אלגוריתמים אקראיים

[עד עכשיו דיברנו על מודל חישובי של מכשיר אחד עובד בצורה של "לקבל קלט, לבצע חישוב מקומי ולתת פלט." בהרצאה זו נדבר על מודל חישובי המשתמש במספר מכשירים שונים.]

פרוטוקולי תקשורת

[<u>הערה</u>: בפרוטוקולי תקשורת, הפרוטוקול ידוע לשני הצדדים מראש.]
נתחיל ב:

דוגמה [אנסטסיה ובוריס]

יש שני סטודנטים, ~~Alice~~ אנסטסיה ~~Bob-ו~~ ובוריס, שמגישים יחד עבודות ב-PPL.

אנסטסיה יצאה לחופשה, ובזמן שהייתה לה בנופש בדקה את העבודה שבוריס היה אמור להגיש, וגילתה באמצע הקובץ קללה יפה ועסיסית. היא לא בטוחה אם בוריס רוצה להגיע לוועדת משמעת או אם היה שיבוש בתקשורת והיא קיבלה קובץ שהמידע בו השתבש.
היא רוצה לדעת אם העבודה שיש אצלה והעבודה שיש אצל בוריס זהות או לא.

לאנסטסיה יש קלט $x \in \{0,1\}^n$ [[זה הקובץ שנמצא אצלה]] ולבוריס יש קלט $y \in \{0,1\}^n$ [[זה הקובץ שנמצא אצלו, ויש לו אותו אצלו במחשב בלי קשר למה שהוא יקבל מאנסטסיה בהמשך]].
<u>מטרה</u>:

1. אנסטסיה שולחת הודעה z לבוריס.
2. בוריס מקבל את z, ויש לו את y, ומחליט על סמך z, y האם $x = y$ ושולח לאנסטסיה את התשובה [אורך ההודעה שבוריס שולח הוא תמיד ביט 1 (כן / לא)].

רוצים למזער את מס' הביאים בהודעה z של אנסטסיה [[ושהדבר עדיין יעבוד]].
[הערה: אנו מניחים (קצת בסתירה לסיפור הרקע) שהתקשורת בין אנסטסיה ובוריס עובדת נהדר ואין שיבושים במידע שמועבר.]

פתרון פשוט

אנסטסיה שולחת את x.
$\Leftarrow$ עלות התקשורת $= n$ [או $n+1$, אם מחשיבים את ביט התשובה שבוריס שולח].

<u>טענה</u>:
אם אנסטסיה שולחת פחות מ-n ביטים (בפרוטוקול **כלשהו**) אזי קיים קלט עבורו בוריס יטעה.

<u>הוכחה</u>:

- א' מחשבת פונקציה $z = f(x)$.
- x יכול לקבל 2^n ערכים שונים.
- אם z באורך $\le n-1$ אזי z יכול לקבל $\le 2^{n-1}$ ערכים שונים.
- לפי עיקרון שובך היונים קיימים $x_1, x_2 \in \{0,1\}^n$ כך ש-$x_1 \neq x_2$ כך ש-$f(x_1) = f(x_2)$.
- אם $y = x_1$, בוריס יחזיר את אותה תשובה בין אם $x = x_1$ ובין אם $x = x_2$ (כי הוא מקבל את אותה הודעה מאנסטסיה [[$z = f(x_1) = f(x_2)$]]), ובאחד משני המקרים הוא יטעה.

אלגוריתם אקראי פשוט

אנטסיה מגרילה מיקום אקראי $b \in \{1, \dots, n\}$ ושולחת את המיקום והסיבית:

$$z = b, x_b$$

בוריס בודק את הביט ה- b ומחזיר תשובה לפי $y_b = x_b$.

עלות התקשורת: כ- $\log n$ ביטים, בשביל לקודד את b.

[זאת כי עם k ביטים ניתן לייצג את המספרים $0,1,\dots,2^k - 1$, אז עבור $2^k = n$ נקבל $k = \log_2 n$.]

- אם $x = y$, התשובה תמיד תהיה "כן".
  [מצב זה נקרא **טעות חד-צדדית** (אנחנו יודעים בוודאות שבמקרה הראשון זה תמיד יצליח, אבל במקרה השני לאו דווקא).]
- אם $x \neq y$, במקרה הגרוע רק ביט אחד ב- x, y שונה ובוריס יחזיר תשובה נכונה ("לא") בהסתברות $\frac{1}{n}$.

אלגוריתם Karp-Rabin

הקדמה

נחשוב על x, y כמספרים $\geq 2^n$ [[במקום מחזורות של n ביטים, שכן הם שקולים]].

1. אנטסיה מגרילה מס' ראשוני $p < n^2$ באופן אחיד.
2. אנטסיה מחשבת את $z_1 = x \bmod p$.
3. אנטסיה שולחת את $z_1 = x \bmod p$ ו- $z_2 = p$.
4. בוריס בודק האם $z_1 = y \bmod p$ $\left(x \bmod p =\right)$ ועונה בהתאם.

