שיטות הסתברותיות ואלגוריתמים – הרצאות

אלדר פישר, חדר 625, טלפון 3967, eldar@cs

10 באוגוסט 2024

הקדמה ועניינים טכניים

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).

הספר הבא מכיל מבוא בסיסי לתורת האנטרופיה (הפרק החוברת מסתמך עליו ועל מקורות נוספים):

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

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 בין השאר 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.

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.

סימונים מקובלים

במהלך הקורס יהיה שמוש בסימונים הבאים עבור התנהגות אסימפטוטית של פונקציות $f(n)$ ו־$g(n)$.

- $f(n)=O(g(n))$ פירושו שקיים קבוע $C<\infty$ כך שעבור כל n גדול דיו מתקיים $|f(n)|\le C|g(n)|$. במילים אחרות – $\frac{f(n)}{g(n)}$ חסום בערכו המוחלט החל מ־n מסויים.
- $f(n)=\Omega(g(n))$ פירושו שמתקיים $g(n)=O(f(n))$.
- $f(n)=\Theta(g(n))$ פירושו שמתקיים $f(n)=O(g(n))$ וכן $f(n)=\Omega(g(n))$.
- $f(n)=o(g(n))$ פירושו שמתקיים $\lim_{n\to\infty}\frac{f(n)}{g(n)}=0$. בפרט אם $f(n)=o(1)$ אז הפונקציה $f(n)$ שואפת לאפס.
- $g(n)\gg f(n)$ פירושו בד"כ כלל זהה ל־$f(n)=o(g(n))$. עם זאת במאמרים כיום משתמשים גם במשמעות שניה לביטוי זה: לעיתים כותבים "נבחר $0 \ll \alpha \ll \beta$" כאשר הכוונה היא "נבחר את α להיות קטן מפונקציה מתאימה של β (אבל גדול מ־0), אשר על טיבה נעמוד בהמשך ההוכחה" (שימו לב לכמתים הלוגים המסתתרים במשפט זה). במהלך הקורס נשתדל להימנע מסימון זה.
- $f(n)=\omega(g(n))$ פירושו בספרות $g(n)=o(f(n))$ (במהלך הקורס אעדיף למעט להשתמש בסימון זה).
- בביטוי מתמטי אפשר להחליף חלק מהביטוי בסימון מהסימונים למעלה, שפירושו הוא "פונקציה כל שהיא המקיימת את...". למשל, $f(n)=(1+o(1))g(n)$ פירושו שמתקיים $\lim_{n\to\infty}\frac{f(n)}{g(n)}=1$ (לפעמים מצב זה מסומן בספרות כ־$f(n)\sim g(n)$).

מכיוון שהקורס יכלול דוגמאות רבות מתורת הגרפים, כדאי להזכיר כאן מספר סימונים בנושא זה. נניח שלפנינו גרף לא מכוון ופשוט (חסר לולאות וחסר קשתות מקבילות) G, בעל קבוצת צמתים $V=V(G)$ עם n איברים, ובעל קבוצת קשתות $E=E(G)$ (אלו האותיות שישמשו אותנו בד"כ בהקשר זה). אנו נשתמש במדדים הבאים.

- $\alpha(G)$ יסמן את הגודל המקסימלי של קבוצת צמתים בלתי תלויה ב־G (תת־קבוצה V' של V כך ש־E אינה מכילה קשתות פנימיות ל־V').
- $\omega(G)$ יסמן את מספר הצמתים המקסימלי שיש בקליק של G (ת"ק V' של V עבורה E מכילה את כל הקשתות הפנימיות ל־V' האפשריות).
- $\chi(G)$ יסמן את מספר הצביעה של הגרף (מספר הצבעים המינימלי שבו אפשר לצבוע את צמתי הגרף כך שאף קשת אינה מקשרת בין צמתים מאותו צבע). לכל גרף מתקיים $\chi(G)\ge\max\{\omega(G),\frac{|V(G)|}{\alpha(G)}\}$.
- זיווג מושלם ב־G הוא אוסף של קשתות זרות זו לזו המכסות יחדיו את כל הצמתים של G.
- K_n יסמן את הקליק (הגרף השלם) בעל n הצמתים, ו־$K_{n,m}$ יסמן את הגרף הדו־צדדי השלם בעל מחלקה אחת עם n צמתים ומחלקה אחת עם m צמתים (במילים אחרות, קבוצת הצמתים של $K_{n,m}$ היא איחוד זר של קבוצה V_1 בת n צמתים וקבוצה V_2 בת m צמתים, וקבוצת הקשתות היא קבוצת כל הזוגות האפשריים של צומת מ־V_1 וצומת מ־V_2).
- $G(n,p)$ אינו מסמן גרף מסויים, אלא מרחב הסתברות מעל גרפים בעלי n צמתים. גרף מקרי יוגרל לפי כך שכל קשת אפשרית של G תיבחר באופן בלתי תלוי בקשתות האחרות בהסתברות p. במילים אחרות, עבור קבוצת קשתות E (מעל קבוצה V בת n צמתים) הסיכוי שזו תהיה קבוצת הקשתות של הגרף המוגרל הוא בדיוק $p^{|E|}(1-p)^{\binom{n}{2}-|E|}$. מרחב זה נקרא מודל הגרף המקרי של Erdös-Rényi, למרות שהגרסה הזו שלו הופיעה לראשונה במאמר של Gilbert.

מרחבי ההסתברות שבקורס

בקורס זה (עם כל הצער שבדבר) נמנע ככל האפשר מתיאוריה של תורת המידה. בפרט ברב המקרים מרחבי ההסתברות שלנו יהיו סופיים או לפחות בדידים, והוכחות הסתברותיות נראה למקרה שבו ניתן להשתמש בסכומים רגילים – בד"כ אותם טיעונים יהיו תקפים גם למרחבי הסתברות כלליים (מרחבי מידה עם מידה 1 למרחב כולו) כאשר הסכומים יוחלפו באינטגרלים מוכללים (ואז "משתנים מקריים" יהיו פונקציות מדידות המוגדרות עד כדי הבדל בתת מרחב ממידה 0, וכו').

מבוא לשיטות הסתברותיות

אפילו השיטות ההסתברותיות הבסיסיות ביותר יכולות להביא לתוצאות לא טריביאליות. ראשית נראה ניתוח של אלגוריתם הסתברותי פשוט, ולאחר מכן נראה הוכחה של טענה קומבינטורית המסתמכת על אינטואיציה הסתברותית (למעשה, ההוכחה המסמלת את תחילת הנושא של שיטות הסתברותיות בקומבינטוריקה, אם כי הסטורית היתה הוכחה הסתברותית אחרת לטענה קומבינטורית מספר שנים קודם לכן).

מציאת חתך מינימלי (mincut)

נראה כאן אלגוריתם הסתברותי שניתוחו משתמש בנוסחה הבסיסית להסתברות מותנה. חתך מינימלי בגרף G (לא מכוון, יתכן שעם קשתות מקבילות) הוא חלוקה של הצמתים לשתי קבוצות זרות לא ריקות כך שמספר הקשתות ביניהן מינימלי. נסתכל עתה על האלגוריתם ההסתברותי הבא למציאת גודל החתך המינימלי: נניח שהגרף קשיר, כי אחרת ברור שגודל החתך המינימלי הוא אפס. בכל שלב האלגוריתם בוחר באקראי קשת, ומכווץ אותה – שני צמתי הקשת מוחלפים בצומת יחיד, ולכל קשת שהשתתף בה אחד הצמתים המוחלפים ישתתף בה עתה הצומת החדש. כאן שומרים על הכפילויות בקשתות שיכולות להווצר כתוצאה מהכיווץ, אולם לא שומרים על לולאות. כשנותרים שני צמתים בלבד, פולטים את מספר הקשתות ביניהם.

ננתח את הסיכוי שזהו גודל החתך המינימלי: אם החתך המינימלי הוא בגודל k, אז זהו גם חסם תחתון על הדרגה המינימלית של צמתי הגרף, ולכן מספר הקשתות הכולל הוא לפחות $\frac{kn}{2}$. נסמן ב־E' את קבוצת הקשתות של אחד החתכים המינימלים. האלגוריתם מכווץ $n-2$ פעמים קשת; באיטרציה הראשונה של האלגוריתם, הסיכוי לכווץ קשת מ־E' הוא לא יותר מ־$\frac{2}{n}$ (להוכחת הנ"ל חוסמים את $|E'|/|E|$). באיטרציה ה־i, אם לא כווצה קשת מ־E' קודם לכן, אז הסיכוי שתכווץ עתה קשת מ־E' הוא לא יותר מ־$\frac{2}{n+1-i}$ משיקולים דומים (נשים לב שכל עוד לא כווצה קשת מ־E', הקשתות המתאימות ל־E' יהוו חתך מינימלי גם לאחר הכיווצים). לכן הסיכוי שלא כווצה קשת מ־E' באף שלב הוא לפחות

$$\prod_{i=1}^{n-2}(1-\frac{2}{n+1-i})=\prod_{i=1}^{n-2}(\frac{n-1-i}{n+1-i})=\frac{(n-1-(n-2))(n-1-(n-3))}{(n+1-1)(n+1-2)}=\frac{2}{n(n-1)}$$

מכאן אפשר להראות אלגוריתם עם זמן ריצה פולינומי ב־n שבהסתברות גבוהה (למשל $\frac{2}{3}$, אפשר להגיע גם ל־$1-e^{\Omega(n)}$ ע"י הגדלת מספר ההרצות פי פקטור נוסף של n) מוצא את החתך המינימלי: מריצים את האלגוריתם הנ"ל n^2 פעמים, כל פעם עם הגרלות בלתי תלויות, ובוחרים את החתך (שתי קבוצות הצמתים שכווצו לתוך שני הצמתים הסופיים) שנותן את המינימום מבין ההרצות. ע"י שימוש באי השוויון $1-x<e^{-x}$ (ל־$x>0$), הסיכוי שהחתך אינו מינימלי הוא לכל היותר

$$(1-\frac{2}{n(n-1)})^{n^2}<e^{-(2/n(n-1))n^2}<\frac{1}{3}$$

אגב, כווץ זוגות אקראיים (במקום קשתות אקראיות) לא היה מצליח כאן. הערה נוספת: שימו לב שמההוכחה כאן נובע כי לא יתכנו הרבה חתכים מינימלים לאותו גרף. כל חתך מינימלי חייב להתקבל בסוף הרצה בודדת של האלגוריתם בהסתברות $\frac{2}{n(n-1)}$ לפחות, ולכן אין יותר מ־$\frac{n(n-1)}{2}$ חתכים כאלה, כי סך ההסתברויות לא יכול לעלות על 1. יש דוגמה לגרף שזהו מספר החתכים המדוייק שלו, גרף המעגל על n צמתים.

חסמים תחתונים למשפט רמזי

זהו ישום קומבינטורי של העובדה הבסיסית שהסיכוי לקיום איחוד של מאורעות אינו עולה על סכום הסיכויים של המאורעות הבודדים. משפט רמזי קובע שלכל k,l קיים מספר R (המספר המינימלי הנ"ל מסומן ב־$R(k,l)$), כך שגרף עם R צמתים חייב להכיל או קליק עם k צמתים, או קבוצה ב"ת עם l צמתים. מספרי רמזי המדוייקים אינם ידועים פרט למקרים ספורים, כמו למשל $R(3,3)=6$, אך הוכחת המשפט המקורית נותנת את החסם $R(k,k)\le e^{O(k)}$. אנו נראה עתה $R(k,k)\ge e^{\Omega(k)}$ לפי ההוכחה של Erdös (אגב, המקדם המדוייק שצריך להיות בחזקה גם הוא אינו ידוע).

לשם כך ניקח גרף לפי $G(n,\frac{1}{2})$, ז"א גרף בעל n צמתים שבו כל זוג צמתים נבחר להיות קשת באופן ב"ת בהסתברות $\frac{1}{2}$. לכל קבוצת צמתים מגודל k, הסיכוי שהיא מהווה קליק בגרף או קבוצה חסרת קשתות בגרף הוא $2^{1-\binom{k}{2}}<2^{-k^2/3}$. מספר הקבוצות בנות k צמתים הוא $\binom{n}{k}<n^k$, ולכן למשל עבור $n=2^{\lfloor k/3\rfloor}$ נקבל שבהסתברות חיובית (גדולה מ־0) אף קבוצה בת k צמתים לא תהיה כזו (שימו לב שלא היה נסיון לתת מקדמים טובים כאן; כמו כן לא קשה גם להגיע להסתברות $1-o(1)$ למעלה).

שימושים בלינאריות התוחלת

מבוא ודוגמה ראשונה

התוחלת של משתנה מקרי מעידה על קיומו של מבנה מתאים, מכיוון שבהסתברות חיובית ערך המשתנה הוא לפחות כערך התוחלת. התוחלת היא תמיד לינארית (תוחלת סכום של משתנים מקריים שווה לסכום התוחלות גם אם הם אינם בלתי תלויים), ולכן היא נוחה מאוד לניתוח ומהווה את אחד הכלים השימושיים ביותר בשיטות הסתברותיות במדעי המחשב.

לדוגמה, נניח שנתונה לנו נוסחת 3CNF, עם m פסוקיות ו־n משתנים (כזכור, 3CNF מורכב מפעולת and על הפסוקיות, כשכל פסוקית היא or של משתנים ו/או שלילותיהם, עם שלושה משתנים בדיוק). אנו נטען שללא קשר לאפשרות סיפוק הנוסחה כולה, תמיד ניתן למצוא הצבה שמספקת לפחות $\frac{7}{8}m$ מהפסוקיות.

נגריל לכל משתנה ערך בוליאני באופן אחיד וב"ת בערכי המשתנים האחרים. נסמן ב־P_i את המשתנה המקרי שמקבל 1 אם הפסוקית ה־i מסתפקת, ומקבל 0 אחרת. משתנים כאלו קרויים משתני אינדיקטור. נשים לב עתה ש־$\sum_{i=1}^m P_i$ הוא המשתנה המקרי שערכו הוא מספר הפסוקיות שהסתפקו, והתוחלת שלו היא

$$\mathrm{E}[\sum_{i=1}^m P_i]=\sum_{i=1}^m \mathrm{E}[P_i]=\frac{7}{8}m$$

מכאן נובע שיש הצבה עבורה $\sum_{i=1}^m P_i$ מקבל לפחות את ערך התוחלת, ז"א שההצבה הזו מספקת לפחות $\frac{7}{8}m$ מהפסוקיות. לסיכום, נעיר ש־Håstad הוכיח שזה NP-Hard אפילו להבדיל בין המקרים שבהם כל הפסוקיות ניתנות לסיפוק בו זמנית לבין אלו שבהם לא יותר מ־$(\frac{7}{8}+\epsilon)m$ פסוקיות ניתנות לסיפוק בו זמנית (לכל $\epsilon>0$ קבוע).

דוגמה לישום קומבינטורי

משפט Turán: משפט זה אומר שגרף בעל n צמתים ויותר מ־$(1-\frac{1}{k-1})\frac{n^2}{2}$ קשתות מכיל קליק עם k צמתים. זהו גם חסם מדויק, כפי שניתן לראות מהדוגמה של גרף k־צדדי שלם בעל n צמתים עם בין $\lfloor\frac{n}{k}\rfloor$ ל־$\lceil\frac{n}{k}\rceil$ צמתים בכל מחלקה. נוכיח את המשפט הסתברותית: נסמן את קבוצת הצמתים של הגרף ב־$V=\{v_1,\ldots,v_n\}$, ונזכור שמספר הקשתות נתון ע"י $|E|=\frac{1}{2}\sum_{v\in V}d(v)$ כאשר $d(v)$ יסמן את דרגת הצומת v. עתה נגריל סדר מקרי על הצמתים באופן יוניפורמי מ־$n!$ הסדרים האפשריים. לכל צומת v, הסיכוי שהוא מחובר לכל הצמתים שלפניו הוא $\frac{1}{n-d(v)}$ (זה הסיכוי שהסדר ימקם את הצומת v לפני כל "לא־שכניו"). תוחלת מספר

הצמתים עבורם זה קורה היא

$$\sum_{v\in V}\frac{1}{n-d(v)} \ge \frac{n^2}{\sum_{v\in V}(n-d(v))} = \frac{n^2}{n^2-2|E|} > \frac{n^2}{n^2-(1-1/(k-1))n^2} = k-1$$

(אי שוויון הממוצעים קובע שלכל $\alpha_1,\ldots,\alpha_n$ חיוביים מתקיים $\frac{1}{n}\sum_{i=1}^n \alpha_i \ge (\prod_{i=1}^n \alpha_i)^{1/n} \ge n/\sum_{i=1}^n \alpha_i^{-1}$, ומכאן הצעד הראשון של הפיתוח). לכן בסיכוי חיובי יהיו יותר מ־$k-1$ צמתים המחוברים לכל הצמתים שסודרו לפניהם, ולכן בסיכוי חיובי יהיו לפחות k צמתים כאלו. נבחר סדר כזה וניקח את קבוצת הצמתים המקיימים זאת; אלו הם בהכרח צמתים של קליק.

הגרלה עם תיקונים

נזכור עתה את אי שוויון מרקוב, הקובע שלכל משתנה מקרי אי־שלילי X ולכל $\lambda > 1$ מתקיים החסם $\Pr\big[X \ge \lambda \mathrm{E}[X]\big] \le \lambda^{-1}$. למרות פשטות ההוכחה שלו, זהו אולי אי השוויון ההסתברותי החשוב ביותר.

לפעמים, בעיקר כאשר משתמשים בלינאריות התוחלת בשילוב אי שוויון מרקוב, ניתן להגיע בשיטה הסתברותית למבנה שבו התכונה הלוקלית הרצויה (כגון אי הכלת משולש) רק "כמעט" מתקיימת. כאן נראה איך ניתן לעיתים להתחיל ממבנה "כמעט מושלם" כזה ולהגיע ממנו למבנה הרצוי באמצעות תיקון מתאים.

נראה בשיטה זו שלכל g,k קיים גרף בעל מספר צביעה לפחות k ומותן לפחות g (המותן היא גודל המעגל הפשוט הקטן ביותר בגרף). זוהי תוצאה של Erdös מ־1959 (קודם לכן היתה ידועה בניה אלמנטרית של Mycielski עבור $g=4$, ז"א עבור גרפים חסרי משולשים). אנו נראה קיום גרף עם מותן לפחות g שבו אין קבוצות ב"ת גדולות כלל, ונשתמש בקשר בין גודל הקבוצה הב"ת המקסימלית לבין מספר הצביעה של הגרף על מנת להשלים את ההוכחה.

נתחיל מכך שנסתכל על גרף שנבחר לפי מרחב ההסתברות $G(n,p)$, ז"א גרף בעל n צמתים שבו כל זוג צמתים נבחר להיות קשת בהסתברות p באופן ב"ת בזוגות האחרים, כאשר $p=n^{(1-g)/g}$. ראשית נחסום את תוחלת מספר המעגלים מגודל קטן מ־g עבור n גדול דיו:

$$\sum_{i=3}^{g-1}\frac{n!}{(n-i)!2i}n^{i(1-g)/g} \le \sum_{i=3}^{g-1}\frac{n^{i/g}}{2i} \le (g-1)\frac{n^{(g-1)/g}}{2(g-1)} = o(n)$$

הסבר לביטוי: מספר האפשרויות לבחירה סדורה של i צמתים הוא $(n)_i=\frac{n!}{(n-i)!}$, וכל מעגל מגודל i מתקבל כך ב־$2i$ אופנים שונים. מהביטוי למעלה נובע לפי אי שוויון מרקוב שהסיכוי שיש יותר מ־$\frac{n}{2}$ מעגלים מגודל קטן מ־g הוא $o(1)$.

עתה נראה שבהסתברות גבוהה מתקיים $\alpha(G) < x = \lceil 3\ln(n)/p\rceil$. נחסום את הסיכוי ש־$\alpha(G) \ge x$, ע"י שימוש בכך שהסיכוי לאיחוד של מאורעות אינו עולה על סכום הסיכויים, כאשר כאן לכל קבוצת צמתים ספציפית בגודל x נגדיר את המאורע שהיא ב"ת.

$$\binom{n}{x}(1-p)^{\binom{x}{2}} < n^x e^{-\binom{x}{2}p} = (ne^{-(x-1)p/2})^x = e^{-x(1/2-o(1))\ln(n)} = o(1)$$

מכאן שעבור n גדול דיו קיים G שעבורו מספר המעגלים מגודל קטן מ־g בו הוא לכל היותר $\frac{n}{2}$, ובנוסף מתקיים עבורו $\alpha(G) \le \lceil 3\ln(n)/p\rceil$. ניקח גרף כזה ונסיר ממנו צומת אחד מכל מעגל הקטן מ־g (כשמסירים צומת, מסירים כמובן גם את כל הקשתות המכילות אותו). קיבלנו גרף עם לפחות $\frac{n}{2}$ צמתים, ומותן לפחות g. מכיוון שהסרת צמתים (כאן חשוב שהסרנו צמתים ולא רק קשתות) אינה מגדילה את $\alpha(G)$, מספר הצביעה של הגרף החדש הוא לפחות $\frac{n/2}{\lceil 3\ln(n)/p\rceil} = \Omega(\frac{n^{1/g}}{\ln(n)})$, ועבור n גדול דיו מספר זה גדול מ־k.

למת הבידוד (isolating lemma)

הצגת הלמה

לפני שנמשיך לשיטות הסתברותיות נוספות, נראה כאן ישום אלגוריתמי כללי של עקרונות הסתברותיים בסיסיים. למת הבידוד, אשר תנוסח מייד, מאפשרת רדוקציה של בעית מציאתו של תת מבנה אופטימלי (לפי מדד מתאים) לבעיה שבה המבנה האופטימלי הוא יחיד.

למת הבידוד: נניח ש־A היא קבוצה בת m איברים, וש־$\mathcal{F}$ היא משפחה של תתי־קבוצה של A. אם מגרילים משקלות $w: A \to \{1,\ldots,n\}$ כך ש־$w(a)$ נבחר באופן יוניפורמי וב"ת לכל $a \in A$, ולכל קבוצה $F \subseteq A$ מגדירים $w(F) = \sum_{a \in F} w(a)$. אז בהסתברות לפחות $1 - \frac{m}{n}$ קיים $F \in \mathcal{F}$ יחידי עבורו $w(F)$ מינימלי (מבין איברי $\mathcal{F}$).

הוכחת הלמה: ראשית מניחים שכל $a \in A$ מופיע כאיבר בלפחות אחת מהקבוצות של $\mathcal{F}$ (אחרת מסירים אותו מ־A), ושלכל $a \in A$ קיימת גם קבוצה ב־$\mathcal{F}$ שאינה מכילה אותו (אחרת אפשר לחסר את a מכל הקבוצות ב־$\mathcal{F}$ מבלי שזה ישנה את היחידות של המשקל המינימלי). לכל $a \in A$, נסמן עתה ב־W_a את המשקל המינימלי מבין איברי $\mathcal{F}$ המכילים את a, וב־$\overline{W}_a$ את המשקל המינימלי מבין איברי $\mathcal{F}$ שאינם מכילים את a. נגיד ש־a הוא איבר חד־משמעי כאשר $W_a \neq \overline{W}_a$.

אם כל איברי A הם חד־משמעיים, אז קיים $F \in \mathcal{F}$ יחיד בעל משקל מינימלי: נניח שקיימים F_1, F_2 בעלי משקל מינימלי, ושקיים $a \in F_1 \setminus F_2$. מכך נובע $W_a = w(F_1) = w(F_2) = \overline{W}_a$ (שהרי W_a ו־$\overline{W}_a$ אינם יכולים להיות קטנים יותר מהמינימום על כל איברי $\mathcal{F}$), בסתירה.

