אנטרופיה - הגדרה

יהי x משתנה מקרי המקבל מספר סופי או בן מניה של ערכים $\{a_n\}$ ויהי $p_n = P(x = a_n)$ אז מגדירים את האנטרופיה של x להיות:

$$H(x) = -\sum_n p_n \lg p_n$$

אם x משתנה מקרי רציף עם פונק' צפיפות $f(x)$ אז האנטרופיה מוגדרת כך:

$$H(x) = -\int_{-\infty}^{\infty} f(x) \lg f(x)$$

$$X \sim [p_1 \dots p_n]$$

$$\sum p_i = 1$$

$$\forall i: p_i \geq 0$$

$$H(X) = -\sum_{i=1}^{n} p_i \log(p_i)$$

הלוג יכול להיות בבסיסים שונים, תלוי מאיזה עולם מגיעים. אנו נעבוד בד"כ בלוג לפי בסיס 2, פיזיקאים עובדים בלוג לפי בסיס טבעי.

דרך נוספת לחשוב על האנתרופיה:

$$H(X) = E_p\left(\log\left(\frac{1}{p_i}\right)\right)$$

תרגיל

מטילים 3 מטבעות הוגנים. מה האנטרופיה של מס' הראשים?

פתרון

מס' הראשים	p_i
0	$\frac{1}{8}$
1	$\frac{3}{8}=\binom{3}{1}\frac{1}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$

נסמן את מס' הראשים בx אזי:

$$\begin{aligned} H(x) &= -\left(\frac{1}{8}\lg\frac{1}{8}+\frac{3}{8}\lg\frac{3}{8}+\frac{3}{8}\lg\frac{3}{8}+\frac{1}{8}\lg\frac{1}{8}\right) \\ &= -\left(\frac{1}{4}\lg\frac{1}{8}-\frac{3}{4}\lg\frac{3}{8}\right) \\ &= -\frac{1}{4}\cdot(-3)-\frac{3}{4}(\lg 3-3) \\ &= \frac{3}{4}-\frac{3}{4}\lg 3+\frac{9}{4}=1.81 \end{aligned}$$

תרגיל: חשבו אנטרופיה של משתנה נורמלי

The "normal distribution" or "Gaussian distribution" or Gaussian probability density function is defined by

$$N(x;\ \mu,\ \sigma) = \frac{1}{(2\pi\sigma^2)^{1/2}}\, e^{-(x-\mu)^2/2\sigma^2}\ . \tag{8.1}$$

This density function, which is symmetrical about the line $x = \mu$, has the familiar bell shape shown in Figure 8.1. The two parameters, μ and σ^2, each have special significance; μ is the mean and σ^2 the variance of the distribution. All probability density functions must be normalized to unity, and it is shown in most textbooks on advanced calculus that

$$\int_{-\infty}^{\infty} N(x;\ \mu,\ \sigma)\, dx = 1\ . \tag{8.2}$$

The expectation of x, $E(x)$, is equal to the mean; that is,

$$E(x) = \int_{-\infty}^{\infty} x N(x;\ \mu,\ \sigma)\, dx = \mu\ . \tag{8.3}$$

The expectation of $(x-\mu)^2$, $E(x-\mu)^2$, is equal to the variance; that is,

$$E(x-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 N(x;\ \mu,\ \sigma)\, dx = \sigma^2\ . \tag{8.4}$$

The differential entropy of the normal distribution can be found without difficulty. From the definition of differential entropy given in Chapter 7, and using Equation (8.1),

$$H = -\int_{-\infty}^{\infty} (2\pi\sigma^2)^{-\frac{1}{2}}\, e^{-(x-\mu)^2/2\sigma^2} \ln\left[(2\pi\sigma^2)^{-\frac{1}{2}}\, e^{-(x-\mu)^2/2\sigma^2}\right] dx$$

$$H = \frac{1}{2}\ln(2\pi\sigma^2) \int_{-\infty}^{\infty} (2\pi\sigma^2)^{-\frac{1}{2}}\, e^{-(x-\mu)^2/2\sigma^2}\, dx + \frac{1}{2\sigma^2}\int_{-\infty}^{\infty} (2\pi\sigma^2)^{-\frac{1}{2}}\, (x-\mu)^2\, e^{-(x-\mu)^2/2\sigma^2}\, dx\ .$$

