שיטות הסתברותיות ואלגוריתמים – תרגולים

מחברים: אלדר פישר, יונתן גולדהירש

10 באוגוסט 2024

הקדמה ועניינים 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). ההגשה תהיה ביחידים בלבד. הגשת התרגילים, קבלת המשוב וכו יהיו דרך מערכת Webcourse (במקרים מסויימים ניתן יהיה לקבל אישור להגשה ידנית). פתרונות רשמיים לתרגילים ינתנו בערך בזמן קבלת המשוב לכל תרגיל.

תזכורת מהירה וסימונים בהסתברות

הקורס בהסתברות הוא קדם לקורס זה, אבל בכל זאת נעבור כאן על מספר סימונים וחוקים בסיסיים.

כמעט כל מרחבי ההסתברות שלנו יהיו בדידים. מרחב הסתברות בדיד יכול להיות מוגדר מעל קבוצה S שהיא סופית או בת מניה, והוא מאופיין ע"י פונקציה $\mu : S \to [0,1]$ המקיימת $\sum_{s\in S}\mu(s) = 1$. למשל, מרחב הסתברות שמתאים להטלה של שני מטבעות הוגנים (באופן בלתי־תלוי זה בזה) יוגדר מעל $S = \{0,1\}\times\{0,1\}$ ע"י $\mu(a,b) = \frac{1}{4}$ לכל $a,b \in \{0,1\}$. בד"כ נשתמש בסימון $\mu(s) = \Pr_\mu[s]$, במקרים שאנחנו דנים במרחב אחד שהוגדר מראש נשמיט את ה־μ מהסימון $\Pr[s]$. מצד שני, במקרים שנרצה להיות מפורשים בקשר לבחירה של $s \in S$ לפי μ נשתמש בסימון המורחב $\Pr_{s\sim\mu}[s]$ (בעיקר נשתמש בסימון כזה עבור ביטויים של תוחלות).

עיקר הניתוח ההסתברותי נסוב סביב מאורעות ומשתנים מקריים. במקרה הבדיד, מאורע E הוא פשוט תת־קבוצה של S, ומגדירים $\Pr[E] = \sum_{s\in E}\Pr[s]$. דוגמה למאורע במרחב "שני המטבעות" למעלה היא המאורע "שני המטבעות שווים", ז"א $E = \{(0,0),(1,1)\}$ שעבורו $\Pr[E] = \frac{1}{2}$.

כאשר יש מספר מאורעות, לרב נהוג להשתמש בסימונים לוגים לחיתוכים ואיחודים. למשל המאורע $E \wedge F$ ("E וגם F") הוא זה המתאים לקבוצה $E \cap F$. שני המאורעות יקראו זרים אם מתקיים $\Pr[E \wedge F] = 0$. זה לא אומר שהקבוצות עצמן זרות, יכול להיות שיש איברים בחיתוך $E \cap F$ כל עוד הפונקציה μ מתאפסת עליהם. חישוב ישיר מראה שעבור איחוד ("או") של זוג מאורעות זרים מתקיים $\Pr[E\vee F] = \Pr[E]+\Pr[F]$. אם לא נתון שהם זרים אז מתקיים $\Pr[E \vee F] \le \Pr[E \vee F] + \Pr[E \wedge F] = \Pr[E] + \Pr[F]$.

משתנה מקרי הוא פונקציה מ־S לקבוצת הממשיים $\mathbb{R}$. בד"כ הוא מסומן באות גדולה (לא כמו שמסמנים פונקציות בתחומים מתמטים אחרים). למשל, המשתנה "מספר המטבעות שיצאו 1" יוגדר ע"י $X(a,b) = a+b$. הרבה פעמים, עבור מאורע E, מגדירים מ"מ אינדיקטור לפי $X_E(s) = 1$ אם $s \in E$ ו־$X_E(s) = 0$ אם $s \notin E$. על משתני אינדיקטור נלמד עוד הרבה בקורס.

התוחלת של מ"מ X היא $\mathrm{E}[X] = \sum_{s\in S} X(s)\Pr[s]$ (בסימון יותר מפורש אפשר לכתוב $\mathrm{E}_{s\sim\mu}[X(s)]$, או לפעמים $\mathrm{E}_{X\sim\mu}[X]$). זו תהיה למשל שווה ל־1 עבור המשתנה "מספר המטבעות שיצאו 1" במרחב של הטלת שני מטבעות. אתם מוזמנים לראות שעבור משתנה אינדיקטור מתקיים $\mathrm{E}[X_E] = \Pr[E]$.

ניתן להגדיר מאורעות לפי משתנים מקריים. למשל המאורע "$X = \alpha$" יתאים לקבוצה $\{s \in S : X(s) = \alpha\}$. אתם מוזמנים לבדוק שבמרחבי הסתברות בדידים מתקיים $\mathrm{E}[X] = \sum_{\{\alpha\in\mathbb{R}:\Pr[X=\alpha]>0\}} \alpha \cdot \Pr[X = \alpha]$.

כדוגמה נוכיח את אי־השוויון הבסיסי של מרקוב (Markov): אם X לא מקבל ערכים שליליים, אז לכל $\alpha > 0$ מתקיים $\Pr[X \ge \alpha] \le \mathrm{E}[X]/\alpha$ (זה באמת אומר משהו עבור $\alpha > \mathrm{E}[X]$). אם נסתכל לדוגמה על המשכורות של התושבים בישראל, אי־השוויון קובע שלא יותר מחצי מהם יקבלו משכורת שהיא לפחות כפולה מהממוצעת. לשם הוכחת אי־השוויון כותבים

$$\mathrm{E}[X] = \sum_{s\in S} X(s)\Pr[s] \ge \sum_{\{s:X(s)<\alpha\}} 0\cdot\Pr[s] + \sum_{\{s:X(s)\ge\alpha\}} \alpha\cdot\Pr[s] = \alpha\Pr[X \ge \alpha]$$

ואז מעבירים את α אגפים.

מרחבים מותנים

בהינתן מרחב הסתברות μ מעל S ומאורע E המקיים $\Pr_\mu[E] > 0$, נגדיר את מרחב ההסתברות המותנה $\mu_E : E \to [0,1]$ לפי $\mu_E(s) = \mu(s)/\Pr_\mu[E]$. את ההסתברות המותנה, למשל עבור מאורע A, נסמן לרב לפי $\Pr_\mu[A|E]$ במקום הסימון הצפוי $\Pr_{\mu_E}[A]$. דבר זה יאפשר לנו להמשיך ולהשמיט את μ מהסימונים כל עוד אנחנו מדברים על מרחב הסתברות "מקורי" אחד.

לרב אנחנו גם נרחיב את המרחב המותנה μ_E לכל S ע"י כך שנגדיר $\mu_E(s) = 0$ לכל $s \in S \setminus E$. דבר זה יאפשר לנו להשתמש בסימון $\Pr[A|E]$ גם כאשר $A \not\subseteq E$ (בעצם יתקיים $\Pr[A|E] = \Pr[A \wedge E|E]$).

חישוב ישיר יראה לנו שעבור מאורע E עם הסתברות חיובית מתקיים $\Pr[A\wedge E] = \Pr[A|E]\Pr[E]$. חוק בייס (Bayes) קובע שעבור מאורעות A ו־B בעלי הסתברות חיובית מתקיים $\Pr[A|B] = \Pr[B|A]\Pr[A]/\Pr[B]$, ואפשר להראות אותו ע"י העברת אגפים של $\Pr[B]$ והצבה של נוסחת הקשר לחיתוך מאורעות.

