TSKS21 Signaler, information & bilder

Föreläsning 10

Informationsteori – källkodning

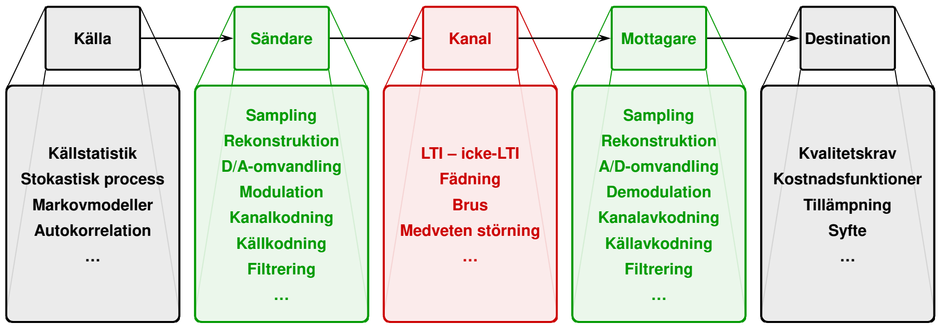
Mikael Olofsson
Institutionen för Systemteknik (ISY)
Ämnesområdet Kommunikationssystem
LINKÖPINGS UNIVERSITET

Dagens planering

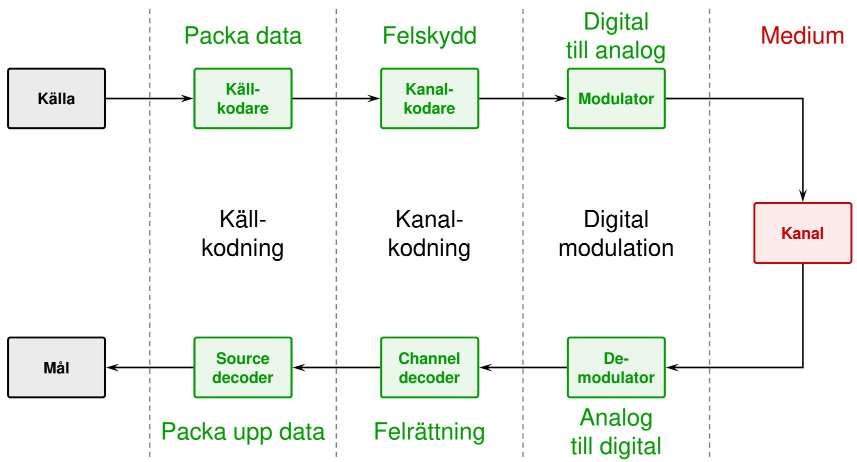
Källkodning – att packa data

- Shannon-information – entropi
- Huffmankoder
- Källutvidgning
- Skurlängdskodning

Kommunikationslänk – översikt



Ett envägstelekommunikationssystem



Entropi

Singla slant m gånger ger en av 2^m lika sannolika följder som naturligt representeras med m bitar. Sannolikheten för följden är 2^{-m}.

Notera: $m = -\log_2(2^{-m})$.

Alltså:
m är mängden information som vi förknippar med sannolikheten 2^{-m}.

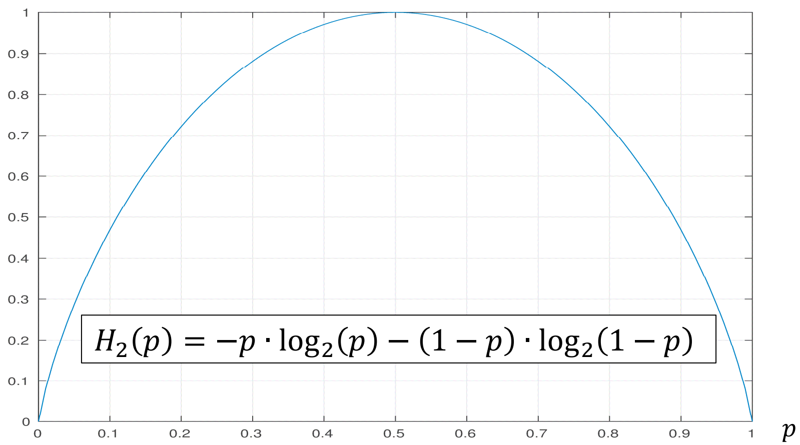
Shannoninformation förknippad med sannolikheten p är därför $-\log_2(p)$.
Behöver inte vara ett heltal.

