילות ו/או לולאות בגרף.

נראה עתה שכל הערכים העצמיים של P חסומים בערכם המוחלט על ידי 1: יהא w וקטור עצמי עם ערך עצמי λ, כלומר $AD^{-1}w = \lambda w$. נניח בה"כ כי $|w_1| \geq |w_i|$ לכל אינדקס i. אז $\lambda w_1 = \frac{1}{d(1)}\sum_{(1,j)\in E} w_j$ לפי הכפל במטריצה, וכך

$$|\lambda w_1| = |w_1|\,|\lambda| = \left|\frac{1}{d\left(1\right)}\sum_{(1,j)\in E} w_j\right| \leq \frac{1}{d\left(1\right)}\sum_{(1,j)\in E} |w_j| \leq \frac{1}{d\left(1\right)}\sum_{(1,j)\in E} |w_1| = |w_1|$$

ולכן $|\lambda| \leq 1$.

כעת נראה שגרף קשיר הוא דו־צדדי אם ורק אם מתקיים $\lambda_n = -1$, כאשר λ_n הוא הע"ע הנמוך ביותר: ראשית, אם הגרף הוא דו־צדדי, אז המטריצה P של ההילוך עליו (עבור סידור מתאים של הצמתים) היא מהצורה $\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix}$, ואז אם $\begin{pmatrix} u \\ v \end{pmatrix}$ וקטור עצמי של λ אז $\begin{pmatrix} u \\ -v \end{pmatrix}$ הוא וקטור עצמי של $-\lambda$ — בפרט זה נכון עבור הווקטור π, בעל הערך העצמי 1. בכוון השני נביט ב־P^2, כאשר נניח שהמטריצה P של ההילוך על הגרף היא בעלת ע"ע של -1. מטריצה זו מתאימה להילוך המתקבל על הגרף שבו יש קשת מ־u ל־v עבור כל מסלול מאורך 2 על הגרף המקורי (בגרף זה בד"כ יהיו קשתות מקבילות ולולאות). למטריצה P^2 יש את 1 כערך עצמי מריבוי גדול מאחת, ולכן מהטענה הקודמת הגרף המתאים אינו קשיר. כלומר, ניתן לחלק את צמתי הגרף לשתי קבוצות צמתים U, W כך שאין קשתות ביניהן. על כן בגרף המקורי אין מסלולים באורך שתיים מצמתי U לצמתי W. נראה שזאת גם חלוקה שמראה שהגרף הוא דו צדדי, כלומר שאין קשתות פנימיות ל־U (או ל־W). נניח בשלילה כי ישנם $u_1, u_2 \in U$ שכנים בגרף. לכל $v \in W$, כיוון שהגרף קשיר ישנו מסלול מ־u_1 ל־v. נסמן ב־w את הצומת הראשון במסלול זה שמקיים כי $w \in W$ וב־z את הצומת הקודם לו במסלול. אם $z = u_1$ אז u_2u_1w הוא מסלול באורך שתיים מ־U ל־W, בסתירה. אחרת, נסמן ב־z' את הצומת הקודם ל z במסלול. נשים לב כי $z' \in U$ לפי הגדרת w, ולכן $z'zw$ הוא מסלול מאורך שתיים מ־U ל־W, ושוב הגענו לסתירה.

לסיכום, אם הגרף שלנו קשיר ולא דו־צדדי אז מתקיים כי 1 הוא ערך עצמי פשוט, וכל הערכים העצמיים האחרים קטנים ממש ממנו בערכם המוחלט.

כעת נסיים את הוכחת ההתכנסות להתפלגות הסטציונרית — תהא p התפלגות התחלתית, ונסמן את הוקטורים העצמיים של P^T ב־$w_1,\ldots,w_n$ כאשר $w_1=\pi$, ונכתוב את ההתפלגות כצירוף לינארי שלהם $p=\sum_{i=1}^n \alpha_i w_i$. כשנכפול נקבל

$$P^T p=\sum_{i=1}^n \alpha_i P^T w_i=\sum_{i=1}^n \alpha_i \lambda_i w_i$$

כאשר λ_i הוא הערך העצמי המתאים לוקטור העצמי w_i . נבצע k צעדים של ההילוך ואז נקבל

$$\left(P^T\right)^k p=\sum_{i=1}^n \alpha_i \left(P^T\right)^k w_i=\sum_{i=1}^n \alpha_i \left(\lambda_i\right)^k w_i$$

מכיוון שלכל $i>1$ מתקיים ש־$|\lambda_i|<1$, אז עבור $\alpha_1,\ldots,\alpha_n$ קבועים נקבל $\lim_{k\to\infty}\left(P^T\right)^k p=\alpha_1\pi$, ומכיוון שכל אלו ווקטורי התפלגות בהכרח $\alpha_1=1$.

