29.03.2023 ***ניתוח מידע אדפטיבי***

הרצאה 3: אורך יצוג חסום

Source: Lecture notes by Aaron Roth and Adam Smith

מרצה: אורי שטמר

בשיעור שעבר הגדרנו את המטרה הבאה: אנחנו רוצים לתכנן מכניזמים (α,β)-מדוייקים-סטטיסטית עבור k שאילתות אדאפטיביות בהינתן מדגם בגודל n, כאשר המטרה שלנו היא שגודל המדגם n יהיה כמה שיותר קטן כפונקציה של k,α,β.

ראינו את ה baseline שמקבלים בעזרת sample splitting. אבל זה לא מאפשר לנו לענות על יותר ממספר לינארי של שאילתות (כלומר לא מאפשר לנו ש k יהיה יותר מלינארי ב n).

איך נוכל לעשות יותר טוב?

לפני שנתחיל לתכנן מכניזמים טובים יותר, נראה את הטענה הבאה שתקל לנו על החיים. באופן כללי, נרצה שההבטחות שלנו יתקיימו גם אם האנליסט הוא אקראי. הטענה הבאה מראה שבלי הגבלת הכלליות נוכל להניח שהאנליסט הוא דטרמניסטי:

<u>טענה 1:</u> כדי להראות שמכניזם M הוא (α,β)-מדוייק, מספיק להראות שהוא מדוייק לכל אנליסט דטרמניסטי, כלומר מספיק להראות שלכל התפלגות $\mathcal{D}$ ולכל אנליסט דטרמניסטי A מתקיים:

$$\Pr_{\substack{S\sim\mathcal{D}^n \\ \mathrm{AG}_{n,k}(A,S,M)}}[\exists i \text{ s.t. } |a_i - q_i(\mathcal{D})| > \alpha] \le \beta$$

<u>הוכחה:</u> נניח שלכל התפלגות $\mathcal{D}$ ולכל אנליסט דטרמניסטי A מתקיים:

$$\Pr_{\substack{S\sim\mathcal{D}^n \\ \mathrm{AG}_{n,k}(A,S,M)}}[\exists i \text{ s.t. } |a_i - q_i(\mathcal{D})| > \alpha] \le \beta$$

יהי A אנליסט אקראי ונסמן ב r את המטבעות האקראיים שלו. אזי

$$\Pr_{\substack{S\sim\mathcal{D}^n \\ \mathrm{AG}_{n,k}(A,S,M)}}[\exists i \text{ s.t. } |a_i - q_i(\mathcal{D})| > \alpha] = \sum_r \Pr[r] \cdot \Pr_{\substack{S\sim\mathcal{D}^n \\ \mathrm{AG}_{n,k}(A,S,M)}}\left[\exists i \text{ s.t. } |a_i - q_i(\mathcal{D})| > \alpha \;\middle|\; r\right]$$

$$\le \sum_r \Pr[r] \cdot \beta = \beta$$

מ.ש.ל.

מסקנה: בהמשך הקורס נוכל להניח בה"כ שהאנליסט (או היריב) הוא דטרמניסטי.

Compression

עכשיו נתחיל לדבר על שיטה נוספת (מעבר ל sample splitting). אנחנו נראה שאם למכניזם יש את התכונה הבאה, שנקראת דחיסה, אז הוא מבטיח דיוק עבור שאילתות אדפטיביות (עם פרמטרים שתלויים באיכות הדחיסה שהוא מבטיח).

<u>הגדרה 2:</u> נאמר שמכניזם M מאפשר <u>דחיסת-טרנסקריפט</u> ל- $b(n,k)$ ביטים אם <u>לכל</u> אנליסט A קיימת קבוצת טרנסקריפטים אפשריים H_A בגודל $|H_A| \leq 2^{b(n,k)}$ כך שלכל מדגם S מתקיים

$$\Pr\left[AG_{n,k}(A,S,M) \in H_A\right] = 1$$

<u>הערה:</u> קבוצת הטרנסקריפטים H_A יכולה להיות תלויה באנליסט A. כלומר זאת יכולה להיות קבוצה שונה עבור אנליסטים שונים.

<u>עוד הערה:</u> מכיוון שאנחנו מניחים (בה"כ) שהאנליסט הוא דטרמניסטי, מספיק לתאר טרנסקריפט $(q_1, a_1, \dots, q_k, a_k)$ על ידי <u>התשובות</u> בלבד, כלומר על ידי $(a_1, \dots, a_k)$ מכיוון שכאשר האנליסט דטרמניסטי אז כל q_i באופן חד משמעי מתוך $(a_1, \dots, a_{i-1})$ ומההגדרה של A.