נראה עתה את נוסחת ההסתברות השלמה עבור סדרת מאורעות סופית. נניח ש־$E_1,\ldots,E_k$ מחלקים את המרחב – ז"א שכל המאורעות זרים זה לזה ומתקיים $\Pr[\bigvee_{i=1}^k E_i]=1$. נניח כאן גם שלכולם הסתברות חיובית. במקרה כזה מתקיים לכל מאורע A:

$$\begin{aligned}\Pr[A] &= \Pr[A\wedge(\bigvee_{i=1}^k E_i)]+\Pr[A\setminus(\bigvee_{i=1}^k E_i)]=\Pr[A\wedge(\bigvee_{i=1}^k E_i)]+0\\ &= \sum_{i=1}^k \Pr[A\wedge E_i]=\sum_{i=1}^k \Pr[A|E_i]\Pr[E_i]\end{aligned}$$

לסיום, נשים לב שההגדרה של מרחבים מותנים ניתנת להכללה למדדים הסתברותיים נוספים. עבור מ"מ X מעל S ומאורע E בעל הסתברות חיובית, נגדיר $\mathrm{E}_\mu[X|E]=\mathrm{E}_{\mu_E}[X]=\sum_{s\in E}X(s)\Pr_\mu[s|E]$. נוסחת התוחלת השלמה קובעת שמתקיים $\mathrm{E}[X]=\sum_{i=1}^k \mathrm{E}[X|E_i]\Pr[E_i]$ עבור $E_1,\ldots,E_k$ שמקיימים את התנאים של נוסחת ההסתברות השלמה. אתם מוזמנים לנסות להוכיח אותה – אפשר באופן ישיר, או ע"י הצבת נוסחת ההסתברות השלמה עבור מאורעות מהצורה "$X=\alpha$".

מרחבים לא בדידים

איך מגדירים הסתברויות מעל, למשל, קטע הממשיים $[0,1]$? כאן התשובה היא יותר מסובכת. על מנת לדעת את התורה המתמטית עליכם לדעת את התחום המתמטי של תורת המידה. כעיקרון, לא מוגדרת פונקציה $\mu:S\to[0,1]$, אלא במקום זאת מוגדרות הסתברויות על מאורעות. למשל, אם $S=[0,1]$ ו־$E=[a,b]\subseteq[0,1]$, אז ההתפלגות היוניפורמית הרציפה תגדיר $\Pr_\mu[E]=b-a$. ההתפלגות חייבת להיות כזו שתקיים את תנאי החיבור על איחוד מאורעות זרים (כולל על איחוד של מספר בן מניה שלהם).

כדי שכזה דבר יתאפשר, לרב אי אפשר להגדיר את ההסתברות לכל תתי־הקבוצות של S! לא כל תת־קבוצה היא "מאורע", אלא רק הקבוצות הקרויות "מדידות" בתורת המידה. במקרה של $S=[0,1]$, אלו יכללו בין השאר את כל הקטעים, ואת כל מה שניתן להשיג מהם באמצעות צעדים של לקיחת משלים ולקיחת איחודים בני מניה. באופן צפוי, גם לא כל פונקציה היא "משתנה מקרי", לשם כך צריך להתקיים שלכל $a\le b$ ממשיים (או מתוך $\pm\infty$) יתקיים ש־$\{s: a\le X(s)\le b\}$ תהיה מאורע (ואז ניתן למשל להגדיר את התוחלת של X כ"אינטגרל" מתאים של הפונקציה).

כעיקרון, אם מגדירים את מרחב ההסתברות בצורה נכונה, אז מרבית המשפטים שנראה בקורס באמצעות סכומים עבור מרחבים בדידים יהיו נכונים גם למרחב הרציף (אבל לא תמיד עם אותה הוכחה).

שימושים בלינאריות התוחלת

כאשר יש לנו שני מ"מ X,Y בעלי תוחלת, תמיד מתקיים $\mathrm{E}[X+Y]=\mathrm{E}[X]+\mathrm{E}[Y]$. 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 $\mathrm{E}[X_i]=\Pr[A_i]$ אז נקבל שסכומם מקיים

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

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 בתוצאה של Erdös משנת 1965: 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 $\mathbb{Z}_p$ (שדה המספרים השלמים מודולו 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 $a\neq0$ 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 שעבורו $iS\cap A$ היא תת הקבוצה המבוקשת של A. לסיום נעיר שתוצאה של Eberhand, Green, Manners, שפורסמה ב־2014, נותנת דוגמאות של קבוצה A כך שאין תת־קבוצה בת יותר מ־$(\frac{1}{3}+o(1))n$ איברים מבלי שאחד מהם יהיה סכום של שניים מהאחרים.

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 $|W'|\leq|\hat{W}|$, 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

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

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

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

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־$\Pr\big[X_I \oplus X_J = 1\big] = \frac{1}{2}$, 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 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 $\gamma(r, k, \frac{b}{r-1})$ ו־$d_0(r, k, \frac{b}{r-1})$ 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־$n \leq (1-\beta)n + (\alpha+\beta)\frac{r-1}{r}n$ ולכן $\beta \leq (r-1)\alpha = b$. 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.

הישום 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 $r\geq 2$ וממשיים $K\geq 1$, $\epsilon>0$, 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 $m,D\geq D_0$ 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

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

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

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

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

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

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

לפני הוכחת הלמה, נראה איך מוכיחים ממנה את המשפט הכללי: נבחר $\epsilon>0$, $\frac{1}{10}>\delta>0$, ומספר שלם t כך ש־$(1+4\delta)(\frac{\epsilon}{1-e^{-\epsilon}}+r\epsilon)<1+a$ וכן $e^{-\epsilon t}<\epsilon$. עתה נבחר סדרה $\delta=\delta_t>\delta_{t-1}>\ldots>\delta_0$ 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 $\prod_{i=0}^{t}(1+\delta_i)\leq 1+2\delta$. 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" D' 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 השנייה משתמש בחסימת טור הנדסי ובהנחה $e^{-\epsilon t}<\epsilon$).

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־$\delta_1 \leq \delta'$ 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 $(1-\frac{\epsilon}{D})^{(1+\delta)D} \leq E[I_v] \leq (1-\frac{\epsilon}{D})^{(1-\delta)D}$. לכן בסיכום קבוצת הצמתים $(1-\delta)(1-\frac{\epsilon}{D})^{(1-\delta)D}m \leq E[|V'|] \leq (\delta+(1-\frac{\epsilon}{D})^{(1+\delta)D})m$, 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

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

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: $\Pr[X_i = -p_i] = 1 - p_i, \ \ \Pr[X_i = 1 - p_i] = p_i$ 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

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

כעת, נחסום את הלוגריתם של המכפלה בצד ימין:

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

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

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

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:

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

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 ניסוי 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:

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

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 $\Pr\left[X>a\right]\leq e^{a-pm\ln(1+a/pm)-a\ln(1+a/pm)}$. 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 $\mu=pm$ 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־$\Pr[X\leq-\delta\mu]<e^{-\delta^2\mu/2}$.

סיכום וטבלה

עבור הניתוחים בחרנו מ"מ עם שני ערכים אפשריים ותוחלת 0. בשימוש הנפוץ יהיו לנו משתנים ב"ת $Y_1,\ldots,Y_m$ כך ש־Y_i מקבל 1 בהסתברות p_i ומקבל 0 בהסתברות $1-p_i$. המעבר ל־X_i כמקודם הוא פשוט ע"י קביעת $X_i=Y_i-p_i$, ולכן החסמים על ההסתברות שהסכום X יהיה רחוק מ־0 מתרגמים לחסמים על ההסתברות שהסכום $Y=\sum_{i=1}^m Y_i$ יהיה רחוק מהתוחלת שלו $\sum_{i=1}^m p_i=pm$. הטבלה הבאה מסכמת את עיקר החסמים השימושיים מהקורס.

