Hírközlés elmélet

dr. Bitó János

Szerda 8^{15} IE220

csütörtök 10^{15} E1C

4 kis ZH van

inkrementális ZH rendszer!

2 legjobb ZH átlaga

10 hallgató felett

össz 15 pont

4 pont 4,5-7 1
7,5-9 - 2
9,5-11 - 3
11,5-13 - 4
13,5-15 - 5

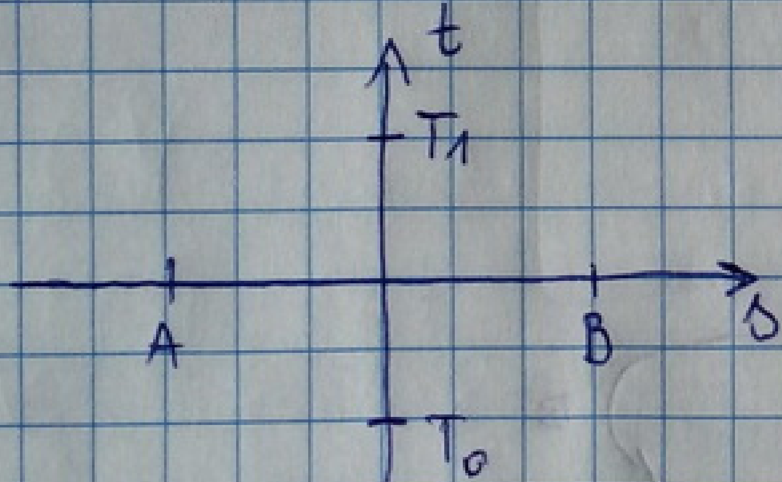
ZH: 3 rész → teszt példa tétel

10 tesztes feleletbefejezős (információ)
tétel kifejtés
példa számolás

1. előadás

információ forrásai üzeneteinek eljuttatása az információ nyelőjéhez = HÍRKÖZLÉS

+ zaj, zavar, interferencia (mesterséges vagy természetes)



időben
vagy elkülönült pontok
térben

forrás kódoló / tömörítés [időben elkülönültkor]

- mi az információ? felkel-e a nap holnap?

  $P(nap) \to 0$ Holnap nem kel fel a nap ⇒ nagy az információ tartalma

- hogyan mérem az információt?

  - kell egy kvantitatív mérték

  ~ Hartly (1928, Bell System Technical Journal) "Transmission of information" cikke

  () az ember alapvető igénye a kommunikáció

ANALÓG! ~ szinuszos jelek reprodukciója volt a feladat (telefon)

↓ ehelyett!

esemény: $\xi \sim \{x_1, x_2, x_3 \dots x_n\}$ infót csak akkor viszünk át, ha az esemény több kimenetelű
– lényeg a döntés a kimenetelek közt

$\#\xi = D$ (# ~ ahány esetet felvehet az esemény) $I(\xi)$

$\bar{\xi} = \{\xi_1, \xi_2, \xi_3 \dots \xi_n\}$ n db val. változó együttese

$\#\bar{\xi} = D^n \quad [D \cdot D \cdot D \dots D]$ $\qquad I(\bar{\xi}) = n \cdot I(\xi)$

információ tartalom van, ha: több kimenetel van!

$$\boxed{\begin{array}{ll} \#\xi = D & I(\xi) = \log_a D \\ \#\bar{\xi} = D^n & I(\bar{\xi}) = n \cdot I(\xi) \\ \multicolumn{2}{c}{\log_a D^n = n \cdot \log_a D} \end{array}}$$

ha $a = 2 \rightsquigarrow \log_2$
⇓
ld [bit]
ha $a = 10 \rightsquigarrow \log_{10}$
⇓
lg [Hartly]

DE EZ MÉG NEM JÓ:

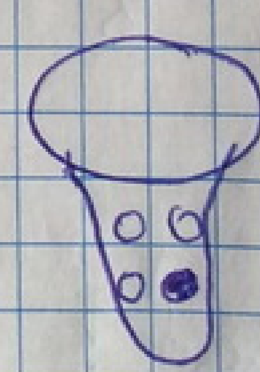
példa:
kalap
4 golyó
1-et húzunk


A

$D = 2$

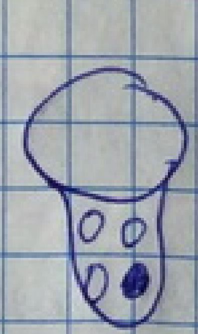
$\mathrm{ld}\, D = 1$ [bit]

binary digit


B

() azt várnám hogy fehér lesz ezért jobban meglepődök! több info

B



● húzok: $\mathrm{ld}\,\frac{4}{1} = 2$ [bit]

○ $\mathrm{ld}\,\frac{4}{3} = 0{,}415$ [bit]

a baj az, hogy a Hartly mérték nem veszi figyelembe a valószínűségi eloszlást!

[bit] az információmennyiség mértéke

$\boxed{\text{BIT}}$

$\bullet = 2$

$\circ = 0,415$

$$\overset{\text{valség}}{\frac{1}{4}} \cdot \underbrace{2}_{\text{info tartalom}} + \overset{\text{valség}}{\frac{3}{4}} \cdot \underbrace{0,415}_{\text{info tartalom}} = \underline{0,811\,[bit]}$$

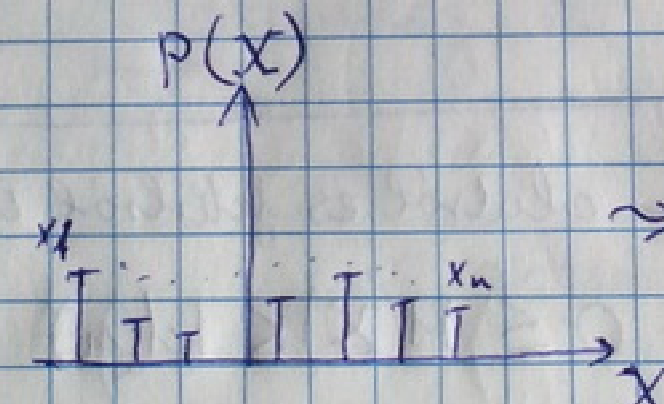
→ kevesebb infó mint egyenletes eloszlásnál

Claude Shannon (1948, BSTJ) – "A Mathematical Theory of Communication"

legyen:

$$\underbrace{X}_{\text{diszkrét értékű val. változó}} = \{x_1 \dots x_n\}; \quad \underbrace{p(X)}_{\text{sűrűség fv.}} = (p_1 \dots p_n)$$

