# מבני נתונים
# תרגיל 3

גלעד אשרוב

3 באפריל 2013

**שאלה 1.**

1. הוכח את חסם האיחוד (*union bound*). כלומר, יהיו $A_1, \ldots, A_n$ מאורעות. אזי מתקיים:

$$\Pr\left[\bigvee_{i=1}^{n} A_i\right] \leq \sum_{i=1}^{n} \Pr\left[A_i\right]$$

**פתרון:** ההוכחה היא באינדוקציה על מספר המאורעות. עבור שני מאורעות $A_1, A_2$, מתקיים:

$$\Pr\left[A_1 \cup A_2\right] = \Pr\left[A_1\right] + \Pr\left[A_2\right] - \Pr\left[A_1 \cap A_2\right] \geq \Pr\left[A_1\right] + \Pr\left[A_2\right]$$

נניח נכונות עבור $k$ מאורעות שמתקיים:

$$\Pr\left[\bigvee_{i=1}^{k} A_i\right] \leq \sum_{i=1}^{k} \Pr\left[A_i\right]$$

נראה עבור $k+1$ מאורעות: נסמן $B = \bigvee_{i=1}^{k} A_i$. אזי מתקיים:

$$\Pr\left[\bigvee_{i=1}^{k+1} A_i\right] = \Pr\left[B \vee A_{k+1}\right] \leq \Pr\left[B\right] + \Pr\left[A_{k+1}\right] \leq \sum_{i=1}^{k} \Pr\left[A_i\right] + \Pr\left[A_{k+1}\right] = \sum_{i=1}^{k+1} \Pr\left[A_i\right]$$

כאשר אי-השיוויון האחרון מתקיים לפי הנחת האינדוקציה.

**שאלה 2.** מגרילים גרף לא מכוון באקראי עם $n$ קודקודים בצורה הבאה: לכל זוג קודקודים בגרף מטילים מטבע השווה 1 בהסתברות $p$ (עבור $0 < p < 1$). אם המטבע יצא 1 - מגדירים קשת בין זוג הקודוקדים בגרף.

1. חשבו את ההסתברות שנקבל גרף עם בדיוק $t$ קשתות, עבור $t$ קבוע כלשהו (מהם הערכים שאותם $t$ יכול לקבל?)
2. חשב את תוחלת מספר הקשתות בגרף.

   **הנחיה:** הגדר משתנה מקרי לכל קשת (איזה?) והשתמש בלינאריות התוחלת.

3. תת קבוצה בלתי תלויה בגודל $k$ - היא תת קבוצה של קודקודים (בגודל $k$), כך שלכל זוג קודקודים בתת הקבוצה הזו - אין קשת בין הקודקודים. מצא חסם עליון להסתברות שקיים בגרף תת קבוצה בלתי תלויה בגודל $k$.
4. קליק בגרף בגודל $k$ היא תת קבוצה של קודקודים בגודל $k$ כך שקיימת קשת בין כל זוג קודקודים בתת קבוצה זו. מצא חסם תחתון להסתברות שלא קיים בגרף קליק בגודל $k$.

**פתרון:**

1. נשים לב שבגרף יכולים להיות בין 0 קשתות, לבין $\binom{n}{2}$ קשתות. נסמן $m = \binom{n}{2}$.

   בכדי שבגרף יהיו בדיוק $t$ קשתות, אנחנו צריכים ראשית לבחור את $t$ הקשתות הנ"ל מתוך $m$ הקשתות האפשריות. לאחר מכן, אנחנו צריכים שעבור $t$ הקשתות שנבחרו - ההגרלה תצא 1, ועבור $m - t$ הקשתות שנותרו - תוצאת ההגרלה היא 0. נקבל אם כן:

$$\Pr\left[\text{exact } t \text{ edges}\right] = \binom{m}{t} \cdot p^t \cdot (1-p)^{m-t}$$

2. נסמן את הקשתות האפשריות כ- $e_1, \ldots, e_m$. נגדיר $X_i$ - משתנה מקרי המקבל 1 במקרה והקשת $e_i$ באמת נוצרה (כלומר, שתוצאת ההגרלה עבורה יצאה 1). נגדיר $X$ להיות משתנה מקרי המציין את מספר הקשתות הכולל בגרף. מתקיים: $X = \sum_{i=1}^{m} X_i$. בנוסף, לכל $i$ מתקיים:

$$E[X_i] = 1 \cdot \Pr\left[X_i = 1\right] + 0 \cdot \Pr\left[X_i = 0\right] 1 \cdot p = p$$