תנאי סטיה	חסם הסתברות	הערות
$Y<pm-a$ או $Y>pm+a$	$\exp(-\frac{2a^2}{m})$	החסם מההרצאה
$Y>pm+a$	$\exp(-\frac{a^2}{2pm}+\frac{a^3}{2(pm)^2})$	יותר מותאם ל־p נמוך ו־m גבוה
$Y>(1+\delta)pm$	$(\frac{e^{\delta}}{(1+\delta)^{1+\delta}})^{pm}$	מאוד כללי
$Y>(1+\delta)pm$ עבור $\delta\geq 1$	$\exp(-\delta pm/3)$	מסקנה שימושית
$Y>(1+\delta)pm$ עבור $\delta\leq 1$	$\exp(-\delta^2 pm/3)$	מסקנה שימושית למקרים רבים
$Y<(1-\delta)pm$	$(\frac{e^{\delta}}{(1-\delta)^{1-\delta}})^{pm}$	מאוד כללי
$Y<(1-\delta)pm$ עבור $\delta\leq 1$	$\exp(-\delta^2 pm/2)$	מסקנה שימושית למקרים רבים

מרטינגלים

כזכור מההרצאה, סדרה $\underline{X}$ של משתנים מקריים היא מרטינגל אם $\mathrm{E}\left[X_{i+1}|X_0,\ldots,X_i\right]=X_i$ לכל i. המדובר בשוויון בין משתנים מקריים (הסבר נוסף על זה נמצא בסוף הפרק), ועבור מרחבים בדידים ניתן לפרשו כך: אם $\Pr\left[X_0=a_0,\ldots,X_i=a_i\right]>0$ אז $\mathrm{E}\left[X_{i+1}|X_0=a_0,\ldots,X_i=a_i\right]=a_i$. 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 $\delta_{i,w} = X_i\left(w+b+i\right) - w$ 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 $z_1,\ldots,z_n$ תחום סופי D. המבנה $C = \left(z_1,\ldots,z_n\right)$ ניתן לזיהוי עם וקטור מתוך D^n, או עם פונקציה מ־$\{1,\ldots,n\}$ ל־D, והחשיפה מתבצעת משתנה־משתנה, כלומר $\mathcal{D}_i = \{1,\ldots,i\}$. אנחנו רוצים, עבור פונקציה $f : D^n \to \mathbb{R}$ ווקטור $C \in D^n$ שנבחר באופן מקרי, את ההסתברות לסטיה של $f(C)$ מהתוחלת המתאימה.

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 $\left(z_1,\ldots,z_n\right)$ נלקחים מתחום אינסופי ועם תומך לאו דווקא סופי. מקרה פרטי זה של אי שוויון אזומה נקרא לעיתים אי שוויון מקדיארמיד (McDiarmid) וניתן לגזור ממנו תוצאות רבות.

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 מההשמה $Z_i = 1$. 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 מההשמה $Z_i = 0$, 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 כל 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 בין שתי קבוצות הפרמוטציות הנ"ל.

נשים לב עתה 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 הפרש שתי התוחלות המותנות שלמעלה.

לסיום, ניזכר בהגדרה של המשתנים המקריים של המרטינגל כפונקציות ממשיות מעל קבוצת הבסיס של מרחב ההסתברות (במקרה זה, קבוצת כל הפרמוטציות מעל n איברים). כזכור, עבור פרמוטציה $\tilde{\sigma}$ קובעים $X_i(\tilde{\sigma}) = \mathrm{E}_\sigma[f(\sigma)|\sigma(1)=\tilde{\sigma}(1),\ldots,\sigma(i)=\tilde{\sigma}(i)]$. לפי זה נכתוב את $X_{i-1}(\tilde{\sigma})$:

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

ראינו כאן ש־$X_{i-1}(\tilde{\sigma})$ הוא ממוצע של ערכים שכל אחד מהם נמצא במרחק של לא יותר מ־c מהערך של $X_i(\tilde{\sigma})$, ולכן גם $X_{i-1}(\tilde{\sigma})$ עצמו נמצא במרחק של לא יותר מ־c מ־$X_i(\tilde{\sigma})$, 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), ומנתחים 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 רק שבהסתברות $o(1)$ 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 בוקטורים קודמים שנבחרו, ואז יחשוף את כל היתר. נגדיר עתה במפורש את הרעיונות הללו. בהגדרה הבאה, מדובר במרחב הסתברות μ מעל קבוצה S של פונקציות $C : \mathcal{D} \to \mathcal{R}$, ואנו מנסים לחסום את הסטיה מהתוחלת של פונקציה ממשית $f : S \to \mathbb{R}$ המוגדרת עבור פונקציות אלו.

הגדרה (סכמת חשיפה): נסמן ב־G את קבוצת כל הפונקציות $g : \mathcal{D}' \to \mathcal{R}$ מתת־קבוצה כל שהיא $\mathcal{D}' \subseteq \mathcal{D}$ ל־$\mathcal{R}$ שמקיימות $\Pr_{C\sim\mu}[C|_{\mathcal{D}'} = g|_{\mathcal{D}'}] > 0$. 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 $\mathcal{D}' = \mathcal{D}$.

יכול להיות שבהגדרה ספציפית לא נגדיר את $\mathcal{D}_g$ לכל $g \in G$, למשל אם יש תתי־קבוצה של $\mathcal{D}$ ש"אי אפשר להגיע אליהן". למשל, במרטינגל חשיפת צמתים רגיל, מגיעים אך ורק לקבוצות מהצורה "קבוצת כל הזוגות מתוך $\{1,\ldots,i\}$". אפשר להניח שכל תתי־הקבוצה שאינם מוגדרים במפורש בסכמת החשיפה שווים ל־$\mathcal{D}$.

הגדרה (מרטינגל חשיפה אדפטיבי): בהנתן $\mathcal{D},\mathcal{R},\mu,f$ כמו למעלה, סכמת חשיפה $\mathrm{D} = \{\mathcal{D}_g : g \in G\}$ ופונקציה (מבנה) $\widetilde{C} : \mathcal{D} \to \mathcal{R}$ מתוך S, נגדיר את $\mathcal{D}_0(\widetilde{C}), \mathcal{D}_1(\widetilde{C}), \ldots$ ואת ערכי המ"מ $X_0(\widetilde{C}), X_1(\widetilde{C}), \ldots$ באופן האינדוקטיבי הבא.

1. מגדירים $\mathcal{D}_0(\widetilde{C}) = \emptyset$ ובהתאמה $X_0(\widetilde{C}) = \mathrm{E}_{C\sim\mu}\left[f(C)\right]$.
2. בהנתן $\mathcal{D}_{i-1}(\widetilde{C})$ מגדירים אינדוקטיבית את $\mathcal{D}_i(\widetilde{C}) = \mathcal{D}_{C|_{\mathcal{D}_{i-1}(\widetilde{C})}}$, ולפיו את ערך המשתנה המקרי $X_i(\widetilde{C}) = \mathrm{E}_{C\sim\mu}\left[f(C)|C|_{\mathcal{D}_i(\widetilde{C})} = \widetilde{C}|_{\mathcal{D}_i(\widetilde{C})}\right]$.

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

ההבדל בין הגדרה זו להגדרה הרגילה של מרטינגל החשיפה של Doob היא שהתחומים $\mathcal{D}_i$ עשויים להיות תלויים גם הם בפונקציה $\widetilde{C}$. גם כאן אפשר להראות שהמדובר במרטינגל, אם כי הוכחה בסגנון של ההרצאה תהיה מסורבלת למדי. את המשפט שיאפשר לנו לפעמים לבצע חסימה נוחה של סטיות גדולות ננסח למען הפשטות רק למקרה שבו כל החסמים שווים ל־1.