הוכחת התכנסות הילוך בשיטת הצימוד

נראה עתה דוגמה לשיטה אחת להוכחת התכנסות מהירה להתפלגות הסטציונרית, שיטת הצימוד. שיטה אחרת, שבה משתמשים לניתוח הילוכים על גרפים מרחיבים (expanders), היא שיטת הערכים העצמיים שלא תילמד כאן. הרעיון הוא זה: בנוסף להילוך המקרי $X_0,X_1,\ldots$ מגדירים הילוך מקרי שני על אותו גרף $Y_0,Y_1,\ldots$ התלוי בו, כך ש־Y_0 מתפלג לפי ההתפלגות הסטציונרית, וכך שההסתברות $\Pr\left[Y_t=X_t\right]$ שואפת מהר ל־1 עם גדילת t.

נמחיש זאת ע"י דוגמה. ננסה לערבב חפיסה בת n קלפים באורך הבא: בכל שלב נבחר באופן אקראי ואחיד קלף מהחפיסה, ונעביר אותו לראש החפיסה (שימו לב שזהו הילוך מקרי על גרף מכוון בעל $n!$ צמתים). נראה שניתן בצורה זו לערבב את החפיסה בזמן סביר. לשם כך, נחסום את המרחק בין $q^{(t)}$ לבין ההתפלגות היוניפורמית על כל סדרי החפיסה האפשריים, שהיא ההתפלגות הסטציונרית של הילוך זה.

אנו נראה שלכל $\epsilon>0$ קבוע מתקיים $|q^{(t)}-\pi|\le\epsilon$ עבור $t=O(n\log n)$, כאשר $q^{(t)}$ מסמן את התפלגות סדר החפיסה בזמן t, ו־$q^{(0)}$ מתאר בחירה דטרמיניסטית של סדר שרירותי כל שהוא. לשם כך נבנה לצד השרשרת $X_0,X_1,\ldots$, המתארת את ערבוב החפיסה, שרשרת שניה $Y_0,Y_1,\ldots$ באופן הבא. נניח שלקחנו חפיסה שניה, אשר סידורה ההתחלתי נבחר באופן מקרי ויוניפורמי מכל הסידורים האפשריים (כלומר ההתפלגות הסטציונרית). בשלב ה־t, בהינתן הערך של X_{t-1} (שהוא סידור אפשרי של החפיסה), הערך של X_t נבחר כזכור ע"י כך שלוקחים קלף שנבחר באופן יוניפורמי ומעבירים אותו להתחלה. לקבלת Y_t מתוך Y_{t-1} ניקח עתה את הקלף בחפיסה השניה עם אותו מספר סידורי (כלומר "אותו קלף"), ונעביר אותו לראש החפיסה השניה.

הדבר לשים לב אליו הוא ש־$Y_0,Y_1,\ldots$ היא שרשרת מרקוב עם אותה מטריצת מעבר כמו $X_0,X_1,\ldots$, ולכן ההתפלגות (הלא־מותנה) של Y_t היא עדיין ההתפלגות הסטציונרית π. עתה נסמן ב־$A_i^{(t)}$ את המאורע שהקלף שמספרו i נבחר והועבר לראש בשתי החפיסות בשלב כל שהוא עד השלב ה־t, ונסמן את $A^{(t)}=\bigwedge_{i=1}^n A_i^{(t)}$. לא קשה להראות שמתקיים $\Pr\left[X_t=Y_t|A^{(t)}\right]=1$. כמו כן עבור $t\ge n\ln(n/\epsilon)$ יתקיים

$$\Pr\left[A^{(t)}\right]=1-\Pr\left[\bigvee_{i=1}^n \neg A_i^{(t)}\right]\ge 1-\sum_{i=1}^n \Pr\left[\neg A_i^{(t)}\right]\ge 1-n\left(1-\frac{1}{n}\right)^t\ge 1-\epsilon$$

מקיום המאורע $A^{(t)}$ בעל התכונות הנ"ל נובע שהמרחק בין התפלגות X_t והתפלגות Y_t (הלא־מותנות) אינו עולה על ϵ בנורמת ה־variation distance, לפי הטענות שהוכחו בפרק על מרחק בין התפלגויות בחוברת התרגילים.