$p_i = p(x_i)$
i-dik esemény



$$\rightsquigarrow \underbrace{F_X(y)}_{\text{eloszlásfüggvény}} = \sum_{y=-\infty}^{\infty} p_X(y) = \underbrace{\underline{\underline{1}}}_{\text{teljes eseménytér!}}$$

Def: $I(x_i) = ld \frac{1}{p(x_i)}$ ⇝ ez egy esemény információja

minél valószínűbb az esemény annál kisebb az info. tartalom

$\to -ld\,(p(x_i))$ [bit]

(Self Information – az esemény saját információ tartalma)

Def: Entrópia, Átlagos információtartalom
(entropy)

$$H(X) = E\{I(x_i)\} = \sum_{x_i \in X} p(x_i) \cdot \overbrace{ld \frac{1}{p(x_i)}}^{I(x_i)}$$

( ) ⇒ Marci és a lájkok

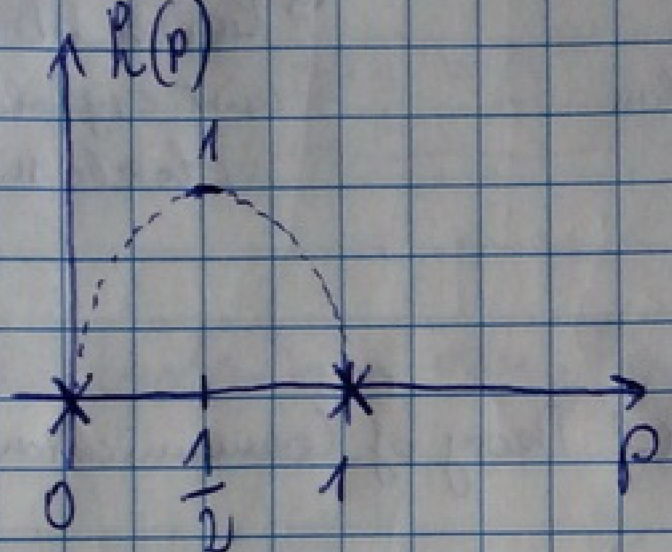
legyen $X = \{x_1, x_2\}$, $p(X) = \{p(x_1), p(x_2)\} \Rightarrow \{\overset{p}{p(x_1)};\ \overset{1-p}{1-p(x_1)}\}$

ezért ez 1 paraméteres

$$H(X) = \sum_{x_i \in X} p(x_i) \cdot ld\,\frac{1}{p(x_i)} = \underbrace{p\, ld\,\frac{1}{p} + (1-p)\cdot ld\,\frac{1}{1-p}}$$

$h(p)$ = bináris entrópia fv

határérték probléma a széleken!



legyen $\boxed{x = \frac{1}{p}}$

$p \cdot ld\,\frac{1}{p}$; $\frac{1}{x} \cdot ld\,x$

$$\lim_{x\to\infty} \frac{1}{x}\cdot ld\,x = \frac{1}{\ln(2)} \cdot \overset{L'H'}{\lim_{x\to\infty} \frac{\ln(x)}{x}} = \frac{1}{\ln(2)} \lim_{x\to\infty}\frac{1}{x}$$

$\Rightarrow 0$

irodalom kérülés ! ! ! szólni Bitokról

Tétel: az entrópia alulról és felülről is korlátos

$$\#X = n;\quad 0 \le H(X) \le ld(n)$$

2. előadás

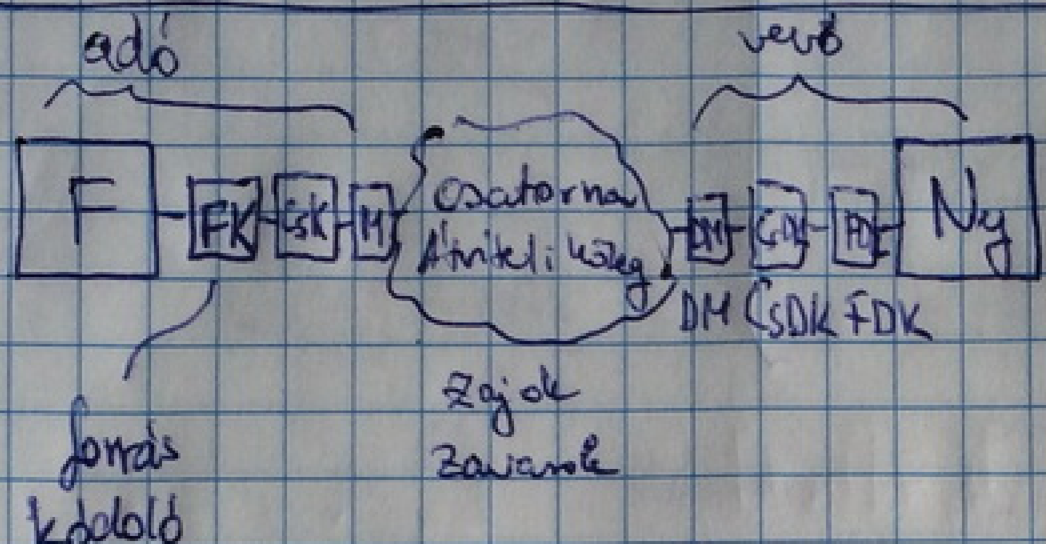
Proakis, Salehi: Communication Systems Engineering

Dallos: Tantárgyi segédlet a hírközléselmélet tárgyhoz

Frigyes I.: Hírközlő rendszerek

Csibi S.: Információ közlése és feldolgozása

IRODALOM



FK: ami redundáns, azt törölni [veszteséges és veszteségmentes]

CsK: a fellépő zavarok csillapítása (hibajavító + redundancia hozzáadó)