   מלינאריות התוחלת נובע:

$$E\left[X\right] = E\left[\sum_{i=1}^{m} X_i\right] = \sum_{i=1}^{m} E\left[X_i\right] = m \cdot p$$

3. נחשב תחילה מה ההסתברות שתת קבוצה מסויימת ונתונה של קודקודים $S = \{u_1, \ldots, u_k\}$ תהיה תת קבוצה בלתי תלויה. בכדי שתהיה בלתי תלויה, כל הקשתות בין הקודקודים צריכות שלא להיבחר, כלומר, שתוצאות ההגרלה עבור $\binom{k}{2}$ הקשתות יהיו 0. נקבל אם כן שעבור קבוצה נתונה $S$ בגודל $k$, ההסתברות שתהיה בלתי תלויה היא: $(1-p)^{\binom{k}{2}}$.

   כעת, כמה תתי קבוצות של קודקודים בגודל $k$ יש לנו? התשובה היא $\binom{n}{k}$. נקבל אם כן:

$$\begin{aligned}
&\Pr\left[\exists \text{ independent set of size } k\right] \\
&= \Pr\left[\bigvee_{S s.t. |S|=k} S \text{ is independent set}\right] \le \sum_{S s.t. |S|=k} (1-p)^{\binom{k}{2}} = \binom{n}{k} (1-p)^{\binom{k}{2}}
\end{aligned}$$

4. נחשב תחילה את ההסתברות שלא קיים בגרף קליק בגודל $k$. בצורה דומה לסעיף הקודם, כדי שקבוצה נתונה בגודל $k$ תהיה קליק, כל הקשתות שלה צריכות להיבחר (קורה בהסתברות $p^{\binom{k}{2}}$). נקבל אם כן:

$$\begin{aligned}
&\Pr\left[\exists \text{ Clique of size } k\right] \\
&= \Pr\left[\bigvee_{S s.t. |S|=k} S \text{ is Clique}\right] \le \sum_{S s.t. |S|=k} p^{\binom{k}{2}} = \binom{n}{k} p^{\binom{k}{2}}
\end{aligned}$$

   ולכן:

$$\Pr\left[\neg\exists \text{ Clique of size } k\right] = 1 - \Pr\left[\exists \text{ Clique of size } k\right] \ge 1 - \binom{n}{k} p^{\binom{k}{2}}$$

**שאלה 3.** נניח קוף שיושב על המקלדת שבה רק אותיות בשפה העברית (27 מקשים בסה"כ, כולל אותיות סופיות), ומקיש אותיות באופן אקראי לחלוטין. בכל שלב הוא בוחר אות באקראי באופן בלתי תלוי מהבחירה הקודמת ומקיש אותה. נניח הקוף מקיש בסה"כ 100,000,000 אותיות. מהי תוחלת מספר הפעמים שנקבל את הצירוף "תוחלת" בטקסט שיצר?

**הדרכה:** השתמשו בלינאריות התוחלת.

**פתרון:** נסמן $m = 99,999,995$. לכל $1 \le i \le m$ נגדיר משתנה מקרי המקבל 1 אם באינדקס $i$ מתחיל מופע של המילה תוחלת. נגדיר משתנה מקרי $X$ - המציין את מספר המופעים של המילה תוחלת בטקסט. נקבל אם כן: $X = \sum_{i=1}^{m} X_i$. ההסתברות שבמקום ה־$i$ יתחיל מופע של המילה תוחלת היא: $\Pr\left[X_i = 1\right] = (1/27)^5$, ולכן:

$$E[X_i] = 1 \cdot \Pr\left[X_i = 1\right] + 0 \cdot \Pr\left[X_i = 0\right] = \left(\frac{1}{27}\right)^5$$

ולפי לינאריות התוחלת:

$$E\left[X\right] = \sum_{i=1}^{m} E\left[X_i\right] = m \cdot \left(\frac{1}{27}\right)^5 \approx 6.96$$

ולכן אנחנו מצפים שיהיו כ־7 מופעים של המילה תוחלת.

