TSKS10 Signaler, information & Kommunikation

Föreläsning 5

Datakompression, entropi, källkodningssatsen

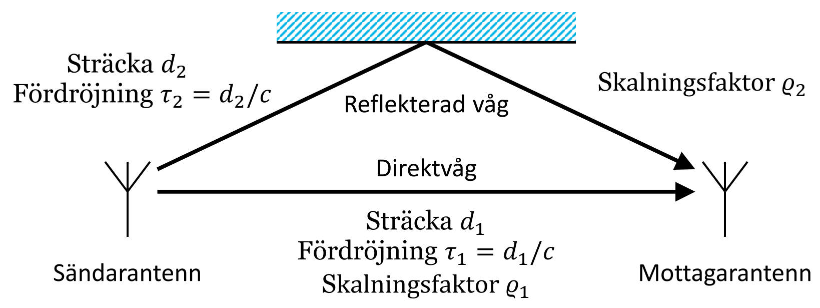
Mikael Olofsson
Institutionen för Systemteknik (ISY)
Ämnesområdet Kommunikationssystem

LINKÖPINGS UNIVERSITET

Dagens planering

- Avsluta signaler (labtips)
- Introduktion informationsteori
- Källkodning – att packa data
- Shannon-information – entropi
- Källkodningssatsen

Radiokanal med reflekterad våg



Informationsteori - introduktion

Källkodning:

- Vad är information?
- Hur kan man mäta information?
- Kan man packa information felfritt?
- Varför då?
- Kan man alltid packa information?
- Gränser för detta?
  - Källkodningssatsen

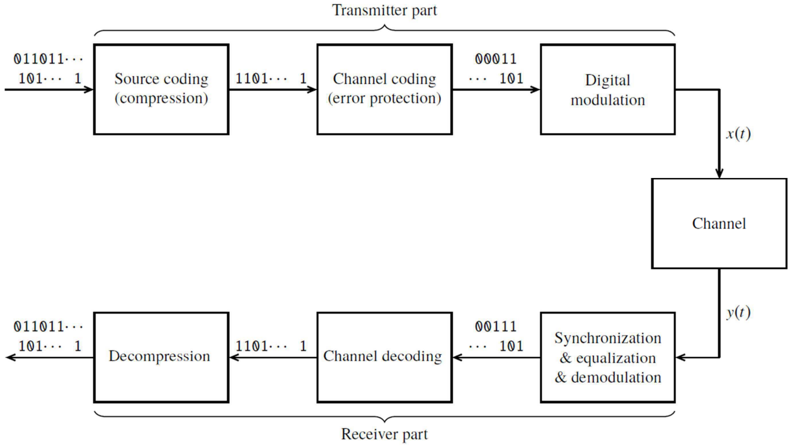
Denna föreläsning

Kanalkodning:

- Kanalen orsakar fel
- Felrättning
- Är felfri kommunikation möjlig?
- Om inte, kan man begränsa felsannolikheten?
  - Kanalkodningssatsen
- Kanalkapacitet

Kommande två föreläsningar

Ett kommunikationssystem



Figur 1.11 ur kursboken

Experiment med ett skevt mynt

$p = P\{1\} = 0.7$ $\qquad N = 20$

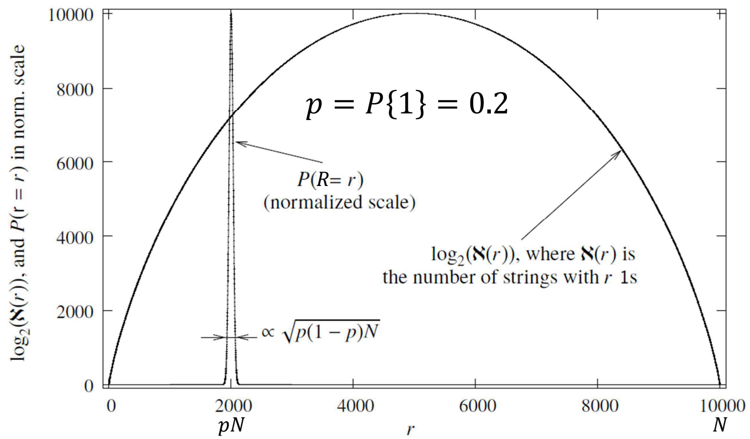
(a) 00000000000000000000 (20 nollor) $\quad (1-0.7)^{20} \approx 3 \cdot 10^{-11}$