<u>משפט 3:</u> יהי M מכניזם המאפשר <u>דחיסת-טרנסקריפט</u> ל- $b(n,k)$ ביטים. אזי <u>לכל</u> אנליסט A ולכל התפלגות $\mathcal{D}$ מתקיים:

$$\Pr_{\substack{S\sim\mathcal{D}^n \\ \mathrm{AG}_{n,k}(A,S,M)}} [\exists i \text{ s.t. } |q_i(S) - q_i(\mathcal{D})| > \alpha] \leq \beta$$

עבור

$$\alpha = O\left(\sqrt{\frac{b(n,k) + \ln(k/\beta)}{n}}\right)$$

<u>הוכחה:</u> נקבע אנליסט A ונקבע התפלגות $\mathcal{D}$. נזכור שלאחר שקבענו את A ישנה קבוצה H_A של לכל היותר $2^{b(n,k)}$ טרנסקריפטים אפשריים באינטראקציה בין A ל- M. בכל טרנסקריפט כזה ישנם לכל היותר k שאילתות סטטיסטיות שונות, ולכן ישנם לכל היותר $w = k \cdot 2^{b(n,k)}$ שאילתות סטטיסטיות אפשריות במהלך הריצה. נסמן קבוצת שאילתות זהו כ- Q_A. אזי, לפי משפט שראינו בהרצאות הקודמות (עבור המקרה הלא-אדפטיבי) מתקיים:

$$\Pr_{S\sim\mathcal{D}}\left[\max_{q\in Q_A}|q(S) - q(\mathcal{D})| > \sqrt{\frac{\ln(2w/\beta)}{2n}}\right] \leq \beta$$

בפרט,

$$\Pr_{\substack{S\sim\mathcal{D}^n \\ \mathrm{AG}_{n,k}(A,S,M)}}\left[\exists i \text{ s.t. } |q_i(S) - q_i(\mathcal{D})| > \sqrt{\frac{\ln(2w/\beta)}{2n}}\right] \leq \beta$$

מ.ש.ל.

הערות:

1. עדיין לא סיימנו. כל מה שהראינו זה שאם מכניזם M מאפשר דחיסת-טרנסקריפט אזי $q_i(S) \approx q_i(\mathcal{D})$. מה שאנחנו באמת רוצים להראות זה $a_i \approx q_i(\mathcal{D})$.
2. בפרט, אם המכניזם שלנו מתעלם מהשאילתות ותמיד מחזיר $a_i = 0$, אזי הוא מאפשר דחיסת-טרנסקריפט ל- 0 ביטים, ולכן $q_i(S) \approx q_i(\mathcal{D})$ לכל i, אבל התשובות שהוא מחזיר כמובן לא מועילות.
3. אנחנו נתעניין במכניזמים המאפשרים דחיסת-טרנסקריפט ובנוסף מבטיחים שהתשובות שהם מחזירים קרובות לערך האמפירי של השאילתות.

<u>הגדרה 4:</u> מכניזם M הוא $(\boldsymbol{\alpha}, \boldsymbol{\beta})$**-מדוייק-אמפירית** עבור k שאילתות אדאפטיביות בהינתן מדגם בגודל n אם <u>לכל</u> מדגם S בגודל n <u>ולכל</u> אנאליסט A מתקיים

$$\Pr_{\mathrm{AG}_{n,k}(A,S,M)}[\exists i \text{ s.t. } |a_i - q_i(S)| > \alpha] \leq \beta$$

<u>משפט 5:</u> יהי M מכניזם אשר

א. M מאפשר דחיסת-טרנסקריפט ל- $b(n,k)$ ביטים
ב. M הוא (α', β')-מדוייק-אמפירית עבור k שאילתות אדאפטיביות בהינתן מדגם בגודל n

