BI019 Bioinformatika
Osnove teorije informacij

A Blejec

3. oktober 2012

Kaj je informacija

- Računalnik je stroj za predelavo informacij
- GIGO

Sistemi dogodkv in izidi

- Gremo v kino ali na žur?
- Izberemo eno od šestih jedi.
- "Joško je naš najbol'š prjatu" ali katera srečka bo zadela?

Sistemi z enakomožnimi stanji in negotovost

$$\alpha = \begin{pmatrix} a_1 & a_2 \\ 1/2 & 1/2 \end{pmatrix}$$

Sistemi z enakomožnimi stanji in negotovost

$$\alpha = \begin{pmatrix} a_1 & a_2 \\ 1/2 & 1/2 \end{pmatrix}$$

$$\beta = \begin{pmatrix} b_1 & b_2 & \cdots & b_6 \\ 1/6 & 1/6 & \cdots & 1/6 \end{pmatrix}$$

Sistemi z enakomožnimi stanji in negotovost

$$\alpha = \begin{pmatrix} a_1 & a_2 \\ 1/2 & 1/2 \end{pmatrix}$$

$$\beta = \begin{pmatrix} b_1 & b_2 & \cdots & b_6 \\ 1/6 & 1/6 & \cdots & 1/6 \end{pmatrix}$$

$$\gamma = \begin{pmatrix} c_1 & c_2 & c_3 & \cdots & c_{100,000} \\ 0.00001 & 0.00001 & 0.00001 & \cdots & 0.00001 \end{pmatrix}$$

Sistemi z enakomožnimi stanji in negotovost

$$\omega = \begin{pmatrix} o_1 \\ 1 \end{pmatrix}$$

$$\alpha = \begin{pmatrix} a_1 & a_2 \\ 1/2 & 1/2 \end{pmatrix}$$

$$\beta = \begin{pmatrix} b_1 & b_2 & \cdots & b_6 \\ 1/6 & 1/6 & \cdots & 1/6 \end{pmatrix}$$

$$\gamma = \begin{pmatrix} c_1 & c_2 & c_3 & \cdots & c_{100,000} \\ 0.00001 & 0.00001 & 0.00001 & \cdots & 0.00001 \end{pmatrix}$$

Merjenje negotovosti

Mera negotovosti

Sistem α_n z n enakomožnimi stanji naj ima negotovost $H(\alpha_n) = H(n)$

Pravila za računanje negotovosti

1. Sistem z enim stanjem je gotov, $H(1) = 0$

Pravila za računanje negotovosti

1. Sistem z enim stanjem je gotov, $H(1) = 0$
2. Sistem z več stanji ima večjo negotovost kot sistem z manj stanji

$$n > m \Leftrightarrow H(\alpha_n) > H(\alpha_m) \Leftrightarrow H(n) > H(m)$$

$$H(2) > H(1) = 0$$

Pravila za računanje negotovosti

1. Sistem z enim stanjem je gotov, $H(1) = 0$
2. Sistem z več stanji ima večjo negotovost kot sistem z manj stanji

$$n > m \Leftrightarrow H(\alpha_n) > H(\alpha_m) \Leftrightarrow H(n) > H(m)$$

$$H(2) > H(1) = 0$$

3. Kakšno negotovost ima sestavljen sistem

$$\delta_{n \times m} = \alpha_n \otimes \beta_m$$

Pravila za računanje negotovosti

1. Sistem z enim stanjem je gotov, $H(1) = 0$
2. Sistem z več stanji ima večjo negotovost kot sistem z manj stanji

$$n > m \Leftrightarrow H(\alpha_n) > H(\alpha_m) \Leftrightarrow H(n) > H(m)$$

$$H(2) > H(1) = 0$$

3. Kakšno negotovost ima sestavljen sistem

$$\delta_{n \times m} = \alpha_n \otimes \beta_m$$

$$H(\alpha_n \otimes \beta_m) = H(n \times m) = H(n) + H(m)$$

Funkcija za računanje negotovosti

Funkcija za računanje negotovosti

Logaritem

$$H(n) = C \log_a n$$

Funkcija za računanje negotovosti

Logaritem

$$H(n) = C \log_a n$$

Dvojiški logaritem

$$H(n) = \log_2 n$$

$$H(2) = 1$$

bit, nit in dit

2	$log_2 2 = 1$	bit
e	$log_e 2 = 0.6931$	nit
10	$log_{10} 2 = 0.301$	dit

Enakomožna stanja: $p = 1/n$

$$\alpha_n = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ p & p & \cdots & p \end{pmatrix}$$

$$\begin{aligned} H(n) &= log_2 n \\ &= -log_2(1/n) = -log_2 p \\ &= -n \cdot (1/n) log_2(1/n) \\ &= -\sum (1/n) log_2(1/n) \\ &= -\sum p \cdot log_2 p \end{aligned}$$

