Teoria Informacji

Miara informacji Hartleya: X - zbiór N-elementowy

$$I_H(X) = \log_2 N$$

- addytywność

$|X| = N \cdot M$ — małe x

$X = \{X_1, X_2, \ldots, X_{NM}\}$

$\tilde{X} = \{X_1, X_2, \ldots X_M\}$ — duże X

$X_1 = \{x_1, \ldots x_N\}$

$X_j = \{x_{(j-1)N+1}, \ldots x_{(j-1)N+N}\}$

$X_M = \ldots$

wtedy:

$I_H(X_j) = \log_2 N$

$I_H(\tilde{X}) = \log_2 M$

$I_H(X) = \log_2 N M = \log_2 N + \log_2 M = I_H(X_j) + I_H(\tilde{X})$

Def: <u>miara ilości informacji</u> $I: X \to \mathbb{R}$, gdzie $X = \{x_1, \ldots x_N\}$ jest dyskretna p-prob: $P(X_i) = p_i$

$$I(X_i) = -\log p_i$$

$// I(x_i) \text{"="} I(p_i) = \log p_i$

Własności:

- $I(1) = 0$, $I(\frac{1}{2}) = 1$
- $I(pq) = I(p) + I(q)$ — zdarzenie $X \cap Y$, gdzie X, Y niezależne $P(X) = p$, $P(Y) = q$
- I jest różniczkowalna
- $\lim_{p \to 0^+} I(p) = \infty$

Def: <u>Entropia Shannona</u> $X = \{x_1 \ldots x_N\}$, $P(x_i) = p_i$ $H: X \to \mathbb{R}$

$$H(X) = \sum_{i=1}^{N} p_i \log \frac{1}{p_i} = \sum_{i=1}^{N} p_i(-\log p_i)$$

$H(X) = \sum I$

Def: <u>Długość kodu</u> (dich.) $L: X \to \mathbb{N}$

$$L(x_i) = len(code(x_i))$$

co istotne: $EL = \sum_{i=1}^{N} p_i \cdot L(x_i)$

$$H(X) \leq EL(X)$$

mega ważne

(P) $X=\{x_1,\dots,x_8\}$

$p_1=\frac{1}{2}$, $p_2=\frac{1}{8}$, $p_3=\dots=p_8=\frac{1}{16}$

$H(X)=\frac{1}{2}\log_2 2+\frac{1}{8}\log_2 8+6\cdot\frac{1}{16}\log_2 16=\frac{1}{2}+\frac{3}{8}+\frac{3}{8}\cdot 4=\frac{19}{8}$

$C_1(x_i)=(i-1)_2$

$EL_1=3$

$C_2(x_1)=0$

$C_2(x_2)=10$

$C_2(x_3)=11\_\,\_\,\_$

$C_2(x_8)=11\_\,\_\,\_$

$EL_2=\frac{1}{2}\cdot 1+\frac{1}{8}\cdot 2+6\cdot\frac{1}{16}\cdot 5=\frac{21}{8}$

TW: Shannona $H(X)\le EL(X)$

Def: Kod $c: X\to\{0,1\}^*$

- c jest iniekcją
- c jest prefiksowe $\iff \forall x,x'\ \ x\neq x' \to c(x)$ nie jest prefiksem $c(x')$

Def: Lemiwości dr Żeberskiego

$$L=L(c)=\sum_{i=1}^{N} p_i\,|c(x_i)|=\sum_{i=1}^{N} p_i l_i$$

Def: Jednoznaczna dekodowalność

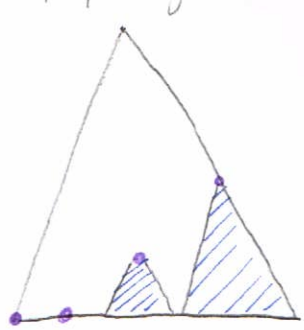
$\forall \sigma\in\{0,1\}^*$ nie istnieje lub istnieje 1 podział

$\sigma=\sigma_1{}^\frown\sigma_2{}^\frown\dots{}^\frown\sigma_k$ t.że $\sigma_i\in C(X)$

Def: Nierówność Krafta

$$\sum_{i=1}^{N} 2^{-l_i}\le 1$$

Kod prefiksowy, wizualizacja dostępności:



Def: Kod Shannona – to kod prefiksowy, spełniający:

$$l_i = \lceil \log_2 \frac{1}{p_i} \rceil$$

jako tako dobre

$$H \le L \le H+1$$

(P) kodów Shannona

$p_1 = \frac{1}{2}$, $p_2 = \frac{1}{8}$, $p_3 = \dots = p_8 = \frac{1}{16}$

$l_1 = \log_2 2 = 1$ $\quad l_2 = \log_2 8 = 3$ $\quad l_3 = \dots = l_8 = 4$

$P_1 = 0 \qquad = 0.\boxed{0}000$

$P_2 = p_1 = \frac{1}{2} \qquad = 0.\boxed{100}0$

$P_3 = p_1 + p_2 = \frac{1}{2} + \frac{1}{8} = \frac{5}{8} \qquad = 0.\boxed{1010}$

$P_4 = \dots = \frac{11}{16} \qquad = 0.\boxed{1011}$

$P_5 = \dots$

$\vdots$

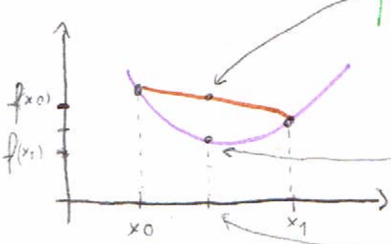
$P_8 = \dots \qquad = 0.\boxed{1111}0$

Odrobina analizy:

Def: wypukłość funkcji $f: [a,b] \to \mathbb{R}$

f wyp. $\iff \forall_{x_0, x_1 \in [a,b]}, \forall_{p \in [0,1]}$

$$p f(x_0) + (1-p) f(x_1) \ge f(\underbrace{p \cdot x_0 + (1-p) x_1})$$



Def: ściśła wypukłość

f ści. wyp. $\iff \forall_{x_0, x_1 \in [a,b]}, x_0 \ne x_1, \forall_{p \in [0,1]}$

$$p f(x_0) + (1-p) f(x_1) > f(p \cdot x_0 + (1-p) x_1)$$

FAKT: jeśli $f''(x) \ge 0 \to f$ jest <u>wypukła</u>

$f''(x) > 0 \to f$ jest <u>ściśle wypukła</u>

TW: Nierówność Jensena

niech: $f:[a,b] \to \mathbb{R}$ wypukła, $x_0, x_1, \dots, x_n \in [a,b]$, $p_0 p_1 \dots p_n \geq 0$, $\sum_{i=1}^{N} p_i = 1$

wtedy:

$$f\left(\sum_{i=0}^{N} p_i x_i\right) \leq \sum_{i=0}^{N} p_i f(x_i)$$

jeśli $|\{x_0, x_1, \dots x_n\}| \geq 2$ i $\forall_i\ p_i > 0$ to... < ...

(P) $f(x) = x\log_2(x)$ $x_i = p_i$

$f'(x) = \log_2 x + \log_2 e$

$f''(x) = \frac{1}{x} \frac{1}{\ln 2} > 0$ dla $x > 0$

uwaga $f:(0,\infty) \to \mathbb{R}$, czasami robimy $f(0)=0$ i wtedy $f:[0,\infty) \to \mathbb{R}$

u nas nierówność Jensena:

$$E(f(X)) \geq f(EX)$$

TW: Złoty lemat

niech $x_0, x_1, \dots x_n > 0$, $y_0, y_1 \dots y_n > 0$, $\sum_{i=0}^{n} x_i = 1 = \sum_{i=0}^{n} y_i$

wtedy:

$$\sum_{i=0}^{n} x_i \log\frac{1}{x_i} \leq \sum_{i=0}^{N} x_i \log\frac{1}{y_i}$$

główny wniosek:

dla $p_i = \frac{1}{N}$, $|X| = N$

$\sum p_i = 1$

$$H(X) = \sum_{i=0}^{N} p_i \log\frac{1}{p_i} = \sum_{i=0}^{N} p_i \log N = \log N$$

Kodowanie słów długości n. kodem Shannona

$$X_n = \underbrace{X \cdot X \cdot \ldots \cdot X}_{n}$$

TW: Shannona 2.