אזי, לכל $\beta'' > 0$ מתקיים ש- M הוא (α, β)-מדוייק-סטטיסטית עבור $\beta = \beta' + \beta''$ ועבור

$$\alpha = \alpha' + O\left(\sqrt{\frac{b(n,k) + \ln(k/\beta'')}{n}}\right)$$

<u>הוכחה:</u>

נסמן $\alpha'' = O\left(\sqrt{\frac{b(n,k)+\ln(k/\beta'')}{n}}\right)$. מכיוון ש- M הוא (α', β')-מדוייק-אמפירית וגם מאפשר דחיסת-טרנסקריפט ל- $b(n,k)$ ביטים, לפי חסם האיחוד נקבל

$$\Pr_{\substack{S\sim\mathcal{D}^n \\ \mathrm{AG}_{n,k}(A,S,M)}}\left[\left\{\exists i \text{ s.t. } |q_i(S) - q_i(\mathcal{D})| > \alpha''\right\} \text{ OR } \left\{\exists i \text{ s.t. } |a_i - q_i(S)| > \alpha'\right\}\right] \leq \beta'' + \beta'$$

כלומר, בהסתברות לפחות $1 - \beta'' - \beta'$, לכל i מתקיים

$$|a_i - q_i(S)| \leq \alpha' \qquad \text{וגם} \qquad |q_i(S) - q_i(\mathcal{D})| \leq \alpha''$$

במקרה זה, לפי אי-שיוויון המשולש, נקבל ש-

$$|a_i - q_i(\mathcal{D})| \leq \alpha'' + \alpha'$$

מ.ש.ל.

בעצם, מסתבר שאפשר להכליל את המשפט האחרון שהוכחנו גם מעבר לשאילתות סטטיסטיות, למקרה של <u>שאילתות עם רגישות נמוכה</u>:

<u>הגדרה 6:</u> תהי $q: X^n \to \mathbb{R}$ פונקציה שממפה דטהבייסים בגודל n מעל X ל- $\mathbb{R}$. לפונקציה q יש <u>רגישות</u> c אם לכל $x_1, \dots, x_n \in X$ לכל $i \in [n]$ ולכל $x_i' \in X$ מתקיים:

$$|q(x_1, \dots, x_n) - q(x_1, \dots, x_{i-1}, x_i', x_{i+1}, \dots, x_n)| \leq c$$

<u>שימו לב:</u> ששאילתות סטטיסטיות הן פונקציות עם רגישות $c = 1/n$

<u>הגדרה 7:</u> עבור שאילתא q שהיא פונקציה עם רגישות c ועבור מדגם $S \sim \mathcal{D}^n$ נסמן ב- $q(S)$ את הערך האמפירי של q על S, ונסמן את הערך "האמיתי" כ-

$$q(\mathcal{D}) = \mathbb{E}_{T \sim \mathcal{D}^n}[q(T)]$$

<u>משפט 8 (הכללה של משפט 5):</u> יהי M מכניזם אשר בהינתן מדגם בגודל n עונה על שאילתות עם רגישות $1/n$ כך ש-

א. M מאפשר דחיסת-טרנסקריפט ל- $b(n,k)$ ביטים

ב. M הוא (α', β')-מדוייק-אמפירית אמפירית עבור k שאילתות אדאפטיביות בהינתן מדגם בגודל n

אזי, לכל $\beta'' > 0$ מתקיים ש- M הוא (α, β)-מדוייק-סטטיסטית עבור $\beta = \beta' + \beta''$ ועבור

$$\alpha = \alpha' + O\left(\sqrt{\frac{b(n,k) + \ln(k/\beta'')}{n}}\right)$$

ההוכחה של ההכללה הזאת דומה מאוד להוכחה שראינו. לא נראה את הפרטים כאן.

(כל מה שצריך לשנות זה שבמקום השימוש שעשינו במשפט צ'רנוף, צריך להשתמש בהכללה של צ'רנוף עבור פונק' עם רגישות נמוכה, שנקראת McDiarmid's inequality)

אוקיי. אז אנחנו יודעים שאם מכניזם גם מאפשר דחיסת-טרנסקריפט וגם מבטיח דיוק-אמפירי, אז אנחנו מקבלים גם דיוק-סטטיסטי. איך נבנה מכניזמים כאלה שמבטיחים את 2 התכונות האלה? נתחיל עם דוגמה פשוטה (ולא מפתיעה במיוחד). בהמשך נדבר על מכניזמים יותר מסובכים (שמשיגים תוצאות טובות יותר).

<u>הגדרה 9:</u> נגדיר את "מכניזם הקיטום", שיסומן כ- M^b_{Tranc}, באופן הבא. המכניזם מקבל כקלט מדגם S. בהינתן שאילתא q, המכניזם מחזיר את הערך האמפירי $q(S)$, קטום ל- b ביטים.

