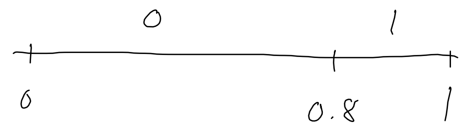
Aritmetisk koding kap 7 i komp

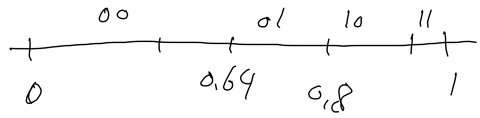
Eks. 7.17. Vi har teksten $x = 00100$

$P(0) = 0.8$, $P(1) = 0.2$.

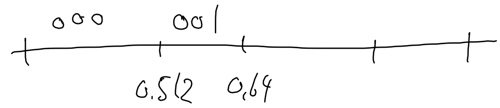
Vi assosierer $[0, 0.8)$ med 0 og $[0.8, 1)$ med 1.



Vi må dele $[0, 0.8)$ på samme måte



Neste tegn ~~er~~ 0 så vi havner i $[0, 0.64)$. Dermed må dette deles i samme forhold



Neste tegn er 1 så vi havner i $[0.512, 0.64)$.

Da må dette deles på samme måte i $[0.512, 0.6144)$ og $[0.6144, 0.64)$

Vi ender opp til slutt med at den aritmetiske koden til teksten må ligge i $[0.512, 0.59392)$.

Den aritmetiske er et tall i dette intervallet som kan representeres eksakt i 2-tallsystemet, dvs. et tall på formen $i/2^k$ der $k \in \mathbb{N}$ og $i \in \mathbb{N}$. Dessuten: k minst mulig siden $i \in [0, 2^k)$ vi trenger k binære siffere til i.

$k=1$: $1/2$, nei $\quad$ $k=3$, $1/8, 3/8, 3/8, 7/8$ Nei

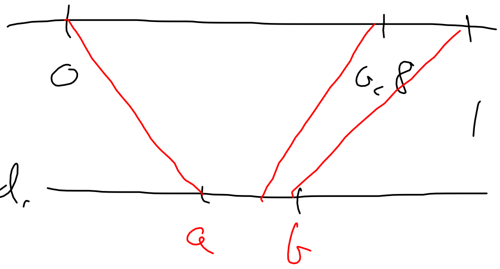
$k=2$: $1/4, 3/4$, nei $\quad$ $k=4$, $9/16$ - OK.

$9/16 = 0.5625 = 0.1001_2$. Vi lagrer 1001.

Generell utfordring: Avbilde $[0,1]$ på et eller annet intervall $[a,b]$

Bruk funksjonen

$$g(z) = a + z(b-a)$$

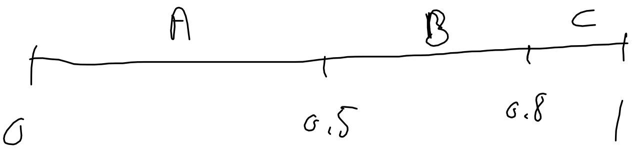


Kumulativ sannsynlighetsford.

$$F(\alpha_j) = \sum_{i=1}^{j} P(\alpha_i), \quad j = 1, \ldots, n$$

$$L(\alpha_j) = F(\alpha_j) - P(\alpha_j)$$

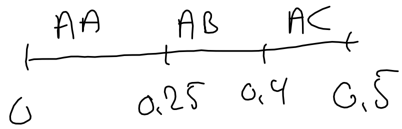
Ex. 7.21. $X = ACBBCAABAA$, $P(A) = 0.5$



$P(B) = 0.3$

$P(C) = 0.2$

Videre med $[a_1, b_1) = [0, 0.5)$



For å finne delintervallene bruker vi $q_2(z) = 0.5z$

Vi ser at $q_2(0.5) = 0.25$

$q_2(0.8) = 0.4$

Siden tegn to er C blir

$[a_2, b_2) = [q_2(0.8), q_2(1))$

$= [0.4, 0.5)$

Vi må avbilde $[0,1]$ på $[0.4, 0.5]$ med

$q_3(z) = a_2 + z(b_2 - a_2) = 0.4 + 0.1z$

$q_3(0.5) = 0.4 + 0.1 \cdot 0.5 = 0.45$, $q_3(0.8) = 0.48$

Siden 3. tegn er B blir

$[a_3, b_3) = [q_3(0.5), q_3(0.8)) = [0.45, 0.48)$

$\vdots$

$[a_{10}, b_{10}) = [0.472425, 0.4724875)$

Aritmetisk kode = midtpunktet til $[a_{10}, b_{10})$ rundet av til

$\lceil -\log_2 (P(A)^5 P(B)^3 P(C)^2) \rceil + 1 = 16$

Midtpkt $= 0.472441875 = 0.01111\ldots$

Egenskaper ved AC

Bredden av det siste intervallet er

$$b_m - a_m = P(x_1)\cdot P(x_2) \cdots P(x_m).$$

Aritmetisk kode i $[a_m, b_m)$, $i/2^k$

Finn $\mu \in \mathbb{R}$ slik at

$$\frac{1}{2^\mu} = b_m - a_m \quad \text{og sett} \quad k = \lceil \mu \rceil$$

Altså $k = \lceil \mu \rceil = \lceil -\log_2 (b_m - a_m) \rceil$

$$= \lceil -\log_2 P(x_1) \cdots P(x_m) \rceil$$

$$= \left\lceil -\log_2 \left(P(\alpha_1)^{f(\alpha_1)} \cdots P(\alpha_n)^{f(\alpha_n)} \right) \right\rceil$$

$$= \left\lceil -\sum_{i=1}^{n} f(\alpha_i) \log_2 P(\alpha_i) \right\rceil$$

Antall bits :

Antall bits pr. tegn

$$\frac{1}{m} \left\lceil -\sum_{i=1}^{n} f(\alpha_i) \log_2 P(\alpha_i) \right\rceil$$

$$\le -\frac{1}{m} \sum_{i=1}^{n} f(\alpha_i) \log_2 P(\alpha_i) + \frac{1}{m}$$

$$\xrightarrow[m \to \infty]{} -\sum_{i=1}^{n} P(\alpha_i) \log_2 P(\alpha_i) = H(P_1, \dots, P_n)$$

AC er optimalt.