$$\lim_{n\to\infty} \frac{L(X_n)}{n} = H(X)$$

- $L(X_n) = n L(X)$
- $H(X_n) = n H(X)$

$$nH(X) \leq nL(X) \leq nH(X) + 1$$

$$H(X) \leq L(X) \leq H(X) + \frac{1}{n}$$

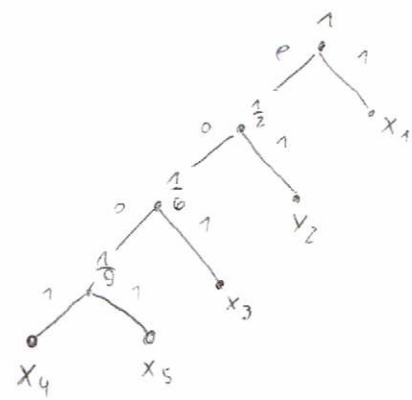
Def: Kodowanie Huffmana mamy: $X=\{x_1,x_2,\dots x_N\}$ $p_1 \geq p_2 \geq \dots \geq p_N$

Kody Huffmana prefiksowe

własności:

- $l_1 \leq l_2 \leq \dots \leq l_N$ $l_i = |c_{Huff}(x)|$
- $l_{N-1} = l_N$ ostatnie dwa kody różnią się na ostatnim bicie
- inaczej $c(x_N)$ i $c(x_{N-1})$

Algorytm: P $X=\{x_1,\dots,x_5\}$ $p_1=\frac{1}{2}, p_2=\frac{1}{3}, p_3=\dots=p_5=\frac{1}{18}$



$c(x_1) = 1$ $c(x_4) = 0001$

$c(x_2) = 01$ $c(x_5) = 0000$

$c(x_3) = 001$

TW: Kody Huffmana są optymalne

Inny pogląd na entropię:

kiedy piszemy $p(x)$ mamy na myśli $P(X=x)$, ogólnie $p(x,y) = P(X=x \wedge Y=y)$

$X: \Omega \to \mathcal{X}$
$Y: \Omega \to \mathcal{Y}$

Def: entropia łączna:

$$H(X,Y) = \sum_{x\in\mathcal{X}, y\in\mathcal{Y}} p(x,y) \cdot \log \frac{1}{p(x,y)}$$

// $p(x) = \sum_y p(x,y)$

jeżeli X i Y są niezależne to: $(P(X=x, Y=y) = P(X=x) \cdot P(Y=y))$

$$H(X,Y) = H(X) + H(Y)$$

TW: $H(X,Y) \leq H(X) + H(Y)$

dowód przez złoty lemat.

ponadto $H(X,Y) = H(X) + H(Y) \iff p_n = q_n$ bo wtedy $p(x,y) = p(x)p(y)$

Def: Informacja wzajemna

$$I(X,Y) = H(X) + H(Y) - H(X,Y)$$

$$I(X,Y) = H(X) - H(X|Y)$$

$$I(X,Y) = \sum_{x,y} p(x,y) \log \frac{1}{p(x)p(y)} - \sum_{x,y} p(x,y) \log \frac{1}{p(x,y)} = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x) \cdot p(y)}$$

- $I(X,Y) \geq 0$
- $I(X,Y) = 0 \iff X$ i Y niezależne

Def: Odległość Kullbacha-Leiblera

$$D(p\|q) = \sum_a p(a) \log \frac{p(a)}{q(a)}$$

- $D(p\|q) \geq 0$
- $D(p\|q) = 0 \iff p = q$

Def: entropia warunkowa X od Y „mierzy ilość informacji potrzebną do odkodowania X'a jeśli znamy Y" - dr. Żebroski

$$H(X|Y) = \sum_{y\in\mathcal{Y}} p(y) H(X|y)$$

gdzie: $H(X|y) = \sum_{x\in\mathcal{X}} p(x|y) \log \frac{1}{p(x|y)}$

bardziej użytecznie: $H(X|Y) = \sum_{x,y} p(y) P(x|y) \log \frac{1}{p(x|y)} = \sum_{x,y} p(x,y) \log \frac{1}{p(x|y)}$

bo: $p(y) \cdot p(x|y) = p(y) \cdot \frac{p(x,y)}{p(y)}$

FAKT: $H(X,Y) = H(X|Y) + H(Y)$

$$H(X|Y) = H(X,Y) - H(Y)$$

najogólniej $H(X|Y,Z) = H(X,Y,Z) - H(Y,Z)$?