הגדרה (תנאי ליפשיץ ביחס לסכמת חשיפה): נאמר כי f היא ליפשיץ ביחס ל־D אם לכל $C_1, C_2 \in S$ המקיימות כי הן מזדהות על $(\mathcal{D} \setminus \mathcal{D}_{i+1}(C_1)) \cup \mathcal{D}_i(C_1)$ עבור i כלשהו מתקיים כי $|f(C_1) - f(C_2)| \le 1$.

נוח יותר להשתמש בתנאי החזק יותר: לכל $g : \mathcal{D}' \to \mathcal{R}$ ששייכת ל־G ולכל $C_1, C_2 \in S$ המסכימות עם g על $\mathcal{D}'$ ומסכימות זו עם זו על $\mathcal{D} \setminus \mathcal{D}_h$ מתקיים $|f(C_1) - f(C_2)| \le 1$.

בדומה למקרה של מרטינגל חשיפה רגיל, גם כאן ניתן להראות שאם μ הוא כזה שכל ערך של הפונקציה נבחר מבלי תלות באחרים ו־f היא ליפשיץ ביחס ל־D, אז גם המרטינגל מקיים את תנאי ליפשיץ $|X_i - X_{i-1}| \le 1$ לכל i.

נחזור לשאלה שאיתה התחלנו. אנו מעוניינים להראות חסם מהצורה $\Pr\left[|\dim(v_I) - \rho| > \beta\sqrt{d}\right] < e^{-\Omega(\beta^2)}$ (לא ננסה לתת ערך אופטימלי למקדם של סימן ה־Ω).

נניח בלי הגבלת הכלליות כי $n > 10d$, כי אחרת ניתן להשתמש באי שוויון אזומה ומרטינגל חשיפה רגיל. כזכור ההתפלגות μ היא ההתפלגות היוניפורמית מעל ת"ק של $\{1,\ldots,n\}$. נקבע בהתאמה $\mathcal{D} = \{1,\ldots,n\}$

ו־$\mathcal{R}=\{0,1\}$, ואז כל ת"ק $I \subseteq \{1,\ldots,n\}$ תתאים לפונקציה האופיינית שלה, ו־μ תהיה ההתפלגות שבה כל ערך נבחר באופן יוניפורמי וב"ת. על מנת להגדיר את סכמת החשיפה D, תהא $g:\mathcal{D}'\to\{0,1\}$ פונקציה עבורה אנחנו רוצים להגדיר את $\mathcal{D}_g$, ונסמן ב־$J_g=\{i\in\mathcal{D}':g(i)=1\}$ את האינדקסים החברים בקבוצה המתאימה לה. נבחין בין שני מקרים:

- אם $\{1,\ldots,n\}\setminus\mathcal{D}'$ מכיל אינדקס j_g עבורו v_{j_g} בלתי תלוי ב־v_{J_g}, נבחר j_g כזה ונקבע $\mathcal{D}_g=\mathcal{D}'\cup\{j_g\}$.
- אם אין j_g כזה, אז בהכרח $\dim\left(v_{J_g}\right)=\dim\left(v_{J_g\cup(\mathcal{D}\setminus\mathcal{D}')}\right)$. במקרה הזה נקבע $\mathcal{D}_g=\{1,\ldots,n\}$.

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

במקרה הראשון ליצירת $\mathcal{D}_g$ מתקיים $|\mathcal{D}_g\setminus\mathcal{D}'|=1$, וכן אם I_1,I_2 נבדלות לכל היותר בקואורדינטה אחת, אז מימדי קבוצות הוקטורים המתאימות v_{I_1},v_{I_2} יבדלו גם כן לכל היותר ב־1. במקרה השני נשים לב כי אם I_1,I_2 מזדהות עם g על $\mathcal{D}'$ אז המימד של שתי קבוצות הוקטורים המתאימות שווה ל־$\dim\left(v_{J_h}\right)$ בכל מקרה.

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

תהא I קבוצה הנבחרת באופן אקראי כבהגדרת מרטינגל החשיפה. לכל אינדקס $0\le i\le 10d$ נסמן $J_i=I\cap\mathcal{D}_i(I)$ ו־$d_i=\dim(v_{J_i})$. כך, אם $\mathcal{D}_i(I)=\mathcal{D}$ אז $X_i(I)=X_n(I)$. אם $X_i(I)\neq X_n(I)$, אז מהאופן שבחרנו את D אנו יודעים כי $\mathcal{D}_i(I)\setminus\mathcal{D}_{i-1}(I)$ מכילה איבר בודד, שנסמן ב־j_i. כיוון ש־j_i נבחר להיות ב־I (או במושגים של פונקציות, $I(j_i)$ נבחר להיות שווה ל־1) באופן בלתי תלוי ב־$I\cap\mathcal{D}_{i-1}$, האיבר הנ"ל ייכנס ל־J_i 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}.

בנוסף, מאי שוויון אזומה מתקיים $\Pr\left[|X_0-X_{10d}|>\beta\sqrt{d}\right]<2e^{-\beta^2/20}$. מחסם האיחוד מעל שני המאורעות ה"רעים" (שהמרטינגל לא מתקצר או ש־X_{10d} אינו קרוב מספיק ל־X_0) אנחנו מקבלים חסם מהצורה $\Pr\left[|\dim(v_I)-\rho|>\beta\sqrt{d}\right]<e^{-\Omega(\beta^2)}$, שכן $X_0=\rho$ ו־X_n מתפלג כמו $\dim(v_I)$ (עבור $\beta\le\sqrt{d}$ זוהי התוצאה של החיבור של שתי ההסתברויות, ועבור $\beta>\sqrt{d}$ ההסתברות היא 0 בכל מקרה).

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

עוד על ההגדרה הפורמלית של התניה על מ"מ

נתמקד כעת בשימוש בסימון מסוג "$\mathrm{E}[X|Y]$" כסימון מקוצר לביטויים מהצורה "$\mathrm{E}[X|Y=\beta]$". לביטוי הזה יש משמעות מתמטית מדוייקת, ואנחנו נראה כאן את משמעותו עבור מרחבי הסתברות בדידים. הדיון כאן יהיה עבור מרחב הסתברות בדיד μ מעל קבוצת הבסיס S.

התוצאה של $\mathrm{E}[X|Y]$ היא למעשה משתנה מקרי מעל מרחב ההסתברות. לכל β המקיים $\Pr[Y=\beta]>0$, נגדיר את המאורע ש־Y קיבל את הערך הזה: $E_\beta=\{s\in S:Y(s)=\beta\}$. שימו לב שאלו מאורעות זרים שמכסים את מרחב ההסתברות (ז"א שההסתברות לאיחודם שווה ל־1). עתה נגדיר את המ"מ Z לפי $Z(s)=\mathrm{E}[X|E_{Y(s)}]$. הערך $Z(s)$ על s עבורם $\mu(s)=0$ אינו חשוב. במילים אחרות, אנחנו "מחלקים" את S לתתי־קבוצה לפי הערכים של Y, ועל כל תת־קבוצה כזו Z תקבל את הערך של התוחלת המותנה המתאימה של X. הביטוי $\mathrm{E}[X|Y]$ מוגדר להיות המשתנה המקרי Z, ואז השוויונים המשתמשים בביטויים מעין זה בהקשר של מרטינגלים יכולים להתפרש כשוויונים בין משתנים מקריים (כאשר לא מחייבים את השוויון על איברים בהסתברות 0).

לסיום נשים לב לבעיה המתעוררת מעל מרחבי הסתברות לא־בדידים: במקרה כזה יכול להיות שאין ל־Y ערכים בהסתברות חיובית. לדוגמה, Y יכול להתפלג יוניפורמית מעל הקטע $[0,1]$. בתורת המידה יש משפטים "כבדים" שמאפשרים להגדיר את המ"מ $\mathrm{E}[X|Y]$ גם עבור מקרים אלו.

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 $A_i \subseteq R$.

