תרגיל ראשון   הגשה: 20.4.2021

לבלוע את החוכמה (6 נקודות)

נתונה משפחה $\mathcal{F}$ של תתי־קבוצה של $\{1,\ldots,n\}$. עבור $0 \le q \le n$ ועבור $0 \le p \le 1$, נגדיר שני מרחבי הסתברות להגרלת תת־קבוצה $Q \subseteq \{1,\ldots,n\}$.

- במרחב μ, נגריל את Q באופן יוניפורמי מכל $\binom{n}{q}$ האפשרויות לתת־קבוצה בגודל q של $\{1,\ldots,n\}$.
- במרחב ν, נבחר כל אינדכס $1 \le i \le n$ להיות איבר ב־Q בהסתברות p, באופן ב"ת לחלוטין בתוצאות ההגרלה עבור אינדכסים אחרים.

נסמן ב־A את המאורע "קיימת $F \in \mathcal{F}$ שעבורה מתקיים $F \subseteq Q$". הראו שמתקיים $\Pr_\mu[A] > \Pr_\nu[A] - \frac{pn}{q}$.

הדרכה: אפשר לפתור את השאלה בלי להסתבך מתמטית. נסו להגדיר מרחב הסתברות על זוגות של תתי־קבוצה $Q \subseteq Q' \subseteq \{1,\ldots,n\}$, כך ש־$Q$ מתפלגת לפי ν ו־Q' מתפלגת לפי התפלגות קרובה ל־μ (ראו את ההגדרה של מרחק־התפלגות בקורס "שיטות הסתברויות ואלגוריתמים").

לחפש את הצדק (6 נקודות)

אנחנו מעוניינים לבצע ϵ־בדיקה עבור התכונה "פרמוטציה" של פונקציות $f : \{1,\ldots,n\} \to \{1,\ldots,n\}$ (פונקציה כזו היא פרמוטציה אם היא חח"ע ועל, ועבור התחום והטווח הנ"ל, זה שקול לתכונה שהפונקציה היא חח"ע). נתון האלגוריתם הבא: בוחרים באופן יוניפורמי (מבין כל תתי־הקבוצה האפשריים) קבוצה $Q \subset \{1,\ldots,n\}$ בגודל q, ובודקים את כל ערכי f מעל Q (סה"כ q שאילתות). דוחים אם מוצאים $i \ne j$ בקבוצה Q שעבורם $f(i) = f(j)$, ואחרת מקבלים.

הראו שהאלגוריתם הנ"ל לא עובד עבור $q = o(\sqrt{n/\epsilon})$. באופן מדוייק: שקיימים קבועים $c, \eta > 0$ כך שאם $\epsilon \le \eta$ וכן $q < c\sqrt{n/\epsilon}$, אז קימת פונקציה f שהיא ϵ־רחוקה מלהיות פרמוטציה, ובכל זאת האלגוריתם יקבל אותה (באופן שגוי) בהסתברות גבוהה מ־$\frac{1}{3}$.

למצוא את הצדק (6 נקודות)

עבור האלגוריתם המתואר בשאלה הקודמת, הוכיחו שהוא באמת מהווה ϵ־בדיקה עבור $q = O(\sqrt{n/\epsilon})$ מתאים. באופן מדוייק: קיים קבוע C כך שאם $q \ge C\sqrt{n/\epsilon}$, אז האלגוריתם ידחה בהסתברות לפחות $\frac{2}{3}$ כל פונקציה שהיא ϵ־רחוקה מלהיות פרמוטציה. מותר להשתמש בתוצאת השאלה "לבלוע את החוכמה" גם אם לא פתרתם אותה (מותר גם לא להשתמש בה אם אתם לא צריכים).

תרגיל שני   הגשה: 1.6.2021

להקיא את החוכמה (6 נקודות)

נתונה לנו המשפחה $\mathcal{F}$ של תתי־קבוצה של $\{1,\ldots,n\}$ כמו בשאלה "לבלוע את החוכמה" מהתרגיל הקודם, וכן מרחבי ההסתברות μ ו־ν שמוגדרים שם. הראו הפעם שאם מתקיים $pn \geq q$, אז מתקיים $\Pr_\nu[A] > (1-e^{-(p-q/n)^2n/2p})\Pr_\mu[A]$.