• $H(X|X) = \sum_{x\in\mathcal{X}} p(x) \sum_{x\in\mathcal{X}} p(x'|x)\log\frac{1}{p(x'|x)} = \sum_{x\in\mathcal{X}} p(x)\cdot 0 = 0$

$p(x'|x) = P(X=x' \mid X=x) = \begin{cases}1 : x'=x\\ 0 : x'\neq 0\end{cases}$

• $H(X,X) = \sum_{x\in\mathcal{X}}\sum_{x'\in\mathcal{X}} p(x,x')\cdot\log\frac{1}{p(x,x')} = \sum_{x\in\mathcal{X}} p(x)\log\frac{1}{p(x)} = H(X)$

$p(x',x) = P(X=x \wedge X=x') = \begin{cases}0 : x\neq x'\\ p(x) : x=x'\end{cases}$

• $I(X,X) = H(X)+H(X)-H(X,X) = H(X)$

Reguła łańcucha

$H(X,Y) = H(X|Y) + H(Y)$

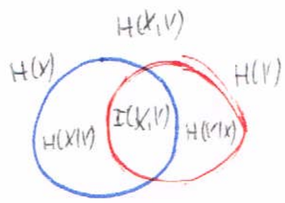
FAKT: • $I(X,Y) = H(\overset{X}{\cancel{Y}}) - H(X|Y)$

• $H(X,Y) = H(Y,X)$

• $I(X,Y) = I(Y,X)$

⟶ wnioski ⟶ $I(X,Y) = H(Y) - H(Y|X)$

$I(X,Y) \le \min\{H(X), H(Y)\}$



3 zmienne losowe i więcej zabawy

$P(X=x \wedge Y=y \wedge Z=z) = p(x,y,z) = P((X,Y)=(x,y) \wedge Z=z) = p((x,y),z)$

$\sum_{x,y,z} p(x,y,z)\log\frac{1}{p(x,y,z)} = H(X,Y,Z) = H(X,Y;Z) = H(X;Y,Z) = \ldots$

$\| $

$\sum_{x,y}\sum_{z} p((x,y),z)\log\frac{1}{p((x,y),z)}$

TW: Reguła łańcucha dla entropii łącznej:

$$H(X_1, X_2, \dots, X_n) = \sum_{i=1}^{n} H(X_i \mid X_{i+1}, \dots, X_n) = \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \dots, X_1)$$

$$\parallel$$

$$\sum_{i=1}^{n} H(X_i \mid (X_{i+1}, \dots, X_n))$$

na przykład:

$$H(X_1, X_2) = H(X_1 \mid X_2) + H(X_2)$$

TW: Reguła łańcucha dla entropii warunkowej:

$$H(X_1, X_2, \dots, X_n \mid Y) = \sum_{i=1}^{n} H(X_i \mid X_{i+1}, \dots, X_n, Y) = \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \dots, X_1, Y)$$

np: $H(X_1, X_2 \mid Y) = H(X_1 \mid X_2, Y) + H(X_2 \mid Y)$

Def: warunkowa informacja wzajemna:

$$I(X,Y \mid Z) = H(X \mid Z) + H(Y \mid Z) - H(X,Y \mid Z)$$

FAKT: $I(X,Y \mid Z) = H(X \mid Z) + H(Y \mid Z) - (H(X \mid Y,Z) + H(Y \mid Z)) =$

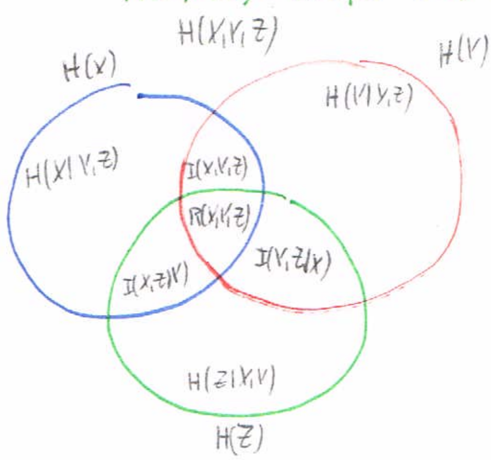
$= H(X \mid Z) - H(X \mid Y, Z)$

TW: Reguła łańcucha dla informacji wzajemnej

$$I(X_1, X_2, \dots, X_n ; Y) = \sum_{i=1}^{n} I(X_i, Y \mid X_{i+1}, \dots, X_n) = \sum_{i=1}^{n} I(X_i, Y \mid X_{i-1}, \dots, X_1)$$