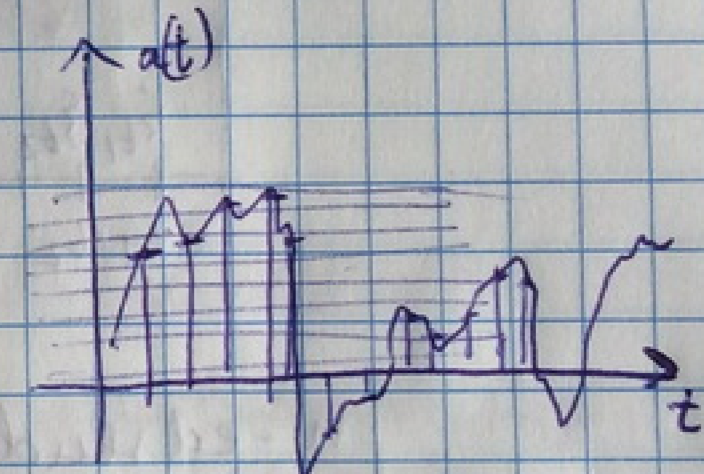
M: modulátor

DM: demodulátor

CsDK: csatorna dekóder

FDK: forrás dekódoló

$\boxed{F}$ analóg $a(t)$

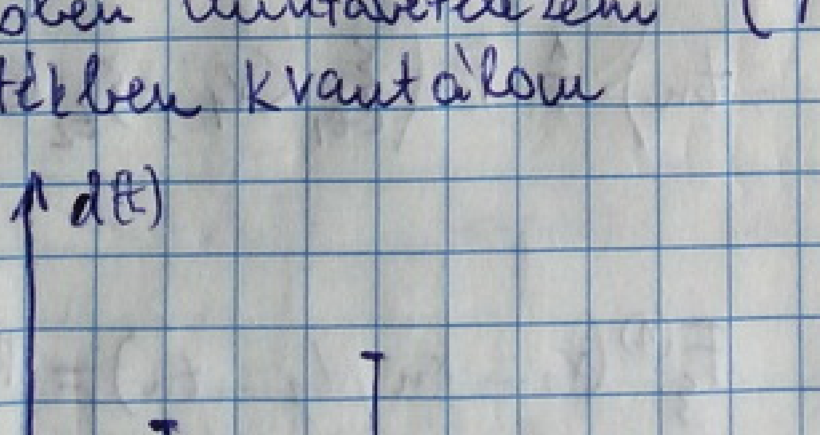


régen Bell idejében

Ha sávkorlátozott $\left(\frac{1}{2T}\right)$
mv. tétel!

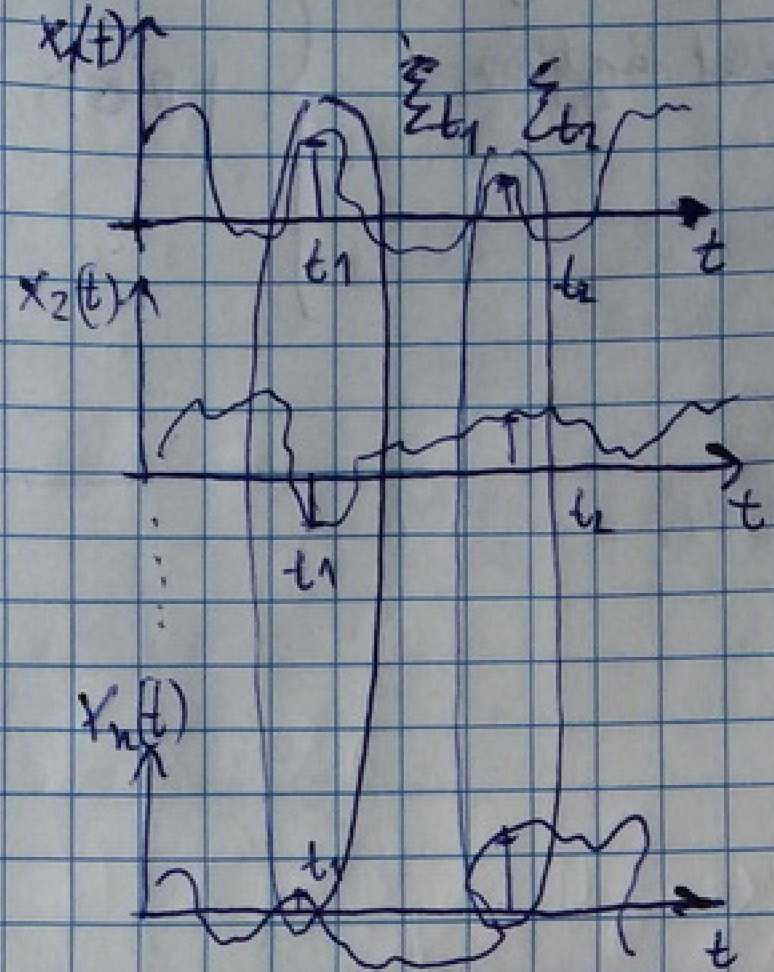
ma digitális jel:

- időben mintavételezem (T)
- értékben kvantálom



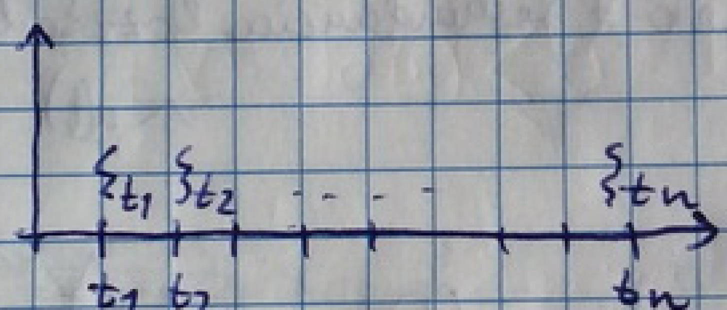
Sztochasztikus folyamatok

① a folyamat realizációjának az együttese



együtt ők a sztochasztikus folyamat!

② Valószínűségi változók rendezett (időben) serege



n-edrendű eloszlás fv (n dimenziós)

$\cancel{F_\xi^{(n)}(\xi_{t_1}, \xi_{t_2}, \ldots, \xi_{t_n}, t_1, t_2 \ldots t_n) \triangleq}$

eloszlás fv.

$$F_\xi^{(n)}(x_1, \ldots, x_n, t_1, t_2, \ldots t_n) \triangleq P(\xi_{t_1} < x_1, \xi_{t_2} < x_2, \ldots \xi_{t_n} < x_n, t_1 \ldots t_n)$$

- Erősen stacionárius folyamat: $F_\xi^{(n)}(x_1 \ldots x_n, t_1 \ldots t_n) = F_\xi^{(n)}(x_1 \ldots x_n, t_1+\tau, t_n+\tau)$

~ időben invariáns

$\forall \tau, \forall n, \forall \{t\}$ időben bármilyen eltolásra érzéketlen

Várható értéke

$$m_\xi(t) \Rightarrow E(\xi_t) = \int_{-\infty}^{\infty} x\, f_\xi(x,t)\, dx$$

erősen stacion. foly. várható értéke időfüggetlen

$$m_\xi(t) = m_\xi \quad \forall\, t \text{ esetén}$$

erősen stac. foly → ergodikus foly vagy ergodikus foly. keveréke

- ergodikus folyamat

bármelyik realizációból megvalósítható

$$\underset{\text{időátlag}}{A(\xi)} = \lim_{T \to \infty} \int_{t_0}^{t_0+T} \xi_t\, dt = m_\xi$$

