Теорија информација

- количина информације $Q(s_i)$ неког догађаја s_i обрнуто је пропорционална вјероватноћи догађаја s_i и дефинишемо је преко log:

$$Q(s_i) = \log\left(\frac{1}{P(s_i)}\right) = -\log P(s_i)$$

- ако су дате двије независне поруке s_i и s_j са одговарајућим вјероватноћама $p(s_i)$ и $P(s_j)$, здружена количина информације једнака је збиру њихових количина информације.

$$Q(s_i, s_j) = \log\left(\frac{1}{p(s_i, s_j)}\right) = \log\left(\frac{1}{P(s_i)\cdot P(s_j)}\right) = \log\left(\frac{1}{P(s_i)}\right) + \log\left(\frac{1}{P(s_j)}\right) =$$

$$= Q(s_i) + Q(s_j)$$

$$Q(s_i) = \log_2\left(\frac{1}{P(s_i)}\right) \quad [sh]$$

$$Q(s_i) = \log\left(\frac{1}{P(s_i)}\right) \quad [\text{Hartley}]$$

$$Q(s_i) = \log_e\left(\frac{1}{P(s_i)}\right) \quad [\text{NAT}]$$

* бацање новчића: $Q(s_i) = ld\frac{1}{0,5} = ld2 = 1\ [sh]$
* бинарни извор: $p(0) = p(1) = 0,5$; $Q(s_i) = ld\frac{1}{0,5} = 1\ [sh]$
* бацање коцке: $Q(s_i) = ld\frac{1}{\frac{1}{6}} = ld6 = 2,583\ [sh]$
* количина информације када се сазна неко слово азбуке под условом да је вјероватноћа појаве неког слова 1/30

$$Q(s_i) = ld\frac{1}{1/30} = ld30 = 4,904\ [sh]$$

* Колико се информација добије када се сазна да је из шпила од 32 карте извучена а) дама
б) дама херц ?

а) у шпилу имамо 4 даме ; вероватноћа да се извуче било која од њих је :
$$p = \frac{4}{32} = \frac{1}{8} = 2^{-3}$$
па је
$$Q = ld\frac{1}{p} = ld\frac{1}{2^{-3}} = ld2^{3} = 3\ [Sh]$$

б) вероватноћа извлачења даме херц: $p = \frac{1}{32} = 2^{-5}$
$$Q = ld\frac{1}{p} = ld\frac{1}{2^{-5}} = ld2^{5} = 5\ [Sh]$$

- за дискретни извор без меморије, ентропија се рачуна као:
$$H(S) = \sum_{i=1}^{2} p(s_i)\, ld\frac{1}{p(s_i)} \quad \left[\frac{Sh}{simb}\right]$$

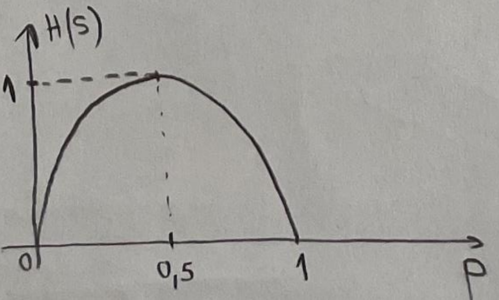
- максималну ентропију има извор код кога су вероватноће емитовања свих симбола исте (тада је неизвесност највећа)
  - ако извор има 2 симбола онда је $p(s_i) = \frac{1}{2}$
$$H(S) = \sum_{i=1}^{2} \frac{1}{2}\, ld\,2 = ld\,2 = H_{max}(S)$$
$$\boxed{H_{max}(S) = Q(s_i)}$$

* Одредити ентропију бинарног извора у функцији вероватноће појаве 0 или 1.

$p(0) = p$

$p(1) = 1 - p(0) = 1 - p$

$$H(S) = \sum_{i=1}^{2} p(s_i)\, ld\frac{1}{p(s_i)} = p\, ld\frac{1}{p} + (1-p)\, ld\frac{1}{1-p}$$



- ако са H_0 означимо ентропију извора чији су симболи независни и имају исту вјероватноћу појављивања, односно

$$H_0 = H_{max},$$

а са H означимо стварну вриједност ентропије, тада однос

$$\frac{H}{H_{max}} = \eta$$ представља степен искориштења симбола или релативну ентропију

- сувишност (редундантност) дефинишемо као:

$$R = 1 - \frac{H}{H_{max}} = 1 - \eta \quad \text{или} \quad R[100\%] = 100 \cdot R$$

- n-то проширење дискретног извора је извор чији су симболи секвенце од по n симбола првобитног извора
- Ако првобитни извор има 2 симбола, онда је његово n-то проширење извор са 2^n симбола

* Дат је дискретни извор без меморије са листом симбола

$$S_i = \{S_1, S_2, S_3\} \qquad p(S_1) = \frac{1}{2}$$
$$p(S_2) = \frac{1}{4}$$
$$p(S_3) = \frac{1}{4}$$

Одредити:

а) Количину информација коју носи сваки симбол.