Introducing Equations (8.2) and (8.4),

$$H = \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2}\ .$$

Writing $\frac{1}{2}$ as $(\frac{1}{2}\ln e)$,

$$H = \frac{1}{2}\ln(2\pi e\sigma^2)\ , \tag{8.5}$$

תכונות

- $H(x) \geq 0$
- $H(x)$ לא תלויה בערכים שx מקבל אלא רק בהסתברויות $\{p_i\}$.
- אם x, y מ"מ בת"ל אז $H(x, y)$ היא האנטרופיה של המשתנה (x, y) והיא:

$$H(x, y) = H(x) + H(y)$$

ניסוי- קובייה עם 8 צדדים. עלינו לדעת איזה מספר יצא בקובייה. אפשר לשאול סדרה של שלוש שאלות כדי לדעת מה נפל בקובייה. שלוש שאלות מאפשרות לפתור את כל אי הוודאות. זה מתאים לאינטואיציה, שכל התוצאות אפשריות באותה מידה. אין רמז לוודאות על אחת התוצאות , ולכן כל אי הוודאות נמחקת בשלוש שאלות כן או לא.

האנתרופיה היא לא בהכרח מספר שלם. קשה להבין את האינטואיציה של מספר שאלות שהוא לא שלם, אבל נדבר על זה בהמשך. אפשר נניח לחזור על הניסוי הרבה מאד פעמים "ולצבור" חלקי שאלות מפעם לפעם.

החסם של האנתרופיה הבדידה היא מה שקורה כאשר ההתפלגות היא אחידה. נניח שאחד מה Pi עולה ושני יורד. מה קורה? האיברים שאנו משנים לא מפצים אחד על השני.. נוכיח בצורה מסודרת.

$$\max_{\bar{p}} \; -\sum p_i \log(p_i)$$

$$\left.\sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0 \right\}$$ אלו האילוצים שלו

<u>הרעיון</u>: כופל לגרנז' - מוסיפים פונ' עונש. אנו מנסים למקסם על פני ה-P אבל לתת עונש אם עוברים על אחד האילוצים. כיון שרוצים למקס' מקסימום העונש שניתן יהיה גדול אם עוברים על אילוץ.

המקסימום מתקבל כאשר :

$$\forall i \; p_i = \frac{1}{n}$$

ולכן $$0 \leq H(\bar{X}) \leq \log(n)$$

אינטואיציה למה בהתפלגות האחידה האנטרופיה מקסימלית: חוסר הודאות הוא מקסימלי כי אין שום הטיה בערכים של המשתנה המקרי (לדוגמא קוביה הוגנת ולא הוגנת)

תרגיל: הוכיחו שהאנטרופיה המקסימלית מתקבלת עבור ההתפלגות האחידה:

$$L = -\sum p_i \log p_i + \lambda\left(\sum p_i - 1\right)$$

$$\frac{\partial L}{\partial p_1} = -(1 \cdot \log p_1 + 1) + \lambda$$

$$\lambda - 1 - \log p_1 = 0$$

$$\log p_1 = \lambda - 1$$

$$p_1 = e^{\lambda - 1}$$

$$p_i = e^{\lambda - 1}$$

ובאותו הדבר
עבור i כלשהו

$$\frac{\partial L}{\partial \lambda} = \sum p_i - 1 = 0 \Rightarrow \sum p_i = 1$$

כל ההסתברויות שוות וישנן n אפשרויות

לכן $p_i = \frac{1}{n}$

וזה הערך המקסימלי.

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i = -\sum \frac{1}{n} \log \frac{1}{n} = \log n$$

בהתפלגות אחידה:

$$X \sim U(n) \quad p_1, \ldots p_n$$

$$H(X) = -\sum p_i \log p_i = -\sum_{i=1}^{n} \frac{1}{n} \cdot \log\left(\frac{1}{n}\right) =$$

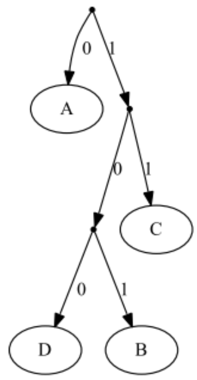
$$n \cdot \frac{1}{n} \cdot \log(n) = \log_2(n)$$

מוטיבציה לעץ למטה: יש ארבע אותיות, באופן רגיל שלכל האותיות אותו אורך קוד נדרש ל-4 ביט. ניצור קידוד באורך משתנה כך שלאותיות שכיחות יותר יש אורך קוד קצר יותר וכך נדחוס את האינפורמציה

אלגוריתם הופמן

מקודדים מחרוזת ע"י עץ קידוד.