- legyen

$$\cancel{E\xi(t) = M\xi} \qquad \underset{\text{energia}}{\overbrace{e}^{\text{jel energiája}}}\xi(t) = \underset{\text{várható érték}}{E}\{\xi_t^2\} = \int_{-\infty}^{\infty} x^2 \cdot f_\xi(x,t)\, dx$$

- Autokorreláció

$$R_\xi\{t_1, t_2\} = E\{\xi_{t_1} \cdot \xi_{t_2}\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 \cdot x_2 \cdot \underbrace{f_\xi}_{\text{sűrűségfv.}}(x_1, x_2, t_1, t_2)\, dx_1\, dx_2$$

$$R_\xi(t_1, t_1+\Delta T) = R_\xi(t_2, t_2+\Delta T) \rightsquigarrow \text{független az abszolút időtől}$$

csak az időkülönbségtől függ (ΔT)

$\Downarrow$

peremfelt: $\forall \Delta T$ és $\forall (t_1, t_2)$

- Gyengén stacionárius (Wide-Sense-Stationary)

(másodrendben stacionárius → gyengén stacionárius)

① $\Rightarrow m_\xi(t) = m_\xi \quad \forall t$ időfüggetlen

② $\Rightarrow R_\xi(\Delta T)$ korreláció csak a különbségtől függ

- ha ~~n~~ n rendben stac → $n-1$

de n rendben stac $\not\Rightarrow n+1$

• Memória mentes: $\mathcal{X}$

– forrás

$$P(\xi_n = x_n \mid \overbrace{\xi = x_1, \ldots, \xi_{n-1} = x_{n-1}}^{\text{előző betűkészlet}}) = P(\xi_n = x_n)$$

diszkrét

sűrűség eloszlás

folytonosnál (f)

példa:

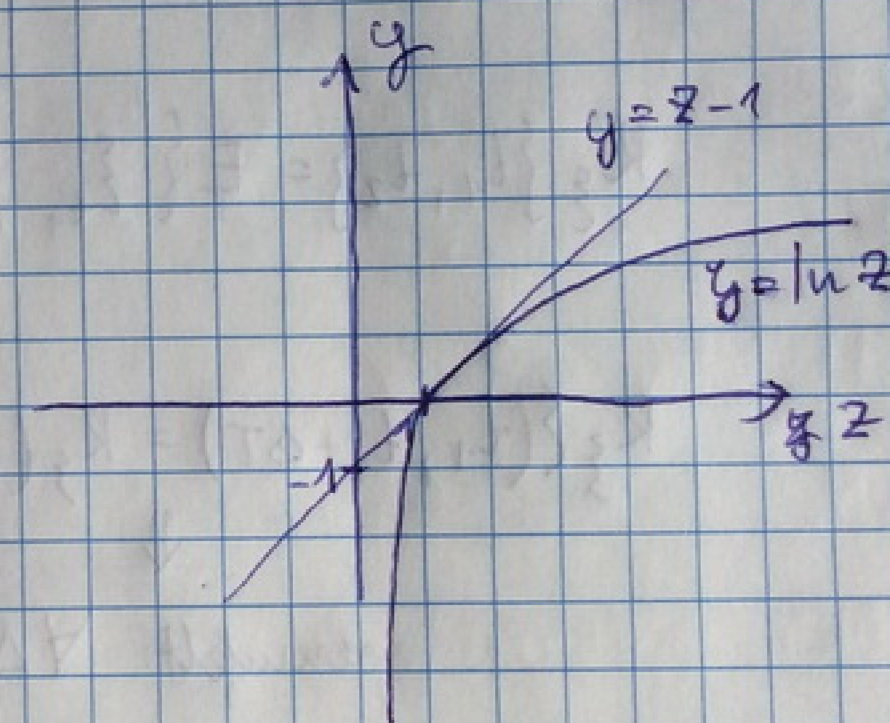
$$\# \mathcal{X} = 26$$

$$H_0(\mathcal{X}) = \sum_{x_i \in \mathcal{X}} p x_i \, ld \frac{1}{p x_i} = \sum \frac{1}{26} \cdots = 4,7 \, [bit]$$

entrópia

$$\log_a x = \frac{\ln(x)}{\ln(a)}$$

(Tétel: $0 \leq H(\mathcal{X}) \leq ld(n)$



$$H(\mathcal{X}) - ld(n) \leq 0$$

$$\sum_{x_i \in \mathcal{X}} p x_i \cdot ld \frac{1}{p x_i} - \overbrace{\sum_{x_i \in \mathcal{X}} p x_i}^{\text{mivel} = 1} \cdot ld\, n \leq 0$$

$$\frac{1}{n \cdot p x_i} = 1$$

$$\Rightarrow p(x_i) = \frac{1}{n}$$

$$\sum p x_i \cdot ld \frac{1}{n \cdot p x_i} \leq 0$$

↓

$$\sum p x_i \cdot \ln \underbrace{\left(\frac{1}{n p x_i}\right)}_{z} \leq 0$$

$$\frac{1}{\ln 2} \cdot \sum_{x_i \in X} p x_i \cdot \ln \frac{1}{p x_i \cdot n} \leq \underbrace{\frac{1}{\ln(2)} \sum_{x_i \in X} p x_i \left[\overbrace{\frac{1}{n \cdot p x_i} - 1}^{z-1}\right]}_{} =$$

$$\underbrace{\frac{1}{\ln(2)} \left[\underbrace{\sum_{x_i \in X} \frac{1}{n}}_{1} - \underbrace{\sum p x_i}_{1}\right]}_{0}$$

német nyelvre

$$H(\mathcal{X}) = 4,1 \text{ bit}$$

párok $H_2(\mathcal{X}) = 3 \text{ bit}$

$$H_\infty(\mathcal{X}) = 1,6 \text{ bit}$$

$R(X) = H_0(X) - H(X)$ — entrópia a feltehető kérdések átlagos száma

redundancia — érdemes annyi bitet kódolni, amelyre gyakran fordul elő.