Entropi är förväntad shannoninformation:

Binärt: $\quad H_2(p) = -p \cdot \log_2(p) - (1-p) \cdot \log_2(1-p)$

Allmänt: $\quad H(p_0, p_1, \ldots, p_{N-1}) = -\sum_{i=0}^{N-1} p_i \log_2(p_i)$

Binära entropifunktionen



En källkodares uppgift

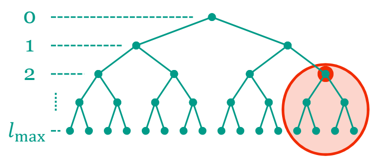
En källkodare översätter meddelanden till bitar.

Krafts olikhet 1(2)

Påstående: Det finns en trädkod med längder $l_0, \ldots, l_{N-1}$ om och endast om

$$\sum_{i=0}^{N-1} 2^{-l_i} \le 1 \quad \text{gäller.}$$

Bevis: Först "endast om". Antag att vi har en trädkod med givna längder. Definiera $l_{\max} = \max\{l_0, \ldots, l_{N-1}\}$. Fullständigt binärt träd, djup $l_{\max}$:



Ett kodord av längd l_i (djup l_i) har $2^{l_{\max}-l_i}$ löv under sig på djup $l_{\max}$. Totalt $2^{l_{\max}}$ löv. Om inget kodord är onödigt långt:

$$\sum_{i=0}^{N-1} 2^{l_{\max}-l_i} = 2^{l_{\max}}$$

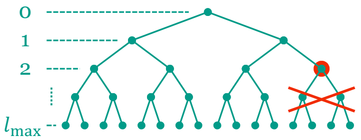
Kodord kan vara onödigt långa. Vissa löv är i det fallet inte under något kodord:

$$\sum_{i=0}^{N-1} 2^{l_{\max}-l_i} \le 2^{l_{\max}} \quad \Rightarrow \quad \sum_{i=0}^{N-1} 2^{-l_i} \le 1$$

Krafts olikhet 2(2) (fortsatt bevis)

Sedan "om". Antag att $\sum_{i=0}^{N-1} 2^{-l_i} \le 1$ gäller. Sortera längderna $l_0 \le \cdots \le l_{N-1}$

Samma träd:



Algoritm för att konstruera en kod

För $i = 0,1,\dots,N-1$:

- Välj en punkt på djup l_i.
- Markera den som använd.
- Ta bort delträdet under den punkten.

Fråga: Finns det löv kvar för alla kodord?

I iteration k finns det så här många löv kvar:

$$\underbrace{2^{l_{\max}} - \sum_{i=0}^{k-1} 2^{l_{\max}-l_i}}_{\text{Heltal}} = 2^{l_{\max}}\left(1 - \underbrace{\sum_{i=0}^{k-1} 2^{-l_i}}_{<1}\right) > 0$$

Det finns alltså löv kvar i varje iteration.

Vi kan konstruera en trädkod med längder $l_0, \dots, l_{N-1}$. Då finns det en!

Entropin är en undre gräns för trädkoder

Källstatistik: $p_i = \Pr\{A = a_i\}$ **Entropi av A:** $H(A) = -\sum_{i=0}^{N-1} p_i \log_2(p_i)$

Kodordslängder: $A = a_i \Rightarrow L = l_i$ **Påstående:** $m_L \ge H(A)$

Bevis: Betrakta $H(A) - m_L$. Borde vara ≤ 0.

$$H(A) - m_L = -\sum_{i=1}^{N} p_i \cdot \log_2(p_i) - \sum_{i=1}^{N} p_i \cdot l_i = \sum_{i=1}^{N} p_i \cdot (-\log_2(p_i) - l_i)$$