לכל a, נגדיר את $V_a = W_a - w(a) = \min_{a \in F \in \mathcal{F}} w(F \setminus \{a\})$. גם $\overline{W}_a$ וגם V_a אינם תלויים ב־$w(a)$ (הם תלויים רק במשקלות האיברים האחרים), וברור ש־a אינו חד־משמעי אם ורק אם $w(a) = \overline{W}_a - V_a$. נובע מכך שהסיכוי ש־$a$ אינו חד־משמעי חסום ע"י

$$\sum_{i=1}^{n} \Pr\left[w(a)=i \wedge \overline{W}_a - V_a = i\right] = \sum_{i=1}^{n} \Pr\left[w(a)=i\right]\Pr\left[\overline{W}_a - V_a = i\right] = \sum_{i=1}^{n} \frac{1}{n}\Pr\left[\overline{W}_a - V_a = i\right] \leq \frac{1}{n}$$

כאשר סוכמים על כל איברי A נובע מכך שבסיכוי לפחות $1 - \frac{m}{n}$ כל האיברים הם אכן חד־משמעיים, ומכך נובע לפי הדיון למעלה שבסיכוי זה קיים $F \in \mathcal{F}$ יחיד המשיג את המינימום ל־$w(F)$, כנדרש.

ישום בתורת הסיבוכיות

נניח שבידינו אלגוריתם אשר מוצא קליק מקסימלי בגרף, במידה וקליק זה הוא יחיד. נראה מכך שקיים אלגוריתם הסתברותי שמוצא קליק מקסימלי בגרף גם ללא הנחה זו (המדובר למעשה ברדוקציה הסתברותית – randomized reduction); מכך נובע שלא סביר שקיים אלגוריתם יעיל למציאת קליק מקסימלי בגרף אפילו תחת ההנחה שהוא יחיד, שהרי גרסת הבעיה ללא יחידות היא NP-Hard.

נניח שמספר צמתי הגרף G הוא $n > 4$. ראשית, לכל צומת v ב־G נגריל יוניפורמית ובאופן ב"ת מספר $w(v)$ בין 1 ל־$3n$. אם $\mathcal{F}$ מסמנת את קבוצת הקליקים המקסימלים בגרף המקורי, אז לפי למת הבידוד (שעוברת גם למקסימום של משקלות), בסיכוי לפחות $\frac{2}{3}$ יהיה קליק מכסימלי יחיד שעבורו סכום משקלות הצמתים הוא מקסימלי. עתה נעבור לגרף G' ע"י כך שנחליף כל צומת v של הגרף המקורי ב־$4n^2 + w(v)$ צמתים של G', שיסומנו $\{v_1, \ldots, v_{4n^2+w(v)}\}$. ב־$G'$ תהיה קשת בין v_i ל־u_j אם $u = v$ ו־$i \neq j$, או אם היתה קשת ב־G בין u ל־v. במילים אחרות: G' נוצר ע"י "ניפוח" כל צומת v של G לקליק בעל $4n^2 + w(v)$ צמתים, ושכפול הקשתות של G בהתאם.

נניח שגודל הקליק המקסימלי בגרף המקורי הוא $k \leq n$. כל קליק מקסימלי ב־G' יכיל איחוד של קבוצות צמתים שמתאימות לצמתי קליק ב־G (אם הוא מכיל צומת v_i אז ניתן להוסיף לו גם את שאר ה־v_j והוא ישאר קליק). כמו כן, לקליק K בעל l צמתים ב־G יתאים קליק בעל $4ln^2 + \sum_{v \in K} w(v)$ צמתים ב־G'. לכן, קליק מקסימלי ב־G' בהכרח יתאים לקליק מקסימלי ב־G (ההבדלים ב־$\sum_{v \in K} w(v)$ אף פעם לא יגיעו

ל־$4n^2$). אם מוסיפים לכך את העובדה שבהסתברות לפחות $\frac{1}{3}$ יהיה קליק מקסימלי יחידי ב־G עם ערך מקסימלי ל־$\sum_{v\in K} w(v)$, אז בהסתברות זו יהיה ב־G' קליק מקסימלי יחיד. מגודל הקליק המקסימלי ב־G' אפשר בקלות לחשב את k, גודל הקליק המקסימלי ב־G.

ישום אלגוריתמי

מסתבר שבהרבה ישומים של למת הבידוד יש גם שימוש באלגברה. נראה עתה אלגוריתם הסתברותי ניתן למקבול (ב־RNC) לבדיקת קיום זיווג מושלם בגרף. האלגוריתם משתמש בכך שאפשר למקבל (ב־NC) חישוב דטרמיננטה של מטריצת שלמים. על מנת לבצע רדוקציה, לגרף נתון G בעל n צמתים נבנה מטריצה $n \times n$ מתאימה. ראשית נסמן את הצמתים $V = \{v_1, \ldots, v_n\}$, ולכל קשת $v_iv_j \in E$ עם $i < j$ נבחר באופן יוניפורמי וב"ת משקל $w_{ij} = w(v_iv_j) \in \{1, \ldots, n^2\}$. נשים לב שבסיכוי לפחות $\frac{1}{2}$ יהיה זיווג מושלם יחידי עם סכום משקלים מינימלי על קשתותיו (אלא אם כן לא היה זיווג מושלם מלכתחילה).

עתה נגדיר את המטריצה A בצורה הבאה. $a_{ij} = 0$ אם v_iv_j אינו קשת של הגרף, $a_{ij} = 2^{w_{ij}}$ אם $i < j$ ו־v_iv_j קשת של G, ו־$a_{ij} = -2^{w_{ji}}$ אם $j < i$ ו־v_iv_j קשת. נשתמש עתה באלגוריתם מקבילי לחישוב $\det A$. הטענה היא שאם אין זיווג מושלם ב־G אז $\det A = 0$ (בהסתברות 1), ומצד שני אם יש זיווג מושלם יחיד בעל משקל מינימלי אז $\det A \neq 0$, וזה קורה כזכור בהסתברות לפחות $\frac{1}{2}$ אם יש זיווג מושלם כל שהוא בגרף G.

ניזכר בנוסחת הדטרמיננטה, $\det A = \sum_{\sigma\in S_n}(-1)^{\mathrm{sgn}(\sigma)}\prod_{i=1}^n a_{i\sigma(i)}$, כאשר S_n מציינת את קבוצת כל הפרמוטציות מעל $\{1, \ldots, n\}$, ו־$\mathrm{sgn}(\sigma)$ מציינת את זוגיות הפרמוטציה המתאימה. בסכום על הפרמוטציות, כל פרמוטציה $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ עבורה $\prod_i a_{i\sigma(i)} \neq 0$ מתאימה לאיחוד זר־צמתים של מעגלים מכוונים לאורך קשתות ב־G (העגילים של σ), שכולם פשוטים פרט לאלו מגודל 2 (במעגלים מאורך 2 עוברים על אותה קשת של G הלוך ושוב).

יתרה מזו, הפרמוטציות המערבות עגילים אי־זוגיים (בפירוק שלהן לעגילים זרים) מקזזות זו את זו: לכל פרמוטציה σ בעלת עגיל אי־זוגי ניקח את העגיל האי־זוגי המכיל את האינדכס הקטן ביותר, ונהפוך את כיוונו לקבלת פרמוטציה σ' כך שהמחוברים המתאימים ל־σ ו־σ' בנוסחת הדטרמיננטה מקזזים זה את זה (חשוב לשים לב ש־$\sigma = \sigma''$ לכל פרמוטציה כזו). מכאן נובע בפרט שאם אין זיווג מושלם אז $\det A = 0$ (בהסתברות 1), כי אם קיימת פרמוטציה המתאימה לאיבר שאינו מתאפס ושבה כל המעגלים זוגיים, אז בהכרח קיים זיווג מושלם ב־G.

לכל פרמוטציה σ המכילה מעגלים זוגיים בלבד ושעבורה $\prod_i a_{i\sigma(i)} \neq 0$, נסמן ב־$M_1, M_2$ פירוק שלה לשני זיווגים מושלמים ב־G (יתכן שפירוק זה אינו יחיד; לצורך ההוכחה מספיק לקחת פירוק כל שהוא). אז מתקיים $|\prod_i a_{i\sigma(i)}| = 2^{w(M_1)+w(M_2)}$. נסמן עתה ב־$M_0$ את הזיווג המושלם היחיד עבורו $w(M_0)$ מינימלי, וב־σ_0 את הפרמוטציה המתקבלת ממעבר הלוך ושוב על כל קשת של M_0. עבור פרמוטציה זו מתקיים $|\prod_i a_{i\sigma_0(i)}| = 2^{2w(M_0)}$.

נחזור לנוסחת הסכום של $\det A$: ניתן לראות עתה שסכום זה מכיל איבר יחיד שערכו המוחלט הוא $2^{2w(M_0)}$ (כי רק σ_0 ניתנת לפירוק לשני עותקים של M_0), ושלכל האיברים האחרים שאינם מתאפסים או מתקזזים בסכום זה יש ערך מוחלט שמתחלק ב־$2^{2w(M_0)+1}$, כי לכל M_1, M_2 שאינם שניהם זהים ל־M_0 מתקיים $w(M_1) + w(M_2) > 2w(M_0)$. מכאן נובע ש־$\det A$ אינו יכול להתחלק ב־$2^{2w(M_0)+1}$, ובפרט $\det A \neq 0$ כנדרש.

שיטת המומנט השני

מבוא

שיטת המומנט השני היא השיטה הראשונה שנלמד מבין מספר שיטות המבוססות על "ריכוז". הכוונה היא לאי שוויונים הסתברותיים שפירושם הוא שבהסתברות גבוהה משתנה מקרי מסויים (למשל משתנה שהוא סכום של משתנים "קטנים" אחרים) יקבל ערך קרוב לתוחלת שלו. אלו מועילים במקרים רבים שבהם לא מספיק לצורך המשך ההוכחה לדעת שבסיכוי חיובי הערך יהיה לפחות זה של התוחלת.

נסמן את השונות $\mathrm{V}[X] = \mathrm{E}\big[(X-\mathrm{E}[X])^2\big] = \mathrm{E}\big[X^2\big] - (\mathrm{E}[X])^2$. בספרות משתמשים גם בסימון $\sigma = \sqrt{\mathrm{V}[X]}$ (זוהי סטיית התקן, standard deviation), וב־$\mu = \mathrm{E}[X]$. אי שוויון צ'בישף (Chebyschev) אומר שלכל $\lambda > 0$ מתקיים $\Pr\big[|X-\mathrm{E}[X]| \geq \lambda\sqrt{\mathrm{V}[X]}\big] \leq \lambda^{-2}$. תזכורת מהירה של ההוכחה: משתמשים באי שוויון מרקוב עבור $Y = (X-\mathrm{E}[X])^2$ לקבלת:

$$\Pr\big[|X-\mathrm{E}[X]| \geq \lambda\sqrt{\mathrm{V}[X]}\big] = \Pr\big[Y \geq \lambda^2\mathrm{E}[Y]\big] \leq \lambda^{-2}$$

זה משפט הריכוז הראשון שלנו. אנו נשתמש בחסימת השונות על מנת להוכיח שמשתנה מקרי מקבל בהסתברות גבוהה ערך קרוב לתוחלת – זוהי שיטת המומנט השני. כזכור השונות אינה חיבורית כמו התוחלת, אבל ניתן לעיתים לחשב או לחסום אותה בקלות יחסית באמצעות הנוסחה

$$\mathrm{V}\big[\sum_{i=1}^n X_i\big] = \sum_{i=1}^n\sum_{j=1}^n \mathrm{Cov}[X_i,X_j] = \sum_{i=1}^n \mathrm{V}[X_i] + 2\sum_{i>j}\mathrm{Cov}[X_i,X_j]$$

כאשר מגדירים את הקוואריאנס לפי

$$\mathrm{Cov}[Y,Z] = \mathrm{E}\big[(Y-\mathrm{E}[Y])(Z-\mathrm{E}[Z])\big] = \mathrm{E}[Y\cdot Z] - \mathrm{E}[Y]\cdot\mathrm{E}[Z]$$

נשים לב בפרט שעבור זוג מ"מ בלתי תלויים הקווריאנס הוא אפס, ולכן אם $X_1,\ldots,X_n$ ב"ת בזוגות (אפילו אם הם לא ב"ת כ־n־יה של משתנים) אז $\mathrm{V}\big[\sum_{i=1}^n X_i\big] = \sum_{i=1}^n \mathrm{V}[X_i]$, ז"א שבמקרה זה השונות היא חיבורית.

כדוגמה בסיסית ראשונה נניח שאכן $X_1,\ldots,X_n$ הם משתנים מקריים בלתי תלויים בזוגות שכל אחד מהם מתפלג יוניפורמית מעל $\{0,1\}$, ונבחן את $X = \sum_{i=1}^n X_i$. עתה נראה שבהסתברות $1-o(1)$ מתקיים $X = (\frac{1}{2}+o(1))n$ (אם המשתנים ב"ת כ־n־יה אז אפשר לקבל כאן חסמים מספריים חזקים מאשר בשיטה זו, אותם נראה בקרוב). מתקיים $\mathrm{E}[X] = \frac{n}{2}$, וכן $\mathrm{V}[X] = \frac{n}{4}$ (מכיוון ש־$\mathrm{V}[X_i] = \frac{1}{4}$). לכן לפי אי שוויון צ'בישף מתקיים $\Pr\big[|X-\frac{n}{2}| \geq n^{3/4}\big] \leq \frac{1}{4}n^{-1/2}$, ז"א שבסיכוי של לפחות $1-\frac{1}{4}n^{-1/2} = 1-o(1)$ מתקיים $X = \frac{n}{2}+O(n^{3/4}) = (\frac{1}{2}+o(1))n$ כנדרש.

ישום קומבינטורי־מספרי

עתה נסתכל על השאלה: מהו הגודל המקסימלי $f(n)$ של ת"ק של $\{1,\ldots,n\}$, כך שכל הסכומים של תתי הקבוצות שלה שונים זה מזה? אפשר למצוא קבוצה כזו שגודלה הוא $1+\lfloor\log n\rfloor$, ע"י לקיחת חזקות עוקבות של 2. שאלה פתוחה של Erdös היא האם הגודל המקסימלי חסום ע"י $\log n + C$ עבור קבוע מתאים. מכיוון שלתת־קבוצה בת k איברים כל הסכומים חסומים ע"י $kn-1$, קל לראות שאם 2^k הסכומים האפשריים שונים זה מזה אז $2^k < nk$, ולכן $f(n) \leq \log n + \log\log n + O(1)$.

בשיטת המומנט השני נשפר זאת מעט. נסמן את איברי A ב־$\{\alpha_1,\ldots,\alpha_k\}$, ויהיו $X_1,\ldots,X_k$ מ"מ ב"ת כך ש־X_i בסיכוי $\frac{1}{2}$ שווה ל־0 ובסיכוי $\frac{1}{2}$ שווה ל־α_i. עבור X המוגדר ע"י הסכום $X = \sum_{i=1}^k X_i$ מתקיים $\mathrm{V}[X] = \sum_{i=1}^k \mathrm{V}[X_i] = \frac{1}{4}\sum_{i=1}^k \alpha_i^2 \leq \frac{n^2k}{4}$ ולכן $\Pr\big[|X-\mathrm{E}[X]| \leq \lambda n\sqrt{k}/2\big] \geq 1-\lambda^{-2}$ (מאי שוויון צ'בישף), לכל $\lambda > 1$. מכאן שבסיכוי לפחות $1-\lambda^{-2}$ המשתנה $X-\mathrm{E}[X]$ מקבל אחד מקבוצה בת לא יותר מ־$\lambda n\sqrt{k}+1$ ערכים אפשריים (מכיוון ש־X הוא תמיד מספר שלם).

מצד שני, כל ערך אפשרי של $X-\mathrm{E}[X]$ מתקבל בהסתברות 2^{-k} (רק סכום חלקי יחיד של $\{\alpha_1,\ldots,\alpha_k\}$ יכול לתרום להסתברות), ולכן $2^{-k}(\lambda n\sqrt{k}+1) \geq 1-\lambda^{-2}$. עבור $\lambda = 2$ למשל (כאשר מניחים $k \geq 2$) נובע מכך $n \geq (\frac{3}{4}2^k-1)/2\sqrt{k} \geq 2^k/4\sqrt{k}$ ולכן מתקיים בהכרח $f(n) \leq \log n + \frac{1}{2}\log\log n + O(1)$, שיפור מסויים של החסם הקודם.

פונקציית סף לקיום קליק בגרף

נראה כאן שהפונקציה $f(n) = n^{-2/3}$ היא פונקצית סף עבור התכונה של קיום קליק בעל 4 צמתים בגרף מקרי: אם $p(n) = \omega(n^{-2/3})$ אז בהסתברות $1 - o(1)$ הגרף $G(n, p(n))$ מכיל עותק של K_4 (הקליק בעל 4 הצמתים), ואם $p(n) = o(n^{-2/3})$ אז בהסתברות $1-o(1)$ הגרף $G(n, p(n))$ אינו מכיל עותק כזה. החלק השני של הטענה אינו קשה: אם $p(n) = o(n^{-2/3})$, אז הסיכוי לקיומו של קליק כזה חסום ע"י $\binom{n}{4}(p(n))^6 = o(1)$ ע"י החסם הבסיסי על איחוד מאורעות. עבור החלק הראשון עלינו להשתמש בשיטת המומנט השני.

עתה נניח שמתקיים $p(n) = \omega(n^{-2/3})$, ונוכיח מכך שמתקיים $\mathrm{V}[X] = o((\mathrm{E}[X])^2)$, כאשר X יסמן את מספר העותקים של K_4 שהתקבלו בגרף. בזאת נסיים כי מהאחרון נובע לפי אי שוויון צ'בישף החסם הנדרש:

$$\Pr[X = 0] \le \mathrm{V}[X]/(\mathrm{E}[X])^2 = o(1)$$

מלינאריות התוחלת ניתן לראות שמתקיים $\mathrm{E}[X] = \binom{n}{4}(p(n))^6$, ועתה נפנה ל־$\mathrm{V}[X]$. נסמן ב־$X_{i,j,k,l}$ את משתנה האינדיקטור עבור המאורע ש־v_i, v_j, v_k, v_l מהווים קליק ב־G, ועבור משתנים אלו מתקיים $X = \sum_{1 \le i<j<k<l \le n} X_{i,j,k,l}$.

נחשב את הקוואריאנס בין המשתנים, לפי גודל החיתוך בין $\{i, j, k, l\}$ לבין $\{i', j', k', l'\}$:

- אם $|\{i, j, k, l\} \cap \{i', j', k', l'\}| \le 1$, אז $X_{i,j,k,l}$ ו־$X_{i',j',k',l'}$ הם ב"ת זה בזה והקוואריאנס הוא 0.
- אם גודל החיתוך הוא 2 אז $\mathrm{Cov}[X_{i,j,k,l}, X_{i',j',k',l'}] = (p(n))^{11} - (p(n))^{12} < (p(n))^{11}$.
- אם $|\{i, j, k, l\} \cap \{i', j', k', l'\}| = 3$ אז $\mathrm{Cov}[X_{i,j,k,l}, X_{i',j',k',l'}] < (p(n))^9$.
- המקרה האחרון הוא $\mathrm{Cov}[X_{i,j,k,l}, X_{i,j,k,l}] = \mathrm{V}[X_{i,j,k,l}] < (p(n))^6$.

מכל אלו מתקבל:

$$\begin{aligned}
\mathrm{V}[X] &= \sum_{\substack{1 \le i<j<k<l \le n \\ 1 \le i'<j'<k'<l' \le n}} \mathrm{Cov}[X_{i,j,k,l}, X_{i',j',k',l'}] \\
&< \binom{n}{4}(p(n))^6 + \binom{n}{3}(n-3)(n-4)(p(n))^9 + \binom{n}{2}\binom{n-2}{2}\binom{n-4}{2}(p(n))^{11} \\
&= O(n^4(p(n))^6 + n^5(p(n))^9 + n^6(p(n))^{11})
\end{aligned}$$

כדי לסיים את ההוכחה, נשים לב שמתקיים $(\mathrm{E}[X])^2 = \Omega(n^8(p(n))^{12})$. מכך נובע שאם $p(n) = \omega(n^{-2/3})$, אז $(p(n))^{-1} = o(n^{2/3})$, ואז

$$\frac{\mathrm{V}[X]}{(\mathrm{E}[X])^2} = O(n^{-4}(p(n))^{-6} + n^{-3}(p(n))^{-3} + n^{-2}(p(n))^{-1}) = o(1 + n^{-1} + n^{-4/3}) = o(1)$$

כנדרש.

חסימת סטיות גדולות (large deviation inequalities)

למות החסימה

שימוש בחסמים "אקספוננציאלים" על סטיות גדולות (הידועים גם בשם "חסמים מסוג Chernoff") היא אחת השיטות ההסתברותיות הנפוצות ביותר, אולם עם זאת חסמים אלו בהרבה מקרים מובאים ללא התעכבות על

הוכחתם. כאן נוכיח פורמלית מספר חסמים מטיפוס זה, כאשר הפרק על המרטינגלים יכלול הוכחה של חסם דומה נוסף.

בתחילת הקורס נעשה שימוש בעובדה שבהינתן m משתנים מקריים בינארים אחידים וב"ת (כ־m־יה), הסיכוי שכולם יהיו 0 הוא 2^{-m}. הרעיון בחסימת סטיות גדולות הוא להראות שגם הסיכוי שמספר המשתנים המקבלים 0 יהיה שונה בהרבה מ־$\frac{m}{2}$ הוא קטן באופן אקספוננציאלי ב־m. ההוכחה עוברת (שוב) דרך אי שוויון מרקוב, אשר מופעל במקרה זה על מ"מ מהצורה $e^{\lambda X}$ (פונקציה זו של X קרויה גם "הפונקציה יוצרת המומנטים", בגלל שטור החזקות שלה כולל את כל החזקות של X) כאשר X הוא הסכום שאנו רוצים לחסום, וזוהי עוד דוגמה לחשיבותו של אי שוויון בסיסי זה.

נניח ש־$X_1, \ldots, X_m$ הם מ"מ ב"ת המקבלים יוניפורמית את ערכם מ־$\{1, -1\}$, ונראה עבור $X = \sum_{i=1}^m X_i$ שלכל $a > 0$ מתקיים $\Pr\big[X > a\big] < e^{-a^2/2m}$. שימו לב שמסימטריה נובע שגם $\Pr\big[X < -a\big] < e^{-a^2/2m}$. בהרבה שימושים a יהיה פרופורציונלי ל־m ומכאן נקבל חסם על ההסתברות שהוא קטן אקספוננציאלית ב־m. מצד שני, אי אפשר לתת חסם שהוא יותר טוב מקבוע על ההסתברות לסטיה בגודל $O(\sqrt{m})$: השונות של X היא $\Omega(m)$, ואכן ניתן להראות שבהסתברות קבועה יש סטיה מגודל $\Omega(\sqrt{m})$.