**שאלה 4.** נניח שיש לי שני פולינומים $A(x)$, $B(x)$, שניהם מדרגה $n$. שני הפולינומים מיוצגים בצורה שונה: הראשון נתון בצורה קנונית, כלומר - $A(x) = \sum_{k=0}^{n} a_k x^k$. השני מיוצג בצורה הבאה: $B(x) = \Pi_{k=1}^{n}(x - c_k)$ (הפרמטרים $a_i, c_i$ הם קבועים ונתונים). אנחנו מעוניינים לדעת האם $B(x) \equiv A(x)$, כלומר, האם הפולינומים הם שקולים.

1. נתבונן באלגוריתם הבא לפתירת הבעיה: בהיתן הקלט $A(x), B(x)$, האלגוריתם ימיר את $B(x)$ לצורה קנונית וייחשב בפירוש האם הפולינומים הם זהים. מהי עלות האלגוריתם? הסבירו.

   נשים לב שהאלגוריתם מקבל כקלט שני מערכים: במקרה של $A(x)$ מערך בגודל $n+1$ של מקדמים $a_0, \dots, a_n$, ובמקרה של $B(x)$ מערך בגודל $n$ של $c_1, \dots, c_n$.

   **פתרון:** נקבל זמן ריצה של $\Theta(n^2)$.
   בצורה הכי נאיבית, נרוץ בצורה איטרטיבית על $k$ (כלומר, באיטרציה ה$\ell$ יהיה לנו ביטוי קנוני של $\prod_{k=1}^{\ell}(x - c_k)$). בשלב ה-$\ell$, הביטוי הקנוני הוא למעשה פולינום מדרגה $\ell - 1$ ומכיל בסה"כ $\ell$ מקדמים, ואנו מכפילים אותו במונום $(x - c_\ell)$, ומקבלים כתוצאה פולינום בצורה קנונית מדרגה $\ell$. כלומר, עם כל איטרציה דרגת הפולינום הנוכחי גדלה ב-1. בכל שלב כזה, עלות העבודה היא $\Theta(\ell)$. נקבל $n$ איטרציות, כאשר באיטרציה ה-$\ell$-ית אנחנו מבצעים $\Theta(\ell)$ עבודה. לפיכך בסה"כ עלות האלגוריתם היא־ $\Theta(n^2)$ (עלות העבודה היא: $\sum_{\ell=1}^{n} \Theta(\ell)$).

2. נניח איבר $\alpha$ כלשהו. מהי עלות חישוב הנקודה $B(\alpha)$?

   **פתרון:** יש לנו פה $n$ מכפלות, ו-$n$ חיסורים. נקבל בסה"כ $\Theta(n)$.

3. מהי עלות חישוב הנקודה $A(\alpha)$?

   **פתרון:** בצורה הכי נאיבית, אנחנו צריכים לחשב את כל החזקות $\alpha, \alpha^2, \dots, \alpha^n$ (בהינתן החזקות מגיעים לתשובה תוך $n$ מכפלות וחיבורים, כלומר, תוך $2n$ עבודה). בצורה הכי נאיבית, מחשבים

כל חזקה מחדש, ללא שמירת הערך הקודם. עלות חישוב $\alpha^k$ הוא $k-1$ מכפלות, ולכן סה"כ כמות המכפלות לחישוב החזקות הוא:

$$0+1+2+\ldots+(n-1)=\Theta(n^2)$$

נקבל אם כן עלות האלגוריתם $\Theta(n^2)$.

4. בכדי להשיג זמן ריצה מהיר יותר, נניח מחשבים $A(\alpha)$ בצורה הבאה:

$$A(\alpha)=a_0+\alpha\cdot(a_1+\alpha\cdot(a_2+\cdots+\alpha\cdot(a_{n-1}+a_n\cdot\alpha)\cdots)).$$

מהי עלות חישוב $A(\alpha)$ כעת?

**פתרון:** פה יש לנו רק $n$ מכפלות, ו-$n$ חיבורים. נקבל אם כן- $\Theta(n)$.

5. הסעיף הבא נועד להדגיש את הכוח באלגוריתמים המשתמשים ב**אקראיות** (אלגוריתם הסתברותי). האלגוריתם ירוץ בזמן מהיר יותר מהאלגוריתם שהוצג לעיל, אך עלול לטעות ולהחזיר תשובה **לא נכונה** בהסתברות נמוכה. נחזור לבעיה של שקילות פולינומים, ונתבונן באלגוריתם הבא. האלגוריתם מקבל כקלט $A(x),B(x)$, **ובוחר באקראי** איבר כלשהו $\alpha\in\{1,2,\ldots,c\cdot n\}$ ($c$ קבוע חיובי כלשהו שנבחר אותו בהמשך). האלגוריתם מחשב $A(\alpha)$ ו- $B(\alpha)$. אם הנ"ל שווים, האלגוריתם מחזיר "הפולינומים שקולים". אם התשובות שונות, האלגוריתם מחזיר "הפולינומים שונים".