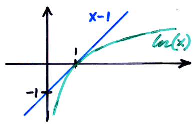
$$= \sum_{i=1}^{N} p_i \cdot (\log_2(2^{-l_i}) - \log_2(p_i)) = \sum_{i=1}^{N} p_i \cdot \log_2\left(\frac{2^{-l_i}}{p_i}\right) = \Big/ \log_2(x) = \frac{\ln(x)}{\ln(2)} \Big/$$

$$= \frac{1}{\ln(2)} \sum_{i=1}^{N} p_i \cdot \ln\left(\frac{2^{-l_i}}{p_i}\right) \le \Big/ \ln(x) \le x - 1 \ \forall x \Big/$$

$$\le \frac{1}{\ln(2)} \sum_{i=1}^{N} p_i \cdot \left(\frac{2^{-l_i}}{p_i} - 1\right) = \frac{1}{\ln(2)}\left(\underbrace{\sum_{i=1}^{N} 2^{-l_i}}_{\text{Kraft: } \le 1} - \underbrace{\sum_{i=1}^{N} p_i}_{=1}\right)$$

$$\le \frac{1}{\ln(2)} \cdot (1 - 1) = 0$$

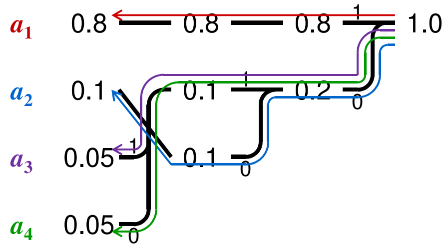
Done!



Likhet? Exakt här. Alltså $\ln\left(\frac{2^{-l_i}}{p_i}\right) = 0 \Rightarrow l_i = -\log_2(p_i)$

$$\Rightarrow 2^{-l_i} = 2^{\log_2(p_i)} = p_i \quad \Rightarrow \sum_{i=1}^{N} 2^{-l_i} = \sum_{i=1}^{N} p_i = 1.$$

Huffmankodning



$c^{(1)} = 1$

$c^{(2)} = 00$

$c^{(3)} = 011$

$c^{(4)} = 010$

p_i	$c^{(i)}$	l_i	$p_i l_i$
0.8	1	1	0.8
0.1	00	2	0.2
0.05	011	3	0.15
0.05	010	3	0.15

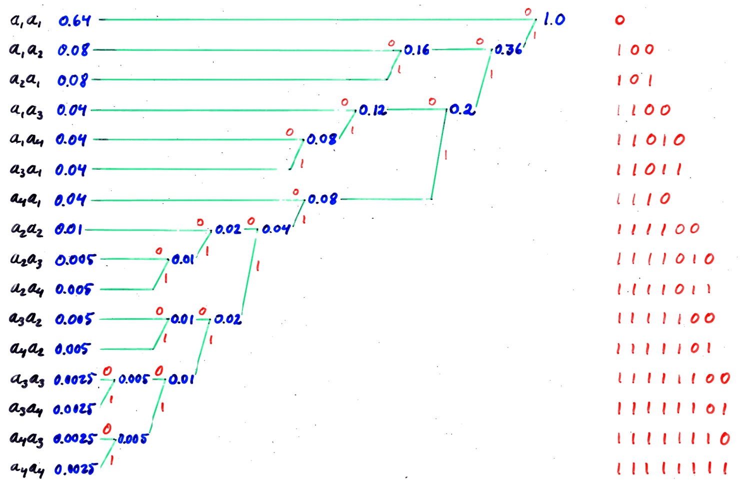
$m_L = 1.3$

Entropi: $H(A) = -\sum_{i=1}^{N} p_i \log_2 p_i \approx 1.022$

Redundans: $m_L - H(A) \approx 0.278$

Kompressionskvot: $\frac{\log_2 N}{m_L} \approx 1.54$

Förenklad huffmankod för en utvidgad källa



Resultat av källkodning

Nästan lika sannolika
nästan okorrelerade bitar.

Ju närmare entropin vi kommer, desto närmare kommer vi lika sannolika och okorrelerade bitar.

Mikael Olofsson
ISY/KS

www.liu.se