עתה נגדיר גרף תלויות D עבור המאורעות. קבוצת הצמתים של D תהיה I, ולכל $i, j \in I$ שונים זה מזה נגיד ש־ij היא קשת של הגרף אם $A_i \cap A_j \neq \emptyset$. בפרט, אם $i \neq j$ אבל ij אינה קשת של D, אז B_i, B_j הם מאורעות בלתי תלויים. יתרה מזו, אם $J \subset I$ ו־$i \in I \setminus J$ הוא צומת שאין קשתות בינו לבין 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. מכן נובע שאם קבוצת צמתים J היא חסרת קשתות, אז המאורעות המתאימים לה הם ב"ת לחלוטין.

נגדיר את "הערך שהיה ל־$\Pr\left[\bigwedge_{i \in I} \neg B_i\right]$ לו היו ה־B_i בלתי תלויים", $M = \prod_{i \in I} \Pr\left[\neg B_i\right]$, ומדד לתלות המאורעות $\Delta = 2\sum_{ij \in E(D)} \Pr\left[B_i \wedge B_j\right]$ (הסכום הוא על קבוצת הקשתות של הגרף D). 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־$M \leq e^{-\mu}$: נשים לב כי $\Pr\left[\neg B_i\right] = 1 - \Pr\left[B_i\right] \leq e^{-\Pr[B_i]}$ וכך נקבל את החסם $M = \prod_{i \in I} \Pr\left[\neg B_i\right] \leq \prod_{i \in I} e^{-\Pr[B_i]} = \exp\left(-\sum_{i \in I} \Pr\left[B_i\right]\right) = \exp\left(-\mu\right)$.

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

הוכחה: ראשית עלינו לנצל את אי השוויון הבא: לכל תת קבוצה $J \subset I$ כך ש־$i \notin J$ מתקיים כי $\Pr\left[B_i | \bigwedge_{j \in J} \neg B_j\right] \leq \Pr\left[B_i\right]$. 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 $\Pr\left[\neg B_i | \bigwedge_{1 \leq j < i} \neg B_j\right] \geq \Pr\left[\neg B_i\right]$. החסם התחתון מתקבל לפי נוסחת ההסתברות המותנה:

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

נשים לב שאי השוויון שהשתמשנו בו קודם (זה שעוד לא הוכחנו) תקף גם את נתנה את שני צדדיו במאורע B_k עבור $k \notin J$ כך שבגרף D אין קשתות בין k ל־J, מכיוון שהמאורע B_k יהיה בלתי תלוי בכל הצירופים של המאורעות $\{B_j\}_{j \in J}$. בזאת מקבלים לכל $J \subset I$ כך ש־$i \in J$, $k \notin J$ ואין קשתות בין k ל־J, 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־$D_i = N_D(i) \cap \{1, \ldots, i-1\}$ את קבוצת ה j־ים הקטנים מ־i שהם שכנים שלו, ונסמן את $\bar{D}_i = \{1, \ldots, i-1\} \setminus D_i$. מנוסחת ההסתברות המותנה (כאשר מתנים את שני צידי הנוסחה על מאורע נוסף), לכל שלושה מאורעות A, B, C מתקיים $\Pr\left[A | B \wedge C\right] \geq \Pr\left[A \wedge B | C\right]$. אצלנו נציב את $A = B_i$, את $B = \bigwedge_{j \in D_i} \neg B_j$ ואת $C = \bigwedge_{j \in \bar{D}_i} \neg B_j$, ונשים לב כי A לא תלוי ב C.

מתקיים:

$$\begin{aligned}\Pr\left[B_i\middle|\bigwedge_{1\le j<i}\neg B_j\right] &= \Pr\left[B_i\middle|\bigwedge_{j\in D_i}\neg B_j\wedge\bigwedge_{k\in\bar{D}_i}\neg B_k\right]\\ &\ge \Pr\left[B_i\wedge\bigwedge_{j\in D_i}\neg B_j\middle|\bigwedge_{k\in\bar{D}_i}\neg B_k\right]\\ &= \Pr\left[B_i\middle|\bigwedge_{j\in\bar{D}_i}\neg B_j\right]\Pr\left[\bigwedge_{j\in D_i}\neg B_j\middle|B_i\wedge\bigwedge_{k\in\bar{D}_i}\neg B_k\right]\\ &= \Pr\left[B_i\right]\Pr\left[\bigwedge_{j\in D_i}\neg B_j\middle|B_i\wedge\bigwedge_{k\in\bar{D}_i}\neg B_k\right]\end{aligned}$$

נחסום עתה את ההסתברות המותנה בביטוי:

$$\Pr\left[\bigwedge_{j\in D_i}\neg B_j\middle|B_i\wedge\bigwedge_{k\in\bar{D}_i}\neg B_k\right]\ge 1-\sum_{j\in D_i}\Pr\left[B_j\middle|B_i\wedge\bigwedge_{k\in\bar{D}_i}\neg B_k\right]\ge 1-\sum_{j\in D_i}\Pr\left[B_j|B_i\right]$$

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

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

נעבור למאורעות המשלימים:

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

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

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

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 $j \sim i$ 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 $3(n-3)p^5$, 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

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

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 $3\binom{n}{3}p^5(n-3)/2\left(1-p^3\right) \le c$, 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 $B_1, \ldots, B_m$ ורשימת תלויות עבורם $D_1, \ldots, D_m$, כך שלכל i מתקיים $\Pr[B_i] < \frac{1}{2}$ וכן $\sum_{j \in D_i} \Pr[B_j] \leq \frac{1}{4}$, אז מתקיים $\Pr[\bigwedge_{i=1}^m \neg B_i] > 0$.

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

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

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 $\Pr[\bigwedge_{i=1}^m \neg B_i] \geq \prod_{i=1}^m (1 - 2\Pr[B_i]) > 0$ 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" $A_1, \ldots, A_n$ וידוע כי לכל אחד מהם $\Pr[A_i] \leq p_i < 1$, 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, $\Pr[A_i] \leq 2^{-k}$. על כן תנאי הלמה הלוקלית הסימטרית יוצא $2^{-k}\left(2^k e^{-1}\right) \leq e^{-1}$ 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 (כולל C עצמה):
   (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 נוכיח שהאלגוריתם עוצר בהסתברות גבוהה עבור מקרה פשוט יחסית בו כל פסוקית נחתכת עם לכל היותר $2^{k/2-b}$ פסוקיות אחרות, כאשר b הוא קבוע שנבחר בקרוב. כיום ידועה גרסה אלגוריתמית שעובדת עבור כל הפרמטרים שבהם הלמה הלוקלית הלא־קונסטרוקטיבית עובדת.

ניתוח האלגוריתם של מוזר

אנחנו נראה כאן חסם על זמן הריצה שמתקיים בהסתברות לפחות $\frac{1}{2}$. לא קשה לתרגם את זה לחסם על תוחלת זמן הריצה גם. ניתוחי האלגוריתם המצויים בספרות בד"כ משתמשים או במושג של אקראיות קולמוגורוב (Kolmogorov), שעבורו למרבה הצער אין זמן בקורס זה, או בניתוח מבוסס אנטרופיה (לקראת סוף הקורס נלמד על אנטרופיה, אולם לא נגיע לניתוח של האלגוריתם של מוזר). למתעניינים בסיבוכיות קולמוגורוב מומלץ להסתכל בספר

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

עבור ההוכחה של המקרה הפשוט כאן ננסח טענה פשוטה ספציפית שתספיק עבור השימוש שלנו: עבור n קבוע, נניח ש־$f:\{0,1\}^* \to \{0,1\}^*$ היא פונקציה נתונה מראש ל"תרגום" של מחרוזות סופיות, וש־$x \in \{0,1\}^{\mathbb{N}}$ היא מחרוזת אינסופית של ביטים, שנבחרת ע"י כך שכל $x_i \in \{0,1\}$ נבחר באופן יוניפורמי וב"ת בביטים האחרים. בהסתברות לפחות $\frac{1}{2}$, לא קיימת שום מחרוזת y מאורך $n+t$ עבור $t \geq 2$, שעבורה $f(y)$ תהיה מאורך לפחות $n+2t$ וגם תסכים עם $x_1, \ldots, x_{n+2t}$ ב־$n+2t$ התווים הראשונים.

