

סיבוכיות תקשורת ואינפורמציה

תרגיל בית 4: סיבוכיות אינפורמציה ודחיסה

הגשה: 31/5 עד 18:00, לתא של הותם (או באימייל)

1. בכיתה הוכחנו ש- $IC_{\mu,\epsilon}^{int}(And) = \Omega(1)$ עבור שגיאה ϵ קבועה והתפלגות μ אחידה על הקלטים $(0,0), (0,1), (1,0)$.

- א. תהא η התפלגות על $\{0,1\}^2$ כך ש- $\eta(0,0) \neq 0$ ו- $\eta(1,1) \neq 0$. האם ייתכן שקיים $\epsilon < 1/2$ כך ש- $IC_{\eta,\epsilon}^{int}(And) = 0$? מה לגבי $IC_{\eta,\epsilon}^{ext}(And)$? הוכיחו את תשובתכם.
- ב. עבור ההתפלגות המקורית μ ושגיאה קטנה אך לא קבועה $\epsilon < 1/10$, מצאו חסם תחתון טוב ככל שתוכלו על $IC_{\mu,\epsilon}^{int}$ כפונקציה של השגיאה ϵ .
- ג. מצאו חסם עליון טוב ככל שתוכלו על $IC_{\mu,\epsilon}^{int}$ כפונקציה של השגיאה ϵ .

(התעלמו מקבועים והתמקדו בתלות ב- ϵ)

2. שאלה זו עוסקת ב- divergence tree, אותו הגדרנו באופן הבא. יהי Π פרוטוקול, ויהיו x, y קלטים ל- Π ו- r מחרוזת אקראיות פומבית. יהי $\Pi(x, y, r)$ משתנה אקראי המייצג את התמליל של Π בריצתו על הקלטים x, y עם אקראיות פומבית r . (כלומר, האקראיות ב- $\Pi(x, y, r)$ נובעת מהאקראיות הפרטית של השחקנים בלבד).

העץ $T(x, y, r)$ הינו עץ בו כל צומת הוא רישא של תמליל אפשרי של Π בריצתו על הקלטים x, y עם אקראיות פומבית r . כל צומת u בעץ "שייך" לשחקן אשר תורו לדבר בצומת זה, והבנים של u הם התמלילים um_1, \dots, um_k כאשר m_1, \dots, m_k הן ההודעות שעשויות להשלח ע"י שחקן זה. בכל צומת t ישנה התפלגות "נכונה" על הבנים, המוגדרת ע"י

$$\alpha_t(um_i) := \Pr[um_i \leq \Pi(x, y, r) \mid u \leq \Pi(x, y, r)]$$

כאשר ההסתברות היא על האקראיות הפרטית של השחקן שלו שייך צומת u והסימון " $a \leq b$ " משמעותו " a הוא רישא של b ".

- א. הוכיחו שההתפלגות של $\Pi(x, y, r)$ היא בדיוק ההתפלגות המתקבלת מבחירת עלה בעץ $T(x, y, r)$ ע"י ניווט במורד העץ מהשורש מטה, כאשר בכל צומת u בוחרים בן לפי α_u .
- ב. הגדירו פורמלית (באותו אופן בו הוגדרה α_u לעיל) את ההתפלגות β_u המייצגת את ה"אמונה" (prior) של השחקן השני, שצומת u אינה בבעלותו, בצורה שמתאימה לאינפורמציה הפנימית של הפרוטוקול.
- ג. הוכיחו שעבור ההגדרה שנתתם בסעיף ב' מתקיים:

$$IC^{int}(\Pi) = \mathbb{E}_{x,y,r,t}[C(t)]$$

כאשר x, y נבחרים לפי ההתפלגות של הקלט, r נבחר לפי ההתפלגות של האקראיות הפומבית של Π , ו- $t \sim \Pi(x, y, r)$. $C(t)$ הוא ה- divergence cost של המסלול t כפי שהוגדר בכיתה:

$$C(v_0, \dots, v_\ell) = \sum_{i=1}^{\ell} \log \frac{\alpha_{v_{i-1}}(v_i)}{\beta_{v_{i-1}}(v_i)}$$

ד. בדרך-כלל נהוג להניח שפרוטוקולי תקשורת נתונים ב"צורה נורמלית" בה כל שחקן בתורו שולח בדיוק ביט אחד (כלומר, אורך כל הודעה הוא ביט), והשחקנים מדברים לסרוגין. הנחה זו היא בד"כ ללא הגבלת הכלליות. אולם בדחיסה שהצגנו כדי להראות ש-

$$\lim_{n \rightarrow \infty} \frac{CC(f^n, \epsilon)}{n} = IC(f)$$

לא הנחנו הנחה זו – איפשרנו הודעות שאורכן גדול מ-1 (לא הנחנו שהעץ $T(x, y, r)$ הוא בינארי, אלא איפשרנו יותר משני בנים). מדוע?

3. בשאלה זו אין צורך לחשב קבועים מדויקים, תנו את תשובותיכם בסימון O , אך תנו ביטויים פשוטים ככל האפשר.

נתבונן בפרוטוקול Π הבא, עם פרמטר C : אליס מקבלת קלט $X \in \{-1, +1\}$ מפולג אחיד, ובוב אינו

מקבל קלט. במשך C סיבובים, אליס מטילה בכל סיבוב מטבע שהתפלגותו $Bernoulli\left(\frac{1}{2} + \frac{X}{\sqrt{C}}\right)$

(באופן בלתי-תלוי במטבעות שהיטלה בסיבובים הקודמים), ושולחת את ערך המטבע לבוב.

א. מה סיבוכיות האינפורמציה הפנימית של Π ? (העזרו בשאלה 1 בתרגיל בית 3)

ב. עבור הדחיסה ל- $\sqrt{I \cdot C}$ שראינו בכיתה, שבה אליס ובוב דוגמים מסלול בעץ הפרוטוקול

באופן מתואם, מהי תוחלת מס' הטעויות (מקומות במסלול בהם אליס ובוב אינם מסכימים)?

ג. הציעו דחיסה טובה יותר, כלומר, מצאו פרוטוקול Π' בעל סיבוכיות תקשורת קטנה מזו

שמתקבלת בסעיף ב', המאפשר לשחקנים לדגום תמלילים של Π לפי ההתפלגות הנכונה (של

Π) לכל קלט. (תזכורת: איננו סופרים את המחיר של כתיבת הפלט.)