סדרות הילוך אוניברסליות

נניח כעת כי הגרף הנתון G הוא גרף d־רגולרי. נקבע $v_0 \in V(G)$ ונניח כי עבור כל צומת v בגרף יש התאמה בין קבוצת שכניו לבין הקבוצה $[d]$. סדרת הילוך עבור גרף זה, צומת זה, וזיהוי שכנים זה היא סדרה $(h_1, h_2, \ldots, h_t) \in [d]^t$, כך שאם נתחיל סיור בגרף בצומת v_0 ובצעד ה־i נעזוב את הצומת הנוכחי לשכן שמספרו h_i אז נבקר בכל צמתי הגרף. סדרת הילוך נקראת (d,n)־אוניברסלית אם היא סדרת הילוך לכל גרף d־רגולרי על n צמתים, לכל בחירת זיהוי לשכנים וכל צומת התחלה. ב־1979 הוכיחו Aleliunas, Karp, Lipton, Lovasz, Rackoff כי סדרות אלה קיימות, ואף אינן ארוכות מאוד.

נבחר סדרה מקרית $H = (h_1, \ldots, h_t) \in [d]^t$ עבור $t = 16dmn^2 \log n$ (m הוא מספר הקשתות בגרף, במקרה שלנו $m = \frac{1}{2}dn$). נשים לב שעבור G נתון, הסיור בגרף שמוגדר על ידי הסדרה הוא פשוט הילוך מקרי על G. לכן עלינו לבדוק מה ההסתברות שהילוך באורך t יבקר בכל הצמתים. זמן הכיסוי של הילוך מקרי על גרף הוא תוחלת מספר הצעדים שידרשו על מנת לבקר בצמתי הגרף כולם. אם כך עלינו לחסום כמות זו. ראינו בהרצאה ש־$k_{st} = 2mR_{st}$, ובפרט אם (s,t) קשת בגרף אז $k_{s,t} \leq 2m$. נביט בעץ פורש T לגרף G. נכפיל כל קשת ב־T, וקיבלנו גרף בו יש מעגל אוילר C. נביט במעגל זה כאשר הוא מתחיל מצומת ההתחלה של ההילוך. עבור כל קשת (u,v) במעגל, מתקיים גם $k_{u,v} \leq 2m$ ולכן תוחלת מספר הצעדים שנדרשים על מנת להגיע מ־u ל־v הוא לכל היותר $2m$. יש $2(n-1)$ קשתות במעגל זה, ולכן (מלינאריות התוחלת) לאחר תוחלת של $4mn$ צעדים לכל היותר נכסה את כל קשתות C, ומכאן גם את כל קשתות T וצמתי G (נעיר שידועים חסמים טובים יותר, לדוגמה Feige הראה חסם של $2n^2$ לזמן הכיסוי). לכן, מאי שוויון מרקוב, ההסתברות שלאחר $8mn$ צעדים לא כיסינו את כל הצמתים היא לכל היותר $1/2$. מכיוון שניתן להביט ב־$8mn$ הצעדים לאחר מכן כהילוך מקרי חדש, נקבל כי ההסתברות שלא ראינו את כל הצמתים לאחר t צעדים היא לכל היותר $2^{-t/8mn} = n^{-2dn}$.

כעת, ישנם לכל היותר n^{dn} גרפים d־רגולרים עם שכנים מתוייגים, ולכן ההסתברות ש־H אינה סדרת הילוך עבור אחד מגרפים אלה, עבור נקודת התחלה כלשהי, היא פחות מ־1 $nn^{nd}n^{-2nd} < 1$. לכן בהכרח קיימת סדרת הילוך אוניברסלית מאורך $O(dmn^2 \log n)$.