3. előadás

- eddig amiket tanultunk

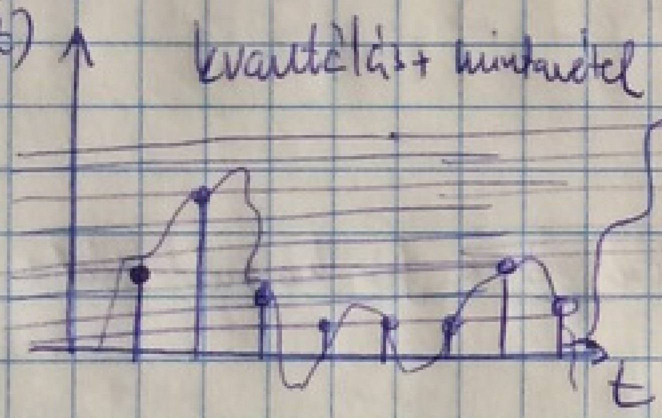
$$F_\xi^{(n)}(\xi_{t_1} < x_1 \dots \xi_{t_n} < x_n, t_1 \dots t_n) \overset{\text{röviden}}{=} F_\xi^{(n)}(\bar{x}, \bar{T}) \quad \forall \Delta T \ \forall \{\bar{T}\}$$

n-ed rend. stacionárius: $F_\xi^{(n)}(\bar{x}, t_1+\Delta T, t_2+\Delta T \dots t_n+\Delta T) = F_\xi^{(n)}$

WSS $m_\xi(t) = m_\xi$ $\forall$ t-re, $R_\xi(\Delta T)$

erősen stac, ha $\forall$ n-re!

gyengén, ha nem $\forall$n-re ~

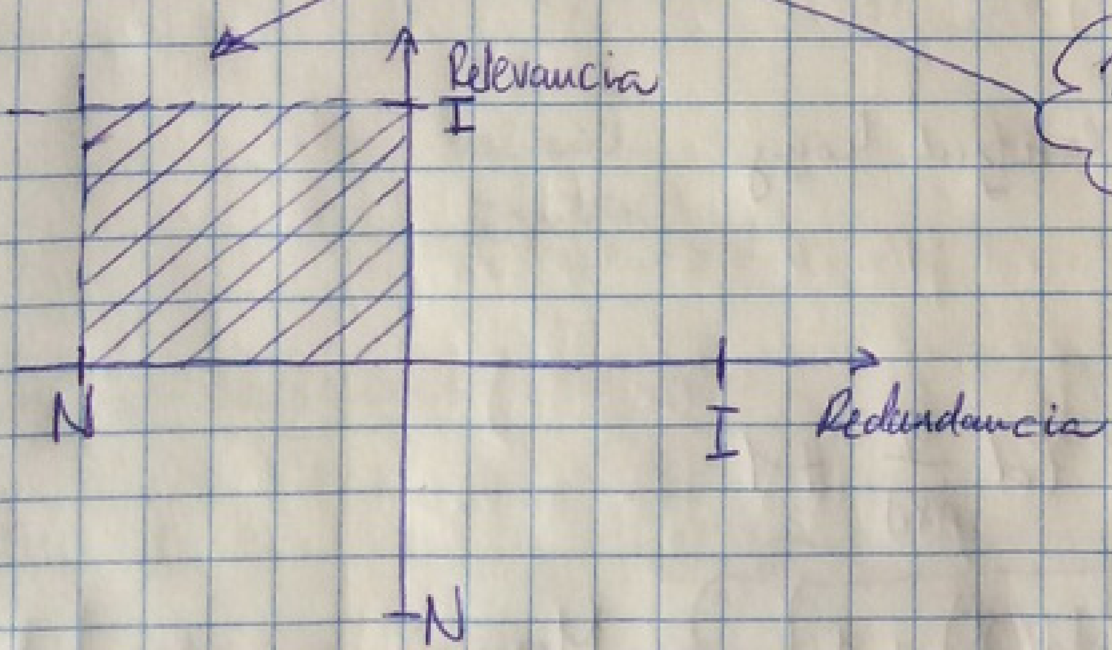


de! ha elcsúszik jel felfelé pl. akkor a kvantáló telítésbe megy mindig

ha változik időben a várható érték → nem stacionárius

VALÓS VILÁG FOLYAMATAI

releváns információ átvitel kevés redundanciával



$$R(X) = \underbrace{\overset{\text{egyenletes eloszlás}}{\overbrace{H_0(X)}} - H(X)}_{\text{cél ennek a minimalizálása}}$$

redundancia

német szöveg:

$\# X = 26$ $\quad H_0(X) = ld\, 26 = 4,7$ [bit]

$p(X)$ $\quad H_1(X) = 4,1$ [bit] ← növelt a redundancia

$H_\infty(X) = 1,6$ bit ← teljes szövegre (redundánsabb)

Forráskódolás

- diszkrét, DMS, D+M

DMS: discrete memoryless source

D+M: discrete memory source

- kódolási szabály (Q)

$Q(x_i) = c_i$; dekódolhatóság!, fix hosszú kódolás [5 bit]

a → 00000

b → 00001

példa: $X = \{a, b, c, d\}$ $\quad C = \{\overset{a}{00}, \overset{b}{01}, \overset{c}{10}, \overset{d}{11}\}$

$p(X) = \{\frac{1}{4}; \frac{1}{4}; \frac{1}{4}; \frac{1}{4}\}$ egyenletes eloszlás esetén jó a fix hosszú kód

- változó hosszú kódolás:

$p(X)$ $\quad l_i = l(x_i)$ $\quad \underset{\text{átlagos kódhossz}}{L_X} = \sum p(x_i) \cdot l(x_i)$

- entrópia kódolás ⇝ $l(x_i)$-t úgy választjuk hogy

$$\cancel{l(x_i) \geq ld \frac{1}{p(x_i)}}$$

$$ld \frac{1}{p(x_i)} \leq l(x_i) \leq ld \frac{1}{p(x_i)} + 1$$

$$\boxed{H(X) \leq L_X < H(X) + 1}$$

Shannon I.

Memória-mentes forrásra!!