רמז: את הרב אתם כבר בוודאי מבינים איך לעשות לאור נסיונכם מהתרגיל הקודם. איבר החזקה שמופיע בביטוי כאן מתקבל מחסם צ'רנוף מתאים.

לא למצוא את הצדק (8 נקודות)

נסתכל שוב על התכונה "פרמוטציה" של פונקציות $f:\{1,\ldots,n\}\to\{1,\ldots,n\}$.

- הראו שאם קיים אלגוריתם אשר מבצע ϵ־בדיקה לתכונה באמצעות q שאילתות, אז קיים אלגוריתם כזה שבוחר את השאילתות שלו ע"י בחירה יוניפורמית של תת־קבוצה $Q\subset\{1,\ldots,n\}$ באופן יוניפורמי מבין כל תתי־הקבוצה האפשריים.
- הראו שבנוסף, אפשר גם להניח שהאלגוריתם הזה תמיד ידחה קלט אם הוא גילה שני אינדכסים ב־Q שעבורם f מקבלת ערך זהה, ובמידה שהאלגוריתם לא גילה שני אינדכסים כאלה, הסתברות הדחיה אינה תלוייה כלל בערכים הספציפים שהתגלו.

הערה: מהסעיפים למעלה נובע שהאלגוריתם הפשוט לבדיקת פרמוטציה של התרגיל הראשון הוא בעצם האלגוריתם הטוב ביותר האפשרי.

עולם של קליקות (8 נקודות)

גרף $G=(V,E)$ הוא איחוד של קליקות (זרות־צמתים) אם יש חלוקה של כל קבוצת הצמתים שלו לקבוצות זרות $V=V_1\cup\cdots\cup V_k$ עבור $k\leq n$ (כאשר n הוא מספר הצמתים), כך שכל שני צמתים u,v הם מחוברים בקשת אם ורק אם הם לא משתייכים לאותו V_i בחלוקה (ז"א שכל קשתות הגרף הן קשתות הקליקות $(V_1,\ldots,V_k$).

כתבו (והוכיחו) אלגוריתם בדיקה עם שגיאה חד־כיוונית במודל הגרפים הצפוף, ומספר שאילתות פולינומי ב־$1/\epsilon$ (ולא תלוי n), עבור התכונה שהגרף הוא איחוד של קליקות.

רמז: אפשר להשתמש בשיטות שמזכירות את הבדיקה של דו־צדדיות. שימו לב שצומת ללא קשתות בכלל נחשב ל"קליקה" בת צומת יחיד.

תרגיל סופי – חלק א   הגשה: 1.8.2021

בערך מונוטוניות (8 נקודות)

פונקציה $f:\{1,\ldots,n\}\to\mathbb{N}$ תיקרא k־בערך מונוטונית אם לכל $i,j\in\{1,\ldots,n\}$ שעבורם $i+k\leq j$ מתקיים $f(i)\leq f(j)$. אין שום דרישה על ערכי הפונקציה במקרה שמתקיים $i<j<i+k$. הראו שקיים אלגוריתם בעל $O(\log(n)/\epsilon)$ שאילתות שייקבל כל פונקציה שהיא k־בערך מונוטונית, אבל ידחה כל פונקציה שהיא ϵ־רחוקה מלהיות $10k$־בערך מונוטונית (אפשר לשפר את המקדם "10", אבל ממש לא חייבים).

הערה: זה לא בדיוק בדיקת תכונה, כי התנאי לדחיה הוא לגבי תכונה אחרת שמכילה את התכונה המקורית. כזה סוג של אלגוריתם נקרא bi-criteria testing, כי יש שני קריטריונים המעורבים בפסילה של הפונקציה: הראשון הוא המרחק, והשני הוא הפרמטר k.