ההוכחה היא לפי חסם על הסתברות האיחוד של מספר בן מניה של מאורעות: לכל t ספציפי ישנן 2^{n+2t} אפשרויות עבור $x_1,\dots,x_{n+2t}$, אבל רק 2^{n+t} אפשרויות עבור $f(y)$ כאשר y היא מחרוזת מאורך $n+t$. על כן הסיכוי שבחרנו את x כך שקיים y מאורך $n+t$ שסותר את הטענה הוא 2^{-t}. כל שנותר הוא לחסום את איחוד המאורעות ע"י $\sum_{t=2}^{\infty}2^{-t}=\frac{1}{2}$.

נחזור לניתוח שלנו: נניח ש־x היא המחרוזת המספקת את כל הביטים המקריים שהאלגוריתם שלנו משתמש בהם, ז"א שכל פעם שהאלגוריתם צריך ערך מקרי, הוא משתמש בביט הבא של x. נסמן את מספר הקריאות (כולל הרקורסיביות) לפונקציית התיקון המקומי ב־s, ונראה שעבור s גדול מדי אפשר לבנות פונקציה שתתנהג כמו f שבטענה למעלה.

האלגוריתם משתמש סה"כ ב־$n+sk$ ביטים מתוך המחרוזת x: צריך n ביטים עבור ההשמה המקרית הראשונה, ועוד k ביטים לכל קריאה לפונקציית התיקון המקומי. מצד שני, אם נדע מה היא הפסוקית שמתוקנת, נדע שהיא הייתה מופרת ולכן נדע בדיוק את ערכי הביטים מ־x שהיו בה לפני שתוקנה, שכן לכל פסוקית יש השמה לא־מספקת יחידה. כך אפשר יהיה לתאר את $x_1,\dots,x_{n+sk}$ באופן אלטרנטיבי, על ידי המזהים של סדרת הפסוקיות שהתיקון המקומי מתקן, ולאחריהם 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 על מנת שנוכל לזהות אותם ע"י מספר סידורי. יש לכל היותר $2^{k/2-b}$ פסוקיות כאלה, ולכן נדרש ל־$k/2-b+c$ ביטים (עבור c קבוע גדול דיו) על מנת לתאר כל אחת מהן, יחד עם סימן מיוחד לסוף הרקורסיה. נשים לב שכאשר קריאת LocalFix על פסוקית C מסתיימת, אז הפסוקית מסופקת. זאת מכיוון שאנחנו 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. בפרט נובע מכך ש־SolveLovasz תריץ בעצמה את LocalFix לא יותר מפעם אחת על כל פסוקית של F.

על כן, תיאור מלא של $x_1,\dots,x_{n+sk}$ ידרש ל־$m\log m$ ביטים עבור רשימת הפסוקיות שעליהן נקרא התיקון המקומי מתוך הלולאה הכללית (המדובר בסדרה ללא חזרות של ערכים ב־$1,\dots,m$, שיש עבורה פחות מ־$\frac{1}{2}m^m$ אפשרויות), $s(k/2-b+c)$ ביטים לרשימת הפסוקיות שעליהן נקרא התיאור המקומי רקורסיבית, ו־n ביטים לתיאור ההשמה הסופית. אפשר לכתוב פונקציה f שבהינתן תיאור כזה בן $n+m\log m+s(k/2-b+c)$ ביטים היא תכתוב את $x_1,\dots,x_{n+sk}$.

מכוון שהמחרוזת x היא אקראית יוניפורמית, בהסתברות לפחות $\frac{1}{2}$ היא תקיים את הטענה למעלה עבור n והפונקציה f שתוארה, ואז בהכרח מתקיים $n+sk<n+2(m\log m+s(k/2-b+c))$. מהעברת אגפים מתקבל $m\log m>s(b-c)$. עבור בחירה של $b=c+1$ נקבל $s<m\log m$, ז"א שאם אנחנו מגבילים את מספר הקריאות ל־LocalFix ל־$m\log m$ (דבר אשר יבטיח זמן ריצה פולינומי עבור האלגוריתם), בהסתברות לפחות $\frac{1}{2}$ האלגוריתם יעצר.

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 $f:\mathcal{P}(S)\to\mathbb{R}$ אי־שלילית ומונוטונית לא־יורדת, $h:\mathcal{P}(S)\to\mathbb{R}$ אי־שלילית ומונוטונית לא־עולה, ו־$\mu:\mathcal{P}(S)\to\mathbb{R}$ אי־שלילית ולוג־סופר־מודולרית. נגדיר $\alpha=\max_{A\subseteq S}h(A)$ ואז $g(A)=\alpha-h(A)$ 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 $p=(p_1,\ldots,p_n)$ 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 $\Pr_p\left[A\right]=\prod_{i\notin A}(1-p_i)\prod_{j\in A}p_j$. 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 $\mu_p:P([n])\to\mathbb{R}^+$ על ידי $\mu_p(A)=\Pr_p\left[A\right]$. זאת פונקציה לוג־סופר־מודולרית, שכן מתקיים $\mu_p(A)\,\mu_p(B)=\mu_p(A\cup B)\,\mu_p(A\cap B)$, 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־$\mu_p(A),\mu_p(A\cup B)$ ו־$1-p_i$ ל־$\mu_p(B),\mu_p(A\cap B)$, 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 $\Pr_p\left[\mathcal{A}\cap\mathcal{B}\right]=\Pr\left[B_i\wedge\bigwedge_{j\in J}\neg B_j\right]$ וכן מתקיים $\Pr_p\left[\mathcal{A}\right]=\Pr\left[B_i\right],\Pr_p\left[\mathcal{B}\right]=\Pr\left[\bigwedge_{j\in J}\neg B_j\right]$, ואז מההכללה שהוכחנו למשפט קלייטמן מתקיים כי

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

כעת,

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

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 $\mathrm{H}[X,Y]\leq\mathrm{H}[X]+\mathrm{H}[Y]$, ובאינדוקציה נוכל לקבל כי אם $X=\langle X_1,\ldots,X_n\rangle$ משתנה מקרי המקבל ערכים ב־$S=S_1\times\ldots\times S_n$ אז מתקיים $\mathrm{H}[X]\leq\sum_{i=1}^{n}\mathrm{H}[X_i]$. אי שוויון המכליל זאת הוכח ב־1986 על ידי Shearer.

משפט שירר: תחת הסימונים לעיל, אם $\mathcal{A}$ משפחה של תתי קבוצות של $\{1,\ldots,n\}$ וכל $1\leq a\leq n$ שייך לפחות k איברים של $\mathcal{A}$, אז $k\mathrm{H}[X]\leq\sum_{A\in\mathcal{A}}\mathrm{H}[X_A]$, כאשר X_A הוא המ"מ שמקבל ערכים מהקבוצה $\prod_{a\in A}S_a$ לפי $X_A=\langle X_a:a\in A\rangle$ (ז"א ש־X_A הוא סדרת הערכים המתאימים ל־X_a עבור $a\in A$).