עבור הוכחת החסם שלנו, ראשית נשים לב שמתקיים לכל $\lambda > 0$ ו־$1 \leq i \leq m$

$$\mathrm{E}\big[e^{\lambda X_i}\big] = (e^{\lambda} + e^{-\lambda})/2 = \cosh(\lambda) \leq e^{\lambda^2/2}$$

(אי השוויון הימני ניתן להוכחה למשל ע"י השוואת טורי החזקות המתאימים לפונקציות). מכאן שניתן לחסום את התוחלת של $e^{\lambda X}$ ע"י

$$\mathrm{E}\big[e^{\lambda X}\big] = \mathrm{E}\big[e^{\lambda \sum_{i=1}^m X_i}\big] = \mathrm{E}\big[\prod_{i=1}^m e^{\lambda X_i}\big] = \prod_{i=1}^m \mathrm{E}\big[e^{\lambda X_i}\big] \leq e^{m\lambda^2/2}$$

כאשר הוצאת המכפלה אל מחוץ לסימן התוחלת מתאפשרת מאי תלות המ"מ המעורבים. לכן לכל $\lambda > 0$ ו־$a > 0$ מתקיים לפי אי שוויון מרקוב

$$\Pr\big[X > a\big] = \Pr\big[e^{\lambda X} > e^{\lambda a}\big] < \mathrm{E}\big[e^{\lambda X}\big]/e^{\lambda a} \leq e^{\lambda^2 m/2 - \lambda a}$$

ובפרט אם נבחר $\lambda = \frac{a}{m}$ אז נקבל $\Pr\big[X > a\big] < e^{-a^2/2m}$, כנדרש (הבחירה דווקא ב־$\lambda$ הזה נעשתה ע"פ חיפוש מינימום של $\lambda^2 m/2 - \lambda a$).

עתה נראה דוגמה להכללה של הטענה הקודמת. נניח עתה שכל X_i מקבל ערך $1 - p_i$ בהסתברות p_i וערך $-p_i$ בהסתברות $1 - p_i$ (הסיבה לערכים אלו של X_i במקום הערכים $\{0, 1\}$ היא שיתקיים $\mathrm{E}[X_i] = 0$). נראה שעבור $X = \sum_{i=1}^m X_i$ מתקיים אי השוויון $\Pr\big[X > a\big] < e^{-2a^2/m}$; לא קשה לראות שעבור $p_i = \frac{1}{2}$ טענה זו שקולה לטענה הקודמת. כמסקנה אפשר להראות עבור מ"מ ב"ת Y_i המקבלים 1 בהסתברות p_i ו־0 בהסתברות $1 - p_i$ שמתקיים החסם $\Pr\big[\sum_{i=1}^m Y_i - \mathrm{E}[\sum_{i=1}^m Y_i] > a\big] < e^{-2a^2/m}$, ע"י הסתכלות על המשתנים $X_i = Y_i - p_i$. חסימת ההסתברות לסטיה הסימטרית בכיוון השני גם אפשרית.

להוכחת החסם עבור X נרשום $\mathrm{E}\big[e^{\lambda X_i}\big] = p_i e^{\lambda(1-p_i)} + (1 - p_i)e^{-\lambda p_i}$. זה חסום ע"י $e^{\lambda^2/8}$ (מוכיחים זאת ע"י מציאת אקסטרימים של פונקציה בת שני משתנים; הנכם מוזמנים לקרוא את ההוכחה בנספח של הספר של Alon, Spencer). מכאן נובע בדומה למה שנעשה קודם שמתקיים לכל $\lambda > 0$ ו־$a > 0$

$$\Pr\big[X > a\big] = \Pr\big[e^{\lambda X} > e^{\lambda a}\big] < \mathrm{E}\big[e^{\lambda X}\big]/e^{\lambda a} \leq e^{\lambda^2 m/8 - \lambda a}$$

ובחירה של $\lambda = \frac{4a}{m}$ תחסום את הביטוי ע"י $e^{-2a^2/m}$, כנדרש.

דוגמאות לישומים

ראשית נראה דוגמה קומבינטורית: אנו נראה שבסיכוי $1 - o(1)$, הגרף המקרי $G(n, \frac{1}{2})$ כאשר n זוגי מכיל זיווג מושלם. לשם כך אנו נסתכל על הגרף שלנו כעל איחוד של שני גרפים מקריים על אותה קבוצת צמתים

(האיחוד הוא של קבוצות הקשתות), G_1 שנבחר לפי $G(n,\frac{1}{3})$ ו־G_2 שנבחר לפי $G(n,\frac{1}{4})$ באופן ב"ת ב־G_1. לא קשה לראות שבאיחוד כל זוג צמתים אכן יהיה קשת בהסתברות $\frac{1}{2}$ באופן ב"ת בזוגות האחרים.

בשלב הראשון נראה שבסיכוי $1-o(1)$ יש ב־G_1 קבוצה בת $l=\lceil\frac{9}{19}n\rceil$ קשתות זרות. לשם כך לא צריך חסימת סטיות גדולות, ואפשר להיזכר בהוכחה של החסם התחתון על משפט רמזי. איחוד מאורעות יתן לנו שבסיכוי $1-o(1)$ לא תהיה ב־G_1 קבוצה בת $\lceil\frac{1}{19}n\rceil$ צמתים שאין בתוכה קשת. כאשר זה קורה, ניתן לבחור בזו אחר זו את הקשתות הזרות, כי בכל שלב עדיין תהיה קשת בקבוצת הצמתים הנותרים.

עתה נמקד את שאר הדיון בהנחה שאכן יש ב־G_1 את הקשתות הנ"ל, ונסמן אותן ב־$u_1v_1,\ldots,u_lv_l$. מכיוון שקשתות G_2 נבחרות באופן ב"ת ב־G_1, ננתח עתה את התכונות שלהן ביחס לקשתות u_iv_i. נסמן את הצמתים הנותרים ב־$u'_1,\ldots,u'_k,v'_1,\ldots,v'_k$ (באופן שרירותי) כאשר $k=\frac{n}{2}-l$, ונטען שבסיכוי $1-o(1)$ מתקיים הדבר הבא: לכל $1\leq i\leq k$ קיימים לפחות k ערכים אפשריים של j, כך שקיימות שתי קשתות זרות בקבוצה $\{u'_i,v'_i,u_j,v_j\}$. לשם כך עבור i קבוע ו־j קבוע נחסום את הסיכוי ש־j הוא ערך אפשרי עבור i, כאשר נתעלם מהאפשרות ש־u'_i,v'_i תהיה קשת בעצמה (אנו נרצה מאורעות ב"ת בשביל המשך הניתוח). הסיכוי ש־u'_i תהיה לפחות מחוברת לאחד מ־u_j,v_j בעוד v'_i תהיה מחוברת לשני הוא $\frac{31}{256}=\frac{15}{256}+\frac{1}{16}$. המאורע שזה יקרה עבור j מסויים הוא ב"ת בכל המאורעות הנ"ל עבור ערכי j אחרים, ולכן ניתן לחסום את הסיכוי שיהיו פחות מ־$k\leq\frac{1}{9}l$ ערכי j כאלו ע"י $e^{-2(31/256-1/9)^2l}=e^{-\Theta(k)}$. זהו הסיכוי שעבור i בודד לא יהיו מספיק ערכי j, והסיכוי שזה יקרה ל־$1\leq i\leq k$ כל שהוא עתה ניתן לחסימה לפי איחוד מאורעות ע"י $ke^{-\Theta(k)}=o(1)$.

עד עתה ראינו שבסיכוי $1-o(1)$ גם יש l קשתות זרות ב־G_1 וגם מתקיים התנאי השני ביחס ל־G_2 עבור הצמתים הנותרים. נראה עתה כיצד ניתן למצוא את הזיווג המושלם כאשר שני התנאים מתקיימים: נתחיל מקבוצה הקשתות $\{u_1v_1,\ldots,u_lv_l\}$, ולכל $1\leq i\leq k$ נבחר j_i יחודי עבורו כך שיש שתי קשתות זרות ב־$\{u'_i,v'_i,u_{j_i},v_{j_i}\}$. ניתן למצוא את ה־$j_i$ הנ"ל כי יש k אפשרויות עבור כל i. לסיום, פשוט לכל i נמיר את הקשת $u_{j_i}v_{j_i}$ בשתי הקשתות הזרות ב־$\{u'_i,v'_i,u_j,v_j\}$.

הערה: למעשה ידועה תוצאה חזקה בהרבה של Erdös ו־Rényi, שמשמעה הוא שבעיקרון אם p גדול דיו להבטיח שבסיכוי $1-o(1)$ אין ב־$G(n,p)$ צמתים מבודדים, אז הוא יבטיח כבר בסיכוי $1-o(1)$ את קיומו של זיווג מושלם. הנכם מוזמנים לקרוא בספר Random Graphs של Bollobás את הסקירה המלאה בנושא.

עתה נראה בקצרה ישום אלגוריתמי פשוט של חסימת סטיות גדולות. נניח שיש לנו כקלט קבוצה בת l מחרוזות בינאריות מאורך n, וברצונינו לבצע פעולה הקשורה במרחקי Hamming היחסיים ביניהם, המוגדרים ע"י $d(x,y)=\frac{1}{n}|\{k|x(k)\neq y(k)\}|$. אם n גדול דיו, אז ניתן לחסוך ע"י "קירוב" המחרוזות באמצעות תתי מחרוזות באורך $O(\log(l))$. לשם כך נבחר קבוצה $I=\{i_1,\ldots,i_m\}$ בת $m=\lceil\frac{2\log(l)}{\epsilon^2}\rceil$ קורדינטות. לפשטות הניתוח נאפשר חזרות ב־I, כך שהמדובר בעצם בסדרה מקרית של קורדינטות $i_1,\ldots,i_m$, שכל אחת מהן נבחרה באופן יוניפורמי וב"ת מתוך $\{1,\ldots,n\}$. נראה עתה חסמים על $d_I(x_i,x_j)=\frac{1}{m}|\{k|x(i_k)\neq y(i_k)\}|$.

המרחק $d_I(x_i,x_j)$ הוא ממוצע של $\frac{2\log(l)}{\epsilon^2}$ מ"מ ב"ת שכל אחד מהם מקבל 1 בהסתברות $d(x_i,x_j)$ ו־0 בהסתברות $1-d(x_i,x_j)$. לכן הסיכוי עבור $|d_I(x_i,x_j)-d(x_i,x_j)|>\epsilon$ הוא הסיכוי שסכום המ"מ יסטה מתוחלת הסכום ביותר מ־$\frac{2\log(l)}{\epsilon}$, וזה חסום ע"י $2e^{-4\log(l)}=o(1/\binom{l}{2})$. מכאן שבהסתברות גבוהה לכל $1\leq i<j\leq l$ מתקיים $|d_I(x_i,x_j)-d(x_i,x_j)|\leq\epsilon$. הקירוב (או "הפחתת המימד") שנעשה כאן הוא טריביאלי למדי, אולם באופן כללי שאלת הקירובים של קבוצת נקודות במרחבים נורמיים ע"י נקודות במרחבים פשוטים יותר היא נושא מחקרי פעיל למדי. הערה נוספת היא על סימן הריבוע על ϵ במספר המשתנים הדרוש. שימו לב שהריבוע חייב להיות שם (ראו את ההערה קודם על השונות של סכום מ"מ ב"ת).

מרטינגלים

מבוא

מרטינגל (הקרוי על שם הרצועה בריתמת הסוס, ומהווה אחת מהתרומות של תעשיית ההימורים על מרוצי הסוסים למתמטיקה) הוא סדרה של מ"מ ממשיים $\underline{X}=X_0,X_1,\ldots$, בד"כ לא בלתי תלויים, שמקיימת לכל i את השוויון "$\mathrm{E}[X_{i+1}|X_0,\ldots,X_i]=X_i$". בתורת המידה אפשר להגדיר את צד שמאל כמשתנה מקרי, ז"א פונקציה מקבוצת הבסיס S ל־$\mathbb{R}$, ואז דורשים שהשוויון מתקיים בהסתברות 1. תנאי זה נקרא "חוסר זיכרון".

אצלנו נתמקד בעיקר בסדרות סופיות של מ"מ מעל מרחבי הסתברות סופיים, ועבור אלו אפשר להבין את השוויון הזה בצורה פשוטה: לכל סדרה של ערכים $a_0,\ldots,a_i$ עבורם $\Pr\big[X_0=a_0,\ldots,X_i=a_i\big]>0$, מתקיים $\mathrm{E}\big[X_{i+1}|X_0=a_0,\ldots,X_i=a_i\big]=a_i$. עוד דקות שלא ניכנס אליה קשורה בהגדרות עבור מרטינגלים שהם סדרה אינסופית של מ"מ (יכול להיות שמרחבי ההסתברות מוגדרים רק עבור "רישות" של הסדרה). אצלנו בד"כ תהיה סדרה סופית $X_1,\ldots,X_m$, ותמיד מרחב ההסתברות יוגדר עבור כל הסדרה.

קצת אינטואיציה עבור תנאי חוסר הזיכרון: ניתן להתיחס אל מרטינגל כאל "סכום מצטבר". למשל, אם $Y_1,Y_2,\ldots$ הם מ"מ ב"ת עם תוחלת אפס, אז ההגדרה $X_i=\sum_{j=1}^{i}Y_j$ תתן מרטינגל. אולם עבור מרטינגל כללי אין זה חובה שמשתני ההפרשים $Y_j=X_j-X_{j-1}$ אכן יהיו ב"ת כדי שתנאי חוסר הזיכרון יתקיים, כל שאנו דורשים הוא מעין "אי תלות של התוחלת".

ניתן להראות באינדוקציה שעבור מרטינגל מתקיים $\mathrm{E}[X_i]=\mathrm{E}[X_0]$ לכל i; לשם המחשה נראה זאת עבור מרחבי ההסתברות הסופיים שבהם ענינינו, ונכנס הפעם לפרטים הכי קטנים (כל הסכומים הם סכומים סופיים מעל קבוצות הערכים היכולות להתקבל בהסתברות חיובית; עבור מרחבי הסתברות כלליים משתמשים בטענות מתורת המידה באשר לתוחלות, במקום "לפרוט" אותן לסכומים מפורשים):

$$\begin{aligned}
\mathrm{E}[X_i] &= \sum_{\substack{a_0,\ldots,a_i\\ \Pr[X_0=a_0,\ldots,X_i=a_i]>0}} a_i\Pr\big[X_0=a_0,\ldots,X_i=a_i\big]\\
&= \sum_{\substack{a_0,\ldots,a_{i-1}\\ \Pr[X_0=a_0,\ldots,X_{i-1}=a_{i-1}]>0}} \Bigg(\sum_{\substack{a_i\\ \Pr[X_i=a_i|X_0=a_0,\ldots,X_{i-1}=a_{i-1}]>0}} a_i\Pr\big[X_i=a_i|X_0=a_0,\ldots,X_{i-1}=a_{i-1}\big]\cdot\\
&\qquad\cdot\Pr\big[X_0=a_0,\ldots,X_{i-1}=a_{i-1}\big]\Bigg)\\
&= \sum_{a_1,\ldots,a_{i-1}} \mathrm{E}\big[X_i|X_0=a_0,\ldots,X_{i-1}=a_{i-1}\big]\cdot\Pr\big[X_0=a_0,\ldots,X_{i-1}=a_{i-1}\big]\\
&= \sum_{a_1,\ldots,a_{i-1}} a_{i-1}\Pr\big[X_0=a_0,\ldots,X_{i-1}=a_{i-1}\big]=\mathrm{E}[X_{i-1}]=\cdots=\mathrm{E}[X_0]
\end{aligned}$$

הסבר למעלה: השורה הראשונה משתמשת בסכום הסתברויות של מאורעות זרים. אח"כ השתמשנו בנוסחת ההסתברות המותנה ל־$X_i=a_i$, אח"כ בהגדרת התוחלת המותנה של X_i, ואח"כ בתנאי חוסר הזיכרון כדי להגיע ל־a_{i-1}.

דוגמה ראשונה למרטינגל: נגדיר את X_i להיות סכום הזכיה (או ההפסד) המצטבר לאחר i משחקי הטלת מטבע הוגנת. באופן פורמלי: יהיו $Y_1,Y_2,\ldots$ מ"מ המקבלים ערך מ־$\{1,-1\}$ באופן יוניפורמי וב"ת, ומגדירים את הסכום המצטבר $X_i=\sum_{j=1}^{i}Y_j$ לכל $i\geq 0$.

כדוגמה שניה נניח שלפנינו מהמר בעל האסטרטגיה הבאה: "המשך לשחק עד אשר סכום הזכיה הוא $1+$". על מנת לנתח את סדרת הזכיות המצטברות כאן, נגדיר את $\underline{X}'=X'_0,X'_1,\ldots$ בצורה הבאה: $X'_0=X_0=0$, ולכל $i\geq 0$ נגדיר את X'_{i+1} בהתאם לערכי $X'_0,\ldots,X'_i$: אם קיים $j\leq i$ כך ש־X'_j קיבל את הערך 1, אז $X'_{i+1}=X'_i=1$, ואחרת $X'_{i+1}=X_{i+1}$. הנקודה לשים לב כאן היא שגם $\underline{X}'$ הוא מרטינגל, ולכן אם המהמר הנ"ל מוגבל בזמן N אז האסטרטגיה הזו אינה תורמת לו כלום, כי גם כאן מתקיים $\mathrm{E}[X'_N]=\mathrm{E}[X'_0]=0$ לכל N. בניה דומה תראה תוצאה זו גם לכל אסטרטגית הימור אחרת שהמהמר ה"חכם" יוכל לחשוב עליה. אגב, במשחקי הימורים אמיתיים המצב רע אף יותר, כי שם סכום הזכיה המצטבר אינו מהווה מרטינגל, והתוחלת המותנה $\mathrm{E}[X_{i+1}|X_0,\ldots,X_i]$ אף קטנה ממש מהערך ש־X_i קיבל.

מרטינגלים וחסימת סטיות

נעבור עתה לענינינו, חסימת סטיות גדולות. חסם כזה אינו מפתיע במיוחד כאשר הגדרת המרטינגל משתמשת בסכום מצטבר של מ"מ ב"ת, אולם בהמשך נראה מרטינגלים שבאמצעותם נחסום גם סטיות של פונקציות

ללא הצגה כזו, כגון מספר צביעה של גרף מקרי. נניח שלפנינו מרטינגל $X_0, X_1, \ldots, X_m$ עבורו $X_0 = 0$ (בהסתברות 1) ו־$|X_i - X_{i-1}| \le \alpha_i$ לכל $0 < i \le m$ (בהסתברות 1). אי שוויון Azuma (קצת מוכלל) קובע שלכל $\lambda > 0$ מתקיים החסם $\Pr\left[X_m > \lambda\sqrt{\sum_{i=1}^m \alpha_i^2}\right] < e^{-\lambda^2/2}$. אי השוויון הזה בפרט נותן הכללה ישירה של חסם הסטיות הגדולות הראשון שהראינו (לפי ראיית הסכומים המצטברים של מ"מ ב"ת כמרטינגל), אבל זה לא המקדם שהיה מתקבל אם היינו מנסים להציב מ"מ לפי חסם הסטיות הגדולות השני. אם רוצים "הכללה מלאה" גם עבורו, אפשר להשתמש בשאלה "לסחוט את הלימון" מתוך חוברת התרגילים התפתורים.

ההוכחה דומה לחסימת סטיות גדולות רגילות: ראשית נראה שלכל מ"מ Z עבורו $\mathrm{E}[Z] = 0$ אשר חסום בערכו המוחלט ע"י $\beta > 0$ מתקיים $\mathrm{E}\left[e^Z\right] \le e^{\beta^2/2}$. לשם כך נשים לב שלכל מספר z עבורו $|z| \le \beta$, מתקיים $e^z \le \cosh(\beta) + z\beta^{-1}\sinh(\beta)$. זאת מכיוון ש־$e^z$ היא פונקציה קמורה, ולכן ערכיה בתחום $[-\beta, \beta]$ נמצאים מתחת לערכי הפונקציה הלינארית

$$f(z) = \cosh(\beta) + z\beta^{-1}\sinh(\beta)$$

המתארת את הישר העובר דרך הנקודות $(-\beta, e^{-\beta})$ ו־(β, e^{β}). מזאת נובע עבור המשתנה המקרי Z:

$$\mathrm{E}\left[e^Z\right] \le \mathrm{E}[\cosh(\beta) + Z\beta^{-1}\sinh(\beta)] = \cosh(\beta) < e^{\beta^2/2}$$

עתה נגדיר $\alpha = \lambda/\sqrt{\sum_{i=1}^m \alpha_i^2}$, ולכל $0 < i \le m$ נגדיר $Y_i = X_i - X_{i-1}$. מהנתונים שלנו מתקיים $\mathrm{E}[Y_i|X_0, \ldots, X_{i-1}] = 0$ וכן $|Y_i| \le \alpha_i$. מכאן ניתן לקבל באינדוקציה על m:

$$\mathrm{E}\left[e^{\alpha X_m}\right] = \mathrm{E}\left[e^{\alpha Y_m} e^{\alpha X_{m-1}}\right] \le e^{\alpha^2\alpha_m^2/2}\mathrm{E}\left[e^{\alpha X_{m-1}}\right] \le \cdots \le e^{\alpha^2\sum_{i=1}^m \alpha_i^2/2} \cdot \mathrm{E}\left[e^{\alpha X_0}\right] = e^{\lambda^2/2}$$

עבור המעבר מ־$\mathrm{E}\left[e^{\alpha Y_m} e^{\alpha X_{m-1}}\right]$ ל־$e^{\alpha^2\alpha_m^2/2}\mathrm{E}\left[e^{\alpha X_{m-1}}\right]$ משתמשים בכך שתכונות Y_m מתקיימות גם בהתניה על כל סידרת ערכים אפשרית של $X_0, \ldots, X_{m-1}$, ז"א $\mathrm{E}[e^{\alpha Y_m}|X_0 = a_0, \ldots, X_{m-1} = a_{m-1}] \le e^{\alpha^2\alpha_m^2/2}$. לכן לכל פונקציה אי־שלילית f מתקיים:

$$\begin{aligned}
\mathrm{E}\left[e^{\alpha Y_m} f(X_0, \ldots, X_{m-1})\right] &= \sum_{a_0, \ldots, a_{m-1}} \Big(\mathrm{E}[e^{\alpha Y_m}|X_0 = a_0, \ldots, X_{m-1} = a_{m-1}] \\
&\qquad \cdot f(a_0, \ldots, a_{m-1})\Pr[X_0 = a_0, \ldots, X_{m-1} = a_{m-1}]\Big) \\
&\le e^{\alpha^2\alpha_m^2/2} \sum_{a_0, \ldots, a_{m-1}} f(a_0, \ldots, a_{m-1})\Pr[X_0 = a_0, \ldots, X_{m-1} = a_{m-1}] \\
&= e^{\alpha^2\alpha_m^2/2}\mathrm{E}[f(X_0, \ldots, X_{m-1})]
\end{aligned}$$

לסיום ההוכחה של חסם הסטיה הגדולה משתמשים כרגיל באי שוויון מרקוב:

$$\Pr\left[X_m > \lambda\sqrt{\sum_{i=1}^m \alpha_i^2}\right] = \Pr\left[e^{\alpha X_m} > e^{\alpha\lambda\sqrt{\sum_{i=1}^m \alpha_i^2}}\right] = \Pr\left[e^{\alpha X_m} > e^{\lambda^2}\right] < e^{\lambda^2/2}/e^{\lambda^2} = e^{-\lambda^2/2}$$

עבור מרטינגלים בהם X_0 הוא קבוע שאינו 0, פשוט חוסמים את $X_m - X_0$ ע"י הסתכלות על המרטינגל המוגדר לפי $X_i' = X_i - X_0$. כמובן שמתקיים כאן גם החסם הסימטרי מסביב ל־$X_m - X_0$.