Shannon-Fano kód → a forrás minden szimbóluma (2^{-x})

$\left(\frac{1}{4}, \frac{1}{8}, \frac{1}{16} \dots\right)$

példa: $p(a) = \frac{1}{2}$ $p(b) = \frac{1}{4}$ $p(c) = \frac{1}{8}$ $p(d) = \frac{1}{16}$ $p(e) = \frac{1}{16}$

2^{-1} 2^{-2} 2^{-3} 2^{-4} 2^{-4}

$l_i \Rightarrow$ 1 2 3 4 4 [bit]

kompr. kód hossz

D+M forrás:

→ A memória:

$$\{X\} \quad p(X_n = x_n \mid X_1 = x_1, X_2 = x_2, \dots X_{n-1} = x_{n-1}) = p(x_n \mid x_1 \dots x_{n-1})$$

feltételes valószínűség

→ lehetne feltételes entrópia is!

$$H(X_n \mid X_1 \dots X_{n-1}) = \sum_{x_1 \dots x_n \in X^{(n)}} p(x_1, \dots, x_n) \cdot ld \frac{1}{p(x_n \mid x_1 \dots x_{n-1})}$$

FELTÉTELES ENTRÓPIA

együttes entrópia:

$$H(\bar{X}_1^n) = H(X_1 \dots X_n) = \sum_{\bar{X}^n} p(x_1, \dots, x_n) \cdot ld \frac{1}{p(x_1 \dots x_n)}$$

n darab val. vált. együttese

együttes valószínűség: $p(x_1, x_2 \dots x_n) = p(x_1) \cdot p(x_2 \mid x_1) \cdot p(x_3 \mid x_2, x_1) \dots =$

$$\prod_{i=1}^{n} p(x_i \mid x_1, \dots x_{i-1})$$

$$H(\bar{X}^n) = -\sum_{X} \prod_{i=1}^{n} p(x_i \mid x_1 \dots x_{i-1}) \cdot ld \prod_{i=1}^{n} p(x_i \mid x_1 \dots x_{i-1}) =$$

$$\Rightarrow -\sum_{X} \underbrace{\left(\prod_{i=1}^{n} p(x_i \mid x_1 \dots x_{i-1})\right)}_{}^{\text{együttes val.ség}} \underbrace{ld\, p(x_1)}_{\text{súlyozva } x_1\text{-gyel}} - \sum p(x_1 \dots x_n) \cdot ld\, p(x_2 \mid x_1) - \dots -$$

$$\underbrace{\sum_{X} p(x_1 \dots x_n)\, ld\, p(x_n \mid x_1 \dots x_{n-1})}_{\text{ezek feltételes entrópiák}}$$

$$H(\bar{X}^n) = H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1, X_2) \dots + H(X_n \mid X_1, \dots X_{n-1}) =$$

$$H(\bar{X}^n) = \sum_{i=1}^{n} H(X_i \mid X_1 \dots X_{i-1})$$

<u>1 szimbólumra eső entrópia</u>

$$\boxed{H_n(\bar{X}^n) = \frac{1}{n} H(\bar{X}^n)}$$

– Sztochasztikus folyamat entrópiája

ha van $\Rightarrow$ $\boxed{H_\infty(X) = \lim_{n\to\infty} \frac{1}{n} \cdot H(\bar{X}^n)}$ — együttes entrópia

vagy $\Rightarrow$ $H_\infty(X) \overset{\exists}{=} \lim_{n\to\infty} H(X_n \mid X_1 \dots X_{n-1})$ ha <u>WSS</u> legalább a folyamat

DMS esetben

$\hookrightarrow$ $p(x_1, x_2 \dots x_n) = \prod_{i=1}^{n} p(x_i)$

ftlen esemény el

$H(\bar{X}^n) = \sum_{i=1}^{n} H(X_i)$ $\rightsquigarrow$ mert ftlenek az entrópiák?

ha WSS a forrás (időftlen $\Rightarrow$ index ftlen)

$H(X_i) = H(X_{i+1}) = \cancel{H(X_{i-1})}\ H(X_{i-1}) \dots$

$\Downarrow$

$H(\bar{X}^n) = n \cdot H(X)$

Forrás-kódolás:

átl. kódhossz

Shannon I. $H(X) \le L_X \le H(X)+1$

Shannon Fano $p(x_i) = 2^{-k}$ $k \in \mathbb{N}^+$

$L_X = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + \underbrace{\frac{1}{8}}_{\text{„élése”}}\cdot 4$

[abcd példára]

~ kedden ZH

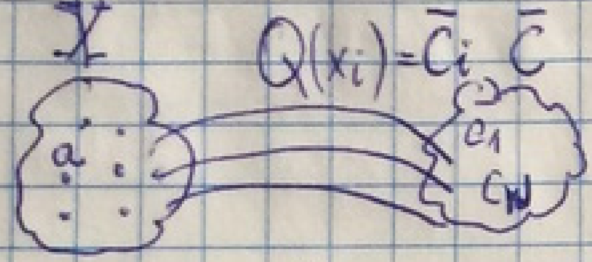
E1B $18^{\underline{15}}$

---

4. előadás

teszt, példa, tétel

Forrás kódolás



forrás ABC diszkrét

kód ABC diszkrét

kölcsönösen egyértelmű kapcsolat

DMS + memória

pillanat kódolónak kell lenni ⇝ azonnal el kell tudjam dönteni a forrás ABC elemét!

l_i : $\bar{C}_i = (c_{i1}, \dots, c_{il_i})$ legyen $L_X = \sum_{x_i \in X} p(x_i) \cdot l_i$

egy kód hossza

kódszimbólumok akár [0,1] de lehet más is!

[Shannon I] $H(X) \le L_x < H(X)+1$

ha $p(x_i) < p(x_j) \Rightarrow l_i > l_j$

entrópia kódolás elve ez [Shannon-Fano]

ha $\forall p(x_i) \in 2^{-k_i}$

$\Downarrow$

$l_i = k_i = ld \frac{1}{p(x_i)}$

ez teljesíti-e a Shannon I-et?

$\Downarrow$

$H(X) = \sum_{x_i} p(x_i) \cdot ld \frac{1}{p(x_i)}$ → egyenlőség van "alulról"

ez alá nem tudok menni

$L_x = \sum p(x_i) l_i$

$l_i = ld \frac{1}{p(x_i)}$

$h_Q = \frac{H(X)}{L_X}$ → mindig kisebb, mint 1

hatékonysága a forráskódnak

Shannon-Fano → minimális átlagos kódhossz

Példa:

X	$p(x_i)$	fix hosszú A
x_1	1/2	00
x_2	1/4	01
x_3	1/8	10
x_4	1/8	11

$HX = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{4}\cdot 3 \Rightarrow 1,75$ [bit]   ez pillanatkód!!!

A eset

$L(X) = 2$ fix hosszú kód miatt!

$h_{Q_A} = \frac{1,75}{2} = 0,875$ miért nem 100% → nem használtuk ~~az~~ az apriori ismeretet (a valószínűségeket)

→ fix hosszú kódot mindig lehet dekódolni!

0 1|0 0|1 1|0 1|1 0|0 0|1

legyen
$x_1 \to 0$
$x_2 \to 1$
$x_3 \to 00$
$x_4 \to 11$

B eset

$L_{X_B} = 1\cdot\frac{3}{4} + 2\cdot\frac{1}{4} \Rightarrow 1,25$

kisebb az entrópiánál ????

nem dekódolható!