על מנת להוכיח את המשפט, ראשית נשים לב שאם כותבים $A=\{a_1,\ldots,a_l\}$ עבור $1\leq a_1<\cdots<a_l\leq n$, מהפעלות חוזרות של כלל השרשרת מתקיים

$$\mathrm{H}[X_A]=\mathrm{H}[X_{a_l}]+\mathrm{H}[X_{\{a_2,\ldots,a_l\}}|X_{a_1}]=\cdots=\sum_{i=1}^{l}\mathrm{H}[X_{a_i}|X_{\{a_1,\ldots,a_{i-1}\}}]$$

עתה מאי השוויון $\mathrm{H}[X|Y,Z]\leq\mathrm{H}[X|Y]$ מתקיים $\mathrm{H}[X_{a_i}|X_{\{a_1,\ldots,a_{i-1}\}}]\geq\mathrm{H}[X_{a_i}|X_{\{1,\ldots,a_i-1\}}]$, ולכן

$$\mathrm{H}[X_A]=\sum_{i=1}^{l}\mathrm{H}[X_{a_i}|X_{\{a_1,\ldots,a_{i-1}\}}]\geq\sum_{i=1}^{l}\mathrm{H}[X_{a_i}|X_{\{1,\ldots,a_i-1\}}]=\sum_{a\in A}\mathrm{H}[X_a|X_{\{1,\ldots,a-1\}}]$$

באמצעות סכימה מעל $\mathcal{A}$ נקבל את אי השוויון של משפט שירר:

$$\sum_{A\in\mathcal{A}}\mathrm{H}[X_A]\geq\sum_{A\in\mathcal{A}}\sum_{a\in A}\mathrm{H}[X_a|X_{\{1,\ldots,a-1\}}]\geq k\sum_{a=1}^{n}\mathrm{H}[X_a|X_{\{1,\ldots,a-1\}}]=k\mathrm{H}[X]$$

אי השוויון באמצע נובע מההנחה שכל אינדכס a מופיע בלפחות k מהאיברים של $\mathcal{A}$.

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 $\Pr_{A,x}\left[A\left(\mathrm{C}\left(x\right),i\right)=x_i\right]\geq1/2+\epsilon$. 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. הדרישה שלנו היא חלשה למדי, שכן אנחנו אפילו לא מחייבים ש־C תהיה חח"ע.

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 $\Pr_{A,x}\left[A\left(\mathrm{C}\left(x\right),i\right)=x_i\right] \geq 1/2+\epsilon$, כאשר ההסתברות נלקחת גם על האקראיות של A וגם על בחירה אקראית של מחרוזת x. אז מתקיים $\log|R| \geq \left(1-H\left(1/2+\epsilon\right)\right)n$.

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

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

בנוגע לאיבר האחרון נשים לב לכך שבהנתן קידוד $\mathrm{C}\left(x\right)$, אם הגרלנו את האקראיות של האלגוריתם אז x_i יהיה שווה לערך $A\left(\mathrm{C}\left(x\right),i\right)$ בהסתברות לפחות $1/2+\epsilon$, ולכן אפשר לראות בו משתנה מקרי המקבל בהסתברות מסויימת את הערך $A\left(\mathrm{C}\left(x\right),i\right)$ ובהסתברות המשלימה את הערך ההפוך. מכוון שהאנטרופיה גדלה ככל שההתפלגות קרובה יותר ליוניפורמיות, מתקיים $\mathrm{H}\left[x_i|\mathrm{C}\left(x\right)\right] \leq H\left(1/2+\epsilon\right)$ וכך (יחד עם $\mathrm{H}\left[x\right]=n$) 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 $C\left(x\right)$ הוא רישא של $C\left(y\right)$ 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 $\ell=\mathbb{E}_{x\sim p}\left[|C\left(x\right)|\right]$. 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 $l_1 \leq l_2 \leq \ldots \leq l_m$. 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 $\sum_{i=1}^{m}2^{(l_m-l_i)} \leq 2^{l_m}$. 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 $\sum_{i=1}^{m} 2^{-l_i} \leq 1$ 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: $\ell \geq \mathrm{H}(p)$.

הוכחה: נכתוב במפורש, תחת הסימונים הקודמים:

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

כאשר השתמשנו באי השוויון $\ln x \leq x - 1$ 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 $\sum_{i=1}^{m} 2^{-\lceil \log \frac{1}{p_i} \rceil} \leq \sum_{i=1}^{m} 2^{-\log \frac{1}{p_i}} = \sum_{i=1}^{m} p_i = 1$ ולכן קיים קוד חסר רישות שזה אורך המילים בו. נבחן את אורך המילה הממוצע בעיני האנטרופיה: $\ell = \sum_{i=1}^{m} p_i \lceil \log \frac{1}{p_i} \rceil \leq \sum_{i=1}^{m} p_i \left(\log \frac{1}{p_i} + 1\right) = H(p) + 1$.
שימו לב שאם כל ה־p_i הם חזקות שלמות של 2 אז אנחנו נשיג במדויק את האנטרופיה.

אינפורמציה משותפת ואי־שוויון פאנו עבור שרשראות מרקוב קצרות

עבור שלושה משתנים מקריים, X, Y ו־Z, נגדיר את האינפורמציה המשותפת של X ו־Y המותנה על Z לפי האנטרופיות המותנות המתאימות, הגדרה שהיא שקולה להגדרה לפי התוחלת של הביטויים המותנים על הערכים האפשריים של Z: $\mathrm{I}[X,Y|Z] = \mathrm{H}[X|Z] + \mathrm{H}[Y|Z] - \mathrm{H}[X,Y|Z] = \mathrm{E}_{\gamma\sim Z}\big[\mathrm{I}[X,Y|Z=\gamma]\big]$. נזכיר שהסימון בצד ימין פירושו $\sum_{\gamma:\Pr[Z=\gamma]>0} \Pr[Z=\gamma]\mathrm{I}[X,Y|Z=\gamma]$, כאשר משתמשים שם בחישובים של האינפורמציה המשותפת במרחבים מותנים על המאורעות $Z=\gamma$.

חישוב ישיר מאפשר לנו לנסח כלל שרשרת עבור חישוב המידע המשותף המותנה. בסימונים הבאים נשתמש בסימון $\mathrm{I}[X,(Y,Z)] = \mathrm{H}[X] + \mathrm{H}[Y,Z] - \mathrm{H}[X,Y,Z]$ לתאר את המידע המשותף בין X לבין המ"מ שמוגדר כ"שרשור" שני המשתנים Y ו־Z. נקבל:

$$\begin{aligned}
\mathrm{I}[X,Y|Z] &= \mathrm{H}[X|Z] + \mathrm{H}[Y|Z] - \mathrm{H}[X,Y|Z] \\
&= \mathrm{H}[X,Z] + \mathrm{H}[Y,Z] - \mathrm{H}[Z] - \mathrm{H}[X,Y,Z] \\
&= \mathrm{I}[X,(Y,Z)] - \mathrm{I}[X,Z]
\end{aligned}$$

עבור המשך הדיון כאן נתייחס לשלשת משתנים מקריים בדידים X,Y,Z שמהווה שרשרת מרקוב קצרה: זה אומר שלכל שלשת ערכים α,β,γ שמקיימת $\Pr[X=\alpha\wedge Y=\beta\wedge Z=\gamma]>0$, מתקיים תנאי חוסר הזיכרון $\Pr[Z=\gamma|X=\alpha\wedge Y=\beta]=\Pr[Z=\gamma|Y=\beta]$. אתם מוזמנים לבדוק ש־$X,Y,Z$ היא שרשרת מרקוב אם ורק אם ה"היפוך" שלה Z,Y,X הוא גם שרשרת מרקוב.

שרשרת מרקוב מדמה סדרה של תהליכים מקריים: אפשר לראות את Y כתוצאה של ביצוע עיבוד הסתברותי של הערך של X, שלאחריו מבצעים עוד הליך הסתברותי לקבלת Z מהערך של Y (לאחר ש"שכחו את ההסטוריה" הגלומה ב־X עצמו). נוכיח עתה שעבור שרשרת כזו מתקיים $\mathrm{I}[X,Z]\leq\mathrm{I}[X,Y]$, טענה שנקראת אי־שוויון עיבוד המידע, ומתאימה לאינטואיציה שככל שאנחנו מפעילים תהליכים על ערכו של מ"מ נתון, אנחנו לא יכולים "להוסיף לקשר" שלו עם משתנים שאת ערכם אנחנו לא רואים ישירות.