מרטינגל החשיפה ושימושיו

על מנת לחסום את הסטיה מהממוצע של אינווריאנטים קומבינטוריים, כגון מספר צביעה של גרף מקרי, בד"כ משתמשים במרטינגל "חשיפה" של Doob (זהו גם האיש הראשון שהכיר בחשיבות הנושא של מרטינגלים, אם

כי מרטינגלים הופיעו קצת קודם בעבודה של Ville). בניית מרטינגלים אלו אינה משתמשת בסכימה מצטברת כמו בדוגמאות הקודמות, אלא בחשיפה מצטברת של מידע על מבנה קומבינטורי אקראי ושימוש בתוחלות המותנות המתאימות.

להמחשת האינטואיציה כאן נחשוב על הדוגמה הבאה: נניח שאנו מעונינים בתמחור של אופציה עתידית של מוצר. המחיר ה"נכון" עבור אופציה זו הוא תוחלת מחיר המוצר בתאריך פקיעת האופציה. אולם, המדובר הוא בעצם בתוחלת מחיר המוצר כאשר מתנים אותו על המידע המצטבר עד תאריך התמחור. אם מסתכלים על סידרת מחירי האופציה מתאריך תחילתה ועד תאריך פקיעתה, מקבלים סדרה של מ"מ, שהראשון בהם הוא קבוע (וזהה לתוחלת הלא־מותנה של מחיר המוצר הסופי), והאחרון שבהם הוא המ"מ שערכו הוא המחיר הסופי שנתקבל עבור המוצר ביום הפקיעה. מכיוון שמדובר בסדרה של תוחלות המותנות על כמות הולכת וגדלה של אינפורמציה, זה יהיה מרטינגל. הגדרה פורמלית של מרטינגל חשיפה וההוכחה שהוא אכן מרטינגל ינתנו עתה.

ראשית נציין את הנתונים שאנחנו צריכים לקראת הגדרה של מרטינגל חשיפה:

- מרחב הסתברות μ מעל קבוצה S של מבנים הסתברותיים. הקבוצה S היא קבוצה של פונקציות מתחום $\mathcal{D}$ לטווח $\mathcal{R}$ (אצלנו שני אלו יהיו סופיים). לדוגמה, המרחב $G(n,\frac{1}{2})$ הוא ההתפלגות היוניפורמית מעל הפונקציות $C:\binom{V}{2}\to\{0,1\}$, כאשר $\binom{V}{2}$ היא קבוצת הזוגות מתוך $V=\{1,\ldots,n\}$, ו־$\{0,1\}$ מייצגת את האפשרויות "לא־קשת" ו־"קשת". הקבוצה S לא בהכרח כוללת את כל הפונקציות האפשריות, אפשר למשל להגדיר מרטינגל חשיפה של פרמוטציה מקרית.

- פונקציה ממשית $f:S\to\mathbb{R}$ (או במילים אחרות, משתנה מקרי), שבד"כ נרצה להוכיח לגביה משפטי חסימה. לדוגמה, אפשר להשתמש עבור $G(n,\frac{1}{2})$ במספר הצביעה של הגרף.

- סידרת תחומים חלקיים $\emptyset=\mathcal{D}_0\subseteq\mathcal{D}_1\subseteq\cdots\subseteq\mathcal{D}_m=\mathcal{D}$, אשר יתארו לנו את אופן החשיפה של המרטינגל. לדוגמה, בהרבה יישומים עבור $G(n,p)$, מגדירים את "חשיפת הצמתים", $\mathcal{D}_i=\binom{\{1,\ldots,i\}}{2}$ (קבוצת כל הזוגות מתוך i הצמתים הראשונים). בפרט $\mathcal{D}_1=\mathcal{D}_0=\emptyset$ ו־$m=n$ במקרה זה. עוד אופן חשיפה נפוץ, לא רק לגרפים, הוא חשיפה איבר־איבר, שעבורו מגדירים סדר $(d_1,\ldots,d_m)$ מעל איברי $\mathcal{D}$, ואז קובעים $\mathcal{D}_i=\{d_1,\ldots,d_i\}$ לכל $1\le i\le m=|\mathcal{D}|$ (חשיפה איבר־איבר במקרה של גרף נקראת גם "חשיפת הקשתות").

לפי הנתונים שלנו נגדיר משתנים מקריים $X_0,\ldots,X_m$ מעל מרחב ההסתברות μ. נגדיר אותם במפורש כפונקציות ממשיות מעל קבוצת הבסיס S: עבור כל $\widetilde{C}\in S$ שמקיים $\mu(\widetilde{C})>0$, נגדיר את $X_i(\widetilde{C})$ להיות שווה ל־$\mathrm{E}_{C\sim\mu}\big[f(C)\big|C|_{\mathcal{D}_i}=\widetilde{C}|_{\mathcal{D}_i}\big]$. זוהי התוחלת המותנה של $f(C)$, כאשר C נבחר לפי μ ואנחנו מתנים על המאורע ש־C שווה ל־$\widetilde{C}$ מעל $\mathcal{D}_i$. נשים לב שבפרט מתקיים $X_0(\widetilde{C})=\mathrm{E}_{C\sim\mu}[f(C)]$ ללא תלות ב־$\widetilde{C}$, כי מאורע השוויון מעל $\mathcal{D}_0=\emptyset$ מתקיים תמיד. כמו כן תמיד מתקיים $X_m(\widetilde{C})=f(\widetilde{C})$, ז"א ש־$X_m$ שווה למ"מ המוגדר ע"י f, מכיוון שכאן בעצם ההתניה היא על $C=\widetilde{C}$. בהרבה מהיישומים שלנו, הגדרה מוצלחת של מרטינגל החשיפה תאפשר את חסימת ההסתברות להפרש גדול בין X_m ו־X_0 באמצעות משפט Azuma.

נחזור רגע לדוגמה הקונקרטית של מרטינגל חשיפת הצמתים עבור פונקצית מספר הצביעה של הגרף המקרי $G(n,p)$. כזכור, X_i הוא המ"מ המחושב כך שערך $X_i(\widetilde{G})$ הוא התוחלת המותנה של $\chi(G)$ (במרחב ההסתברות $G(n,p)$) על הגרף המושרה $\widetilde{G}[\{1,\ldots,i\}]$. במרטינגל הזה נהוג להשמיט את X_0 כי הוא זהה ל־X_1. על מנת להבין את משמעות ההגדרה, נחשב את ההתפלגות המלאה על ערכי (X_1,X_2,X_3) עבור מרטינגל חשיפת הצמתים של מספר הצביעה של $G(3,\frac{1}{2})$: חישוב ישיר (יש 8 אפשרויות עבור הגרף G) יתן לנו $X_1=\mathrm{E}_{G\sim G(3,\frac{1}{2})}[\chi(G)]=2$. לקראת חישוב X_2, חישוב ישיר נותן לנו את התוחלות המותנות $\mathrm{E}_{G\sim G(3,\frac{1}{2})}[\chi(G)|\{1,2\}\notin E(G)]=\frac{7}{4}$ ו־$\mathrm{E}_{G\sim G(3,\frac{1}{2})}[\chi(G)|\{1,2\}\in E(G)]=\frac{9}{4}$. מאלו נקבל שהשלשה (X_1,X_2,X_3) תקבל את הערכים $(2,\frac{7}{4},1)$ בהסתברות $\frac{1}{8}$ (המקרה שמגרילים את הגרף חסר הקשתות מעל $\{1,2,3\}$), את הערכים $(2,\frac{7}{4},2)$ בהסתברות $\frac{3}{8}$ (המקרה שמגרילים גרף לא־ריק שעבורו $\{1,2\}$ אינה קשת), את $(2,\frac{9}{4},2)$ בהסתברות $\frac{3}{8}$ (כשמגרילים גרף לא־מלא שמכיל את הקשת $\{1,2\}$), ואת $(2,\frac{9}{4},3)$ בהסתברות $\frac{1}{8}$ (הגרף המלא).

לפני שנמשיך, נגדיר מספר הגדרות שיעזרו לנו לפשט את ההוכחות הבאות. אנחנו נגביל את עצמנו למקרה שבו S היא סופית (או לפחות בדידה), ונניח ש־S מכילה רק מבנים $C:\mathcal{D}\to\mathcal{R}$ בעלי הסתברות חיובית, ז"א שמתקיים עבורם $\mu(C)>0$ (כזכור S לא חייבת להכיל את כל הפונקציות האפשריות). לכל $0\le i\le m$ ולכל $\widetilde{C}\in S$ נסמן ב־$S_{i,\widetilde{C}}$ את תת־הקבוצה $\{C\in S: C|_{\mathcal{D}_i}=\widetilde{C}|_{\mathcal{D}_i}\}$, ונזהה אותה עם המאורע המתאים "$C|_{\mathcal{D}_i}=\widetilde{C}|_{\mathcal{D}_i}$". כמו כן נסמן ב־$\mu_{i,\widetilde{C}}:S_{i,\widetilde{C}}\to[0,1]$ את ההתפלגות המותנה המתאימה, ז"א את הפונקציה המקיימת $\mu_{i,\widetilde{C}}(C)=\mu(C)/\mu(S_{i,\widetilde{C}})$ לכל $C\in S_{i,\widetilde{C}}$. בשימוש בסימונים שהגדרנו, מתקיים $X_i(\widetilde{C})=\mathrm{E}_{C\sim\mu}[f(C)|S_{i,\widetilde{C}}]=\mathrm{E}_{C\sim\mu_{i,\widetilde{C}}}[f(C)]$ לכל $\widetilde{C}\in S$.

עתה נראה שאכן תנאי חוסר הזיכרון מתקיים עבור סדרת המ"מ $X_0,\dots,X_m$, עבור מרחבי הסתברות בדידים, למרות שהטענה נכונה גם למרחבים כלליים יותר. ראשית, נשים לב שעבור כל $\widehat{C}\in S$ מתקיים:

$$\begin{aligned}
\mathrm{E}_{C\sim\mu}[X_i(C)|S_{i-1,\widehat{C}}] &= \sum_{C\in S_{i-1,\widehat{C}}} X_i(C)\cdot\mu_{i-1,\widehat{C}}(C)=\sum_{C\in S_{i-1,\widehat{C}}}\left(\sum_{\widetilde{C}\in S_{i,C}} f(\widetilde{C})\cdot\mu_{i,C}(\widetilde{C})\right)\mu_{i-1,\widehat{C}}(C)\\
&=\sum_{\widetilde{C}\in S_{i-1,\widehat{C}}} f(\widetilde{C})\left(\sum_{C\in S_{i,\widetilde{C}}}\mu_{i,C}(\widetilde{C})\cdot\mu_{i-1,\widehat{C}}(C)\right)
\end{aligned}$$

הסברים לפיתוח: השוויון הראשון הוא הצבת ההגדרה של התוחלת המותנה כסכום המתאים (עבור מרחבים בדידים), בשוויון השני השתמשנו ב־$X_i(C)=\mathrm{E}_{\widetilde{C}\sim\mu}[f(\widetilde{C})|S_{i,C}]$ (החלפנו את סימוני המשתנים C ו־$\widetilde{C}$), ושוב הצבנו את ההגדרה של התוחלת כסכום, והשוויון השלישי הוא שינוי של סדר הסכימה. שימו לב שאותם זוגות $C,\widetilde{C}$ מקיימים את תנאי האינדכסים של הסכימה בשני המקרים. ספציפית אלו כל הזוגות שמקיימים $C|_{\mathcal{D}_{i-1}}=\widetilde{C}|_{\mathcal{D}_{i-1}}=\widehat{C}|_{\mathcal{D}_{i-1}}$ וגם $C|_{\mathcal{D}_i}=\widetilde{C}|_{\mathcal{D}_i}$.

נסתכל עכשיו על הביטוי $\mu_{i,C}(\widetilde{C})\cdot\mu_{i-1,\widehat{C}}(C)$. הוא מופעל רק על $C,\widetilde{C},\widehat{C}$ שמקיימים $C|_{\mathcal{D}_i}=\widetilde{C}|_{\mathcal{D}_i}$, ולכן $S_{i,C}=S_{i,\widetilde{C}}$, וכמו כן $C|_{\mathcal{D}_{i-1}}=\widetilde{C}|_{\mathcal{D}_{i-1}}=\widehat{C}|_{\mathcal{D}_{i-1}}$ (השתמשנו כאן ב־$\mathcal{D}_{i-1}\subseteq\mathcal{D}_i$). עבור אלו נפתח ונקבל:

$$\begin{aligned}
\mu_{i,C}(\widetilde{C})\cdot\mu_{i-1,\widehat{C}}(C) &= \left(\mu(\widetilde{C})/\mu(S_{i,C})\right)\left(\mu(C)/\mu(S_{i-1,\widehat{C}})\right)\\
&= \left(\mu(\widetilde{C})/\mu(S_{i,\widetilde{C}})\right)\left(\mu(C)/\mu(S_{i-1,\widehat{C}})\right)\\
&= \left(\mu(\widetilde{C})/\mu(S_{i-1,\widehat{C}})\right)\left(\mu(C)/\mu(S_{i,\widetilde{C}})\right)=\mu_{i-1,\widehat{C}}(\widetilde{C})\cdot\mu_{i,\widetilde{C}}(C)
\end{aligned}$$

בחזרה לפיתוח המקורי, נקבל מזה:

$$\begin{aligned}
\mathrm{E}_{C\sim\mu}[X_i(C)|S_{i-1,\widehat{C}}] &= \sum_{\widetilde{C}\in S_{i-1,\widehat{C}}} f(\widetilde{C})\cdot\mu_{i-1,\widehat{C}}(\widetilde{C})\left(\sum_{C\in S_{i,\widetilde{C}}}\mu_{i,\widetilde{C}}(C)\right)\\
&= \sum_{\widetilde{C}\in S_{i-1,\widehat{C}}} f(\widetilde{C})\cdot\mu_{i-1,\widehat{C}}(\widetilde{C})=X_{i-1}(\widehat{C})
\end{aligned}$$

על מנת להוכיח מזה את חוסר הזיכרון, נשים לב ש־$\mathrm{E}_{C\sim\mu}[X_i(C)|X_0(C)=a_0,\dots,X_{i-1}(C)=a_{i-1}]$ שווה לסכום $\sum_{\widehat{C}:X_0(\widehat{C})=a_0,\dots,X_{i-1}(\widehat{C})=a_{i-1}}\mathrm{E}_{C\sim\mu}[X_i(C)|S_{i-1,\widehat{C}}]\mathrm{Pr}_\mu[\widehat{C}|X_0(C)=a_0,\dots,X_{i-1}(C)=a_{i-1}]$, שבו ביטוי התוחלת לפי מה שראינו למעלה תמיד יהיה שווה ל־$X_{i-1}(\widehat{C})=a_{i-1}$, ואז סכום ההסתברויות המותנות יהיה שווה בו ל־1.

עתה נראה שיטה מקובלת לחסימת $|X_i - X_{i-1}|$ עבור מרטינגל חשיפה. נניח ש־S מכילה את כל הפונקציות האפשריות מ־$\mathcal{D}$ ל־$\mathcal{R}$, וההתפלגות μ על C היא כזו שלכל $d \in \mathcal{D}$ הערך $C(d)$ נבחר באופן ב"ת בערכים האחרים של C. למשל, ההתפלגות על גרפים המוגדרת ע"י $G(n,p)$ היא התפלגות כזו. נניח גם שלכל C_1, C_2 הנבדלים ביניהם רק בתוך תת־הקבוצה $\mathcal{D}_i \setminus \mathcal{D}_{i-1}$ מתקיים $|f(C_1) - f(C_2)| \leq \alpha_i$.

נראה עתה שמרטינגל החשיפה, כאשר ההתפלגות μ והפונקציה f מקיימים את התנאים הכתובים למעלה, יקיים $|X_i - X_{i-1}| \leq \alpha_i$ בהסתברות 1. גם כאן נראה את ההוכחה עבור מרחבי הסתברות בדידים בלבד. נעיר שהמקרה של $\alpha_1 = \cdots = \alpha_m = 1$ נקרא גם תנאי ליפשיץ (גם בהתייחסות ל־f וגם בהתייחסות למרטינגל). לא קשה לראות שתנאי ליפשיץ מתקיים בפרט עבור פונקצית הצביעה של גרף ביחס לחשיפת הצמתים.

נראה אם כן שלכל $\widetilde{C} \in S$ מתקיים $|X_i(\widetilde{C}) - X_{i-1}(\widetilde{C})| \leq \alpha_i$. לשם כך נגדיר מרחב הסתברות חדש ν מעל זוגות של מבנים $C_1, C_2 \in S$. לשם קביעתם נגריל את C בהתאם ל־μ, נגדיר את C_1 לפי $C_1(d) = \widetilde{C}(d)$ אם $d \in \mathcal{D}_{i-1}$, ו־$C_1(d) = C(d)$ אם $d \notin \mathcal{D}_{i-1}$, ובאופן דומה נגדיר את C_2 לפי $C_2(d) = \widetilde{C}(d)$ אם $d \in \mathcal{D}_i$ ו־$C_2(d) = C(d)$ אם $d \notin \mathcal{D}_i$.

בגלל תכונת אי־התלות שדרשנו מ־μ, התפלגות C_1 זהה להתפלגות המותנה $\mu_{i-1,\widetilde{C}}$ שהגדרנו קודם (לקראת הוכחת תכונת חוסר הזיכרון של מרטינגל החשיפה). לכן $\mathrm{E}_\nu[f(C_1)] = \mathrm{E}_{C \sim \mu_{i-1,\widetilde{C}}}[f(C)] = X_{i-1}(\widetilde{C})$. בעזרת נימוק דומה נקבל $\mathrm{E}_\nu[f(C_2)] = X_i(\widetilde{C})$. מצד שני C_1 ו־C_2 לפי הגדרתם יכולים להיבדל רק על איברי $\mathcal{D}_i \setminus \mathcal{D}_{i-1}$, ולכן מלינאריות התוחלת נקבל:

$$|X_i(\widetilde{C}) - X_{i-1}(\widetilde{C})| = \left|\mathrm{E}_\nu[f(C_1)] - \mathrm{E}_\nu[f(C_2)]\right| = \left|\mathrm{E}_\nu[f(C_1) - f(C_2)]\right| \leq \mathrm{E}_\nu[|f(C_1) - f(C_2)|] \leq \alpha_i$$

מסקנה מיידית של Shamir, Spencer בהקשר מרטינגל חשיפת הצמתים היא קיום $u_{n,p}$ לכל n ו־p, כך שעבור $G = G(n,p)$ מתקיים $\Pr\left[|\chi(G) - u_{n,p}| > \lambda\sqrt{n}\right] < 2e^{-\lambda^2/2}$. להוכחה פשוט מגדירים את $u_{n,p}$ להיות תוחלת מספר הצביעה $\chi(G)$. שיטת ההוכחה אינה אומרת מהי התוחלת; Bollobás הראה עבור $G(n, \frac{1}{2})$ שמספר הצביעה הוא כמעט תמיד $(1+o(1))n/2\log_2 n$.

דוגמאות שימושים נוספות

דוגמה מיידית אחרת לשימוש במרטינגל חשיפה היא זו: בהינתן פונקציה מקרית $g : \{1,\ldots,n\} \to \{1,\ldots,n\}$ שנבחרת יוניפורמית, נרצה לחסום את גודל $A = \{k | \forall_i g(i) \neq k\}$, קבוצת כל האיברים שאינם נמצאים בתמונת g (שימו לב שכאן הטווח של g הוא לא $\{0,1\}$). לכל k מתקיים $\Pr\left[k \in A\right] = (1 - \frac{1}{n})^n$, ולכן התוחלת של $|A|$ היא $n(1-\frac{1}{n})^n$, ומכך ניתן להסיק (עם חשבון אינפיניטיסמלי) ש־$|\mathrm{E}[|A|] - \frac{n}{e}| < 1$. עתה ניתן להשתמש במרטינגל החשיפה עבור $\mathcal{D}_i = \{1,\ldots,i\}$ כדי להסיק את $\Pr\left[\left||A| - \frac{n}{e}\right| \geq \lambda\sqrt{n} + 1\right] < 2e^{-\lambda^2/2}$ לכל $\lambda > 0$. חשוב לזכור דבר אחד: X_i אינו מתאר את גודל קבוצת האיברים שאינם נמצאים בתמונת $g|_{\{1,\ldots,i\}}$ לאחר שזו נקבעה (ערכים אלו לא היו יוצרים מרטינגל). הוא מתאר את תוחלת גודל קבוצת האיברים שאינם נמצאים בתמונת g כולה, לאחר ש־$g|_{\{1,\ldots,i\}}$ כבר נקבעה.

נראה עתה שעבור $G(n, n^{-\alpha})$ כאשר $\alpha > \frac{5}{6}$, קיים $u(n,\alpha)$ כך שכמעט תמיד (ז"א בהסתברות $1 - o(1)$) מתקיים $u \leq \chi(G) \leq u + 3$. תוצאות דומות (יותר חזקות) ניתנו בעבודתם של łuczak ב־1991 ושל Alon, Krivelevich ב־1997. על מנת להוכיח זאת נראה שלכל $\epsilon > 0$ קיים $u(n,\alpha,\epsilon)$ עבורו $u \leq \chi(G) \leq u + 3$ בהסתברות לפחות $1 - \epsilon$, לכל n גדול דיו. ממנו אפשר לבנות את $u(n,\alpha)$ ע"י טיעון שרשור בדומה להוכחות שרואים בחשבון אינפיניטיסמלי.

למה ראשונה: לכל c קבוע מתקיים כ"ת (כמעט תמיד) שכל הקבוצות מגודל $\lfloor c\sqrt{n} \rfloor$ ב־G הן 3־צביעות. הוכחה: יהי t הגודל המינימלי של קבוצת צמתים שאינה 3־צביעה, ותהי T קבוצה כזו. מכיוון שלכל $v \in T$ הקבוצה $T \setminus \{v\}$ היא 3־צביעה, נובע מכך שהדרגה המינימלית (בין כל הצמתים) של תת הגרף המושרה היא לפחות 3: אחרת היה אפשר לצבוע את $T \setminus \{v\}$ ב־3 צבעים ואז לצבוע את v בצבע שלא מופיע בשכניו,

בסתירה. עקב הדרגה המינימלית, מספר הקשתות בין איברי T הוא לפחות $\frac{3}{2}t$. הסיכוי לקיום קבוצה כזו עם t קטן מספיק חסום (כאשר c_1 ו־c_2 קבועים מתאימים) ע"י

$$\sum_{t=4}^{\lfloor c\sqrt{n}\rfloor}\binom{n}{t}\binom{\binom{t}{2}}{\frac{3}{2}t}n^{-3\alpha t/2} \le \sum_{t=4}^{\lfloor c\sqrt{n}\rfloor}(\frac{ne}{t})^t(\frac{te}{3})^{3t/2}n^{-3\alpha t/2} = \sum_{t=4}^{\lfloor c\sqrt{n}\rfloor}(c_1 n^{1-\frac{3\alpha}{2}}t^{\frac{1}{2}})^t \le \sum_{t=4}^{\lfloor c\sqrt{n}\rfloor}(c_2 n^{\frac{5}{4}-\frac{3\alpha}{2}})^t$$

ומכיוון שחזקת n בסוגריים הימניים היא שלילית, הסכום הוא $o(1)$. שימו לב שהיינו צריכים לעשות סכום על כל הגדלים האפשריים ולא רק על קבוצות מגודל $\lfloor c\sqrt{n}\rfloor$, בגלל שתנאי הדרגה נכון רק לקבוצות לא 3־צביעות מינימליות.