דוגמה



נקודד את $ACBBD$:

$$011101101100$$

אפשר גם לפענח - את 10110100011 נפענח כ:

$$BBAAAC$$

בניית קוד הופמן

הופמן פיתח אלגוריתם חמדן שבונה קוד תחיליות אופטימלי הנקרא **קוד הופמן** (Huffman code). האלגוריתם בונה את העץ T המייצג את הקוד האופטימלי "מלמטה למעלה". הוא מתחיל עם קבוצה של $|C|$ עלים ומבצע סדרה של $|C| - 1$ פעולות "מיזוג" ליצירת העץ הסופי.

בפסאודו־קוד שלהלן אנו מניחים כי C היא קבוצה של n תווים, וכל תו $c \in C$ הוא עצם בעל שכיחות מוגדרת $f[c]$. תור קדימויות Q, עם השכיחויות f כמפתחות, משמש לזיהוי שני העצמים בעלי השכיחות הנמוכה ביותר כדי למזגם. תוצאת מיזוגם של שני העצמים היא עצם חדש, אשר שכיחותו היא סכום השכיחויות של שני העצמים שמוזגו.

```
HUFFMAN(C)
1  n ← |C|
2  Q ← C
3  for i ← 1 to n − 1
4       do z ← ALLOCATE-NODE()
5          x ← left[z] ← EXTRACT-MIN(Q)
6          y ← right[z] ← EXTRACT-MIN(Q)
7          f[z] ← f[x] + f[y]
8          INSERT(Q, z)
9  return EXTRACT-MIN(Q)
```

פעולתו של האלגוריתם של הופמן על האלפבית בדוגמה שלנו מוצגת באיור 17.5. מאחר שהאלפבית מכיל 6 אותיות, הגודל ההתחלתי של התור הוא $n = 6$, ולבנייתו של העץ נדרשים 5 שלבי מיזוג. העץ הסופי מייצג את קוד התחיליות האופטימלי. מילת הקוד עבור אות היא סדרת התוויות על הקשתות המרכיבות את המסלול מן השורש עד לאות זו.

שורה 2 מאתחלת את תור הקדימויות Q על־ידי כך שהיא מציבה בו את התווים המרכיבים את C. לולאת ה־**for** בשורות 8-3 שולפת מן התור, בכל איטרציה, את שני הצמתים x ו־y בעלי השכיחויות הנמוכות ביותר, ומכניסה לתור במקומם צומת חדש z המייצג את מיזוגם. השכיחות של z היא סכום השכיחויות של x ו־y והיא מחושבת בשורה 7. x הוא הבן השמאלי של הצומת z ו־y הוא הבן הימני. (סדר זה הוא שרירותי; אם נחליף בין הבן הימני והבן השמאלי בצומת כלשהו נקבל קוד שונה בעל אותה עלות.) לאחר $n - 1$ פעולות מיזוג, הצומת האחד שנותר בתור — השורש של עץ הקידוד — מוחזר על־ידי השגרה בשורה 9.

תרגיל

נניח שיש התפלגות על אותיות בשפה:

אות	הסתברות
A	0.13
B	0.14
C	0.15
D	0.16
E	0.17
F	0.18
G	0.07

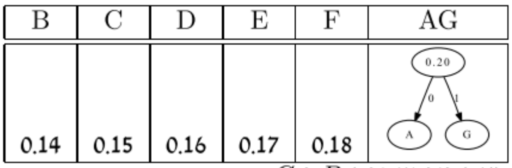
מה עץ הקידוד האופטימלי? מה שיעור הדחיסה שלו לעומת האופטימום (האנטרופיה)?