C prefix kód $x_1 \rightsquigarrow 0$ $x_3 \rightsquigarrow 110$

eset (változóhosszú kód) $x_2 \rightsquigarrow 10$ $x_4 \rightsquigarrow 111$

$\bar{C}_i = (c_{i1} \dots c_{il_i})$ akkor $\cancel{\exists}$ olyan $\bar{C}_j$ ahol $l_j > l_i$, $C_j = (\underbrace{c_{i1} \dots c_{il_i}}_{\text{az eleje olyan mint a másik kód}}, c_{jl_i+1} \dots c_{l_j})$

$$L_x = \frac{1}{2} + \frac{1}{4} \cdot 2 + \frac{1}{4} \cdot 3 = \underline{\underline{1{,}75}}$$

$$h_{Q_C} = \frac{1{,}75}{1{,}75} = 100\%$$

ez prefix mentes kód $\rightsquigarrow$ dekódolható

(Shannon-Fano kód)

<u>pillanatkód</u>

D szeparátort vezetünk be: pl egy vesszőt! „," [comma-code]

$x_1 \rightsquigarrow 0$ elválasztó

$x_2 \rightsquigarrow 01$

$x_3 \rightsquigarrow 011$

$x_4 \rightsquigarrow 0111$

$$L_x = \frac{1}{2} + \frac{1}{4} \cdot 2 + \frac{3}{8} + \frac{1}{2} \Rightarrow 1{,}875$$

$$h_{Q_D} = \frac{1{,}75}{1{,}875} \rightsquigarrow$$

E prefix és comma-code

$x_1 \rightsquigarrow 0$

$x_2 \rightsquigarrow 10$ szeparátor

$x_3 \rightsquigarrow 110$

$x_3 \rightsquigarrow 1110$

$L_x = 1{,}875$

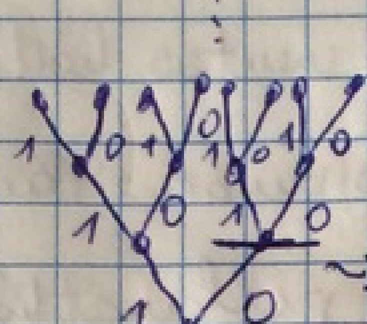
Prefix-kód generálása:

egy kód prefix $\Longleftrightarrow$ ha teljesül a Kraft egyenlőtlenség.

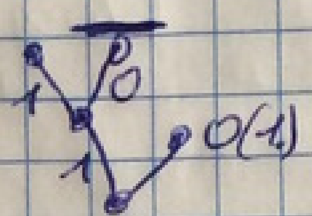
i-edik kódhossza

Kraft egyenlőtlenség: $\sum_{i=1}^{N} 2^{-l_i} \leq 1$ [bináris kódolásra igaz]

bin. fa:



ha elkezdem 0-val → lefűrészelem az ágat



$N - l_i$ szintű részfákat vágok ki egy l_i-hosszú kód szerkesztésekor

$$\sum_{i=1}^{N} 2^{N-l_i} \leq \underbrace{2^N}_{\text{teljes fa!}} \rightsquigarrow \boxed{\sum_{i=1}^{N} 2^{-l_i} \leq 1} \checkmark$$

Kraft.

(a teljes fánál többet nem vághatok ki)

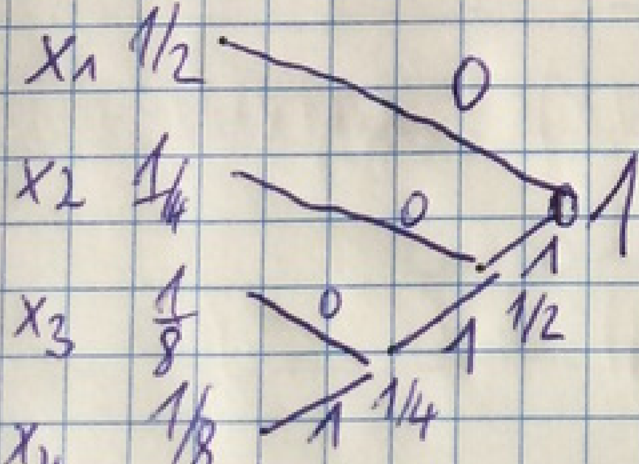
Huffmann:

① ⇝ előfordulási valség szerint sorba rendezed!

② ⇝ bináris fa, 2 legkisebb valségű eseményt rakd össze!

→ rekurzívan!

bináris fa amíg nem érek 1-hez!



$X_1 \rightsquigarrow 0$ ✓ levágom

$X_2 \rightsquigarrow 10$ ✓ prefix + Shannon Fano!

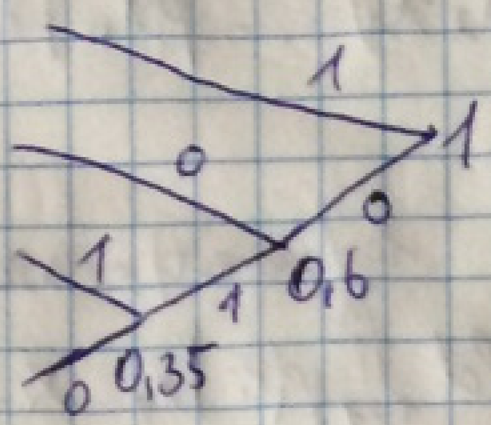
$X_3 \rightsquigarrow 110$ ✓

$X_4 \rightsquigarrow 111$ ✓

DE!!! ⇝ akkor is működik ha a val.ségek nem 2^{-k_i} szerint oszlanak el!

Pl:

	$p(x_i)$
x_1	0,4
x_2	0,25
x_3	0,2
x_4	0,15



	C
x_1	→ 1
x_2	→ 00
x_3	→ 011
x_4	→ 010

HUFFMANN

0,5 0,6 0,45

$$L_x = 0,4 + \frac{1}{4}\cdot 2 + \frac{1}{5}\cdot 3 + 0,15\cdot 3 = 1,95$$

$$H(X) = 0,4\cdot \mathrm{ld}\frac{1}{0,4} + 0,25\cdot \mathrm{ld}\frac{1}{0,25} + 0,2\cdot \mathrm{ld}\frac{1}{0,2} + 0,15\cdot \mathrm{ld}\frac{1}{0,15} =$$

HUFFMANN-hoz kell a valószínűség!

nem jó kevés kódnál, és nagy valószínűség differenciánál!

használjuk a forráskiterjesztés módszerét!

$x_1 x_1$

$x_1 x_2$ — kell ismerni az eseménypárok val.ségét

$x_1 x_3$ — 4^2 esetem van!