<u>אבחנה 10:</u> המכניזם M^b_{Tranc} מאפשר דחיסת-טרנסקריפט ל- $b(n,k) = bk$ ביטים.

<u>הסבר:</u> כפי שאמרנו, בה"כ אנחנו מניחים שהאנליסטים שלנו הם דטרמניסטיים. לכן, לאחר שקבענו אנליסט, נוכל לייצג את הטרנסקריפט ע"י רשימת התשובות שהמכניזם מחזיר. מכיוון שכל תשובה מיוצגת ע"י b ביטים ומכיוון שישנם לכל היותר k תשובות בטרנסקריפט, נקבל שכל טרנסקריפט מיוצג ע"י bk ביטים לכל היותר. לכן ישנם לכל היותר 2^{bk} טרנסקריפטים אפשריים.

<u>אבחנה 11:</u> המכניזם M^b_{Tranc} הוא $\left(\frac{1}{2^b}, 0\right)$-מדוייק-אמפירית אמפירית.

<u>הסבר:</u> בהינתן שאילתא q, המכניזם מחזיר את הערך האמפירי $q(S)$, קטום ל- b ביטים, מה שיכול לעוות את התשובה בלכל היותר $1/2^b$.

<u>מסקנה:</u> יהי $\beta > 0$ ויהי $b = \log\left(\sqrt{\frac{n}{k}}\right)$. מכניזם הקיטום M^b_{Tranc} הוא (α, β)-מדוייק-סטטיסטית עבור

$$\alpha = O\left(\frac{1}{2^b} + \sqrt{\frac{bk + \ln(k/\beta)}{n}}\right) = \tilde{O}\left(\sqrt{\frac{k + \ln(1/\beta)}{n}}\right)$$

דיון: השגיאה כאן נראית כמו $\sqrt{k/n}$ מה זה אומר לי על הקשר בין k ל- n? איך זה ביחס למה שקיבלנו בעזרת sample splitting? מה הרווחנו "מחשבתית"?

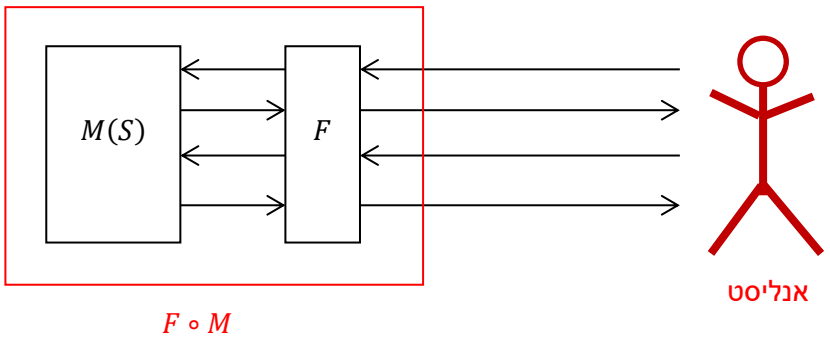
המטרה הבאה: להתחיל לדבר על דוגמאות יותר רציניות. בשביל זה, נרצה להיות מסוגלים להציג מכניזמים פשוטים המאפשרים דחיסת-טרנסקריפט, ולאחר מכן לתכנן מכניזמים מסובכים יותר שמשתמשים במכניזמים הקודמים שהצגנו כ subroutines. לשם כך נצטרך את התכונות הבאות.

Transcript Compressibility: Composition and Post-processing

Post-processing

נניח שיש לנו מכניזם M שמאפשר דחיסת-טרנסקריפט. יהי F מכניזם (דטרמניסטי) כלשהו אשר מאפשר "לעבד" תשובות ש- M פולט (לפני שהתשובות האלה ניתנות לאנליסט) ומאפשר "לעבד" את שאילתות שהאנליסט בוחר (לפני שהשאילתות האלה ניתנות ל- M). מה נוכל לומר על השילוב של M ו- F, שנסמנו $F \circ M$?

בציור:



<u>משפט 12:</u> אם M מאפשר דחיסת-טרנסקריפט ל- b ביטים אזי לכל F כנ"ל מתקיים ש- $F \circ M$ מאפשר דחיסת-טרנסקריפט ל- b ביטים.

הערה: משפט 12 נשאר נכון גם אם ל- F יש מצב, כלומר גם אם העיבוד שהוא מבצע בשלב ה i תלויי בכל הדברים שהוא ראה בשלבים הקודמים.