Def: informacja wzajemna trójki zmiennych:

$$R(X,Y,Z) = I(X,Y) - I(X,Y \mid Z)$$



Słodki Panie to chyba koniec

KANAŁY KOMUNIKACYJNE

Def: kanał komunikacyjny - $(\mathcal{A}, \mathcal{B}, P)$ gdzie:
- $\mathcal{A}, \mathcal{B}$ - skończone niepuste zbiory
- $P: \mathcal{A} \times \mathcal{B} \to [0,1]$

$$(\forall a \in \mathcal{A}) \left(\sum_{b \in \mathcal{B}} \vec{P}(a,b) = 1 \right)$$

$$\vec{P}(a,b) = p(a \to b)$$ ← p-p otrzymano symbolu b pod warunkiem, że na wejściu był symbol a.

generalnie w kanałach:
$$p(y) = \sum_{x \in X} p(y|x) \cdot p(x)$$

in: symbole z $\mathcal{A}$ $\to$ $\boxed{\Gamma}$ $\to$ out: symbole z $\mathcal{B}$

konwencja:
- $a \xrightarrow{x} b \quad : \vec{p}(a,b) = x$
- $[\vec{p}(a,b)]_{\substack{a \in \mathcal{A} \\ b \in \mathcal{B}}}$ wiersze to $\mathcal{A}$, kolumny to $\mathcal{B}$

Def: przepustowość kanału Γ to C_Γ

$$C_\Gamma = \max_A I(A,B)$$ gdzie: A - z.l. o wart. w $\mathcal{A}$

B - z.l. ~~[illegible]~~ opisująca wyjście z Γ.

A daje $P(A=a) = p(a)$ dla $a \in \mathcal{A}$

Γ daje $P(a \to b) = P(B=b \mid A=a) = p(b|a)$

innymi słowy $p(a,b) = p(b|a) \cdot p(a)$

Uwaga: $C_\Gamma \leq \min\{\log|\mathcal{A}|, \log|\mathcal{B}|\}$

(P) kanał wierny

$\mathcal{A} = \mathcal{B} = \{0,1\}$ $\quad 0 \xrightarrow{1} 0$, $\quad 1 \xrightarrow{1} 1$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$ $\vec{p}(0,0) = 1$, $\vec{p}(0,1) = 0$

$$C_\Gamma = \max_A I(A,B) = \max_A (H(A) - H(A|B)) \overset{B=A}{=} \max_A (H(A) - \overset{0}{H(A|A)}) = \max(H(A))$$
$$\overset{2}{=} \log_2 |\mathcal{A}| = 1$$

(P)

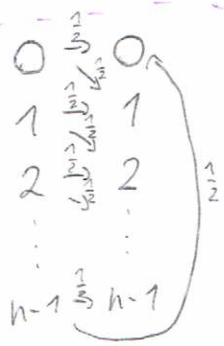
$0 \xrightarrow{\frac{3}{4}} 0$, $\quad 0 \xrightarrow{\frac{1}{4}} -1$

$1 \xrightarrow{\frac{2}{3}} 1$, $\quad 1 \xrightarrow{\frac{1}{3}} 2$

$$C_\Gamma = \max_A I(A,B) = \max_A (H(A) - \overset{0}{H(A|B)}) = \max(H(A)) = 1$$

A jest zdeterminowane przez B zatem $H(A|B) = 0$

(P)



$p(i|i) = \frac{1}{2}$

$p((i+1) \bmod n \,|\, i) = \frac{1}{2}$

jak liczyć pojemność?

$I(A,B) = H(B) - H(B|A)$

$H(B|A) = \sum_a p(a) H(B|a) = \sum_a p(a) \sum_b p(b|a) \cdot \log \frac{1}{p(b|a)} = \sum_a p(a) = 1$

nie zależy od a

$\frac{1}{2}\log 2 + \frac{1}{2}\log 2$

$H(B) = ?$

$p(B=i) = P(A=i)\cdot\frac{1}{2} + P(A=i-1)\cdot\frac{1}{2}$ ← dla rozk. jednostajnego A

$P(A=k) = \frac{1}{n}$ dostajemy rozk. jednostajny B.