(א) מהו זמן הריצה של האלגוריתם?

**פתרון:** כאמור, עלות חישוב $A(\alpha)$ הוא $\Theta(n)$, ועלות חישוב $B(\alpha)$ הוא $\Theta(n)$. נקבל אם כן-$\Theta(n)$.

(ב) אם מתקיים $A(x)\equiv B(x)$. מהי ההסתברות שהאלגוריתם יחזיר "הפולינומים שקולים"?

**פתרון:** במקרה ו-$A(x)\equiv B(x)$, נקבל כי לכל $\alpha$: $A(\alpha)=B(\alpha)$. לפיכך, האלגוריתם יחזיר הפולינומים שקולים בהסתברות 1, שכן לא משנה מהו הערך $\alpha$ שנבחר.

(ג) **(בונוס)** אם מתקיים $A(x)\not\equiv B(x)$, מצא חסם עליון להסתברות שהאלגוריתם **יטעה**. כלומר, מהי ההסתברות שהאלגוריתם יחזיר "הפולינומים שקולים?" מה צריך להיות ערכו של $c$ בכדי שהסתברות השגיאה של האלגוריתם תהיה קטנה מ-$1/1000$?

**הנחיה:** אם $A(x)\not\equiv B(x)$, כמה נקודות לכל היותר מקיימות $A(\alpha)=B(\alpha)$?

**פתרון:** נתבונן בפולינום $C(x)\stackrel{\text{def}}{=}A(x)-B(x)$. פולינום זה מדרגה $n$, ולפיכך יש לו לכל היותר $n$ שורשים (כלומר, $n$ ערכי $\alpha$ עבורם $C(\alpha)=0$). נשים לב ש- $C(\alpha)=0$ אם ורק אם $A(\alpha)=B(\alpha)$. לפיכך, ישנם לכל היותר $n$ "נקודות חיתוך" בין $A(x)$ ו- $B(x)$. נקבל אם כן שבמקרה ו-$A(x)\not\equiv B(x)$, ישנם לכל היותר $n$ ערכי $\alpha$ "רעים" שעבורם האלגוריתם טועה. מהי ההסתברות ש-$\alpha$ שבחרנו ייפול באותה קבוצה רעה? נשים לב ש$\alpha$ נבחר מתוך קבוצה מגודל $c\cdot n$, לפיכך:

$$\Pr\left[A(\alpha)=B(\alpha)\right]\le\frac{n}{c\cdot n}=\frac{1}{c}$$

לפיכך, אם נבחר $c=1000$, נקבל כי ההסתברות לטעות קטנה מ-$1/1000$.

(ד) **(בונוס)** הציעו דרך **להקטין** את שגיאת האלגוריתם תוך שימוש באותו סדר גודל של זמן ריצה, וללא שינוי של הפרמטר $c$.

**פתרון:** נריץ את האלגוריתם שראינו - $k$ פעמים, עבור $k$ קבוע כלשהו. אם בכל הריצות קיבלנו הפולינומים שקולים, נחזיר הפולינום שקולים. אם יש ריצה אחת שעבורה הפלט היה הפולינומים שונים, נחזיר הפולינומים שונים.
נקבל שאם $A(x) \equiv B(x)$, הפלט תמיד יהיה הפולינומים שקולים.
אם $A(x) \not\equiv B(x)$, אזי האלגוריתם יטעה ויחזיר בטעות - הפולינומים שקולים אם בכל הריצות נקבל הפולינומים שקולים, כלומר, אנחנו צריכים לטעות בכל הריצות. זה קורה בהסתברות $(1/1000)^k$.
נקבל אם כן שבעוד שהגדלנו את זמן הריצה בצורה ליניארית ($\Theta(k \cdot n)$ במקום $\Theta(n)$), הסתברות השגיאה ירדה בצורה אקספוננציאלית ($(1/1000)^k$ במקום $(1/1000)$).

**פסח כשר ושמח!**