# התרגיל האפס (לא להגשה)

### 1. קרבה בין התפלגויות

עבור שתי מידות הסתברות $P,Q$ מעל $[n]=\{1,\ldots,n\}$ נגדיר את המרחק ביניהם לפי $|P-Q|=\frac{1}{2}\sum_{i=1}^{n}|\Pr_P(i)-\Pr_Q(i)|$. בספרות מרחק זה קרוי ה־Variation Distance.

א. הראו שלכל מאורע $E$ מתקיים $|\Pr_P(E)-\Pr_Q(E)|\leq|P-Q|$.

ב. הראו שקיים $E$ עבורו מתקיים $|\Pr_P(E)-\Pr_Q(E)|=|P-Q|$.

### 2. התפלגויות מותנות

נניח ש־$X$ ו־$Y$ הם שני משתנים מקריים (לא ב"ת) מעל אותו מרחב הסתברות, אשר מקבלים ערכים ב־$\{1,\ldots,n\}$. נניח שקיים מאורע $E$ כך ש־$\Pr[E]\geq 1-\epsilon$, ושאם $E$ מתקיים אז $X=Y$ (ז"א $\Pr[X=Y|E]=1$). הראו שהמרחק בין ההתפלגות על ערכי $X$ לבין ההתפלגות על ערכי $Y$ הוא לא יותר מ־$\epsilon$, ז"א $\frac{1}{2}\sum_{i=1}^{n}|\Pr[X=i]-\Pr[Y=i]|\leq\epsilon$.

### 3. התפלגויות מותנות - הכיוון השני

נניח ש־$p_1,\ldots,p_n$ ו־$q_1,\ldots,q_n$ הם ווקטורי התפלגות (סדרות של מספרים אי־שליליים שסכומן 1) שעבורם מתקיים $\frac{1}{2}\sum_{i=1}^{n}|p_i-q_i|=\epsilon$. הראו שקיים מרחב ההסתברות, ושני מ"מ $X$ ו־$Y$, כך שלכל $i$ מתקיים $\Pr[X=i]=p_i$ ו־$\Pr[Y=i]=q_i$, וכן המאורע $X=Y$ מתקיים בהסתברות $1-\epsilon$ בדיוק.

### 4. מכפלה של הסתברויות

יהי $\mu$ מרחב הסתברות על קבוצת המחרוזות הבינאריות מאורך $k$, המוגדר כך שלכל $i$ הביט ה־$i$ נבחר להיות 1 בהסתברות $\alpha_i$ באופן בלתי תלוי בבחירת הביטים האחרים. יהי $\nu$ מרחב הסתברות על אותה קבוצה, המוגדר באופן זהה, פרט לכך שלכל $i$ הביט ה־$i$ נבחר להיות 1 בהסתברות $\beta_i$ (באופן בלתי תלוי באחרים). הראו שהמרחק (variation distance) בין $\mu$ ו־$\nu$ הוא לכל היותר $\sum_{i=1}^{k}|\alpha_i-\beta_i|$.

# פתרונות לתרגיל האפס

## 1. קרבה בין התפלגויות

**א.** נסמן ב־$\overline{E}$ את המאורע המשלים ל־$E$. מתקיים $\Pr_P[\overline{E}] - \Pr_Q[\overline{E}] = \Pr_Q[E] - \Pr_P[E]$, כי $\Pr_P[\overline{E}] + \Pr_P[E] = \Pr_Q[\overline{E}] + \Pr_Q[E] = 1$. מכאן מתקבל:

$$\begin{aligned}
|\Pr_P[E] - \Pr_Q[E]| &= \frac{1}{2}|\Pr_P[E] - \Pr_Q[E]| + \frac{1}{2}|\Pr_P[\overline{E}] - \Pr_Q[\overline{E}]| \\
&= \frac{1}{2}|\sum_{i\in E}(\Pr_P(i) - \Pr_Q(i))| + \frac{1}{2}|\sum_{i\notin E}(\Pr_P(i) - \Pr_Q(i))| \\
&\leq \frac{1}{2}\sum_{i\in E}|\Pr_P(i) - \Pr_Q(i)| + \frac{1}{2}\sum_{i\notin E}|\Pr_P(i) - \Pr_Q(i)| \\
&= \frac{1}{2}\sum_{i\in[n]}|\Pr_P(i) - \Pr_Q(i)| = d(P,Q)
\end{aligned}$$