עבור ההוכחה, ראשית נשים לב שתנאי חוסר הזיכרון של שרשרת מרקוב שקול לתנאי אי־התלות המותנה $\Pr[X=\alpha\wedge Z=\gamma|Y=\beta]=\Pr[X=\alpha|Y=\beta]\cdot\Pr[Z=\gamma|Y=\beta]$ עבור כל α,β,γ שמקיימים $\Pr[X=\alpha\wedge Y=\beta\wedge Z=\gamma]>0$. מכאן נובע שמתקיים $\mathrm{I}[X,Z|Y=\beta]=0$ לכל β שעבורו $\Pr[Y=\beta]>0$, ולכן $\mathrm{I}[X,Z|Y]=0$. נותר לפתח תוך שימוש בכלל השרשרת:

$$\mathrm{I}[X,Z]\leq\mathrm{I}[X,Z]+\mathrm{I}[X,Y|Z]=\mathrm{I}[X,(Y,Z)]=\mathrm{I}[X,Y]+\mathrm{I}[X,Z|Y]=\mathrm{I}[X,Y]$$

לבסוף נראה את אי־שוויון פאנו Fano עבור שרשרת מרקוב X,Y,Z. אי־השוויון קובע שאם ל־X ו־Z יש m ערכים אפשריים (הכוונה היא לאיחוד קבוצות הערכים האפשריים של שני המ"מ), אז מתקיים $\mathrm{H}[X|Y]\leq\Pr[X\neq Z]\log(m-1)+H(\Pr[X\neq Z])$, כאשר ה־"$H$" בצד ימין הוא פונקצית האנטרופיה מעל $[0,1]$ שהוגדרה בהרצאות. הרעיון מאחורי הניסוח הוא זה: על X מסתכלים כעל נעלם, על Y כעל תוצאת ניסוי שנועד למצוא את ערך X, ועל Z מסתכלים כעל "פירוש" של אותה תוצאה. המשפט חוסם את כמות האינפורמציה הכלולה ב־X שאינה כלולה ב־Y, במושגים של "סיכויי ההצלחה" של Z.

הגרסה ה"מקוצרת" של אי־שוויון פאנו מתקבלת כאשר עבור שני משתנים X ו־Z כל שהם בוחנים את השרשרת X,Z,Z. הוא קובע בפשטות שמתקיים $\mathrm{H}[X|Z]\leq\Pr[X\neq Z]\log(m-1)+H(\Pr[X\neq Z])$, כאשר ל־$X$ ו־Z יש m ערכים אפשריים. על מנת להוכיח את הגרסה המלאה מספיק להוכיח את הגרסה המקוצרת, כי מאי־השוויון $\mathrm{I}[X,Z]\leq\mathrm{I}[X,Y]$ נובע מיידית $\mathrm{H}[X|Y]\leq\mathrm{H}[X|Z]$.

עבור הוכחת הגרסה המקוצרת, נגדיר משתנה אינדיקטור חדש F עבור המאורע "$X\neq Z$", ונפתח את $\mathrm{H}[X,F|Z]$ לפי כלל השרשרת בשתי דרכים. מצד אחד מתקיים $\mathrm{H}[X,F|Z]=\mathrm{H}[F|Z]+\mathrm{H}[X|F,Z]$, מכיוון שלכל γ שעבורו $\Pr[Z=\gamma]>0$ מתקיים $\mathrm{H}[X,F|Z=\gamma]=\mathrm{H}[F|Z=\gamma]+\mathrm{H}[X|F,Z=\gamma]$ לפי כלל השרשרת ה"רגיל" (המעבר מזה ל"התניה על Z" דומה למה שנעשה בהרצאה בהוכחה של אי־השוויון מהצורה "$\mathrm{H}[X|Y,Z]\leq\mathrm{H}[X|Y]$"). מצד שני מתקיים $\mathrm{H}[X,F|Z]=\mathrm{H}[X|Z]+\mathrm{H}[F|X,Z]$. לסיום ההוכחה ננתח את השוויון שהתקבל, $\mathrm{H}[X|Z]+\mathrm{H}[F|X,Z]=\mathrm{H}[F|Z]+\mathrm{H}[X|F,Z]$.

ראשית נשים לב שמתקיים $\mathrm{H}[F|X,Z]=0$ כי F כאן הוא פונקציה של שני המשתנים X ו־Z. על כן מתקיים $\mathrm{H}[X|Z]=\mathrm{H}[F|Z]+\mathrm{H}[X|F,Z]$. עבור המחובר הראשון נכתוב $\mathrm{H}[F|Z]\leq\mathrm{H}[F]=H(\Pr[X\neq Z])$ (נזכור ש־F הוא משתנה אינדיקטור של המאורע הנ"ל).

עבור המחובר השני נשתמש בהגדרת האנטרופיה המותנה (כשמשאירים את ההתניה לפי Z ו"מפרקים" את זו לפי F), ונקבל $\mathrm{H}[X|Z,F]=\mathrm{H}[X|Z,F=0]\Pr[F=0]+\mathrm{H}[X|Z,F=1]\Pr[F=1]$. נשים לב שמתקיים $\mathrm{H}[X|Z,F=0]=0$ (כי בהתניה על $F=0$ מתקיים $X=Z$), ומהחסם על האנטרופיה לפי מספר הערכים האפשריים $\mathrm{H}[X|Z,F=1]\leq\log(m-1)$ (לכל ערך אפשרי γ של Z, המשתנה X מתפלג על לא יותר מ־$m-1$ הערכים השונים מ־γ). בהצבת $\Pr[F=1]=\Pr[X\neq Z]$ נקבל $\mathrm{H}[X|F,Z]\leq\Pr[X\neq Z]\log(m-1)$, וסה"כ נקבל את אי־השוויון הנדרש $\mathrm{H}[X|Z]\leq\Pr[X\neq Z]\log(m-1)+H(\Pr[X\neq Z])$.

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 $P^T\pi = \pi$. 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, $P^Td = AD^{-1}d = A(1,1,\dots,1)^T = d$, 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 נרמל 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־$(wP)_i = \sum_{k=1}^{m} w_k P_{k,i} = w_i = 0$, 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 $|w_1| \ge |w_i|$ לכל אינדקס i. אז $\lambda w_1 = \frac{1}{d(1)} \sum_{(1,j)\in E} w_j$ לפי הכפל במטריצה, וכך

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

ולכן $|\lambda| \le 1$.

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 $u_2 u_1 w$ 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 וקטורי 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 $|q^{(t)} - \pi| \leq \epsilon$ 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 $t \geq n \ln(n/\epsilon)$ יתקיים

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

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 $(h_1, h_2, \dots, h_t) \in [d]^t$, 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 $H = (h_1, \dots, h_t) \in [d]^t$ 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 $k_{s,t} \le 2m$. 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 $k_{u,v} \le 2m$ 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־$nn^{nd}n^{-2nd} < 1$. לכן בהכרח קיימת סדרת הילוך אוניברסלית מאורך $O(dmn^2 \log n)$.

לסיום נעיר שב־2008 פורסמה תוצאה של Reingold שמוכיחה שוויון בין מחלקת הסיבוכיות של אלגוריתמים דטרמיניסטים עם זיכרון לוגריתמי לבין זו של אלגוריתמים אקראיים עם זיכרון לוגריתמי. דרך ההוכחה כללה בניית אלגוריתם דטרמיניסטי שבונה הילוך אוניברסלי מאורך פולינומי בפרמטרים d, m, n.