Neenakomožna stanja

$$\alpha_n = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ p_1 & p_2 & \cdots & p_n \end{pmatrix}$$

$$H(n) = -\sum p \cdot log_2 p$$

nadomestimo z

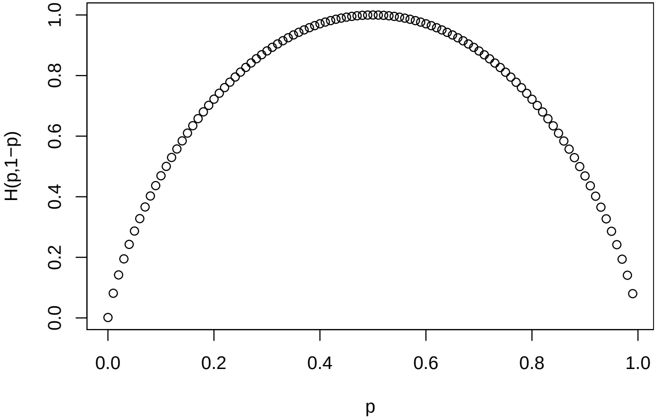
Shannon-Wienerjeva formula

$$H(n) = -\sum_{i=1}^{n} p_i \cdot log_2 p_i$$

Shanon-Wiener (Weaver?) indeks diverzitete

Sistem z dvemi stanji

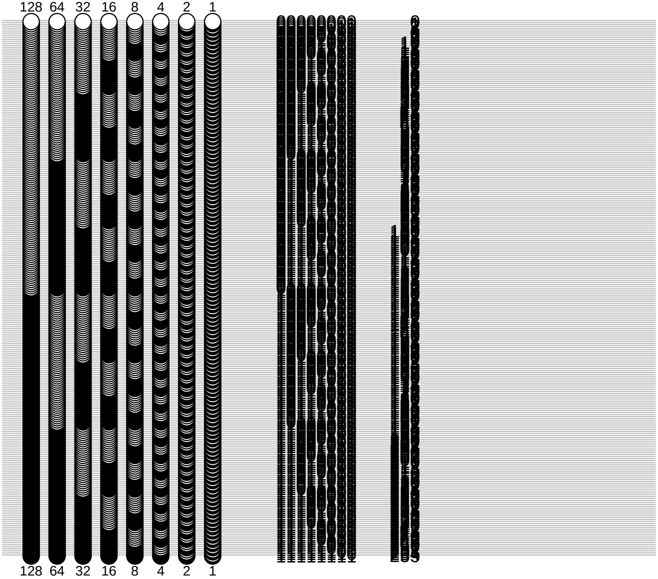
```
> p <- seq(0.0001,0.9999,0.01)
> x <- cbind(p,1-p)
> H <- function(x) -sum(x*log(x,2))
> par(mar=c(4,4,1,0))
> plot(p,apply(x,1,H),ylab="H(p,1-p)")
```



bit ... 4 biti: $2^4 = 16$ stanj

8	4	2	1		
○	○	○	○	0000	0
○	○	○	●	0001	1
○	○	●	○	0010	2
○	○	●	●	0011	3
○	●	○	○	0100	4
○	●	○	●	0101	5
○	●	●	○	0110	6
○	●	●	●	0111	7
●	○	○	○	1000	8
●	○	○	●	1001	9
●	○	●	○	1010	10
●	○	●	●	1011	11
●	●	○	○	1100	12
●	●	○	●	1101	13
●	●	●	○	1110	14
●	●	●	●	1111	15
8	4	2	1		

byte, ... 8 bitov: $2^8 = 256$ stanj



Število bitov (H) in število stanj (n)

bit	stanj
1	2
2	4
3	8
4	16
5	32
6	64
7	128
8	256
9	512
10	1024
11	2048
12	4096
13	8192
14	16384
15	32768
16	65536

$$H = log_2 n$$

$$n = 2^H$$

Kodna tabela ASCII

Bits												
b_7 →					0	0	0	0	1	1	1	1
b_6 →					0	0	1	1	0	0	1	1
b_5 →					0	1	0	1	0	1	0	1
b_4 ↓	b_3 ↓	b_2 ↓	b_1 ↓	Column → / Row ↓	0	1	2	3	4	5	6	7
0	0	0	0	0	NUL	DLE	SP	0	@	P	`	p
0	0	0	1	1	SOH	DC1	!	1	A	Q	a	q
0	0	1	0	2	STX	DC2	"	2	B	R	b	r
0	0	1	1	3	ETX	DC3	#	3	C	S	c	s
0	1	0	0	4	EOT	DC4	$	4	D	T	d	t
0	1	0	1	5	ENQ	NAK	%	5	E	U	e	u
0	1	1	0	6	ACK	SYN	&	6	F	V	f	v
0	1	1	1	7	BEL	ETB	'	7	G	W	g	w
1	0	0	0	8	BS	CAN	(	8	H	X	h	x
1	0	0	1	9	HT	EM	)	9	I	Y	i	y
1	0	1	0	10	LF	SUB	*	:	J	Z	j	z
1	0	1	1	11	VT	ESC	+	;	K	[	k	{
1	1	0	0	12	FF	FC	,	<	L	\	l	|
1	1	0	1	13	CR	GS	-	=	M	]	m	}
1	1	1	0	14	SO	RS	.	>	N	^	n	~
1	1	1	1	15	SI	US	/	?	O	_	o	DEL

Nukleotidna zaporedja

Znaki: A T C G

1. Koliko bitov informacije nosi en nukleotid?
2. Zakaj aminokisline kodirjo tripleti?