(b) 00100110011101011100 (10 nollor, 10 ettor) $\quad (1-0.7)^{10} \cdot 0.7^{10} \approx 2 \cdot 10^{-7}$

(c) 11100111101100110111 (6 nollor, 14 ettor) $\quad (1-0.7)^{6} \cdot 0.7^{14} \approx 5 \cdot 10^{-6}$

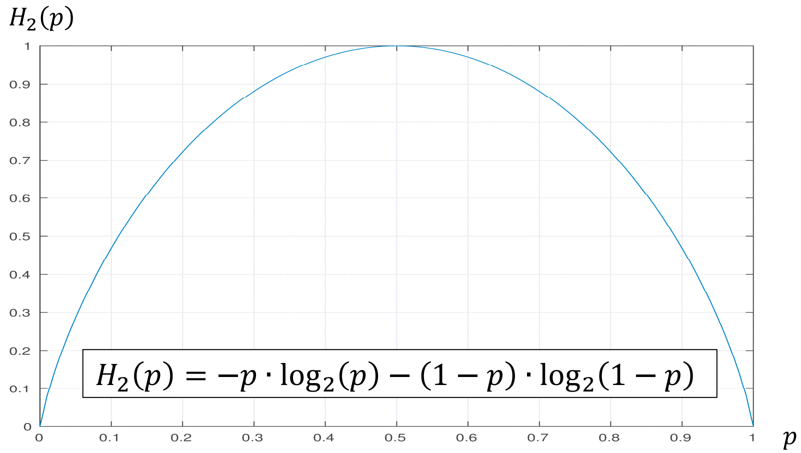
(d) 11111111111111111111 (20 ettor) $\quad 0.7^{20} \approx 8 \cdot 10^{-4}$

Typiska strängar



Figur 6.1 ur kursboken

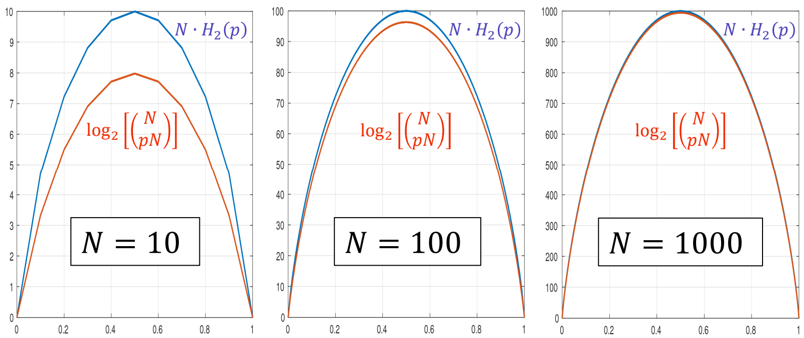
Binära entropifunktionen



Motsvarar figur 6.2 i kursboken

Användbar approximation

$$\log_2\left[\binom{N}{pN}\right] \approx N \cdot H_2(p)$$



Källkodningssatsen (för binära strängar)

Givet $\delta > 0$ och $\varepsilon > 0$ så finns det ett N, sådant att sannolikheten att observera en binär sträng av längd N som inte kan representeras med $N(H_2(p) + \varepsilon)$ bitar är mindre än δ.

Detta betyder:
Det går att packa ner till godtyckligt nära $H_2(p)$ kodordsbitar per informationsbit.