**ב.** נגדיר את $E = \{i \in [n] : \Pr_P(i) > \Pr_Q(i)\}$. במקרה זה מתקיימים

$$\begin{aligned}
|\sum_{i\in E}(\Pr_P(i) - \Pr_Q(i))| &= \textstyle\sum_{i\in E}(\Pr_P(i) - \Pr_Q(i)) &&= \sum_{i\in E}|\Pr_P(i) - \Pr_Q(i)| \\
|\sum_{i\notin E}(\Pr_P(i) - \Pr_Q(i))| &= -\textstyle\sum_{i\notin E}(\Pr_P(i) - \Pr_Q(i)) &&= \sum_{i\notin E}|\Pr_P(i) - \Pr_Q(i)|
\end{aligned}$$

ולכן מתקבל שוויון בפיתוח למעלה, ז"א $|\Pr_P[E] - \Pr_Q[E]| = d(P,Q)$.

## 2. התפלגויות מותנות

משתמשים במשפט Bayes ואח"כ באי שוויון המשולש:

$$\begin{aligned}
\frac{1}{2}\sum_{i=1}^{n}|\Pr[X=i] - \Pr[Y=i]| &= \frac{1}{2}\sum_{i=1}^{n}\Big|\Pr[(X=i)\cap E] + \Pr[(X=i)\cap\neg E] \\
&\qquad -\Pr[(Y=i)\cap E] - \Pr[(Y=i)\cap\neg E]\Big| \\
&= \frac{1}{2}\sum_{i=1}^{n}\Big|\Pr[X=i|E]\Pr[E] + \Pr[X=i|\neg E]\Pr[\neg E] \\
&\qquad -\Pr[Y=i|E]\Pr[E] - \Pr[Y=i|\neg E]\Pr[\neg E]\Big| \\
&\leq \frac{1}{2}\sum_{i=1}^{n}\Big|\Pr[X=i|E]\Pr[E] - \Pr[Y=i|E]\Pr[E]\Big| \\
&\quad +\frac{1}{2}\sum_{i=1}^{n}\Big|\Pr[X=i|\neg E]\Pr[\neg E] - \Pr[Y=i|\neg E]\Pr[\neg E]\Big| \\
&= \frac{1}{2}\sum_{i=1}^{n}\Big|\Pr[X=i|\neg E] - \Pr[Y=i|\neg E]\Big|\Pr[\neg E] \\
&\leq \frac{1}{2}\left(\sum_{i=1}^{n}\Pr[X=i|\neg E] + \sum_{i=1}^{n}\Pr[Y=i|\neg E]\right)\Pr[\neg E] \\
&\leq \frac{1}{2}\cdot 2\cdot\epsilon = \epsilon
\end{aligned}$$

## 3. התפלגויות מותנות ־ הכיוון השני

לכל $1 \leq i \leq n$ נגדיר $m_i = \min\{p_i, q_i\}$. מכיוון שמתקיים $|p_i - q_i| = p_i + q_i - 2m_i$, מתקיים גם

$$\sum_{i=1}^{n} m_i = \frac{1}{2}(\sum_{i=1}^{n} p_i + \sum_{i=1}^{n} q_i - \sum_{i=1}^{n} |p_i - q_i|) = 1 - \epsilon$$

עתה צריך לשים לב ששלושת הסדרות המוגדרות באופן הבא הן וקטורי התפלגות (ז"א שהערכים כולם אי שליליים וסכומם הוא 1).

$$\begin{aligned} m'_i &= m_i/(1-\epsilon) \\ p'_i &= (p_i - m_i)/\epsilon \\ q'_i &= (q_i - m_i)/\epsilon \end{aligned}$$