<u>הוכחה:</u> נקבע אנליסט A. עלינו להראות שקיימת קבוצת טרנסקריפטים H בגודל $|H| \leq 2^b$ כך שלכל מדגם S מתקיים

$$\Pr\left[AG_{n,k}(A, S, F \circ M) \in H\right] = 1$$

לצורך כך, נזכר בהגדרת המשחק $\text{AdaptiveGame}_{n,k}(A, S, F \circ M)$, כאשר נרשום בצורה מפורשת את הפעולה של F:

$\textbf{AdaptiveGame}_{n,k}(A, S, F \circ M)$
1. המכניזם M מקבל את המדגם S (האנליסט A לא מקבל את המדגם S) 2. עבור $i = 1,2, \dots, k$: • האנליסט A קובע שאילתא סטטיסטית q_i • המכניזם F מקבל את השאילתא q_i ופולט שאילתא $\hat{q}_i$. נסמן $\hat{q}_i = F(q_i)$ • המכניזם M מקבל את השאילתא $\hat{q}_i$ ומחזיר תשובה a_i • המכניזם F מקבל את התשובה a_i ופולט $\hat{a}_i$. נסמן $\hat{a}_i = F(a_i)$ • האנליסט A מקבל את $\hat{a}_i$ 3. החזר את <u>הטרנסקריפט</u> של האינטראקציה עם האנליסט: $T = (q_1, \hat{a}_1, q_2, \hat{a}_2, \dots, q_k, \hat{a}_k)$

אנחנו רוצים להראות שהטרנסקריפט $(q_1, \hat{a}_1, q_2, \hat{a}_2, \dots, q_k, \hat{a}_k)$ ניתן לדחיסה, כלומר הטרנסקריפט בשיחה בין האנליסט A לבין המכניזם $F \circ M$. כפי שאמרנו, מכיוון שהאנליסט דטרמניסטי, מספיק להראות זאת עבור החלק של הטרנסקריפט שמכיל את התשובות, כלומר עבור $(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_k)$.

נקודת מבט נוחה: לצורך הניתוח אנחנו יכולים לחשוב על F <u>כחלק מהאנליסט</u> במקום כחלק מהמכניזם (נסמן ב A' את האנליסט הזה שמשלב בין A ו- F). כלומר אנו מסתכלים על הטרנסקריפט $T' = (\hat{q}_1, a_1, \hat{q}_2, a_2, \dots, \hat{q}_k, a_k)$ שיושב בתווך בין F לבין M.

מכיוון ש- M מאפשר דחיסת טרנסקריפט ל- b ביטים, ומכיוון שההגדרה של דחיסת-טרנסקריפט מדברת על <u>כל</u> אנליסט, אנחנו יודעים שבתווך בין M לבין F ישנם לכל היותר 2^b טרנסקריפטים אפשריים. נסמן קבוצה זו ע"י $H_{A'}$ כאשר $|H_{A'}| \leq 2^b$.

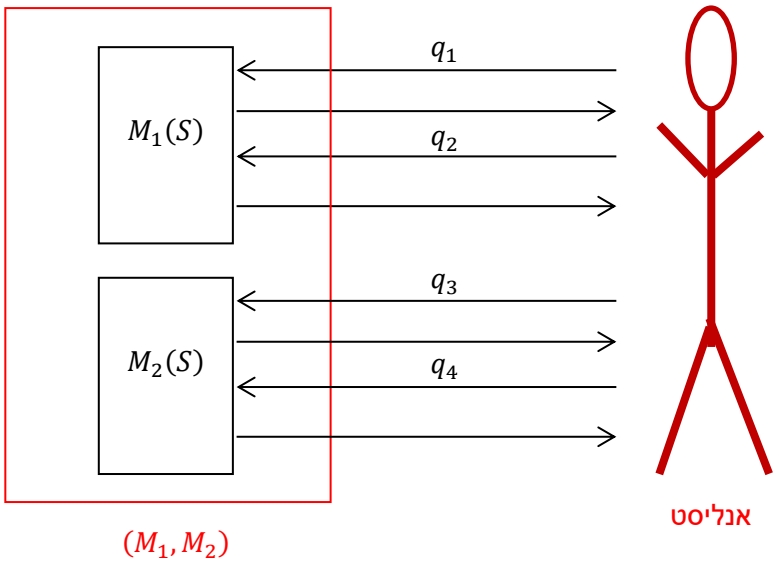
מכיוון ש- F דטרמניסטי, כל טרנסקריפט מ $H_{A'}$ מתורגם לטרנסקריפט אחד של תשובות מתורגמות $(\hat{a}_1, \hat{a}_2, \dots, \hat{a}_k)$. לכן מספר האפשרויות לטרנסקריפטים של תשובות מתורגמות הוא לכל היותר 2^b.