רמז: אפשרות אחת לפתרון היא קודם כל לבדוק מה קורה באינדכסים שהם כפולות של k. אפשרות אחרת (שהופיעה במאמר שהוכיח לראשונה טענה כזו) משתמשת בשיטה השניה שלמדתם לבדיקת מונוטוניות "רגילה", זו שמבוססת על גרף ההפרות (עבור האפשרות השניה אפשר בתור התחלה להגדיר את הסדר החלקי של "כותבים $x\prec y$ אם $x<y-10k$" ואז לנתח את גרף ההפרות של "מונוטוניות מעל $\prec$"). לדעתי האישית האפשרות הראשונה היא הפחות מסובכת.

לומדים במסלול המהיר (12 נקודות)

נניח שיש לנו גישה במודל הגרפים הדליל לגרף שדרגתו חסומה ע"י $d=2$. הראו שיש עבורו אלגוריתם למידה: האלגוריתם מקבל את ϵ, את מספר הצמתים n ואת קבוצת הצמתים V (ע"מ שיוכל לבצע שאילתות לפי "שמות" הצמתים), מבצע מספר שאילתות שתלוי ב־ϵ בלבד, ואז פולט גרף בעל n צמתים שבהסתברות לפחות $\frac{2}{3}$ יהיה ϵ־קרוב להיות איזומורפי לגרף הקלט. אין צורך להגיע למספר השאילתות הכי טוב האפשרי (כל עוד הוא לא תלוי ב־n).

רמז: קודם הראו שעבור k שתלוי ב־ϵ בלבד, הגרף יהיה ϵ־קרוב לגרף שיש בו רק רכיבי קשירות מגודל חסום ע"י k, למעט אולי רכיב קשירות בודד שהוא מסלול מעל כל הצמתים הנותרים.

הערה: אפשר באמצעות אלגוריתם הלמידה הזו לכתוב "אלגוריתם לבדיקת כל תכונה ניתנת לחישוב" (שמבצע מספר שאילתות קבוע, אבל יכול להיות עם זמן חישוב מאוד ארוך): ראשית מבצעים $\frac{\epsilon}{2}$־למידה של גרף הקלט, ואז מקבלים אותו אם גרף הפלט הוא $\frac{\epsilon}{2}$־קרוב לתכונה הרצויה.

המשך יבוא...

תרגיל סופי – חלק ב   הגשה: 1.8.2021

גלגל אותה (8 נקודות)

מחרוזות בינארית $w \in \{0,1\}^n$ תיקרא גלגול של w', אם קיימות מחרוזות u, v כך ש־$w = uv$ וגם $w' = vu$ (מותר ש־v תהיה ריקה, ז"א שבפרט w היא גלגול של עצמה). נרצה לבדוק זוג מחרוזות מאורך n עבור התכונה שאחת מהן היא גלגול של השניה (עבור ההגדרה הפורמלית של בדיקת תכונה, אפשר להסתכל על זוג מחרוזות בינאריות כעל פונקציה $f : \{1, \dots, 2n\} \to \{0,1\}$). הוכיחו שכל אלגוריתם (אדפטיבי או לא) שבודק את התכונה הזו (למשל עבור $\epsilon = \frac{1}{5}$) חייב לבצע לפחות $\Omega(\sqrt{n})$ שאילתות.

הערה: אפשר לקבל 6 נקודות עבור הוכחה נגד אלגוריתמים לא־אדפטיבים בלבד.

ריקוד לשניים (12 נקודות)