$H(B) = \log(n) - 1 = \log\left(\frac{n}{2}\right)$

(P) $C_M = 0 \iff I(A,B) = 0$ niezależnie od rozk. A.

$H(B) - H(B|A) = 0$

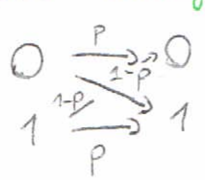
stąd $p(b) = p(b|a)$ dla $\forall a \in A$

na przykład:

$\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \vee \begin{bmatrix} \frac{1}{2}, & \frac{1}{4}, & \frac{1}{4} \\ \frac{1}{2}, & \frac{1}{4}, & \frac{1}{4} \end{bmatrix}$

takie same wiersze!

BSC - binarny kanał symetryczny



$\vec{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$

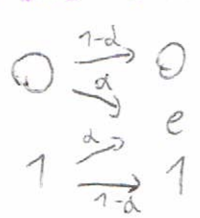
pojemność kanału BSC

$H(B|A) = \sum_a p(a) H(B|a) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} = H(p)$

$H(A) \le H(B) \le \log|B| = 1$

$\max_A H(X) = \log|A|$

(P) BEC - binarny erasure chanel



$\vec{P} = \begin{bmatrix} 1-\alpha & 0 & \alpha \\ 0 & 1-\alpha & \alpha \end{bmatrix}$

$I(A,B) = H(B) - H(B|A)$

tutaj: $H(B|0) = H(B|1) = H(\alpha)$

$H(B|A) = H(\alpha)$

a $H(B) = ?$, $P(A=0) = \pi$, $P(A=1) = 1-\pi$

$H(B) = (1-\alpha)\pi \cdot \log \frac{1}{(1-\alpha)\pi} \ldots$

$= H(\alpha) + (1-\alpha) H(\pi)$

$C_{BEC} = \max_{\pi \in [0,1]} (1-\alpha) H(\pi) = 1-\alpha$

$I(A,B) = H(\alpha) + (1-\alpha) H(\pi) - H(\alpha) = (1-\alpha) H(\pi)$

def: kanał symetryczny − $\Gamma = (\mathcal{X}, \vec{P}, \mathcal{Y})$ jest symetryczny, jeśli każdy wiersz $\vec{P}$ jest permutacją pierwszego wiersza, a każda kolumna permutacją pierwszej kolumny.

wtedy: $C_\Gamma = \log(|\mathcal{Y}|) - H(\vec{r})$ ← wiersz

$$H(\vec{r}) = -\sum_{i=1}^{n} p_i \log p_i$$

$H(Y|X) = \sum_x p(x) \sum_{i=1}^{n} p_i \log \frac{1}{p_i} = H(\vec{r}) \sum p(x) = H(\vec{r})$ bo $\{p(y|x) : y \in \mathcal{Y}\} = \{p_1, \dots, p_n\}$

$\max_x H(Y) = \log|\mathcal{Y}|$ bo kładąc $p(x) = \frac{1}{|\mathcal{X}|}$ ← rozkł. jednost.

dostajemy $p(y) = \sum_x p(x)$, $p(y|x) = \frac{1}{|\mathcal{X}|} \sum_x p(y|x)$

stąd $p(y) = \frac{1}{|\mathcal{Y}|}$ czyli $H(Y) = \log|\mathcal{Y}|$

---

def: kanał słabo symetryczny − jeśli wiersze są swoimi permutacjami oraz suma każdej kolumny #

wtedy $C_\Gamma = \log(|\mathcal{Y}|) - H(\vec{r})$

---

def: Łańcuch Markowa: $X \to Y \to Z$ jeśli $p(x,y,z) = p(x) \cdot p(y|x) \cdot p(z|y)$

równoważnie $p(z|x,y) = p(z|y)$

Wniosek: $X \to Y \to Z \Leftrightarrow Z \to Y \to X$

TW: jeśli $X \to Y \to Z$ to $I(X,Y) \geq I(X,Z)$

TW: Shannona o kodowaniu w kanałach:

1) jeśli $R < C$, istnieje $(2^{nR}, n)$ - kod, dla którego $\lambda^{(n)} \xrightarrow[n\to\infty]{} 0$

2) jeśli $(2^{nR}, n)$-kod ma własność $\lambda^{(n)} \to 0$, to $R \leq C$

$R = \log \frac{M}{n}$ M - liczba słów, n - długość słowa, R - efektywność