б) Ентропију извора.

в) Редундансу извора.

г) Ентропију извора у случају проширења другог реда.

а) $$Q(S_1) = ld\frac{1}{p(S_1)} = ld2 = 1\ [Sh]$$

$$Q(S_2) = Q(S_3) = ld\frac{1}{p(S_2)} = ld\frac{1}{p(S_3)} = ld4 = 2\ [Sh]$$

б) $$H(S) = \sum_{i=1}^{3} p(S_i)\cdot ld\frac{1}{p(S_i)} = \frac{1}{2}ld2 + 2\cdot\frac{1}{4}ld4 = 1{,}5\left[\frac{Sh}{simb}\right]$$

в) $$R = 1-\eta = 1-\frac{H}{H_{max}}$$

$$H_{max} = ld\,q = ld3 = \frac{\log 3}{\log 2} \cong 1{,}585$$

$$R = 1-\frac{1{,}5}{1{,}585} = 1-0{,}946 = 0{,}054$$

$$R[\%] = 5{,}4\%$$

г) $q=3$, $n=2$

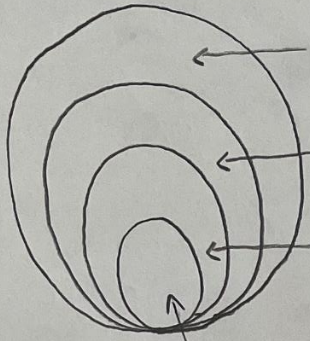
$q^n = 3^2 = 9$ симбола

S_iS_j	$p(S_i,S_j)$
S_1S_1	1/4
S_1S_2	1/8
S_1S_3	1/8
S_2S_1	1/8
S_2S_2	1/16
S_2S_3	1/16
S_3S_1	1/8
S_3S_2	1/16
S_3S_3	1/16

$$H(S^2) = 1\cdot\frac{1}{4}ld4 + 4\cdot\frac{1}{8}ld8 + 4\cdot\frac{1}{16}ld16 =$$
$$= \frac{1}{4}\cdot 2 + \frac{1}{2}\cdot 3 + \frac{1}{4}\cdot 4 = \frac{1}{2}+\frac{3}{2}+1 = \boxed{3\left[\frac{Sh}{simb}\right] = 2\cdot H(S)}$$

$$\boxed{H(S^n) = n\cdot H(S)}$$

- кодовање:
  - статистичко (ентропијско)
  - заштитно
  - шифровање (криптозаштита)



општи кодови

несингуларни кодови – различитим симболима извора одговарају различите кодне речи

једнозначно декодиви кодови – било која секвенца кодних речи кореспондира једнозначно само једној поруци информационог извора

тренутни кодови = префиксни кодови

- код је префиксан ако у њему ниједна кодна реч није префикс неке друге кодне речи
- код је тренутан ако и само ако је префиксан

1. Испитати једнозначну декодивост скупа кодних речи:

a → 1

b → 00

c → 10

X_0	X_1	X_2	X_3	...
1	0	0	0	...
00				
10				

- у вишим сегментним класама нема кодне речи из нулте сегментне класе, дакле код јесте једнозначно декодив, али је кашњење бесконачно

2. Испитати да ли су дати кодови једнозначно декодибилни и тренутни.

а) код A

x_i	код A
x_1	10
x_2	100
x_3	11
x_4	111

X_0	X_1	X_2
10	0	0
100	1	00
11		1
111		11

- код није једнозначно декодибилан
- код није префиксан, па самим тим није ни тренутан

б) код Б

X_i	код Б
X_1	0
X_2	01
X_3	011
X_4	0111

X_0	X_1
0	1
01	11
011	111
0111	1
	11
	1

– код јесте једнозначно декодибилан

– код није префиксан, стога није ни тренутан

в) код В

X_i	код В
X_1	1
X_2	01
X_3	001
X_4	0001

– тренутан је, (јер је префиксан) па је самим тим и једнозначно декодибилан

г) код Г

X_i	код Г
X_1	00
X_2	01
X_3	10
X_4	11

– исто као под в)

3. Испитати једнозначну декодивост скупа кодних речи: TO, AM, TOM, ALI, MAT, ANALI, ATAMAN.

X_0	X_1	X_2	X_3	X_4	X_5	
TO	M	AT	AMAN	AN	ALI	
AM						
TOM						
ALI						
MAT						
ANALI						
ATAMAN						