עתה להוכחת המשפט, מגדירים את u להיות השלם המינימלי עבורו $\Pr\big[\chi(G)\le u\big] > \frac{1}{3}\epsilon$, ומסתכלים על מרטינגל חשיפת הצמתים עבור הפונקציה $Y(G)$ המוגדרת כגודל המינימלי של קבוצה S כך ש־$G\setminus S$ הוא u־צביע. נסמן $\eta = \mathrm{E}[Y]$, ונבחר λ עבורו $e^{-\lambda^2/2} = \frac{1}{3}\epsilon$; לפי משפט אזומה (נשים לב שהפונקציה $Y(G)$ מקיימת את תנאי ליפשיץ ביחס לחשיפת הצמתים), מתקיים $\Pr\big[Y\le \eta-\lambda\sqrt{n-1}\big] < e^{-\lambda^2/2} = \frac{1}{3}\epsilon$, ולכן (מבחירת u כך ש־$Y=0$ מתקיים בהסתברות גדולה מ־$\frac{1}{3}\epsilon$) מתקיים בהכרח $\eta\le\lambda\sqrt{n-1}$.

מצד שני, $\Pr\big[Y\ge 2\lambda\sqrt{n-1}\big] \le \Pr\big[Y\ge \eta+\lambda\sqrt{n-1}\big] \le \frac{1}{3}\epsilon$, ולכן בסיכוי $1-\frac{1}{3}\epsilon$ לפחות יש u־צביעה לכל הצמתים פרט ללא יותר מ־$c\sqrt{n}$ מתוכם (עבור c מתאים). לפי הלמה הקודמת, בסיכוי $1-\frac{1}{3}\epsilon$ לפחות (ל־n גדול דיו) אפשר לצבוע את קבוצת הצמתים הנותרים בלא יותר מ־3 צבעים נוספים (כי הלמה קובעת שניתן לעשות זאת לכל קבוצת צמתים בגודל הזה), ומכאן ש־$\chi(G)\le u+3$ מתקיים בהסתברות $1-\frac{2}{3}\epsilon$ לפחות. לבסוף נזכור שבחירת u מבטיחה שמתקיים $u\le\chi(G)$ בהסתברות $1-\frac{1}{3}\epsilon$ לפחות, להשלמת ההוכחה.

הלמה הלוקלית

הלמה הלוקלית הכללית

הלמה הלוקלית של Lovász (הופיעה לראשונה במאמר של Erdös ו־Lovász מ־1975) היא טענה בתורת ההסתברות שנוסחה והוכחה במיוחד עבור שימושיה הקומבינטוריים. הרעיון: אם יש בידינו סדרה של מאורעות ב"ת (לחלוטין) כך שלכל אחד מהם סיכוי קטן מאחד לקרות, אז ברור שבסיכוי חיובי (אם כי קטן) אף אחד מהמאורעות לא יקרה. הלמה הלוקלית מאפשרת להכליל את הנימוק הזה גם כאשר אין אי־תלות מושלמת. כאן, אם יש קבוצה של מאורעות נדירים מספיק, שכל אחד מהם תלוי במעט מהמאורעות האחרים, אז שוב בהסתברות חיובית אף אחד מהם לא יקרה.

עם זאת החסם התחתון על ההסתברות אינו גדול, כך שבניגוד לרב השיטות האחרות שיטה זו אינה נותנת בניה קונסטרוקטיבית באופן אוטומטי (אפילו לא הסתברותית). לאחר שנים שבהם היתה רק בניה קונסטרוקטיבית של Beck למקרה פרטי, לאחרונה Moser מצא בניה כללית יותר שתינתן בתרגול. כאן נראה את הגרסה המקורית הלא־קונסטרוקטיבית – ראשית ננסח את הגרסה הכללית ביותר, ונוכיח אותה באינדוקציה. לאחריה ננסח את המקרה הפרטי הסימטרי שבו משתמשים בד"כ (אם כי ישנן גם דוגמאות המצריכות את המקרה הכללי). בתרגול תוצג גם "גרסת ביניים" של הלמה הלוקלית, אשר מאוד נוחה לשימוש בחלק מהמקרים שבהם הגרסה הסימטרית אינה מספקת.

עבור ניסוח הלמה נשתמש בסימון מעט שונה מזה של הספר, ונשתמש ברשימות במקום בגרף מכוון על מנת לציין את התלויות (למעשה איבר ברשימת התלויות שלנו מתאים לקבוצת הקשתות היוצאות מצומת מסויים של גרף התלויות בסימון של הספר). נניח שלפנינו סדרה $B_1,\ldots,B_m$ של מאורעות. רשימת תלויות עבור סדרה זו היא סדרה $D_1,\ldots,D_m$ כך שכל D_i היא תת־קבוצה של $\{1,\ldots,m\}\setminus\{i\}$, ולכל i המאורע B_i אינו תלוי באלגברה הנוצרת ע"י המאורעות $\{B_j | j\notin D_i\cup\{i\}\}$. חשוב לשים לב שהכוונה היא ש־B_i אינו תלוי באף מאורע המתקבל מחיתוכים וקומבינציות בוליאניות אחרות של המאורעות בקבוצה הנ"ל. לא קשה לבנות מקרים שבהם איברי הרשימה המותרים אינם יחידים. למשל: המקרה שבהטלת שני מטבעות הוגנים ב"ת B_1 מציין את המאורע שהמטבע הראשון יצא "עץ", B_2 מציין מאורע זה עבור המטבע השני, ו־B_3 מציין את המאורע ששני המטבעות קבלו תוצאות שונות זו מזו. במקרה זה, כל רשימה של שלוש תתי־קבוצה לא ריקות מתאימות היא קבילה עבור מרחב ההסתברות הנ"ל.

הלמה הלוקלית הכללית: אם עבור המאורעות ה"רעים" $B_1,\dots,B_m$ קיימת רשימת תלויות $D_1,\dots,D_m$ ומספרים ממשיים $x_1,\dots,x_m$, עבורם $0 \le x_i < 1$ ולכל i מתקיים $\Pr[B_i] \le x_i \prod_{j\in D_i}(1-x_j)$, אז מתקיים בהכרח $\Pr\big[\bigwedge_{i=1}^m \neg B_i\big] \ge \prod_{i=1}^m (1-x_i) > 0$.

הוכחת הלמה מתבססת על שימוש מושכל (ומאסיבי) בנוסחת ההסתברות המותנה. ראשית, מוכיחים שלכל $S \subset \{1,\dots,m\}$ ולכל $i \notin S$ מתקיים $\Pr\big[B_i | \bigwedge_{j\in S}\neg B_j\big] \le x_i$. מטענת עזר זו הלמה המלאה נובעת באמצעות הפעלות חוזרות של נוסחת ההסתברות:

$$\Pr\big[\bigwedge_{i=1}^m \neg B_i\big] = \Pr\big[\neg B_1 | \bigwedge_{i=2}^m \neg B_i\big]\cdot \Pr\big[\bigwedge_{i=2}^m \neg B_i\big] \ge (1-x_1)\Pr\big[\bigwedge_{i=2}^m \neg B_i\big] \ge \cdots \ge \prod_{i=1}^m (1-x_i)$$

עתה נוכיח את טענת העזר, באמצעות אינדוקציה על $|S|$. עבור $|S|=0$, הטענה נובעת ישירות מתנאי הלמה: $\Pr[B_i] \le x_i \prod_{j\in D_i}(1-x_j) \le x_i$. עתה, אם טענה זו ידועה עבור כל קבוצה מגודל $s-1$ או פחות, נראה אותה עבור כל קבוצה S מגודל s: נסמן ב־S_1 את איברי S הנמצאים ב־D_i, וב־S_2 את כל השאר. מקבלים לפי נוסחת ההסתברות המותנה (בהתניה על $\bigwedge_{l\in S_2}\neg B_l$):

$$\Pr\big[B_i | \bigwedge_{j\in S}\neg B_j\big] = \frac{\Pr[B_i \wedge \bigwedge_{j\in S_1}\neg B_j | \bigwedge_{l\in S_2}\neg B_l]}{\Pr[\bigwedge_{j\in S_1}\neg B_j | \bigwedge_{l\in S_2}\neg B_l]}$$

עקב אי התלות של B_i ב־$\bigwedge_{l\in S_2}\neg B_l$ המונה מקיים:

$$\Pr\big[B_i \wedge \bigwedge_{j\in S_1}\neg B_j | \bigwedge_{l\in S_2}\neg B_l\big] \le \Pr\big[B_i | \bigwedge_{l\in S_2}\neg B_l\big] = \Pr[B_i] \le x_i \prod_{j\in D_i}(1-x_j)$$

באשר למכנה, נסמן $S_1 = \{j_1,\dots,j_r\}$. אם $r=0$ אז המכנה הוא 1 והטענה מוכחת בקלות. אחרת, תוך הסתמכות על הנחת האינדוקציה (ובשימוש חוזר בנוסחת ההסתברות) מקבלים:

$$\Pr\big[\bigwedge_{j\in S_1}\neg B_j | \bigwedge_{l\in S_2}\neg B_l\big] = \prod_{q=1}^r \Pr\big[\neg B_{j_q} | \bigwedge_{t=q+1}^r \neg B_{j_t} \wedge \bigwedge_{l\in S_2}\neg B_l\big] \ge \prod_{t=1}^r (1-x_{j_r}) \ge \prod_{j\in D_i}(1-x_j)$$

לסיכום מציבים את החסמים לקבלת המבוקש:

$$\Pr\big[B_i | \bigwedge_{j\in S}\neg B_j\big] \le x_i \frac{\prod_{j\in D_i}(1-x_j)}{\prod_{j\in D_i}(1-x_j)} = x_i$$

המקרה הסימטרי של הלמה וישומים

בדרך כלל משתמשים במקרה הפרטי הסימטרי של הלמה הלוקלית, שהוא מסקנה פשוטה להפעלה האומרת כך: אם $B_1,\dots,B_m$ מאורעות, כך שכל אחד מהם הוא ב"ת באלגברה הנוצרת ע"י כל האחרים פרט ל־d מהם, והסיכוי לקיום כל מאורע חסום ע"י $p<1$ שעבורו מתקיים $ep(d+1) \le 1$, אז בסיכוי חיובי אף מאורע לא מתקיים. הוכחת המקרה הפרטי הזה היא ע"י לקיחת $x_i = \frac{1}{d+1}$ עבור הלמה הלוקלית הכללית, תוך שימוש באי השוויון $e^{-1} < (1-\frac{1}{d+1})^d$ (אפשר לראות אותו למשל ע"י $e^{-1/d} < 1-\frac{1}{d}+\frac{1}{2d^2} < 1-\frac{1}{d+1}$), וברשימת התלויות המתאימה שכל איבריה בעלי גודל חסום ע"י d. אז מוודאים את תנאי הלמה הכללית:

$$\Pr[B_i] \le p \le \frac{1}{d+1}e^{-1} < \frac{1}{d+1}\Big(1-\frac{1}{d+1}\Big)^d \le x_i \prod_{j\in D_i}(1-x_j)$$

נזכר עתה בשני מקרים קיצוניים: אם מצד אחד $p < \frac{1}{m}$ אבל לא ידוע כלום על התלויות, אז החסם הרגיל על איחוד מאורעות יתן סיכוי חיובי עבור $\Pr\big[\bigwedge_{i=1}^{m} \neg B_i\big] \geq 1 - pm > 0$. הלמה הלוקלית היתה נותנת תוצאה זו רק עבור $p \leq \frac{1}{em}$. אם מצד שני ידוע רק ש־$p < 1$ אבל ידוע בנוסף שכל המאורעות הם ב"ת, אז נובע מכך $\Pr\big[\bigwedge_{i=1}^{m} \neg B_i\big] = (1-p)^m > 0$. במקרה זה הלמה הלוקלית כפי שמנוסחת כאן היתה נותנת זאת עבור $p \leq e^{-1}$. החוזק של הלמה הוא בגישור בין שני המקרים הקיצוניים האלו, והמחיר שמשלמים הוא במקדם הנוסף של e^{-1} (שאכן אי אפשר תמיד להיפטר ממנו עבור d גדול).

בהרבה מקרים יש ללמה הלוקלית ישומים דמויי צביעה. דוגמה מגוחכת: נראה שעבור גרף G בעל דרגה מקסימלית d קיימת צביעה ב־$\lceil e(2d-1) \rceil$ צבעים (הדוגמה מגוחכת כי הוכחה דטרמיניסטית פשוטה מאוד תראה שיש צביעה עבור G אף ב־$d+1$ צבעים, אבל זוהי המחשה טובה לשימוש בלמה הלוקלית).

להוכחה, נגריל לגרף עם דרגה מקסימלית d צביעה ב־$k = \lceil e(2d-1) \rceil$ צבעים, באופן יוניפורמי וב"ת לכל צומת, ולכל קשת uv נגדיר את המאורע B_{uv} כמאורע שקשת זו נצבעה מונוכרומטית. מתקיים $\Pr[B_{uv}] = \frac{1}{k}$. בנוסף, B_{uv} אינו תלוי באלגברה הנוצרת ע"י $\{B_{u'v'} | \{u,v\} \cap \{u',v'\} = \emptyset\}$, ומכאן שאפשר לכתוב עבור המאורעות רשימת תלויות כך שגודל הקבוצות אינו עולה על $2d-2$ (וזהי תהיה למעשה רשימת השכנויות של צמתי ה־line graph של G). לסיכום ההוכחה מוודאים שמתקיים $e\frac{1}{k}(2d-1) \leq 1$, ומכאן שבסיכוי חיובי לא מתקיים אף אחד מהמאורעות הנ"ל. לכן יש עבור צמתי הגרף k־צביעה חוקית.

שימוש דומה מאוד לדוגמה האחרונה יתן עבור היפרגרפים תוצאה לא טריביאלית. היפרגרף יקרא 2־צביע אם יש צביעה ב־2 צבעים שבה כל קשת תכיל צמתים משני הצבעים (אגב, תכונה זו היא NP-Hard עבור כל $r > 2$). נניח שבידינו היפרגרף r־יוניפורמי, ושבו אף קשת אינה חותכת יותר מ־k קשתות אחרות, כך שמתקיים $e(k+1) < 2^{r-1}$. היפרגרף זה הוא בהכרח 2־צביע: מגרילים צבע לכל צומת באופן יוניפורמי וב"ת, ולכל קשת h כותבים את המאורע B_h שקשת זו היא מונוכרומטית תחת הצביעה שהוגרלה. מתקיים בבירור $\Pr[B_h] = 2^{1-r}$, וכן קיימת למאורעות אלו רשימת תלויות עם גודל קבוצות מקסימלי שאינו עולה על k, ומכאן שניתן להפעיל את הלמה הלוקלית לקבלת המבוקש. בפרט נובע מכך שעבור $r \geq 9$, היפרגרף r־יוניפורמי r־רגולרי הוא 2־צביע (וזה אינו נכון ל־$r = 2$; היום ידוע גם שזה אינו נכון ל־$r = 3$ אבל כן נכון כבר ל־$r = 8$). לתוצאה הזו עבור ההיפרגרפים כבר לא ידועה הוכחה פשוטה.

קורלציות (correlation inequalities)

מבוא והצגת משפט ארבע הפונקציות

נניח שבידינו גרף מקרי ע"פ המרחב $G(n, \frac{1}{2})$. יהי E המאורע "G הוא 4־צביע" ו־F המאורע "G חסר משולשים". היינו מצפים שיתקיים $\Pr[E \wedge F] \geq \Pr[E]\Pr[F]$, ז"א ששני המאורעות "יתרמו" זה לזה, כי שניהם קשורים לכך ש"אין יותר מדי קשתות". הניסוח הפורמלי של אינטואיציה זו מהווה משפט של Kleitman. אנו נסיק אותו, וכן הכללה של Fortuin, Kasteleyn, Ginibre, מתוצאה מאוד כללית של Ahlswede, Daykin (היסטורית המשפט של Kleitman הוכח ראשון, כצפוי). לתוצאה הכללית, שנציג עתה, קוראים משפט ארבע הפונקציות.

לשם הצגת התוצאה הכללית דרושים כמה סימונים (הסימונים שיוצגו כאן שונים במעט מהסימונים בספר): נניח שלפנינו קבוצה סופית S. נסמן ב־$\mathcal{P}(S)$ את משפחת תתי הקבוצות של S, ונסתכל על פונקציות מהצורה $\varphi : \mathcal{P}(S) \to \mathbb{R}^+$. בדוגמה למעלה S תהיה קבוצת כל הזוגות של קבוצת הצמתים V, כך ש־$\mathcal{P}(S)$ תהיה בעצם קבוצת כל הגרפים האפשריים מעל V כאשר כל גרף מוגדר לפי קבוצת קשתותיו. עתה, לכל משפחה $\mathcal{A} \subseteq \mathcal{P}(S)$ של ת"ק של S נסמן $\overline{\varphi}(\mathcal{A}) = \sum_{A \in \mathcal{A}} \varphi(A)$, כך שהגדרנו גם את הפונקציה $\overline{\varphi} : \mathcal{P}(\mathcal{P}(S)) \to \mathbb{R}^+$. לשם נוחות, במקרים שבהם אין בלבול אפשר להשתמש בסימון φ גם עבור $\overline{\varphi}$. בספר אין סימון מיוחד ל־$\overline{\varphi}$; המקום היחידי שבו יתכן בלבול הוא ב־$\varphi(\emptyset) = \overline{\varphi}(\{\emptyset\})$ לעומת $\overline{\varphi}(\emptyset) = \sum_{A \in \emptyset} \varphi(A) = 0$.

עבור $\mathcal{A}, \mathcal{B} \subseteq \mathcal{P}(S)$ נגדיר לצורך ענינינו כאן את $\mathcal{A} \sqcup \mathcal{B} = \{A \cup B | A \in \mathcal{A}, B \in \mathcal{B}\}$ (ללא ספירת החזרות) ואת $\mathcal{A} \sqcap \mathcal{B} = \{A \cap B | A \in \mathcal{A}, B \in \mathcal{B}\}$ (בספר משתמשים בסימני איחוד וחיתוך רגילים, אולם אנו נשמור את אלו למובנם המקורי; למשל $\mathcal{A} \cap \mathcal{B}$ יסמן את קבוצת הקבוצות המופיעות גם ב־$\mathcal{A}$ וגם ב־$\mathcal{B}$). על מנת להבין את המשמעות של סימונים אלו, שימו לב לדוגמה שאם $\mathcal{A}, \mathcal{B}$ הן מונוטוניות עולות (ז"א ש־$\mathcal{A}$ מקיימת שאם

$A \in \mathcal{A}$ ו־$A \subset A'$ אז גם $A' \in \mathcal{A}$, ובדומה ל־$\mathcal{B}$), אז מתקיים $\mathcal{A} \sqcup \mathcal{B} = \mathcal{A} \sqcap \mathcal{B}$: מצד אחד, אם $C \in \mathcal{A} \sqcap \mathcal{B}$ אז בוודאי שמתקיים $C = C \cup C \in \mathcal{A} \sqcup \mathcal{B}$. מצד שני, אם $C = A \cup B \in \mathcal{A} \sqcup \mathcal{B}$ אז ממונוטוניות $\mathcal{A}$ ומ־$A \in \mathcal{A}$ נובע $C \in \mathcal{A}$, ובדומה לכך מוכיחים שגם $C \in \mathcal{B}$, ולכן $C \in \mathcal{A} \cap \mathcal{B}$.

עתה ננסח את המשפט הכללי: אם עבור ארבע הפונקציות $\alpha, \beta, \gamma, \delta : \mathcal{P}(S) \to \mathbb{R}^+$ מתקיים לכל שתי קבוצות $A, B \in \mathcal{P}(S)$ ש־$\alpha(A)\beta(B) \leq \gamma(A \cup B)\delta(A \cap B)$, אז מתקיים לכל שתי משפחות של קבוצות $\mathcal{A}, \mathcal{B} \subseteq \mathcal{P}(S)$ ש־$\overline{\alpha}(\mathcal{A})\overline{\beta}(\mathcal{B}) \leq \overline{\gamma}(\mathcal{A} \sqcup \mathcal{B})\overline{\delta}(\mathcal{A} \sqcap \mathcal{B})$. אגב, ניתן גם להכליל משפט זה לפונקציות מעל רשתות פילוג (distributive lattices) כלליות, ולא רק פונקציות מעל $\mathcal{P}(S)$, אולם לא ניכנס לכך כאן.

הוכחת משפט ארבע הפונקציות

ראשית נשים לב שבלי הגבלת הכלליות אפשר להניח כי $\mathcal{A} = \mathcal{B} = \mathcal{A} \sqcup \mathcal{B} = \mathcal{A} \sqcap \mathcal{B} = \mathcal{P}(S)$: אם המצב אינו כך, אז נגדיר את הפונקציות הבאות:

$$\alpha'(C) = \begin{cases} \alpha(C), & C \in \mathcal{A} \\ 0, & C \notin \mathcal{A} \end{cases} \quad \gamma'(C) = \begin{cases} \gamma(C), & C \in \mathcal{A} \sqcup \mathcal{B} \\ 0, & C \notin \mathcal{A} \sqcup \mathcal{B} \end{cases}$$
$$\beta'(C) = \begin{cases} \beta(C), & C \in \mathcal{B} \\ 0, & C \notin \mathcal{B} \end{cases} \quad \delta'(C) = \begin{cases} \delta(C), & C \in \mathcal{A} \sqcap \mathcal{B} \\ 0, & C \notin \mathcal{A} \sqcap \mathcal{B} \end{cases}$$

לא קשה להווכח שאם הפונקציות המקוריות קיימו את תנאי משפט ארבע הפונקציות אז גם הפונקציות האלו יקיימו אותו, וכן שמתקיים עבורן

$$\overline{\alpha'}(\mathcal{P}(S)) = \overline{\alpha}(\mathcal{A}),\ \overline{\beta'}(\mathcal{P}(S)) = \overline{\beta}(\mathcal{B}),\ \overline{\gamma'}(\mathcal{P}(S)) = \overline{\gamma}(\mathcal{A} \sqcup \mathcal{B}),\ \overline{\delta'}(\mathcal{P}(S)) = \overline{\delta}(\mathcal{A} \sqcap \mathcal{B})$$

הוכחת המשפט תיעשה עתה באינדוקציה על $|S|$. ראשית נוכיח את הלמה הבאה, שהיא בעצם שקולה לטענת המשפט עבור המקרה $|S| = 1$: אם עבור המספרים האי־שליליים $\alpha_0, \alpha_1, \beta_0, \beta_1, \gamma_0, \gamma_1, \delta_0, \delta_1$ מתקיים

$$\alpha_0\beta_0 \leq \gamma_0\delta_0,\ \alpha_0\beta_1 \leq \gamma_1\delta_0,\ \alpha_1\beta_0 \leq \gamma_1\delta_0,\ \alpha_1\beta_1 \leq \gamma_1\delta_1$$

אז מתקיים $(\alpha_0 + \alpha_1)(\beta_0 + \beta_1) \leq (\gamma_0 + \gamma_1)(\delta_0 + \delta_1)$.