<u>עלות התקשורת</u>:

$$z: \begin{array}{l} z_1 < p < n^2 \\ z_2 = p < n^2 \end{array}$$

לייצג מס' עד n^2 צריך $\log_2\left(n^2\right) = 2\log_2 n$ ביטים.

לכן בסה"כ אנטסיה שלחה $4\log_2 n = \mathrm{O}\left(\log n\right)$ ביטים.

אם $x = y$, ההסתברות לתשובה נכונה ("כן") היא 1 [כי תמיד נקבל ש- $x \bmod p = y \bmod p$].
כעת נדון במקרה בו $x \neq y$.

- <u>עובדה 1</u>: **משפט המספרים הראשוניים**: מס' הראשוניים $\geq t$ שווה אסימפטוטית ל- $\left(1 + \mathrm{o}(1)\right)\frac{t}{\ln t}$ [שזה $\leq \frac{t}{\ln t}$ עבור t גדול דיו].

- <u>עובדה 2</u>: לכל מספר w, מס' הראשוניים שמחלקים את w הוא לכל היותר $\log_2 w$.

<u>הוכחה</u>: נייצג את w לפי פירוק לראשוניים $w = p_1 p_2 p_3 \cdots p_k$ [$k =$ מס' הגורמים עם ריבוי (בלי ריבוי זה אפילו מספר קטן יותר)].
כל $2 \le p_i$, ולכן $w \ge 2^k$.
$\Leftarrow k \le \log_2 w$.

<u>משפט</u>:
אם $x \neq y$, ההסתברות שבוריס יחזיר "כן" היא $O\left(\frac{\log n}{n}\right)$.
(ההסתברות לתשובה נכונה שואפת ל-1).

<u>הוכחה</u>:
בוריס יחזיר "כן" אם:

$$x \bmod p = y \bmod p$$
$$\Updownarrow$$
$$(x-y) \bmod p = 0$$
$$\Updownarrow$$
$$(y-x) \bmod p = 0$$

נגדיר $w = |x-y| \le 2^n$.
רוצים לדעת מה ההסתברות (על פני הבחירה האקראית של p) ש-$w \bmod p = 0$.
כלומר ש-w מתחלק ב-p.

[ציור של ראשוניים טובים ורעים]

לפי עובדה 1, מס' ה-p-ים שאנסטסיה יכולה לבחור הוא **לפחות** $\frac{n^2}{\ln(n^2)}$.
לפי עובדה 2, מס' ה-p-ים שמחלקים את w הוא **לכל היותר** $\log_2 w \le \log_2(2^n) = n$.
כלומר ההסתברות לטעות [שמתרחשת אם אנסטסיה בחרה p שמחלק את w] היא:

$$\frac{\text{מס' ה-}p\text{-ים הרעים} \le \log_2(2^n) = n}{\text{מס' ה-}p\text{-ים שא' יכולה לבחור} \ge \frac{n^2}{\ln(n^2)} = \frac{n^2}{2\ln n}} \le \frac{n}{\left(\frac{n^2}{2\ln n}\right)} = \frac{2\ln n}{n} = O\left(\frac{\log n}{n}\right)$$

שיפור האלגוריתם [טריק חשוב באלגוריתמים אקראיים]
[בכל אלגוריתם אקראי שמניב הסתברות q לכישלון, ניתן להפעיל את האלגוריתם באופן בלתי-תלוי k פעמים ולקבל הסתברות q^k לכישלון (שזה לרוב הרבה יותר קטן).]

- אנסטסיה תגריל באופן בלתי תלוי (ועם חזרות) t מספרים ראשוניים, $p_1,\dots,p_t < n^2$ ותשלח:

$$p_1,\left(x \bmod p_1\right)$$
$$p_2,\left(x \bmod p_2\right)$$
$$\vdots$$
$$p_t,\left(x \bmod p_t\right)$$

- בוריס מחזיר "כן" <u>רק</u> אם $y \bmod p_i = \left(x \bmod p_i\right)$ **לכל** $1 \le i \le t$.

$\Leftarrow$ עם t חזרות, ההסתברות לתשובה לא נכונה היא לכל היותר:

$$\left(\frac{2\ln n}{n}\right)^t$$

דוגמה: n = 1500

מס' הביטים באיטרציה אחת:

$$4\log_2\left(1500\right) \le 30$$

ההסתברות לשגיאה:

$$\frac{2\ln\left(1500\right)}{1500} < \frac{1}{100}$$

נבצע $t = 5$ איטרציות.
מספר הביטים שנשלחו:

$$5 \cdot \left(4\log_2\left(1500\right)\right) \le 150$$

וההסתברות לשגיאה:

$$\left(\frac{2\ln\left(1500\right)}{1500}\right)^5 < \left(\frac{1}{100}\right)^5 = \left(\frac{1}{10}\right)^{10} \le p_L$$

כאשר p_L היא ההסתברות שברק יפגע במחשב.

[לתקן: ln במקום log]