* המספר יכול להיות יותר קטן, אם 2 טרנסקריפטים מ $H_{A'}$ גוררים את אותו טרנסקריפט של תשובות מתורגמות, אבל זה לא יכול להיות יותר מ $|H_{A'}|$

Composition

נניח שיש לנו שני מכניזמים M_1, M_2 שכל אחד מהם מאפשר דחיסת-טרנסקריפט. נחשוב על המכניזם (M_1, M_2) אשר מקבל מדגם S, מזין אותו גם ל- M_1 וגם ל- M_2, ולאחר מכן עונה על k_1 שאילתות בעזרת M_1 ואז עונה על k_2 שאילתות בעזרת M_2.

בציור:



משפט 13: אם M_1 מאפשר דחיסת-טרנסקריפט ל- $b_1(n, k_1)$ ביטים ו- M_2 מאפשר דחיסת-טרנסקריפט ל- $b_2(n, k_2)$ ביטים, אזי (M_1, M_2) מאפשר דחיסת-טרנסקריפט ל- $b(n, k_1, k_2) = b_1(n, k_1) + b_2(n, k_2)$ ביטים.

הוכחה: כמו קודם, נקבע אנליסט A. עלינו להראות שקיימת קבוצת טרנסקריפטים H בגודל $|H| \leq 2^{b(n,k_1,k_2)}$ כך שלכל מדגם S מתקיים

$$\Pr\left[AG_{n,(k_1+k_2)}\left(A, S, (M_1, M_2)\right) \in H\right] = 1$$

לצורך כך, נזכר בהגדרת המשחק $AG_{n,(k_1+k_2)}\left(A, S, (M_1, M_2)\right)$, כאשר נרשום בצורה מפורשת את הפעולה של (M_1, M_2):

$\mathbf{AdaptiveGame}_{n,(k_1+k_2)}\left(A, S, (M_1, M_2)\right)$

1. כל אחד מהמכניזמים M_1, M_2 מקבל את המדגם S (האנליסט A לא מקבל את המדגם S)
2. עבור $i = 1, 2, \ldots, k_1$:
   - האנליסט A קובע שאילתא סטטיסטית q_i
   - המכניזם M_1 מקבל את השאילתא q_i ומחזיר תשובה a_i
   - האנליסט A מקבל את a_i
3. עבור $i = k_1 + 1, 2, \ldots, k_1 + k_2$:
   - האנליסט A קובע שאילתא סטטיסטית q_i
   - המכניזם M_2 מקבל את השאילתא q_i ומחזיר תשובה a_i
   - האנליסט A מקבל את a_i
4. החזר את הטרנסקריפט של האינטראקציה: $T = \left(q_1, a_1, , \ldots, q_{k_1+k_2}, a_{k_1+k_2}\right)$

ראשית, מכיוון ש- M_1 מאפשר דחיסת-טרנסקריפט ל- $b_1(n,k_1)$ ביטים, אנחנו יודעים שקיימת קבוצת טרנסקריפטים H_A^1 בגודל $\left|H_A^1\right| \le 2^{b_1(n,k_1)}$ כך שלכל S מתקיים

$$\Pr\left[AG_{n,k_1}(A,S,M_1) \in H_A^1\right] = 1$$

- נסמן ב- $T_1 = \left(q_1, a_1, \dots, q_{k_1}, a_{k_1}\right)$ את החלק של הטרנסקריפט שנוצר מהאינטראקציה בין האנליסט A לבין המכניזם M_1.
- נסמן ב- A_{T_1} את האנליסט A במצב הפנימי שמתקבל לאחר סיום האינטרקציה עם M_1 כאשר הטרנסקריפט הוא T_1.