מרחב הסתברות שלנו יוגדר עתה לפי ערכי המ"מ $X$ ו־$Y$ המוגרלים באופן הבא: בהסתברות $1-\epsilon$ אנו נבחר גם ל־$X$ וגם ל־$Y$ ערך המוגרל מתוך $\{1, \ldots, n\}$ לפי וקטור ההתפלגות $m'_1, \ldots, m'_n$. בהסתברות הנותרת $\epsilon$ אנו נבחר באופן בלתי תלוי ל־$X$ ערך המוגרל לפי $p'_1, \ldots, p'_n$, ול־$Y$ ערך המוגרל לפי $q'_1, \ldots, q'_n$.

כדאי עתה לשים לב שלא קיים $i$ עבורו גם $p'_i$ וגם $q'_i$ אינם אפס (מכיוון ש־$m_i$ שווה לאחד מ־$p_i, q_i$), ולכן ההסתברות למאורע $X = i \wedge Y = j$ שווה בדיוק ל־$m_i$ אם $i = j$, ושווה ל־$(p_i - m_i)(q_j - m_j)/\epsilon$ אם $i \neq j$. מכאן נובע שההסתברות של המאורע $X = Y$ היא בדיוק $1 - \epsilon$.

נותר עוד להוכיח שלמשתנים $X$ ו־$Y$ יש את ההתפלגויות (הבלתי־מותנות) הרצויות. נראה זאת לדוגמה עבור $X$ (עבור $Y$ ההוכחה זהה):

$$\begin{aligned} \Pr[X = i] &= m_i + \epsilon^{-1} \sum_{j \neq i} (p_i - m_i)(q_j - m_j) \\ &= m_i + \epsilon^{-1}(p_i - m_i) \sum_{j=1}^{n} (q_j - m_j) \\ &= m_i + \epsilon^{-1}(p_i - m_i)(1 - (1 - \epsilon)) = p_i \end{aligned}$$

## 4. מכפלה של הסתברויות

אפשר לפתור את השאלה באמצעות חישוב ישיר, אולם במקום זאת נראה כאן דרך המשתמשת בתוצאות שהוכחו קודם. נגדיר מרחב הסתברות על זוגות של מחרוזות $x_1, \ldots, x_k, y_1, \ldots, y_k$ באופן הבא: לכל $i$ נבחר את $x_i$ ואת $y_i$ באופן ב"ת בבחירות עבור $i$ אחרים (אולם באופן תלוי זה בזה). אם $\alpha_i > \beta_i$ אז בסיכוי $\beta_i$ נבחר $x_i = y_i = 1$, בסיכוי $\alpha_i - \beta_i$ נבחר $x_i = 1$ ו־$y_i = 0$, ובסיכוי $1 - \alpha_i$ נבחר $x_i = y_i = 0$. אם $\alpha_i < \beta_i$ אז בסיכוי $\alpha_i$ נבחר $x_i = y_i = 1$, בסיכוי $\beta_i - \alpha_i$ נבחר $y_i = 1$ ו־$x_i = 0$, ובסיכוי $1 - \beta_i$ נבחר $x_i = y_i = 0$.

נשים לב עתה שההתפלגות (הלא־מותנה) על $x_1, \ldots, x_k$ זהה ל־$\mu$, ושההתפלגות (הלא־מותנה) על $y_1, \ldots, y_k$ זהה ל־$\nu$. כמו כן, לכל $i$ מתקיים $x_i \neq y_i$ בהסתברות $|\alpha_i - \beta_i|$, ולכן ההסתברות ש־$x_1, \ldots, x_k$ אינה זהה ל־$y_1, \ldots, y_k$ חסומה ע"י $\sum_{i=1}^{k} |\alpha_i - \beta_i|$. מכאן (ע"פ השאלה השניה בתרגיל האפס) שזהו חסם על המרחק בין $\mu$ ל־$\nu$.

לסיום, שימו לב שבחינה יותר מדוקדקת של ההוכחה כאן תניב גם חישוב מדויק של המרחק בין $\mu$ ל־$\nu$. נסו למצוא אותו.