Man kan också visa:
Det går inte att packa ner till färre än $H_2(p)$ kodordsbitar per informationsbit utan förluster.

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}.

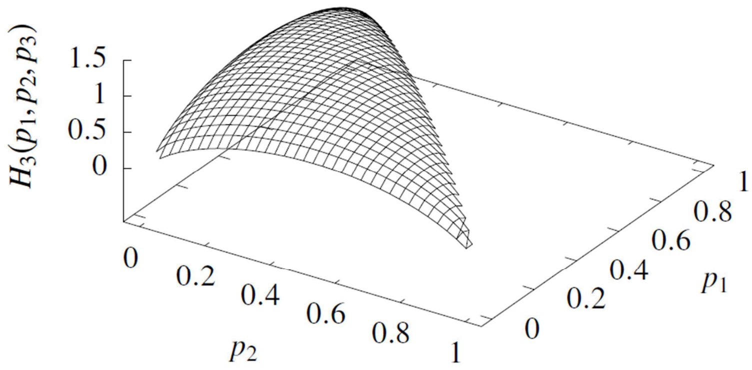
Shannoninformation förknippad med sannolikheten p är därför $-\log_2(p)$. Behöver inte vara ett heltal.

Entropi är förväntad shannoninformation:

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

Allmänt: $H_M(p_1, p_2, \ldots, p_M) = -\sum_{i=1}^{M} p_i \log_2(p_i) \leq \log_2(M)$

Ternära entropifunktionen



Figur 6.3 ur kursboken

En källkodares uppgift

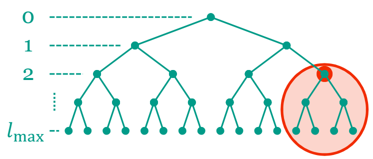
En källkodare översätter meddelanden till bitar.

Krafts olikhet 1(2)

Påstående: Det finns en trädkod med längder $l_1, \ldots, l_M$ om och endast om

$$\sum_{i=1}^{M} 2^{-l_i} \leq 1 \quad \text{gäller.}$$

Bevis: Först ”endast om”. Antag att vi har en trädkod med givna längder. Definiera $l_{\max} = \max\{l_1, \ldots, l_M\}$. 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:

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

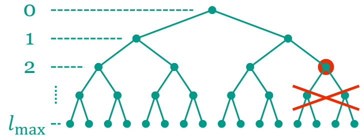
Kodord kan vara onödigt långa. Vissa löv är i det fallet inte under något kodord:

$$\sum_{i=1}^{M} 2^{l_{\max}-l_i} \leq 2^{l_{\max}} \quad \Rightarrow \quad \sum_{i=1}^{M} 2^{-l_i} \leq 1$$

Krafts olikhet 2(2) (fortsatt bevis)

Sedan ”om”. Antag att $\sum_{i=1}^{M} 2^{-l_i} \leq 1$ gäller. Sortera längderna $l_1 \leq \cdots \leq l_M$

Samma träd:



Algoritm för att konstruera en kod

För $i = 1,2, \ldots, M$:
- 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 j finns det så här många löv kvar:

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

Det finns alltså löv kvar i varje iteration.

Vi kan konstruera en trädkod med längder $l_1, \ldots, l_M$. Då finns det en!

Entropin är en undre gräns för trädkoder

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

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

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

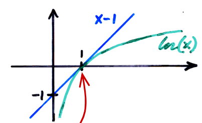
$$H(A) - m_L = -\sum_{i=1}^{M} p_i \log_2(p_i) - \sum_{i=1}^{M} p_i l_i = \sum_{i=1}^{M} p_i(-l_i - \log_2(p_i))$$

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

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

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

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



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

$$\Rightarrow 2^{-l_i} = p_i \Rightarrow \sum_{i=1}^{M} 2^{-l_i} = \sum_{i=1}^{M} p_i = 1$$

Mikael Olofsson
ISY/CommSys