מכיוון ש- M_2 מאפשר דחיסת-טרנסקריפט ל- $b_2(n,k_2)$ ביטים, אנחנו יודעים שקיימת קבוצה $H_{A_{T_1}}^2$ בגודל $\left|H_{A_{T_1}}^2\right| \le 2^{b_2(n,k_2)}$ כך שלכל מדגם S מתקיים

$$\Pr\left[AG_{n,k_2}\left(A_{T_1},S,M_2\right) \in H_{A_{T_1}}^2\right] = 1$$

נגדיר

$$H_A = \left\{(T_1,T_2) \;:\; T_1 \in H_A^1,\; T_2 \in H_{A_{T_1}}^2\right\}$$

מתקיים

$$\Pr\left[AG_{n,(k_1+k_2)}\left(A,S,(M_1,M_2)\right) \in H_A\right] = 1$$

ובנוסף,

$$|H_A| \le \sum_{T_1 \in H_A^1} \left|H_{A_{T_1}}^2\right| \le 2^{b_1(n,k_1)} \cdot 2^{b_2(n,k_2)} = 2^{b_1(n,k_1)+b_2(n,k_2)}$$

מ.ש.ל.

יופי, אז מכניזמים שמאפשרים דחיסת-טרנסקריפט מקיימים post-processing וגם composition. איך נוכל להשתמש בזה כדי לתכנן מכניזמים טובים?

Algorithm AboveThreshold

נתחיל עם הכלי בסיסי הבא, שלא מספק באופן ישיר תשובות מספריות לשאילתות שהוא מקבל. בהמשך נראה איך להשתמש בכלי הזה כ building block כדי לתכנן מכניזמים מורכבים יותר.

AboveThreshold$(S, T, q_1, q_2, \dots)$

1. AllDone ← FALSE
2. While not AllDone do:
   a) Accept the next query q_i
   b) Compute $a_i \leftarrow q_i(S)$
   c) if $a_i < T$ then return $\perp$
   d) else
      * return $\top$
      * AllDone ← TRUE

משפט 14: לכל ערך של T, אלגוריתם $\text{AboveThreshold}(T)$ מאפשר דחיסת-טרנסקריפט ל- $b(n,k)=\log(k+1)$ ביטים.

הוכחה: סדרת התשובות ש- AboveThreshold מחזיר היא מהצורה $\top$ $\perp^i$ עבור $0 \le i \le k-1$, או $\perp^k$. זוהי קבוצת מחרוזות בגודל $k+1$.

למרות שהכלי הזה נראה פשוט, כמו שנראה בהמשך, הוא כלי חזק בצורה מפתיעה. נתחיל משימוש פשוט שלו.

שימש פשוט ל- AboveThreshold

נתכנן מכניזם אשר בכל שלב מקבל שאילתא q_i (עם רגישות $1/n$) וגם "ניחוש" g_i עבור הערך של השאילתא הזאת. כל עוד הניחוש שלנו באמת קרוב לערך של השאילתא, האלגוריתם מאשר זאת וממשיך לסיבוב הבא. בפעם הראשונה שהניחוש שלנו שגוי, האלגוריתם יודיע על כך, ויחזיר קירוב לערך האמיתי של השאילתא הזאת.

OneWrongGuess$(S, \eta, (q_1, g_1), (q_2, g_2), \dots)$

1. Start an instance of AboveThreshold on S with threshold η
2. While AboveThreshold has not halted do
   a. Accept the next query (q_i, g_i)
   b. Feed AboveThreshold the query $\hat{q}_i = |q_i(S) - g_i|$
   c. if AboveThreshold returns $\perp$ then return the answer $a_i = g_i$
3. Return the answer $a_i = M^b_{\text{Tranc}}(S, q_i)$ for $b = \log(1/\eta)$

משפט 15: לכל $0 < \eta \le 1$ מתקיים ש- OneWrongGuess הוא $(\eta, 0)$-מדוייק-אמפירית וגם מאפשר דחיסת-טרנסקריפט ל- $b(n,k)=\log(k+1)+\log(1/\eta)$ ביטים.

הוכחה:
נחשוב על מכניזם "פוסט-פרוססינג" F אשר מחליף שאילתא (q_i, g_i) בשאילתא $\hat{q}_i(S) = |q_i(S) - g_i|$, ומחליף תשובה $a_i = \perp$ בתשובה $a_i = g_i$.
אז אפשר להציג את OneWrongGuess כ-

$$\left(F \circ \text{AboveThreshold},\ M^b_{\text{Tranc}}\right)$$

כלומר כקומפוזיציה של $F \circ \text{AboveThreshold}$ עם M^b_{Tranc}.

מכיוון ש- AboveThreshold מאפשר דחיסת טרנסקריפט ל- $\log(k+1)$ ביטים, לפי משפט הפוסט-פרוססינג, גם $F \circ \text{AboveThreshold}$ מאפשר דחיסת טרנסקריפט ל- $\log(k+1)$ ביטים. לכן, לפי משפט הקומפוזיציה, אלג' OneWrongGuess מאפשר דחיסת טרנסקריפט ל- $b(n,k)=\log(k+1)+\log(1/\eta)$.