פתרון

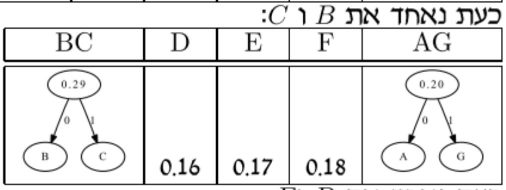
נרשום בטבלה:

A	B	C	D	E	F	G
0.13	0.14	0.15	0.16	0.17	0.18	0.07

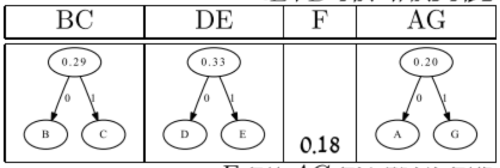
נאחד את הקטנים ביותר, A וG, ונקבל:



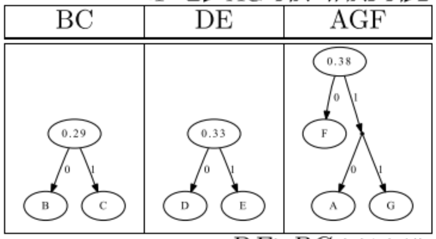
כעת נאחד את B ו C:



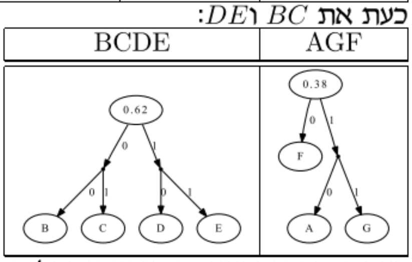
כעת נאחד את D וE:



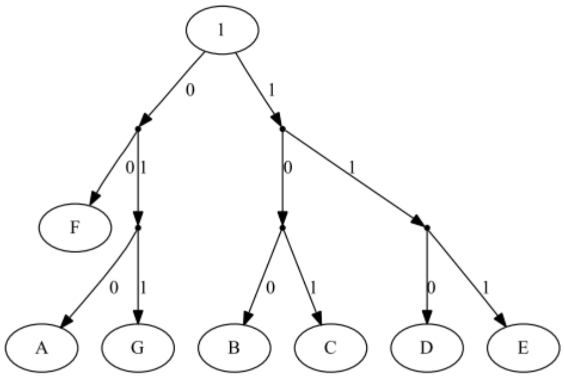
כעת נאחד את AG עם F:



כעת את BC וDE:



כעת נחבר את שניהם ביחד ונקבל את העץ הסופי:



תוחלת כמות הביטית הדרושה לייצוג אות בשפה:

$$\begin{aligned} E(x) &= (0.13+0.14+0.15+0.16+0.17+0.07)\cdot 3+0.18\cdot 2 \\ &= 2.82 \end{aligned}$$

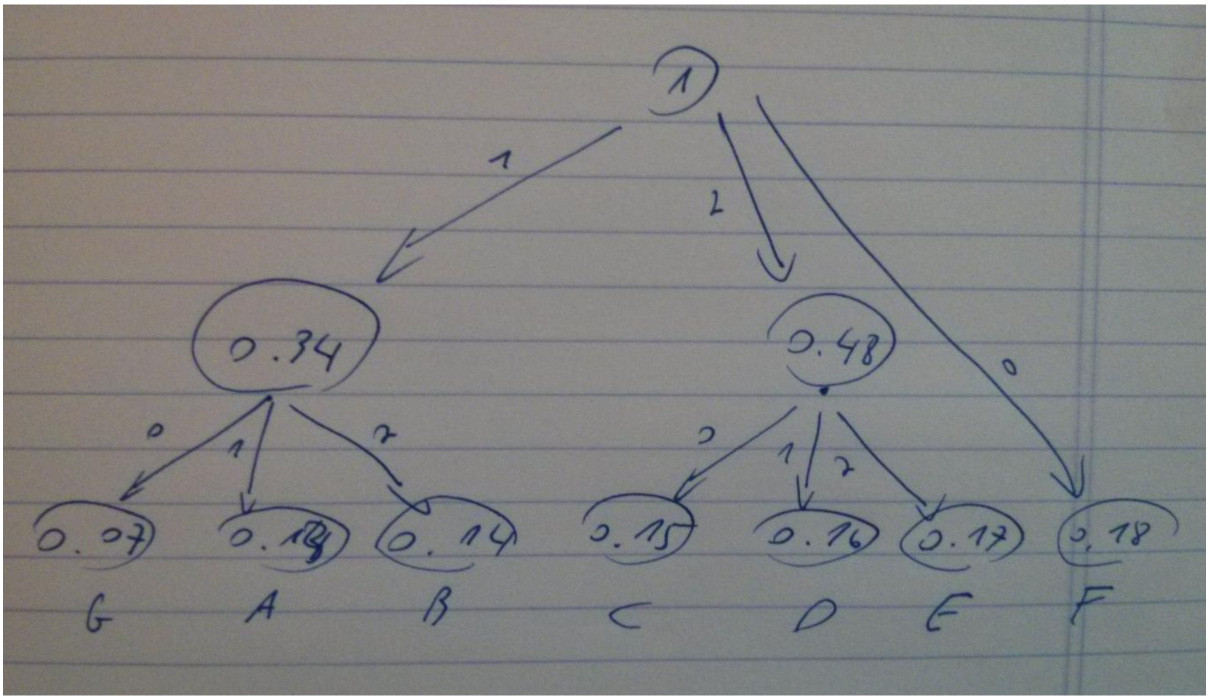
כמות הביטים המינימלית הדרושה לייצוג אות בשפה:

$$H\left(\text{random letter}\right) = -0.13\lg 0.13 - 0.14\lg 0.14 - ...$$

זה יוצא פחות מ2.82 אך במעט.

יוצא 2.76 ביט

ג. מצאו קידוד טרינרי אופטימלי



(להוכיח את 1 בתרגיל הבא באינדוקציה)

תרגיל

1. יהי x מ"מ המקבל ערכים $1, ..., n$ בהסתברויות $p_1, ..., p_n$ בהתאמה ונניח ש$2p_i < p_{i+1}$.
איך יראה עץ קידוד אופטימלי של ערכי המשתנה x?

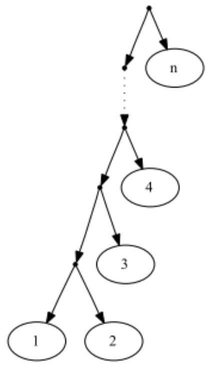
2. נניח שx מקבל את הערכים 1,2 בהסתברויות $\frac{1}{4}, \frac{3}{4}$ בהתאמה.
מה ניתן לעשות כדי לדחוס ביעילות גדולה יותר סדרת ערכים של x?

פתרון

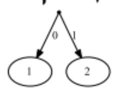
1. נשים לב ש:

$$p_1 + ... + p_k < p_{k+1}$$

לכן באיטרציה ה$k-1$ של אלגוריתם הופמן יהיה עץ עם k ,... ,1 והוא יצורף ל$k+1$.
העץ שנקבל בסוף הוא:



2. עץ הקידוד האופטימלי הוא:



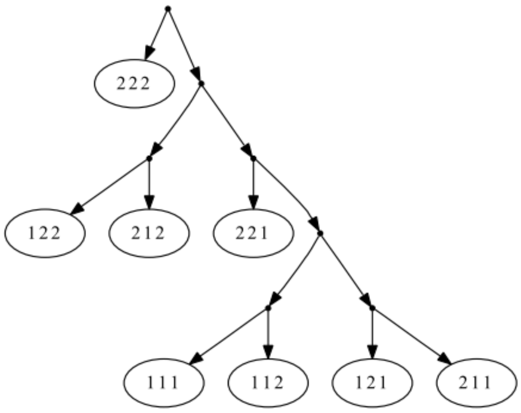
כמות הביטים הממוצעת לאות היא 1, אבל האנטרופיה היא:

$$-\frac{1}{4}\lg\frac{1}{4} - \frac{3}{4}\lg\frac{3}{4} = \frac{1}{2} - \frac{3}{4}\lg\frac{3}{4} = 0.81$$

אפשר להתגבר על הבעיה (כלומר לדחוס ע"י עץ קידוד ביעילות גדולה יותר) ע"י דחיסה זוגות או שלשות של ערכי המשתנה x.
לדוגמה, נדחוס שלשות:

שלשה	הסתברות * 64
111	1
112	3
121	3
211	3
122	9
212	9
221	9
222	27

ואז עץ הקידוד יהיה:



בכמה ביטים דוחסים כל שלשה (תוחלת):

$$E = \frac{1}{64}[27 \cdot 1 + 9 \cdot 3 \cdot 3 + 1 \cdot 5 + 3 \cdot 3 \cdot 5]$$

$$= \quad \frac{158}{64} = 2.46$$

אנחנו מקודדים שלשת ערכים ב2.46 ביטים בממוצע לכן אנו מקודדים ערך בודד בכ-0.82 ביטים (והאנטרופיה היא 0.81).