הוכחת הלמה היא אלמנטרית, והרי התקציר שלה: אם $\gamma_1 = 0$ או $\delta_0 = 0$ אז קל לוודא את טענת הלמה. אחרת מתקיים $(\gamma_0 + \gamma_1)(\delta_0 + \delta_1) \geq (\frac{\alpha_0\beta_0}{\delta_0} + \gamma_1)(\delta_0 + \frac{\alpha_1\beta_1}{\gamma_1})$ (כי $\gamma_0 \geq \frac{\alpha_0\beta_0}{\delta_0}$ ו־$\delta_1 \geq \frac{\alpha_1\beta_1}{\gamma_1}$), ולכן להשלמת ההוכחה של הלמה נותר להוכיח שמתקיים $(\alpha_0+\alpha_1)(\beta_0+\beta_1)\gamma_1\delta_0 \leq (\alpha_0\beta_0+\gamma_1\delta_0)(\gamma_1\delta_0+\alpha_1\beta_1)$; מהעברת אגפים זה שקול ל־$0 \leq \gamma_1\delta_0\gamma_1\delta_0 + \alpha_0\beta_0\alpha_1\beta_1 - \alpha_0\beta_1\gamma_1\delta_0 - \alpha_1\beta_0\gamma_1\delta_0 = (\gamma_1\delta_0 - \alpha_0\beta_1)(\gamma_1\delta_0 - \alpha_1\beta_0)$, ואת הטענה האחרונה ניתן שוב לוודא ישירות מההנחות.

עתה נוכיח את המשפט: אם $|S| = 0$ אז $\mathcal{P}(S) = \{\emptyset\}$, והטענה היא מיידית. אחרת, נבחר איבר $s \in S$ שרירותית, ונגדיר את $S' = S \setminus \{s\}$. לכל φ נגדיר את $\varphi' : \mathcal{P}(S') \to \mathbb{R}^+$ לפי $\varphi'(A) = \varphi(A) + \varphi(A \cup \{s\})$. לכל φ מתקיים

$$\overline{\varphi'}(\mathcal{P}(S')) = \sum_{A \subseteq S'} \varphi'(A) = \sum_{A \subseteq S \setminus \{s\}} \varphi(A) + \sum_{A \subseteq S \setminus \{s\}} \varphi(A \cup \{s\}) = \overline{\varphi}(\mathcal{P}(S))$$

ולכן מספיק להוכיח עתה שמתקיים $\overline{\alpha'}(\mathcal{P}(S'))\overline{\beta'}(\mathcal{P}(S')) \leq \overline{\gamma'}(\mathcal{P}(S'))\overline{\delta'}(\mathcal{P}(S'))$ כדי להוכיח את מסקנת המשפט עבור S. לשם כך ניתן להשתמש בהנחת האינדוקציה, בתנאי שמראים שלכל $A, B \in \mathcal{P}(S')$ מתקיים $\alpha'(A)\beta'(B) \leq \gamma'(A \cup B)\delta'(A \cap B)$. טענה אחרונה זו נובעת מהלמה הקודמת, כאשר מגדירים

$$\begin{array}{llll} \alpha_0 = \alpha(A) & \beta_0 = \beta(B) & \gamma_0 = \gamma(A \cup B) & \delta_0 = \delta(A \cap B) \\ \alpha_1 = \alpha(A \cup \{s\}) & \beta_1 = \beta(B \cup \{s\}) & \gamma_1 = \gamma(A \cup B \cup \{s\}) & \delta_1 = \delta((A \cap B) \cup \{s\}) \end{array}$$

הישום עבור מאורעות מקריים

ננסח עתה את התוצאה של קלייטמן: אם $\mathcal{A},\mathcal{B}$ הן משפחות מונוטוניות עולות (ראו את ההגדרה למעלה) של ת"ק של S, ו־$\mathcal{C},\mathcal{D}$ הן משפחות מונוטוניות יורדות של ת"ק של S (ז"א למשל שאם $C' \subset C \in \mathcal{C}$ אז $C' \in \mathcal{C}$), אז מתקיימים עבור אלו $2^{|S|}|\mathcal{A}\cap\mathcal{B}| \geq |\mathcal{A}|\cdot|\mathcal{B}|$, $2^{|S|}|\mathcal{C}\cap\mathcal{D}| \geq |\mathcal{C}|\cdot|\mathcal{D}|$, ו־$2^{|S|}|\mathcal{A}\cap\mathcal{C}| \leq |\mathcal{A}|\cdot|\mathcal{C}|$. הוכחה: עבור הטענה הראשונה, משתמשים במשפט ארבע הפונקציות, עבור $\alpha=\beta=\gamma=\delta=1$, לקבלת

$$|\mathcal{A}||\mathcal{B}| = \overline{\alpha}(\mathcal{A})\overline{\beta}(\mathcal{B}) \leq \overline{\gamma}(\mathcal{A}\sqcup\mathcal{B})\overline{\delta}(\mathcal{A}\sqcap\mathcal{B}) = |\mathcal{A}\sqcup\mathcal{B}||\mathcal{A}\sqcap\mathcal{B}| = |\mathcal{A}\cap\mathcal{B}||\mathcal{A}\sqcap\mathcal{B}| \leq 2^{|S|}|\mathcal{A}\cap\mathcal{B}|$$

עבור הטענה השלישית (את השניה לא קשה להשלים) נשתמש בטענה הראשונה עבור $\mathcal{A}$ ו־$\hat{\mathcal{C}} = \mathcal{P}(S)\setminus\mathcal{C}$ לקבלת

$$|\mathcal{A}|(2^{|S|}-|\mathcal{C}|) = |\mathcal{A}||\hat{\mathcal{C}}| \leq 2^{|S|}|\mathcal{A}\cap\hat{\mathcal{C}}| = 2^{|S|}(|\mathcal{A}|-|\mathcal{A}\cap\mathcal{C}|)$$

כאשר העברת אגפים תשלים את ההוכחה.

לא קשה לראות עתה למשל שעבור $G(n,\frac{1}{2})$ ועבור תכונות מונוטוניות עולות P,Q של גרפים, כאשר מזהים אותן עם המאורעות המתאימים, מתקיים $\Pr[P\wedge Q] \geq \Pr[P]\Pr[Q]$, ע"י כך שנגדיר את $\mathcal{A}$ להיות קבוצת הגרפים המקיימים את P, ואת $\mathcal{B}$ להיות קבוצת הגרפים המקיימים את Q. כאן S היא קבוצת הזוגות של איברים מ־$\{1,\ldots,n\}$, הגרפים מזוהים עם ת"ק של S המתאימות לקבוצות הקשתות שלהם, וההסתברות לקיום תכונה מסויימת היא מספר הגרפים בעלי n צמתים המקיימים את התכונה מחולק ל־$2^{|S|} = 2^{\binom{n}{2}}$.

על מנת להכליל לגרפים מקריים מהצורה $G(n,p)$ ואחרים, ננסח את המשפט של Fortuin, Kasteley, Ginibre (הנקרא בקיצור משפט FKG). פונקציה $\mu:\mathcal{P}(S)\to\mathbb{R}^+$ תיקרא לוג־סופר־מודולרית אם לכל A,B מתקיים $\mu(A)\mu(B) \leq \mu(A\cup B)\mu(A\cap B)$. דוגמה לפונקציה כזו היא הסיכוי לקבלת גרף מסויים כאשר כל זוג צמתים u,v נבחר להיות קשת באופן ב"ת בהסתברות p (כאשר מזהים כל גרף עם קבוצת הקשתות שלו); אז הפונקציה המתאימה $\mu(E) = p^{|E|}(1-p)^{\binom{n}{2}-|E|}$ היא לוג־סופר־מודולרית, ואף מתקיים עבורה שוויון

$$\mu(E)\mu(F) = p^{|E|+|F|}(1-p)^{2\binom{n}{2}-|E|-|F|} = p^{|E\cup F|+|E\cap F|}(1-p)^{2\binom{n}{2}-|E\cup F|-|E\cap F|} = \mu(E\cup F)\mu(E\cap F)$$

כך שזוהי בעצם פונקציה לוג־מודולרית.

משפט FKG קובע שלכל פונקציה לוג־סופר־מודולרית $\mu:\mathcal{P}(S)\to\mathbb{R}^+$ ולכל שתי פונקציות מונוטוניות לא יורדות $f,g:\mathcal{P}(S)\to\mathbb{R}^+$ מתקיים

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

דוגמת ישום אחת הטובה לענינינו היא כאשר f,g הן הפונקציות האופייניות לתכונות מונוטוניות של הגרף; אז ממשפט FKG תנבע קורלציה חיובית בין שני המאורעות המתאימים עבור $G(n,p)$. אפשר גם להציב בפונקציות אלו אינווריאנטים מונוטונים של הגרף, כגון מספר הצביעה $\chi(G)$ וגודל הקליק המקסימלי $\omega(G)$, ולקבל אי שוויון על התוחלות המתאימות: $\mathrm{E}[\omega(G)]\cdot\mathrm{E}[\chi(G)] \leq \mathrm{E}[\omega(G)\cdot\chi(G)]$. הוכחת משפט FKG ממשפט ארבע הפונקציות אינה קשה: כאן מציבים $\mathcal{A}=\mathcal{B}=\mathcal{P}(S)$, הפעם כאשר $\alpha(A)=\mu(A)f(A)$, $\beta(A)=\mu(A)g(A)$, $\gamma(A)=\mu(A)f(A)g(A)$ ו־$\delta(A)=\mu(A)$. הוכחת התנאי למשפט ארבע הפונקציות נעשית כך:

$$\begin{aligned}\alpha(A)\beta(B) &= \mu(A)\mu(B)f(A)g(B) \leq \mu(A\cup B)\mu(A\cap B)f(A)g(B)\\ &\leq \mu(A\cup B)\mu(A\cap B)f(A\cup B)g(A\cup B) = \gamma(A\cup B)\delta(A\cap B)\end{aligned}$$

אנטרופיה

מבוא והגדרות בסיסיות

במקור מושג האנטרופיה, מידה לאקראיות, משמש בתרמודינמיקה. השימוש במושג האנטרופיה עבור מדעי המחשב (בתחילה בעיקר תורת המידע) נוסד ע"י Shannon.

עבור מרחב הסתברות בדיד μ, נרצה להגדיר מידה שתגיד לנו עד כמה ההתפלגות היא "אקראית", או במילים אחרות, כמה "מטבעות" דרושים בממוצע על מנת להגריל איבר מאותו מרחב. מידה זו שימושית מאוד בניתוח סיבוכיות תקשורת ובתחומים אחרים. למשל, נראה בהמשך שניתן לכתוב בצורה מתמטית טענה שימושית האומרת שתהליך דטרמיניסטי המבוצע על קלט מקרי לא יכול להוסיף אקראיות מעבר לזו שהיתה במקור.

ראשית נראה את ההגדרות ואת ה"אלגברה" (סאב־אדיטיביות וכו') של המידה הזו, כולל טענות שלא נשתמש בהן בדוגמאות אבל עשויות להועיל לכם בעתיד. עבור מרחב הסתברות μ מעל קבוצת הבסיס S, נגדיר את האנטרופיה לפי הנוסחה הבאה:

$$\mathrm{H}[\mu] = \sum_{\{s:\mu(s)>0\}} \mu(s) \log \frac{1}{\mu(s)} = \mathrm{E}_{s\sim\mu}\left[\log \frac{1}{\mu(s)}\right]$$

זהו תמיד מספר אי־שלילי, ושווה ל־0 אם ורק אם המרחב שלנו הוא בעל איבר יחיד בהסתברות 1. בדרך כלל נרצה להשוות מידה זו עבור מספר משתנים מקריים מעל אותו מרחב. בהינתן משתנה מקרי נגדיר את האנטרופיה לפי המרחב הנגזר מעל התוצאות האפשריות של המשתנה:

$$\mathrm{H}[X] = \sum_{\{\alpha:\Pr[X=\alpha]>0\}} \Pr[X=\alpha] \log \frac{1}{\Pr[X=\alpha]}$$

כל הלוגריתמים כאן הם בבסיס 2, כי נרצה למדוד את האקראיות במושגים של "מטבעות". למשל, אם X מתפלג יוניפורמית מעל $\{0,1\}$ אז אכן $\mathrm{H}[X]=1$, ואם X מתפלג יוניפורמית מעל קבוצה מגודל 2^k אז $\mathrm{H}[X]=k$. נהוג גם להגדיר במיוחד פונקציה $H:[0,1]\to[0,1]$ אשר מקבלת את ערך האנטרופיה של מ"מ שיקבל 1 בהסתברות p ו־0 בהסתברות $1-p$, הווה אומר $H(p)=p\log\frac{1}{p}+(1-p)\log\frac{1}{1-p}$ כאשר $H(0)=H(1)=0$.

עבור שני משתנים מקריים X ו־Y מגדירים את $\mathrm{H}[X,Y]$ באופן טבעי, כאנטרופיה על המרחב הנגזר מעל זוגות הערכים המתאימים, ובמפורש $\mathrm{H}[X,Y]=\sum_{\{\alpha,\beta:\Pr[X=\alpha\wedge Y=\beta]>0\}}\Pr[X=\alpha\wedge Y=\beta]\log\frac{1}{\Pr[X=\alpha\wedge Y=\beta]}$. לא קשה לראות שמתקיים $\mathrm{H}[X,Y]=\mathrm{H}[Y,X]$.

אנטרופיה מותנה במאורע תסומן בסימון $\mathrm{H}[X|A]$. למשל, אם המאורע הוא "$Y=\alpha$" (כאשר Y הוא מ"מ אחר מעל אותו מרחב הסתברות, שיכול להיות תלוי ב־X), אז נשתמש בסימון $\mathrm{H}[X|Y=\alpha]$. נגדיר אנטרופיה מותנה של X במ"מ Y לפי הנוסחה $\mathrm{H}[X|Y]=\mathrm{E}_{Y'\sim Y}[\mathrm{H}[X|Y=Y']]$. הסבר לסימון: נגדיר מ"מ חדש Y', שמתפלג בדיוק כמו ההתפלגות הלא־מותנה של Y, אבל אינו תלוי כלל במרחב ההסתברות שלפיו הגרלנו את X ו־Y (בעצם הרחבנו כאן את מרחב ההסתברות שלנו). ההתפלגות המותנה היא התוחלת של האנטרופיה של X על כך ש־Y קיבל ערכים מסויימים, לפי ההתפלגות הנ"ל. נוסחת הסכום המתקבלת:

$$\mathrm{H}[X|Y] = \sum_{\{\beta:\Pr[Y=\beta]>0\}} \Pr[Y=\beta]\mathrm{H}[X|Y=\beta]$$

כלל חשוב עבור האנטרופיה המותנה הוא כלל השרשרת: $\mathrm{H}[X|Y]=\mathrm{H}[X,Y]-\mathrm{H}[Y]$. נוכיח אותו.

$$
\begin{aligned}
\mathrm{H}[X,Y]-\mathrm{H}[Y] &= \\
&= \sum_{\{\alpha,\beta:\Pr[X=\alpha\wedge Y=\beta]>0\}} \Pr[X=\alpha\wedge Y=\beta]\log\frac{1}{\Pr[X=\alpha\wedge Y=\beta]} \\
&\qquad - \sum_{\{\beta:\Pr[Y=\beta]>0\}} \Pr[Y=\beta]\log\frac{1}{\Pr[Y=\beta]} \\
&= \sum_{\{\alpha,\beta:\Pr[X=\alpha\wedge Y=\beta]>0\}} \Pr[X=\alpha\wedge Y=\beta]\left(\log\frac{1}{\Pr[X=\alpha\wedge Y=\beta]}-\log\frac{1}{\Pr[Y=\beta]}\right) \\
&= \sum_{\{\alpha,\beta:\Pr[X=\alpha\wedge Y=\beta]>0\}} \Pr[X=\alpha\wedge Y=\beta]\log\frac{\Pr[Y=\beta]}{\Pr[X=\alpha\wedge Y=\beta]} \\
&= \sum_{\{\beta:\Pr[Y=\beta]>0\}} \Pr[Y=\beta]\left(\sum_{\{\alpha:\Pr[X=\alpha|Y=\beta]>0\}} \Pr[X=\alpha|Y=\beta]\log\frac{1}{\Pr[X=\alpha|Y=\beta]}\right) \\
&= \quad \mathrm{H}[X|Y]
\end{aligned}
$$

תוצאה חשובה של שוויון זה הוא שתמיד מתקיים $\mathrm{H}[X,Y]=\mathrm{H}[Y,X]\geq\mathrm{H}[X]$ (כי האנטרופיה המותנה $\mathrm{H}[Y|X]$ היא אי־שלילית מהגדרתה), תכונה הידועה כמונוטוניות של האנטרופיה. כמו כן אפשר לראות מכלל השרשרת שמתקיים $\mathrm{H}[X,Y]=\mathrm{H}[X]$ אם ורק אם Y הוא פונקציה של X, מכיוון שרק למשתנים קבועים יש אנטרופיה 0, וכאן זה גורם לכך ש־Y יהיה קבוע עבור כל ערך קבוע של X, ז"א שהוא נקבע על ידו.

נגדיר גם את המידע המשותף לשני משתנים, $\mathrm{I}[X,Y]=\mathrm{H}[X]+\mathrm{H}[Y]-\mathrm{H}[X,Y]=\mathrm{H}[X]-\mathrm{H}[X|Y]$. על מנת לקבל אינטואיציה מה זה אומר, נחשוב על הדוגמה הבאה: נניח ש־$X_1,\ldots,X_k,Y_1,\ldots,Y_l,Z_1,\ldots,Z_m$ משתנים בינארים יוניפורמים וב"ת לחלוטין. עתה נניח ש־X הוא פונקציה חח"ע של $X_1,\ldots,X_k,Z_1,\ldots,Z_m$, ו־$Y$ הוא פונקציה חח"ע של $Y_1,\ldots,Y_l,Z_1,\ldots,Z_m$. במקרה כזה ניתן לחשב ולראות ש־$\mathrm{I}[X,Y]=m$, כצפוי. מייד נראה שגם ערך זה לעולם אינו שלילי.

אי־שוויון ינסן ואנטרופיה יחסית (פיצוליות Kullback-Leibler divergence)

אי־שוויון ינסן (Jensen) הוא אי־שוויון מאוד שימושי באנליזה ובניתוח הסתברותי, ועם זאת מאוד נוח לשימוש. ניסוחו ההסתברותי: אם X הוא מ"מ מעל מרחב הסתברות כל שהוא ו־$f:\mathbb{R}\to\mathbb{R}$ היא פונקציה קמורה, אז מתקיים $\mathrm{E}[f(X)]\geq f(\mathrm{E}[X])$. לעומת זאת, אם f היא פונציה קעורה, אז מתקיים $\mathrm{E}[f(X)]\leq f(\mathrm{E}[X])$. יתרה מזו, שוויון אפשרי אך ורק אם X הוא משתנה שמקבל ערך יחיד בהסברות 1, או f היא פונקציה לינארית של X בתחום הערכים הרלוונטי עבור המשתנה (עבור מרחב הסתברות בדיד זה יהיה הקטע בין מינימום ומקסימום הערכים שהמשתנה יכול לקבל בהסתברות חיובית).

נראה עתה שמרחב ההסתברות בעל האנטרופיה המירבית מעל הבסיס S הוא זה בעל ההתפלגות היוניפורמית. במילים אחרות, $\mathrm{H}[\mu]\leq\log|S|$. לשם כך נציב $X(s)=1/\mu(s)$, ומאי־שוויון ינסן עבור הפונקציה הקעורה $f(x)=\log(x)$ נקבל $\mathrm{H}[\mu]=\mathrm{E}_{s\sim\mu}[\log(X)]\leq\log(\mathrm{E}[X])=\log(|\{s:\mu(S)>0\}|)\leq\log|S|$. ההתפלגות היוניפורמית היא הדרך היחידה לקבל שוויון, כי זהו המקרה היחיד שבו גם X הוא קבוע וגם כל איברי S יכולים להתקבל בהסתברות חיובית.

עבור שני מרחבי הסתברות μ ו־ν מעל אותה קבוצת בסיס S, נגדיר פיצוליות Kullback-Leibler divergence (בספרות לפעמים משתמשים בשם המקוצר KL-divergence) לפי הנוסחה הבאה:

$$D(\mu\|\nu)=\sum_{\{s:\mu(s)>0\}}\mu(s)\log\frac{\mu(s)}{\nu(s)}=\mathrm{E}_{s\sim\mu}\left[\log\frac{\mu(s)}{\nu(s)}\right]$$

המידה הזו נקראת גם "אנטרופיה יחסית". השתמשנו בשם "פיצוליות" כי מידה זו אינה מרחק במובן הרגיל של המילה. בפרט היא אינה חילופית, לא תמיד מקיימת את אי שוויון המשולש, ויכולה להיות שווה ל־$+\infty$ כאשר קיים s המתקבל בהסתברות חיובית עבור μ בלבד. מסתבר אבל שמידה זו לעולם אינה שלילית.

עבור ההוכחה, נניח ש־$\mu(s)=0$ כל אימת ש־$\nu(s)=0$ (אחרת ממילא $D(\mu\|\nu)=+\infty>0$), נגדיר מ"מ Z המקבל את $\mu(s)/\nu(s)$ (לא משנה מה הוא מקבל על s עבורם $\mu(s)=\nu(s)=0$ כי בהמשך נתייחס לתוחלת שלו), ונגדיר פונקציה $f:[0,+\infty)\to[0,+\infty)$ לפי $f(z)=z\log(z)$, כאשר מגדירים $f(0)=0$ לפי הגבול מימין בנקודה זו. זוהי פונקציה קמורה בתחום ההגדרה שלה, לפי לקיחת נגזרת שניה. עתה נפתח:

$$
\begin{aligned}
D(\mu\|\nu) &= \sum_{\{s:\mu(s)>0\}} \mu(s)\log\frac{\mu(s)}{\nu(s)} \\
&= \sum_{\{s:\mu(s)>0\}} \nu(s)\frac{\mu(s)}{\nu(s)}\log\frac{\mu(s)}{\nu(s)} \\
&= \sum_{\{s:\mu(s)>0\}} \nu(s) f\left(\frac{\mu(s)}{\nu(s)}\right) \\
&= \sum_{\{s:\nu(s)>0\}} \nu(s) f\left(\frac{\mu(s)}{\nu(s)}\right) \\
&= \mathrm{E}_\nu[f(Z)] \\
&\geq f(\mathrm{E}_\nu[Z]) = f(1) = 0
\end{aligned}
$$

במעבר בסכימה מהאיברים עם $\mu(s)>0$ לאלו עם $\nu(s)>0$ השתמשנו בהנחה מלמעלה על ההתאפסויות של ההסתברויות, וכן ב־$f(0)=0$ (כך שרק הוספנו איברי 0 לסכום). אי השוויון בסוף הוא אי שוויון ינסן, ואת התוחלת של Z קל לוודא. כמו כן, כפי שציינו למעלה, יהיה שוויון רק אם Z הוא קבוע (מכיוון ש־f לא לינארית בשום מקום), ז"א כאשר $\nu=\mu$.

לבסוף, נציין שנהוג גם להגדיר את $D(X\|Y)$ עבור שני משתנים מקריים מתאימים. במקרה כזה לא מתייחסים לשאלה האם הם תלויים או לא, אלא רק לווקטורי ההתפלגות המתאימים. הנוסחה המתאימה היא פשוט $D(X\|Y)=\sum_{\{\alpha:\Pr[x=\alpha]>0\}}\Pr[X=\alpha]\log\frac{\Pr[X=\alpha]}{\Pr[Y=\alpha]}$.

מספר אי שוויונים מועילים

אי־שוויון חשוב אחד שראינו הוא זה שקובע כי מרחב ההסתברות הנותן את מירב האנטרופיה מעל קבוצת בסיס סופית הוא זה בעל ההתפלגות היוניפורמית. בהתאמה לכך, אם קבוצת הערכים האפשריים של מ"מ X היא מגודל k אז מתקיים $\mathrm{H}[X]\leq\log(k)$.