הטענה לגבי הדיוק האמפירי היא טריוויאלית: אנחנו יודעים ש- M^b_{Tranc} הוא $\left(1/2^b, 0\right)$-מדוייק אמפירית, כלומר $(\eta, 0)$-מדוייק אמפירית עבור הבחירה שלנו ל- η. לכל שאר השאילתות שאנחנו עונים, אנחנו מחזירים את התשובה g_i, כאשר לפי הגדרת אלגוריתם AboveThreshold אנחנו יודעים ש- $|g_i - q_i(S)| \le \eta$.

מ.ש.ל.

שאלה: מה אם אנחנו רוצים לעצור רק אחרי הפעם ה-m-ית שבה הניחוש שלנו שגוי?

תשובה: אנחנו יכולים פשוט להריץ m פעמים את OneWrongGuess (פעם אחרי פעם). לפי משפט הקומפוזיציה שהוכחנו, נוכל לנתח את תכונות הדחיסה של האלגוריתם המתקבל.

פורמלית:

GuessAndCheck$(S, \eta, m, (q_1, g_1), (q_2, g_2), \dots)$

1. TimesWrong $\leftarrow 0$
2. While TimesWrong $< m$ do
   a. Start an instance of AboveThreshold on S with threshold η
   b. While AboveThreshold has not halted do
      - Accept the next query (q_i, g_i)
      - Feed AboveThreshold the query $\hat{q}_i(S) = |q_i(S) - g_i|$
      - if AboveThreshold returns $\perp$ then return the answer $a_i = g_i$
   c. Return the answer $a_i = M^b_{\text{Tranc}}(q_i)$ for $b = \log(\mathbf{4}/\eta)$
   d. TimesWrong $\leftarrow$ TimesWrong $+ 1$

<u>משפט 16:</u> לכל η, m, אלגוריתם GuessAndCheck(η, m) הוא $(\eta, 0)$-מדוייק אמפירית וגם מאפשר דחיסת-טרנסקריפט ל- $b(n,k) = m \cdot (\log(k+1) + \log(1/\eta))$ ביטים.

<u>הוכחה:</u> נובע ממשפט הקומפוזיציה, כי אלגוריתם GuessAndCheck הוא קומפוזיציה של m הפעלות של אלגוריתם OneWrongGuess.

<u>הערה:</u> שימו לב שבצעד 2c אנחנו מריצים את מכניזם הקיטום עם פרמטר $b = \log(\mathbf{4}/\eta)$ (ולא $b = \log(\mathbf{1}/\eta)$ כמו קודם). זה אומר שבזמנים שבהם הניחוש שלנו שגוי, אנחנו מחזירים תשובה עם דיוק אמפירי של $\eta/2$ ולא רק η כמו באלגוריתם הקודם. מעבר לכך זה לא באמת משנה משהו בניתוח. השינוי הזה יהיה נח בשבילנו באלגוריתם הבא שנראה.

<u>מסקנה:</u> יהי $m \in \mathbb{N}$ ויהי $\beta > 0$. נסמן $\eta = \sqrt{\frac{m}{n}}$. אזי אלגוריתם GuessAndCheck(η, m) הוא (α, β)-מדוייק-סטטיסטית לכל סדרה של לכל היותר k שאילתות+ניחוש (q_i, g_i), עד שהאלגוריתם עוצר, כאשר כל q_i היא שאילתא עם רגישות $1/n$, עבור

$$\alpha = O\left(\sqrt{\frac{m \cdot \log\left(\frac{kn}{m}\right) + \log\left(\frac{1}{\beta}\right)}{n}}\right)$$

המסקנה נובעת ממשפט 16 הנ"ל, בצירוף העובדה שאם מכניזם הוא גם מבטיח דיוק אמפירי וגם מאפשר דחיסת-טרנסקריפט אז הוא מדוייק-סטטיסטית.

שימו לב שכדי לקבל שגיאה α קטנה, גודל המדגם n צריך לגדול רק לוגריתמית עם k (מספר השאילתות הכולל) שזה מצויין. אבל עדיין, אנחנו חייבים ש- n יהיה גדול משמעותית מ- m (מספר הניחושים השגויים לאורך הריצה) כדי שהשגיאה α תהייה קטנה...