נרצה (באמצעות דגימות) לבדוק התפלגות μ מעל $\{1, \dots, n\} \times \{1, \dots, n\}$ עבור התכונה שהיא בלתי־תלויה. באופן פורמלי, נסמן ב־μ_1 את ההטלה על הקורדינטה הראשונה, ז"א את ההתפלגות המתקבלת מהגרלת (i,j) לפי μ ולקיחת הקורדינטה הראשונה בלבד. בדומה נסמן את μ_2. נסמן ב־$\mu_1 \times \mu_2$ את התוצאה של הגרלה של i לפי μ_1 והגרלה ב"ת (ע"י דגימה נפרדת מתוך μ) של j לפי μ_2. אי־תלות היא התכונה שמתקיים $\mu = \mu_1 \times \mu_2$ – הווה אומר שלכל $i, j \in \{1, \dots, n\}$ מתקיים $\mu(i,j) = \mu_1(i) \cdot \mu_2(j)$. הראו שאפשר לעשות את זה, לכל $\epsilon > 0$ קבוע, באמצעות $\tilde{O}(n)$ דגימות (שימו לב שגודל מרחב ההסתברות של μ הוא n^2, כך שזה יותר יעיל מלקרב את μ בשלמותו).

הדרכה (שימו לב גם למה שלא צריך להוכיח): אפשר, גם עבור μ_1 וגם עבור μ_2, באמצעות $\tilde{O}(n)$ דגימות, לכתוב התפלגות $\tilde{\mu}_1$ והתפלגות $\tilde{\mu}_2$ כך שמתקיים למשל $\tilde{\mu}_1(i) \leq (1+\epsilon)\mu_1(i)$ לכל $1 \leq i \leq n$ וכן מתקיים $(1-\epsilon)\mu(i) \leq \tilde{\mu}_1(i)$ לכל i שעבורו $\mu(i) \geq \epsilon/(1+\epsilon)n$. את זה אתם לא צריכים להוכיח (למתענינים, זה נובע מהפעלה של חסם צ'רנוף כפלי לכל i בתוספת חסם איחוד מאורעות מעל האפשרויות עבור i).

אפשר להשתמש בזה אח"כ בדומה למה שעושים עבור בדיקה מול התפלגות נתונה. במידה ואתם עושים זאת, ניתן לקצר בטיעונים שכבר נמצאים בהוכחה של הבדיקה מול התפלגות נתונה בחוברת הקורס (מותר להגיד "בדומה להוכחה בחוברת. . . ", רק אין לדלג על טיעון מבלי לפחות להזכיר אותו בצורה כזו, וכמובן שזה באמת צריך להיות דומה להוכחה בחוברת).

לא הפעם הגשה: 32.13

מדקדקים בחומרה (12 נקודות)

הראו שעבור ϵ־בדיקה של התפלגות מעל $\{1,\ldots,n\}$ ליוניפורמיות (במודל הדגימות) צריך $\Omega(\sqrt{n}/\epsilon^2)$ דגימות לפחות. זה אומר שצריך להוכיח חסם עם מקדם שאינו תלוי ב־n או ϵ. מותר להניח הנחות בסגנון "n הוא מספר זוגי" על מנת להימנע מסימני עיגול מציקים. מותר גם להניח ש־n עצמו הוא גדול מספיק.

רמז: קודם כל כדאי לבנות התפלגות η שהיא ϵ־רחוקה מיוניפורמיות שעם זאת מקיימת $\|\eta\|_2^2 = (1+\Theta(\epsilon^2))\frac{1}{n}$. בהינתן שלקחנו לא יותר מ־$C\sqrt{n}/\epsilon^2$ דגימות (עבור קבוע C מתאים), שנסמן ב־$A_1,\ldots,A_q$, נשים לב שיש סיכוי קטן למדי שיהיו $i<j<k$ כל שהם שעבורם $A_i = A_j = A_k$, וגם שיהיו ממש הרבה התנגשויות סה"כ. מכאן כבר אפשר לנתח את ההתפלגות על "תמונת ההתנגשויות".

מותר להציץ (6 נקודות)

נניח שמעוניינים לבדוק התפלגות מעל $\{1,\ldots,n\}$ ליוניפורמיות, אבל כאן, בנוסף לקבלת ערכי הדגימות $A_1,\ldots,A_q$, מקבלים גם את ערכי μ לפי הדגימות, ז"א שהאלגוריתם מקבל גם את $\mu(A_1),\ldots,\mu(A_q)$. הראו שעכשיו מספר הדגימות הדרוש עבור ϵ־בדיקה תלוי ב־ϵ בלבד ולא ב־n.