תכונה בסיסית של צירוף משתנים היא המונוטוניות שראינו מכלל השרשרת: $\mathrm{H}[X,Y]\geq\mathrm{H}[Y]$. כפי שהזכרנו שם, שוויון מתקיים אם ורק אם X היא פונקציה של Y (ז"א שלכל ערך אפשרי של Y יהיה ערך יחיד בהסתברות מותנה 1 של X). מכך נובע בפרט שאם X היא פונקציה של Y אז $\mathrm{H}[X]\leq\mathrm{H}[X,Y]=\mathrm{H}[Y]$, אי שוויון המתקשר לאינטואיציה שתהליך דטרמיניסטי (המעבר מ־Y ל־X) לא יכול להוסיף אקראיות.

התכונה החשובה השניה של צירוף משתנים היא סאב־אדיטיביות, $\mathrm{H}[X,Y]\leq\mathrm{H}[X]+\mathrm{H}[Y]$. זה נובע מהטענה שנראה עתה, ש־$\mathrm{I}[X,Y]$ לעולם אינו שלילי. לשם כך נגדיר שני מרחבי הסתברות על זוגות של ערכים. המרחב הראשון, שנסמנו μ, יהיה פשוט התוצאה המתקבלת מלקיחת ערכי שני המשתנים ברצף: $\Pr_\mu[(\alpha,\beta)]=\Pr[X=\alpha\wedge Y=\beta]$. המרחב השני, ν, יוגדר לפי "מה היה קורה אילו שני המשתנים היו בלתי תלויים": $\Pr_\nu[(\alpha,\beta)]=\Pr[X=\alpha]\Pr[Y=\beta]$. עתה נפתח ביטוי אלטרנטיבי ל־$\mathrm{I}[X,Y]$:

$$
\begin{aligned}
\mathrm{I}[X,Y] &= \sum_{\{\alpha:\Pr[X=\alpha]>0\}} \Pr[X=\alpha]\log\frac{1}{\Pr[X=\alpha]} + \sum_{\{\beta:\Pr[Y=\beta]>0\}} \Pr[Y=\beta]\log\frac{1}{\Pr[Y=\beta]} \\
&\quad - \sum_{\{\alpha,\beta:\Pr[X=\alpha\wedge Y=\beta]>0\}} \Pr[X=\alpha\wedge Y=\beta]\log\frac{1}{\Pr[X=\alpha\wedge Y=\beta]} \\
&= \sum_{\{\alpha,\beta:\Pr[X=\alpha\wedge Y=\beta]>0\}} \Pr[X=\alpha\wedge Y=\beta]\log\frac{\Pr[X=\alpha\wedge Y=\beta]}{\Pr[X=\alpha]\Pr[Y=\beta]} \\
&= D(\mu\|\nu)
\end{aligned}
$$

עתה אפשר לסיים, מכיוון שכבר ראינו שפיצוליות KL תמיד מקיימת $D(\mu\|\nu)\geq 0$. יתרה מזו, כזכור שוויון אפשרי רק אם ν ו־μ זהים, ז"א שמתקיים $\mathrm{H}[X,Y]=\mathrm{H}[X]+\mathrm{H}[Y]$ אם ורק אם X ו־Y הם בלתי תלויים.

טענה שקולה לסאב־אדיטיביות (מהצבת כלל השרשרת עבור אנטרופיה מותנה) היא שעבור כל זוג משתנים מתקיים $\mathrm{H}[X|Y]=\mathrm{H}[X,Y]-\mathrm{H}[Y]\leq\mathrm{H}[X]$. אינטואיטיבית המשמעות כאן היא שגילוי חלק מהמידע (המשתנה Y) בממוצע רק יכול לגרוע מהאקראיות הנותרת. עם זאת, יכולים להיות ערכים ספציפים של Y עבורם כן יתקיים $\mathrm{H}[X|Y=\beta]>\mathrm{H}[X]$.

אי השוויון האחרון שנוכיח כאן הוא שלכל שלושה מ"מ מתקיים $\mathrm{H}[X|Y,Z]\leq\mathrm{H}[X|Y]$. ניתן להוכיח את זה ישירות מהסכומים, אולם אנו נגזור את זה מאי השוויון הקודם. שימו לב שפירוש הסימון "$\mathrm{H}[X|Y=\beta,Z]$" המופיע בפיתוח הוא "האנטרופיה המותנה של X על Z, כאשר אנו נמצאים במרחב ההסתברות המותנה על המאורע $Y=\beta$". אנו למעשה משתמשים באי השוויון מקודם מעל אותו מרחב הסתברות מותנה.

$$
\begin{aligned}
\mathrm{H}[X|Y,Z] &= \sum_{\{\beta,\gamma:\Pr[Y=\beta\wedge Z=\gamma]>0\}} \Pr[Y=\beta\wedge Z=\gamma]\mathrm{H}[X|Y=\beta\wedge Z=\gamma] \\
&= \sum_{\{\beta:\Pr[Y=\beta]>0\}} \Pr[Y=\beta] \sum_{\{\gamma:\Pr[Z=\gamma|Y=\beta]>0\}} \Pr[Z=\gamma|Y=\beta]\mathrm{H}[X|Y=\beta\wedge Z=\gamma] \\
&= \sum_{\{\beta:\Pr[Y=\beta]>0\}} \Pr[Y=\beta]\mathrm{H}[X|Y=\beta,Z] \\
&\leq \sum_{\{\beta:\Pr[Y=\beta]>0\}} \Pr[Y=\beta]\mathrm{H}[X|Y=\beta] \\
&= \mathrm{H}[X|Y]
\end{aligned}
$$

דוגמאות ישום

ראשית נראה חסם על גודל משפחה של קבוצות עם מספר מופעים נתון של כל איבר. נניח ש־$\mathcal{F}$ היא משפחה של תתי־קבוצה של $\{1,\ldots,n\}$, ובנוסף לכך נניח שלכל $1\leq i\leq n$ קיימים בדיוק $|\mathcal{F}|p_i$ איברים של $\mathcal{F}$ המכילים את i. הטענה קובעת כי במקרה כזה $|\mathcal{F}|\leq 2^{\sum_{i=1}^n H(p_i)}$ (במעריך יש את הפונקציות הממשיות מעל $[0,1]$ שהוגדרו קודם).

לשם כך נגדיר מרחב הסתברות μ המורכב מבחירה אקראית ויוניפורמית של $F\in\mathcal{F}$, ולכל i נגדיר את X_i להיות משתנה האינדיקטור עבור המאורע "$i\in F$", אשר בפרט יקבל 1 בהסתברות (לא מותנה) p_i ו־0 בהסתברות $1-p_i$. מכיוון שצירוף כל ה־X_i קובע לחלוטין את F, מתקיים $\mathrm{H}[X_1,\ldots,X_n]=\mathrm{H}[\mu]=\log|\mathcal{F}|$.

מצד שני, לפי תת האדיטיביות שהוכחנו למעלה (בתוספת אינדוקציה עבור מספר המשתנים המשתתפים) מתקיים $\mathrm{H}[X_1,\ldots,X_n] \leq \sum_{i=1}^n \mathrm{H}[X_i] = \sum_{i=1}^n H(p_i)$, ובהעברת אגפים נקבל את המבוקש.

לפני הדוגמה הבאה נראה קירוב די טוב לבינום: לכל $0 \leq k \leq n$ מתקיים $\frac{1}{n+1}2^{nH(k/n)} \leq \binom{n}{k} \leq 2^{nH(k/n)}$. החסם העליון ניתן להסקה מיידית מהטענה הקודמת, אולם להוכחת שני החסמים יחדיו נשתמש בשיטות אלמנטריות. שני מקרי הקצה ($k = 0$ או $k = n$) הם טריביאלים. עבור המקרים האחרים נסמן $q = \frac{k}{n}$ ונבדוק את הפיתוח $1 = (q + (1-q))^n = \sum_{i=0}^n \binom{n}{i} q^i (1-q)^{n-i}$.

הסכום בצד שמאל הוא כולו של איברים חיוביים, ואחד מהם הוא $\binom{n}{k} q^k (1-q)^{n-k} = \binom{n}{k} 2^{-nH(k/n)}$, ומכאן (שוב) החסם העליון. על מנת לוודא את חסם התחתון, נסמן $a_i = \binom{n}{i} q^i (1-q)^{n-i}$, ונראה ש־$a_k$ הוא הגדול ביותר מבין $n+1$ האיברים הנ"ל. לשם כך נבדוק את המנה של איברים עוקבים:

$$a_{i+1}/a_i = \frac{q}{1-q} \cdot \binom{n}{i+1} / \binom{n}{i} = \frac{n-i}{i+1} \frac{q}{1-q}$$

ההפרש הזה יהיה חיובי אם ורק אם $\frac{n-i}{i+1}\frac{q}{1-q} > 1$, או בהעברת אגפים $i < k + q - 1$, וזה יקרה בדיוק כל עוד $i < k$, (כי i ו־k הם מספרים שלמים ו־q הוא בין 0 ל־1) מה שגורם לכך ש־a_k הוא האיבר הגבוה ביותר.

עתה נחסום את מספר הביטים הדרושים לקוד תיקון שגיאות. קוד מאורך n להודעות מאורך $k \leq n$ נתון ע"י פונקצית קידוד $f : \{0,1\}^k \to \{0,1\}^n$ ופונקצית קריאה $g : \{0,1\}^n \to \{0,1\}^k$. תנאי תקינות מינימלי הוא שלכל $x \in \{0,1\}^k$ יתקיים $g(f(x)) = x$, אבל אנו נרצה אפשרות לתיקון של עד qn שגיאות (כאשר $q < \frac{1}{2}$ הוא קבוע נתון): אם y ו־$f(x)$ נבדלים בלא יותר מ־qn ביטים, אז גם $g(y) = x$. אנו נראה חסם לקיומו של קוד כזה: $k \leq (1 - H(q) + o(1))n$. ההוכחה שנראה היא למעשה ספירה פשוטה, אבל הכתיבה שלה במושגים של אנטרופיה נותנת אינטואיציה היכולה לשמש גם במקרים מסובכים יותר.

נניח ש־f ו־g הן זוג פונקציות מתאימות. נגריל את x באופן יוניפורמי מ־$\{0,1\}^k$, ועבור y נבחר באופן מקרי, יוניפורמי וב"ת ב־x קבוצה A בת qn איברים בדיוק מתוך $\{1,\ldots n\}$, ונהפוך ב־$f(x)$ את הקורדינטות המתאימות (הטיעון עובד עם עיגול למטה אם qn אינו מספר שלם), ז"א $y = f(x) \oplus 1_A$. נסמן את האנטרופיה של ההתפלגות של y ב־$\mathrm{H}[y]$.

אם זהו אכן קוד תיקון שגיאות, אז $g(y)$ ו־$y \oplus f(g(y))$ מתפלגים באופן ב"ת זה בזה: הראשון זהה ל־x והאנטרופיה שלו היא k, בעוד שהשני זהה ל־1_A, ולכן הוא ב"ת והאנטרופיה שלו (לפי החסמים על הבינום מקודם) היא לפחות $\log(\frac{1}{n+1}2^{nH(q)}) = (H(q) - o(1))n$. מכיוון ששני הערכים האלו הם פונקציות של y, נובע שהאנטרופיה של התפלגות y היא לפחות סכום האנטרופיות של המשתנים הב"ת האלו, ז"א $\mathrm{H}[y] \geq k + (H(q) - o(1))n$.

מצד שני, y הוא בעל 2^n ערכים אפשריים, ולכן האנטרופיה של ההתפלגות עליו אינה יכולה לעלות על n. קיבלנו $n \geq \mathrm{H}[y] \geq k + (H(q) - o(1))n$, ובהעברת אגפים מקבלים את המבוקש.

קצת על דחיסת נתונים

נניח שיש לנו התפלגות μ על קבוצת המחרוזות $\{0,1\}^n$. הרעיון בדחיסה הוא לנצל את "חוסר היוניפורמיות" של μ (אם היא קיימת), על מנת לייצג את המחרוזות שלנו ע"י מחרוזות שיהיו קצרות יותר בממוצע. באופן פורמלי נרצה פונקציה חח"ע $g : \{0,1\}^n \to \{0,1\}^*$ (כאשר $\{0,1\}^*$ היא קבוצת המחרוזות מכל אורך סופי מעל $\{0,1\}$), כך ש־$\mathrm{E}[|g(x)|]$ יהיה קטן ככל שניתן. כאן נראה חסם תחתון על תוחלת זו, שמתאים לאנטרופיה של μ. שימו לב שגם אורך $g(x)$ עצמו יכול לקודד מידע, ואכן נראה כאן דוגמה שבה נוכל "לחסוך" כ־$\log(n)$ ביטים (מספר הביטים הדרוש לכתיבת האורך) מתוך הקידוד עצמו. בתרגול תראו מקרה שבו אין מאפשרים לדלות מידע מאורך הפלט עצמו, זה של קודים חסרי רישות, ושם לא יהיה חיסכון כזה.

כאן נניח שתמיד מתקיים $|g(x)| \leq n$, כי כל עוד ניתן לדעת את אורך הפלט, תמיד ניתן להחליף קידוד של x שגודלו הוא לפחות n ב"קידוד" המחזיר את x עצמו. אם נזכור ש־g חייבת להיות חח"ע ניתן לכתוב את

החסם הבא:

$$\mathrm{H}[\mu]-\log(n+1)\le \mathrm{H}[\mu]-\mathrm{H}[|g(x)|]=\mathrm{H}[g(x)]-\mathrm{H}[|g(x)|]=\mathrm{H}[g(x)\mid |g(x)|]\le \mathrm{E}[|g(x)|]$$

אי השוויון הימני נובע מהדבר הבא: אם Y ו־D הם משתנים מקריים, כאשר D מקבל ערכים של מספרים טבעיים, ולכל k מתקיים $|\{\alpha:\Pr[Y=\alpha|D=k]>0\}|\le k$ (ז"א ש־D למעשה קובע חסם על מספר הערכים האפשרים של Y בהינתן ערך נתון שלו), אז מתקיים $\mathrm{H}[Y|D]\le \mathrm{E}[\log(D)]$. החסם הנ"ל נובע מיידית מכתיבת האנטרופיה המותנה כתוחלת של אנטרופיות, יחד עם החסם הכללי על אנטרופיה של מ"מ עם טווח נתון. במקרה שלנו נסמן $Y=g(x)$ ו־$D=2^{|g(x)|}$. מהפיתוח למעלה קיבלנו את חסם האנטרופיה (בניכוי הקידוד של אורך הפלט) עבור האורך הממוצע של $g(x)$.

לבסוף נראה מקרה שבו אכן אפשר לקבל הפרש קרוב ל־$\log(n+1)$ בין אורך הקידוד לבין האנטרופיה של μ. נגדיר התפלגות מעל $\{0,1\}^{n+1}$ שבה לכל $0\le k\le n$ יש בדיוק 2^k מחרוזות שלכל אחת מהן הסתברות של $\frac{1}{n+1}2^{-k}$. שימו לב שבסה"כ אפיינו $2^{n+1}-1$ מחרוזות; למחרוזת הנותרת נקצה הסתברות 0 (אפשר למשל להגדיר התפלגות כזו כתוצאה של בחירה יוניפורמית של $0\le i\le n$, ואז לקיחת המילה שהיא שרשור של i אחדות שלאחריהם אפס ואחריו מחרוזת המוגרלת יוניפורמית מתוך $\{0,1\}^{n-i}$). בקידוד, נעביר את 2^k המחרוזות שההסתברות שלהן היא $\frac{1}{n+1}2^{-k}$ לקבוצת המחרוזות מאורך k בדיוק.

מצד אחד מתקבל ש־$|g(x)|$ מתפלג יוניפורמית מעל $\{0,\dots,n\}$ ולכן $\mathrm{E}[|g(x)|]=\frac{n}{2}$. מצד שני, אפשר לחשב את האנטרופיה של μ כפי שהוא נתון כאן:

$$\mathrm{H}[\mu]=\sum_{k=0}^{n}\frac{\log(2^k(n+1))}{n+1}=\sum_{k=0}^{n}\frac{k}{n+1}+\sum_{k=0}^{n}\frac{\log(n+1)}{n+1}=\frac{n}{2}+\log(n+1)$$

קיבלנו שהפרש בין תוחלת אורך הקידוד לאנטרופיה כאן הוא $\log(n+1)$, קרוב מאוד לחסם ההפרש מלמעלה, שבמקרה זה הוא $\log(n+2)$.

המשפט של ברגמן

עתה נראה ישום של שיטת האנטרופיה עבור הוכחה של משפט Brégman, כפי שנעשתה ע"י Radhakrishnan. המשפט עצמו נותן חסם על הפרמננט של מטריצה של אפסים ואחדות, או באופן שקול חסם על מספר הזיווגים המושלמים בגרף דו צדדי נתון מעל קבוצת הצמתים $U\cup V=\{u_1,\dots,u_n,v_1,\dots,v_n\}$. החסם קובע שאם d_i היא דרגת הצומת u_i לכל $1\le i\le n$, אז $|\mathcal{M}(G)|\le\prod_{i=1}^{n}(d_i!)^{1/d_i}$.

אנו נייצג זיווג מושלם $M\in\mathcal{M}$ ע"י פרמוטציה $\sigma_M\in S_n$, כך שלכל $1\le i\le n$ הקשת $u_iv_{\sigma_M(i)}$ נמצאת ב־M. נבחר זיווג מושלם $M\in\mathcal{M}$ באופן יוניפורמי, נסמן ב־X את המ"מ שמקבל את הפרמוטציה המתאימה σ_M, ועל מנת להוכיח את המשפט אנו נוכיח שמתקיים $\mathrm{H}[X]\le\sum_{i=1}^{n}\frac{1}{d_i}\log(d_i!)$.

נסמן ב־X_i את המ"מ שמקבל את $\sigma_M(i)$, ונסמן ב־$\tau\in S_n$ פרמוטציה שרירותית כל שהיא (לאו דווקא כזו שמייצגת זיווג מושלם). לפי כלל השרשרת (ואינדוקציה) מתקיים

$$\mathrm{H}[X]=\mathrm{H}[X_1,\dots,X_n]=\sum_{i=1}^{n}\mathrm{H}[X_{\tau(i)}|X_{\tau(1)},\dots,X_{\tau(i-1)}]$$

על מנת לחסום את $\mathrm{H}[X]$ אנו נחסום את ה"מיצוע" עבור τ שנבחר באופן יוניפורמי מתוך S_n, תוך שימוש בשוויון הנובע מלינאריות התוחלת. נשים לב שעכשיו עברנו למרחב הסתברות יותר "רחב", שבו המ"מ הם פונקציות מקבוצת הזוגות של זיווג M ופרמוטציה τ. בפרט צריך לפרש את $X_{\tau(1)},\dots,X_{\tau(n)}$ כסדרה של מ"מ מעל המרחב הזה. עבור אלו מתקיים

$$\mathrm{H}[X] = \mathrm{E}_\tau\left[\sum_{i=1}^{n}\mathrm{H}[X_{\tau(i)}|X_{\tau(1)},\ldots,X_{\tau(i-1)}]\right] = \sum_{i=1}^{n}\mathrm{E}_\tau\left[\mathrm{H}[X_{\tau(i)}|X_{\tau(1)},\ldots,X_{\tau(i-1)}]\right]$$

עתה נרצה לחסום את התוחלת של אותה אנטרופיה מותנה. נסמן ב־D_j את המשתנה המקרי שערכו נתון ע"י מספר השכנים של הצומת u_j שאינם מיוצגים בקבוצה $\{v_{\sigma_M(\tau(1))}, v_{\sigma_M(\tau(2))},\ldots,v_{\sigma_M(\tau(\tau^{-1}(j)-1))}\}$. במילים אחרות, אם מסתכלים על τ כקובע סדר על $u_1,\ldots,u_n$, ודרך בני הזוג שלהם לפי M קובעים לפיו סדר על $v_1,\ldots,v_n$, אז $D_{\tau(i)}$ הוא פונקציה של מיקומו של הצומת ש־M זיווג ל־$u_{\tau(i)}$ בסדר הזה יחסית לקבוצת כל השכנים של $u_{\tau(i)}$ בגרף.

המשתנה המקרי $D_{\tau(i)}$ תלוי ב־σ_M וב־τ, אולם למעשה הוא תלוי אך ורק בערך $\tau(i)$ ובסדרת הערכים $X_{\tau(1)},\ldots,X_{\tau(i-1)}$ (ובגרף הקבוע שלנו). כמו כן $D_{\tau(i)}$ חוסם את גודל הטווח (המותנה) של $X_{\tau(i)}$. כך נקבל עבור כל τ אפשרי (ובחירה מקרית של M):

$$\mathrm{H}[X_{\tau(i)}|X_{\tau(1)},\ldots,X_{\tau(i-1)}]=\mathrm{H}[X_{\tau(i)}|X_{\tau(1)},\ldots,X_{\tau(i-1)},D_{\tau(i)}] \le \mathrm{H}[X_{\tau(i)}|D_{\tau(i)}] \le \mathrm{E}_M[\log(D_{\tau(i)})]$$

אי־השוויון הימני נובע מכל שלכל ערך ספציפי k מתקיים $\mathrm{H}[X_{\tau(i)}|D_{\tau(i)}=k] \le \log(k)$ (לפי החסם על גודל הטווח של $X_{\tau(i)}$), וכאשר לוקחים תוחלת לפי M על שני הצדדים מתקבל אי־השוויון.

נחזור עתה לחישוב האנטרופיה המקורי (עם תוחלת עבור τ מקרי) ונקבל:

$$\mathrm{H}[X] = \sum_{i=1}^{n}\mathrm{E}_\tau\left[\mathrm{H}[X_{\tau(i)}|X_{\tau(1)},\ldots,X_{\tau(i-1)}]\right] \le \sum_{i=1}^{n}\mathrm{E}_{M,\tau}[\log(D_{\tau(i)})] = \sum_{j=1}^{n}\mathrm{E}_{M,\tau}[\log(D_j)]$$

גם כאן השוויון הכי ימני משתמש בכך שלכל M ו־τ ספציפיים מתקיים שוויון בערכי המ"מ המתאימים, $\sum_{i=1}^{n}\log(D_{\tau(i)}(M,\tau)) = \sum_{j=1}^{n}\log(D_j(M,\tau))$, ומכאן משתמשים בלינאריות התוחלת.

הדבר האחרון לשים אליו לב הוא שכל D_j למעשה מתפלג יוניפורמית מעל $\{1,\ldots,d_j\}$, אפילו אם מתנים אותו על ערך מסויים של M. זאת מכיוון שכזכור המ"מ מחזיר את המיקום של $v_{\sigma_M(j)}$ בתוך שכני u_j בסדר המתאים לפרמוטציה המקרית ש־τ משרה דרך הזיווג (ומתפלגת יוניפורמית). בזאת ניתן לסיים:

$$\mathrm{H}[X] \le \sum_{j=1}^{n}\mathrm{E}[\log(D_j)] = \sum_{j=1}^{n}\sum_{k=1}^{d_j}\frac{\log(k)}{d_j} = \sum_{j=1}^{n}\frac{\log(d_j!)}{d_j}$$

הילוכים מקריים

הקדמה וחצי־מבוא לשרשראות מרקוב

בהינתן גרף קשיר לא מכוון $G(V,E)$, כאשר $V=\{1,\ldots,n\}$, הילוך מקרי הוא סדרה של מ"מ $X_0,X_1,\ldots$ המקבלים את ערכיהם ב־V, המוגדרים לפי הקווים הכלליים הבאים: X_0 מתאר בחירה של צומת מסויים של G (באופן דטרמיניסטי, או באמצעות התפלגות התחלתית מסויימת על V). בכל שלב, לאחר שערכו של X_{t-1}

נקבע, בוחרים צומת מקרי באופן יוניפורמי מקבוצת השכנים של X_{t-1}, וקובעים את ערכו של X_t לפי הצומת החדש. בכך מקבלים סדרה של מ"מ $X_0, X_1, \ldots$ שהטווח שלהם הוא קבוצת הצמתים $\{1, \ldots, n\}$. זהו מקרה פרטי של שרשרת מרקוב בת $n = |V|$ מצבים. בקורס לא נעבור על התיאוריה של שרשראות מרקוב (אלו יכולות להוות בסיס לקורס שלם משל עצמן), אלא רק ניתן כאן את ההגדרות הבסיסיות ביותר. בפרט, בפרק זה של הקורס יהיו יותר משפטים ללא הוכחה מאשר בפרקים הקודמים.

שרשרת מרקוב בת n מצבים מתוארת ע"י התפלגות התחלתית $q^{(0)} = (q_1^{(0)}, \ldots, q_n^{(0)})$, אשר לפיה בוחרים את המ"מ X_0 (הרבה פעמים התפלגות זו מתארת בחירה דטרמיניסטית של אחד הצמתים), ומטריצת מעבר $P = (p_{ij})_{i,j\in[n]}$, כאשר p_{ij} מתאר את הסיכוי ש־$X_t = j$ בהינתן ש־$X_{t-1} = i$. התכונה היסודית של שרשרת מרקוב היא שלכל $i_0, \ldots, i_t$ עבורם $\Pr\big[X_0 = i_0, \ldots, X_{t-1} = i_{t-1}\big] > 0$ מתקיימת התכונה הבאה, המתוארת בד"כ כחוסר זיכרון (או "זיכרון קצר"):

$$\Pr\big[X_t = i_t | X_0 = i_0, \ldots, X_{t-1} = i_{t-1}\big] = p_{i_{t-1} i_t} = \Pr\big[X_t = i_t | X_{t-1} = i_{t-1}\big]$$

אם נסמן ב־$q^{(t)}$ את ווקטור ההתפלגות של X_t (ז"א $q_i^{(t)} = \Pr\big[X_t = i\big]$ לכל i, כשההתפלגות אינה מותנה בערכי X_s קודמים) אז נקבל $q^{(t)} = P^T q^{(t-1)}$ עבור $t > 0$ (המכפלה היא במטריצת ה־transpose של P), ובאינדוקציה נקבל $q^{(t)} = (P^T)^t q^{(0)}$. כבר כאן אפשר לראות שמנתונים על הערכים העצמיים של מטריצת המעבר P מקבלים מידע חשוב על השרשרת. למשל, אם ההתפלגות q היא ווקטור עצמי של P^T עם ערך עצמי 1, אז זוהי התפלגות סטציונרית: אם $q^{(0)} = q$ אז $q^{(t)} = q$ לכל t. עבור שרשרת מרקוב עם מספר מצבים סופי תמיד תהיה התפלגות כזו, ולאף שרשרת מרקוב לא יהיו ערכים עצמיים שערכם המוחלט גדול מ־1. אם שאר הערכים העצמיים קטנים ממש בערכם המוחלט מ־1 ואין ריבוי ל־1 עצמו, אז ההתפלגות $q^{(t)}$ של X_t תשאף ל־q (עבור $t \to \infty$) ללא קשר ל־$q^{(0)}$. אפשר להסיק פרטים נוספים מהספקטרום של מטריצת המעבר, אולם בקורס זה לא ננתח את האספקטים האלגבריים של שרשראות מרקוב הרבה מעבר לדרוש למספר הוכחות.

לכל שרשרת מרקוב בת n מצבים אפשר להתאים גרף מכוון: זהו הגרף על $V = \{1, \ldots, n\}$ שבו כל (i, j) מוגדר להיות קשת אם ורק אם $p_{ij} > 0$. אפשר לקבל מידע חשוב מהתכונות הקומבינטוריות של גרף זה. למשל, אם גרף זה הוא קשיר חזק אז לכל הילוך מקרי ארוך דיו יש הסתברות חיובית להגיע בסופו של דבר לכל המצבים, ולא קשה להוכיח אף שהסתברות זו שואפת ל־1 עבור הילוך עם מספר צעדים בלתי מוגבל.

מעתה והלאה נעסוק רק בהילוכים מקריים על גרפים קשירים לא מכוונים, אם כי לפעמים נשתמש בהגדרות וטענות אשר ידועות גם עבור המקרה הכללי. לאלו הרוצים לדעת יותר על שרשראות מרקוב מומלץ להסתכל בפרק העוסק בנושא זה באחד מהספרים הבאים:

W. Feller, An introduction to Probability Theory and its Applications, 3rd edition, Vol. I.

G. R. Grimmett and D. R. Stirzaker, Probability and Random Processes.

הילוך מקרי על גרף מוגדר ע"י כך שערכו של X_t נבחר יוניפורמית מקבוצת השכנים של הצומת המתאים לערכו של X_{t-1}. בניסוח יותר מדוייק, עבור גרף G בעל קבוצת הצמתים $\{1, \ldots, n\}$, הילוך מקרי הוא שרשרת מרקוב בעלת מטריצת המעבר המוגדרת ע"י:

$$p_{ij} = \begin{cases} 1/d(i), & (i,j) \in E \\ 0, & (i,j) \notin E \end{cases}$$

לפני שנמשיך, נשים לב שלהילוך הזה יהיו ערכים עצמיים ממשיים בלבד: אם P תסמן את מטריצת המעבר שלו, ו־D תסמן את מטריצה האלכסונית שערכי האלכסון שלה הם הסדרה $(\sqrt{d_1}, \sqrt{d_2}, \ldots, \sqrt{d_n})$ בהתאמה, אז המטריצה $S = DPD^{-1}$ היא מטריצה סימטרית – s_{ij} יהיה שווה ל־$1/\sqrt{d_i d_j}$ אם (i,j) היא קשת של G, ושווה ל־0 אחרת. לכן גם ל־S וגם ל־P אין ערכים עצמיים מרוכבים.

עבור הילוכים מקריים על גרפים קשירים (סופיים ולא מכוונים) מתקיימות התכונות המתוארות כאן; הן נובעות ממשפטים כלליים על שרשראות המרקוב המתאימות, אולם ניתן גם להוכיח אותם ישירות (ברב המקרים הדבר נעשה באמצעות אלגברה לינארית).

נזכיר שהתפלגות q המקיימת $P^Tq = q$ תיקרא סטציונרית. עבור הילוך מקרי על גרף קשיר קיימת התפלגות סטציונרית יחידה, שהסימון המקובל עבורה הוא π. היא נתונה ע"י $\pi_i = \frac{d(i)}{2m}$, כאשר m יסמן לצורך הדיון את מספר הקשתות בגרף ו־$d(i)$ מסמן את מספר השכנים של הצומת i (עבור גרף לא קשיר התפלגות זו היא גם סטציונרית, אולם היא אינה היחידה). לא קשה לוודא (ע"י הכפלה ב־P^T) ש־π היא אכן התפלגות סטציונרית. להוכחת יחידות מראים שאין ווקטור עצמי נוסף עם ע"ע שווה ל־1 (פרט למכפלות של π בקבוע), ע"י כך שראשית מראים שאין ו"ע כזה עם קורדינטות אי־שליליות שלפחות אחת מהן (אך לא כולן) שווה ל־0. תהיה הוכחת יחידות דומה לזו בהמשך, וההוכחה כאן תינתן בתרגול.

עבור גרף קשיר שבנוסף לכך אינו 2־צביע מתקיימת טענה חזקה יותר – לכל התפלגות התחלתית $q^{(0)}$ יתקיים $q^{(t)} \to \pi$ כאשר $t \to \infty$. לעומת זאת עבור גרף 2־צביע אין הדבר כך, מכיוון שעבור הילוך שהתחיל מצומת הצבוע בצבע מסויים, בהסתברות 1 ההילוך יבקר צומת מאותו צבע בכל צעד זוגי. ניתן לבנות גם ו"ע של מטריצת המעבר עם ע"ע השווה ל־-1, ע"י היפוך הסימן של חלק מהקורדינטות של π בהתאם לצביעה נתונה של הגרף.

זמן פגיעה ושימוש אלגוריתמי בהילוך מקרי

אנו נתעניין בזמן הפגיעה (hitting time) h_{ij}, שהוא מספר הצעדים הממוצע הלוקח להילוך מקרי המתחיל מהצומת i להגיע בפעם הראשונה לצומת j. על מנת לחשב אותו נשתמש בחישוב של זמן הטיול (commute time), $k_{ij} = h_{ij} + h_{ji}$, שהוא התוחלת של זמן החזרה הראשונה ל־i לאחר ביקור ב־j (הערה – במאמרים שונים משתמשים באותיות שונות לסימון ערכים אלו). נראה בהמשך דרך לחשב את k_{ij} באמצעות קשר בין הילוכים מקריים על גרפים לבין זרימה ברשתות חשמליות. הנוסחה של Tetali מאפשרת לבטא את h_{ij} באמצעות k_{ij} וההתפלגות הסטציונרית π באופן הבא: $h_{ij} = \frac{1}{2}(k_{ij} + \sum_{l=1}^{n} \pi_l(k_{lj} - k_{li}))$ (הקונוונציה כאן היא ש־$k_{ii} = h_{ii} = 0$ לכל i).

עבור גרף המורכב ממסלול בודד בן $n+1$ צמתים $\{v_0, \ldots, v_n\}$, ראשית נשים לב שמתקיים $h_{0,n} = h_{0,i} + h_{i,n}$ (כי כל מסלול מ־0 ל־n חייב לעבור דרך הצומת i), ולכן בפרט $h_{0,n} \geq h_{i,n}$. אנו נראה בהמשך שמתקיים $h_{0,n} = n^2$, אבל ראשית נראה שימוש אלגוריתמי בטענה זו: נבנה אלגוריתם הסתברותי (עם שגיאה חד־כיוונית) הבודק קיום פתרון ל־2CNF, ומוצא אותו אם הוא קיים (אומנם יש גם אלגוריתם דטרמיניסטי טוב לפתרון בעיה זו, אולם האלגוריתם כאן ממחיש טוב עקרונות המצויים גם באלגוריתמים מתקדמים יותר). האלגוריתם מתחיל עם הצבה שרירותית של ערכים למשתנים $x_1, \ldots, x_n$ של נוסחת ה־2CNF. בכל שלב שבו יש פסוקיות שאינן מסתפקות, האלגוריתם בוחר שרירותית פסוקית אחת כזו, בוחר באופן אקראי ויוניפורמי את אחד משני המשתנים המופיעים בה, והופך אותו.

נוכיח עתה שאם יש הצבות מספקות, אז האלגוריתם ימצא הצבה כזו בתוחלת של $O(n^2)$ צעדים, כאשר n הוא מספר המשתנים. נקבע הצבה מספקת $\alpha_1, \ldots, \alpha_n$ של משתני הנוסחה, ובכל שלב t נסמן ב־X_t את מספר המשתנים שבשלב זה באלגוריתם קיבלו את הערכים המתאימים להם ב־$\alpha_1, \ldots, \alpha_n$. במידה ולא הגענו בשלב t להצבה מספקת, אז בפסוקית שאינה מסתפקת קיים לפחות משתנה אחד שלא קיבל את הערך המתאים לו ב־$\alpha_1, \ldots, \alpha_n$. על כן, בהסתברות לפחות $\frac{1}{2}$ יתקיים $X_{t+1} = X_t + 1$ (כי זהו הסיכוי שאותו משתנה יבחר וערכו יתוקן), ואם מאורע זה לא יתקיים אז יתקיים $X_{t+1} = X_t - 1$ (יתכן גם מקרה שבו שני המשתנים בפסוקית לא קיבלו את ערכיהם ב־$\alpha_1, \ldots, \alpha_n$, ואז מתקיים $X_{t+1} = X_t + 1$ בהסתברות 1). לכן תוחלת מספר הצעדים עד למציאת הצבה מספקת חסומה ע"י תוחלת מספר הצעדים להגעה בהילוך מקרי על מסלול מ־0 ל־n, ומכאן המבוקש (יתכנו גם מקרים שבהם $X_0 > 0$, ומקרים שבהם האלגוריתם עצר על הצבה מספקת אחרת לפני ש־X_t הגיע ל־n, אולם אלו יכולים רק לקצר את תוחלת זמן הריצה).

פונקציות הרמוניות והקשר לרשתות חשמליות

לחישוב k_{ij} נשתמש במושג של פונקציה הרמונית. בהנתן גרף קשיר $G(V,E)$ וקבוצת צמתים $\emptyset \neq S \subset V$, פונקציה $\phi : V \to \mathbb{R}$ תיקרא הרמונית עם שפה S (לפעמים S נקראת גם "קבוצת הקטבים") אם לכל $v \in V \setminus S$ מתקיים תנאי המיצוע $\phi(v) = \frac{1}{d(v)} \sum_{u \in N(v)} \phi(u)$, כאשר $N(v)$ מסמן את קבוצת השכנים של הצומת v. הערכים ש־ϕ מקבלת ב־S יקראו תנאי השפה של ϕ.

נראה עתה דוגמה ראשונה, ובכך נוכיח בפרט את קיומה של פונקציה כזו לכל קבוצת ערכים אפשרית עבור S. בהינתן $\psi : S \to \mathbb{R}$, לכל $v \in V$ נגדיר את $\phi(v)$ באופן הבא: ניקח הילוך מקרי על הגרף G שמתחיל מהצומת v (ז"א שאם $X_0, X_1, \ldots$ הם המ"מ של ההילוך אז יתקיים $X_0 = v$ בהסתברות 1). עתה נגדיר מ"מ Y, ע"י כך שנבחר את ערכו להיות זהה ל־$\psi(X_t)$, כאשר $t \geq 0$ הוא המספר המינימלי שעבורו התקיים המאורע $X_t \in S$. לבסוף נגדיר את הערך $\phi(v)$ להיות $\mathrm{E}[Y]$ עבור ההילוך המקרי הנ"ל. קל לוודא שכאשר $v \in S$ מתקיים $\phi(v) = \psi(v)$, כי אז $t = 0$ ו־$Y = \psi(v)$ בהסתברות 1. את קיום תנאי המיצוע עבור $v \in V \setminus S$ מראים עתה לפי נוסחת התוחלת השלמה, תוך כדי שימוש בחוסר הזיכרון של הילוך מקרי, ובכך שאם $v \in V \setminus S$ אז $t > 0$ בהסתברות 1:

$$\phi(v) = \mathrm{E}[Y] = \sum_{u \in V} \mathrm{E}\big[Y|X_1 = u\big]\Pr\big[X_1 = u\big] = \sum_{u \in N(v)} \frac{1}{d(v)}\mathrm{E}\big[Y|X_1 = u\big] = \frac{1}{d(v)} \sum_{u \in N(v)} \phi(u)$$

נראה עתה יחידות עבור פונקציות הרמוניות: אם קיימות שתי פונקציות הרמוניות ϕ_1, ϕ_2 עם שפה S עבורן $\phi_1(v) = \phi_2(v)$ לכל $v \in S$, אז $\phi_1(v) = \phi_2(v)$ לכל $v \in V$. לשם כך נבחן את $\phi(v) = \phi_1(v) - \phi_2(v)$, ועתה נראה כי כל פונקציה הרמונית המתאפסת על השפה היא בהכרח פונקצית ה־0. נניח כי אין הדבר כך, ובלי הגבלת הכלליות נניח כי יש ל־ϕ ערכים חיוביים (אחרת נשתמש ב־$-\phi$ במקום ב־ϕ). תהי V' קבוצת הצמתים עבורם $\phi(v)$ מקבלת ערך מקסימלי. זוהי תת־קבוצה לא ריקה של $V \setminus S$, ולכן מקשירות הגרף G קיים צומת v ב־V' שיש לו לפחות שכן אחד w שאינו ב־V'. מכאן מגיעים לסתירה עם תנאי המיצוע $\phi(v) = \frac{1}{d(v)} \sum_{u \in N(v)} \phi(u)$, מכיוון שבסכום הימני כל האיברים אינם גדולים מ־$\phi(v)$, כאשר לפחות אחד מהם, $\phi(w)$, קטן ממש מ־$\phi(v)$.

נשים לב עתה שעבור שפה בת שני איברים $S = \{s, t\}$ ותנאי השפה $\phi(s) = 1$ ו־$\phi(t) = 0$, הפונקציה ההרמונית המתקבלת מתארת לכל צומת את הסיכוי שהילוך מקרי היוצא ממנו יגיע ל־s לפני שיגיע ל־t. עתה נבנה בדרך אחרת פונקציה הרמונית עם אותם תנאי השפה, ואז נשתמש בכך שממשפט היחידות נובע שהיא זהה לפונקצית ההסתברות הנ"ל.

הבניה השניה לפונקציה זו היא באמצעות הגדרה "פיזיקלית" של רשתות חשמליות. לצורך ענייננו רשת חשמלית היא פשוט אוסף של משוואות לינאריות על אותם משתנים ממשיים שבפיזיקה מזהים עם "הפרשי פוטנציאל" ועם "זרמים". עם זאת, בפיזיקה פותחו כלים אשר בהרבה מקרים מפשטים את מציאת הפתרונות למערכת המשוואות המסויימת הנ"ל, בעיקר אלו הקשורים למושג ההתנגדות השקולה (שהיא בעצם המנה של שני פרמטרים של פתרון מערכת המשוואות), ומכאן חוזק הבניה האלטרנטיבית.

נניח שבונים רשת חשמלית לפי G שבה כל קשת היא נגד עם התנגדות 1, והמתח בין s ל־t מוחזק להיות 1. מגדירים את $\phi(v)$ עתה להיות הפרש המתחים בין הצומת v ל־t. ברור שתנאי השפה מתקיימים, ותנאי המיצוע נובע עתה מחוק קירכהוף: סכום הזרמים הנכנסים ל־$v \in V \setminus \{s, t\}$ הוא 0, ולפי חוק אוהם ("חוק אוהם" כאן הוא פשוט ההגדרה של המשוואות הלינאריות המקשרות בין משתני ה"זרם" וה"מתח" השונים) סכום זה שווה ל־$\sum_{u \in N(v)}(\phi(v) - \phi(u))$. מהעברת אגפים מתקבל $\phi(v) = \frac{1}{d(v)} \sum_{u \in N(v)} \phi(u)$ כנדרש.

נוכיח עתה את הקשר הבא בין רשתות חשמליות והילוכים מקריים: אם R_{st} מסמן את ההתנגדות השקולה של המעגל הנ"ל בין s ו־t, אז מתקיים $k_{st} = 2mR_{st}$, כאשר מעתה נסמן $m = |E|$. ראשית אבל נוכיח את הלמה הבאה: ההסתברות שהילוך המתחיל בצומת t יבקר את הצומת s לפני שיחזור שוב ל־t היא בדיוק $\frac{2m}{d(t)k_{st}}$.

הוכחה הלמה: נסמן ב־q את ההסתברות שהילוך היוצא מ־t אכן מבקר את s לפני החזרה ל־t, ב־τ את המ"מ המקבל את זמן החזרה הראשון של הילוך כזה ל־t (ללא קשר לביקור אפשרי ב־s), וב־σ את המ"מ המקבל את זמן החזרה הראשון של ההילוך ל־t לאחר ביקור ב־s. ראשית נציין שמתקיים $\mathrm{E}[\tau] = 1/\pi_t = \frac{2m}{d(t)}$. את השוויון הזה לא נוכיח כאן, והנכם מוזמנים לקרוא עליו בשאלה "בלי הרבה נפנוף ידיים" בחוברת התרגילים; באופן אינטואיטיבי, השוויון נובע מכך שאם נסתכל על הילוך ארוך במיוחד, אז מספר הפעמים שהוא ביקר ב־t מתקרב למספר הצעדים הכולל כפול הסיכוי להמצא ב־t מצד אחד, ומתקרב למספר הצעדים חלקי אורך הזמן הממוצע בין שני ביקורים עוקבים מצד שני. עתה נשים לב שמתקיים $k_{st} - \frac{2m}{d(t)} = \mathrm{E}[\sigma] - \mathrm{E}[\tau] = \mathrm{E}[\sigma - \tau]$. מצד שני, הפרש המ"מ $\sigma - \tau$ מקבל את הערך 0 אם החזרה הראשונה ל־t היתה רק לאחר ביקור ב־s (מאורע המתקיים כזכור בהסתברות q), והתוחלת המותנה של $\sigma - \tau$ בהינתן שהחזרה הראשונה ל־t לא היתה לאחר

ביקור ב־s היא k_{st} (בגלל חוסר הזיכרון של ההילוך המקרי). קיבלנו $k_{st} - \frac{2m}{d(t)} = \mathrm{E}[\sigma - \tau] = (1-q)k_{st}$, ובהעברת אגפים מקבלים $q = \frac{2m}{d(t)k_{st}}$ כנדרש.

נוכיח עתה שמתקיים $k_{st} = 2mR_{st}$: סכום הזרמים היוצאים מ־t ברשת החשמלית שנבנתה עבור G הוא $\sum_{u\in N(t)} \phi(u)$ (כאשר ϕ היא פונקציה הרמונית עם תנאי השפה $\phi(t) = 0$ ו־$\phi(s) = 1$), ומצד שני הסכום מוגדר כשווה ל־R_{st}^{-1} לפי חוק אוהם. עתה, ההסתברות שהילוך המתחיל בצומת t יבקר את s לפני שיחזור ל־t היא $\frac{2m}{d(t)k_{st}}$ לפי הלמה למעלה. אולם הסתברות זו שווה ל־$\frac{1}{d(t)}\sum_{u\in N(t)} \phi(u)$, כי $\phi(u)$ מציין גם את הסיכוי לכך שהילוך מקרי היוצא מ־u יגיע ל־s לפני t, ולכן $\frac{2m}{d(t)k_{st}} = \frac{1}{d(t)}\sum_{u\in N(t)} \phi(u) = \frac{1}{d(t)}R_{st}^{-1}$. מכאן מקבלים את המבוקש.

נסכם עתה את הדיון בשתי דוגמאות. הדוגמה הראשונה היא גרף המסלול על $\{0,\ldots,n\}$ אשר הוזכר קודם. כאן מתקיים $R_{0n} = |E| = n$, ולכן $k_{0n} = 2n^2$. מכיוון שהגרף הוא סימטרי אפשר להסיק כאן שמתקיים $h_{0n} = n^2$, שהוא הערך שהוזכר קודם בדוגמת האלגוריתם ל־2SAT.

הדוגמה השניה היא הגרף הבא על קבוצת הצמתים שתסומן $\{1-n, 2-n, \ldots, 0, 1, \ldots, n\}$: קשתות הגרף יורכבו ממסלול על קבוצת הצמתים $\{0,\ldots,n\}$ (בסדר הזה), וקליק על קבוצת הצמתים $\{1-n,\ldots,0\}$. כאן מתקיים $R_{0n} = n$ ו־$|E| = \binom{n}{2} + n = \frac{n^2+n}{2}$, ולכן $k_{0n} = n^3 + n^2$. לא קשה לראות ש־h_{n0} למעשה זהה לזמן הפגיעה על מסלול בן n קשתות, ולכן מתקיים כאן $h_{0n} = n^3$. הלקח שאפשר ללמוד מדוגמה זו הוא שתוספת קשתות לא בהכרח מקטינה את זמן הפגיעה, ויכולה אף להגדיל אותו. לקח שני הוא ש־h_{ij} יכול להיות אסימטרי למדי.