$x_1 x_4$

$x_2 x_1$

⋮

ha DMS → akkor $p(x_1, x_2) = p(x_1)\cdot p(x_2)$

memória mentes

$$p(x_1 \dots x_n) = \prod_{i=1}^{n} p(x_i)$$

$$H(\underline{X}^n) = H(X_1 \dots X_n) \overset{DMS}{\underset{\text{stacionáris}}{=}} n\cdot H(X)$$

stacionáris időben(n) nem változik $H(X) = H(X_i)\ \forall\, i$-re

kiterjesztett forráshoz rendelt

$$n\cdot H(X) = H(X_1 \dots X_n) \le L_{x_1,\dots x_n} < \overbrace{H(\underline{X}^n)}^{n\cdot H(X)} + 1 \quad \Big|\ n \text{ darabra}$$

n-darab val.változó által generált eseményhez tartozó kód

$$H(X) \le L_x \le H(X)\cdot \frac{1}{n}$$ → ha a kiterjesztést növelem akkor 100% felé tartok!

ez jó DMS esetben!

MS esetén

$$p(X_1 \ldots X_n) = p(X_1) \cdot p(X_2 | X_1)\, p(X_3 | X_2 \cdot X_1) \cdot \ldots \cdot p(X_n | X_1 \cdot \ldots X_{n-1})$$

$$= \prod_{i=1}^{n} p(X_i | X_1 \ldots X_{i-1})$$

log(szorzat) = össz (log)

$$H(X_1 \ldots X_n) = -\sum \prod_{i=1}^{n} p(X_i | X_1 \ldots X_{i-1}) \cdot \mathrm{ld} \prod_{i=1}^{n} p(X_i | X_1 \ldots X_{i-1}) =$$

n val vált. együttes entrópiája

$$= \sum_{i=1}^{n} H(X_i | X_1 \ldots X_{i-1})$$

$$H_n(X) = \frac{1}{n} \cdot H(X_1 \ldots X_n) \quad \text{ha } n \to \infty \quad H_\infty(X) = \lim_{n \to \infty} \frac{1}{n} H(X_1, \ldots X_n)$$

? ha létezik

1 szimbólumra eső entrópia

sztochasztikus folyamat

végtelen rendezett val. vált. sereg!

MS esetén

$$p(x_1 \dots x_n) = p(x_1) \cdot p(x_2 | x_1)\, p(x_3 | x_2 \cdot x_1) \cdot \dots \cdot p(x_n | x_1 \dots x_{n-1})$$
$$= \prod_{i=1}^{n} p(x_i | x_1 \dots x_{i-1})$$

log(szorzat) = összeg (log)

$$H(X_1 \dots X_n) = -\sum \prod_{i=1}^{n} p(x_i | x_1 \dots x_{i-1}) \cdot \operatorname{ld} \prod_{i=1}^{n} p(x_i | x_1 \dots x_{i-1}) =$$

n val vált. együttes entrópiája

$$= \sum_{i=1}^{n} H(X_i | X_1 \dots X_{i-1})$$

$$H_n(X) = \frac{1}{n} \cdot H(X_1 \dots X_n)$$

1 szimbólumra eső entrópia

ha $n \to \infty$

1 változóra eső? ha létezik

$$H_\infty(X) = \lim_{n \to \infty} \frac{1}{n} H(X_1, \dots X_n)$$

sztochasztikus folyamat

végtelen rendezett val. vált. sereg!

~~$H_n(X)$~~ ~~X_n~~

$$H_\infty(X) = \lim_{n \to \infty} H(X_n | X_1 \dots X_{n-1})$$

1 szimbólumra eső entrópia egy sztochasztikus folyamatnak!

④ Gallager bizonyítás:

- $H(X_n | X_1 \dots X_{n-1})$ monoton csökkenő!

$$H(X_n | X_1 \dots X_{n-1}) \leq H(X_n | X_2 \dots X_{n-1}) = H(X_{n-1} | X_1 \dots X_{n-2})$$

kevesebb előismeret nagyobb bizonytalanság! nagyobb entrópia

ha a folyamat legalább n-ed rendben stacionárius legyen (eltolásra érzéketlen)

(B) $H_n(X) \geq H(X_n | X_1 \dots X_{n-1})$

$$H_n(X) \overset{\text{def!}}{=} \frac{1}{n} \cdot \left[H(X_1 \dots X_n) \right] = \frac{1}{n} \cdot \sum_{i=1}^{n} H(X_i | X_1 \dots X_{i-1}) \geq \frac{1}{n} \cdot \overset{\text{n-szer veszem}}{n} H(X_n | X_1 \dots X_{n-1})$$

(A) miatt

(C) $H_n(X)$ mon. csökkenő

$$H_n(X) = \frac{1}{n} \Big[\underbrace{\underbrace{H(X_1 \dots X_{n-1})}_{} + \overset{\text{következő}}{H(X_n | X_1, \dots X_{n-1})}}_{n-1 \text{ val változó}} \Big] = \frac{1}{n} \left[(n-1) \cdot H_{(n-1)}(X) + \right.$$

$$\underbrace{H(X_n | X_1, \dots X_{n-1})}_{\frac{1}{n} \sum_{i=1}^{n} H(X_i | X_1 \dots X_{i-1}) = H_n(X)} \Big] \leq \frac{n-1}{n} \cdot H_{n-1}(X) + \frac{1}{n} \cdot H_n(X) \rightsquigarrow$$

$$\rightsquigarrow \frac{n-1}{n} \cdot H_n(X) \leq \frac{n-1}{n} H_{n-1}(X) \quad \text{mon. csökkenő ✓}$$

$$H_{n+j}(X) = \frac{1}{n+j} \Big[H(X_1 \dots \overset{1 \to (n-1)}{X_{n-1}}) + \sum_{i=n}^{\overset{n \to (n+j)}{n+j}} H(X_i | X_1 \dots X_{i-1}) \Big] \overset{(A)}{\leq} \underbrace{\frac{1}{n-j}}_{n \to \infty \Rightarrow \emptyset} \cdot H(X_1 \dots X_{n-1}) + \overset{n \to \infty \Rightarrow 1}{\frac{j+1}{n+j}} H(X_n | X_1 \dots X_{n-1})$$

$$\lim_{j \to \infty} H_{n+j}(X) \leq H(X_n | X_1 \dots X_{n-1})$$

$$\lim_{\substack{n \to \infty \\ j \to \infty}} H_{n+j}(X) \leq \lim_{n \to \infty} H(X_n | X_1 \dots X_{n-1})$$

+ sztochasztikus folyamat!

ezek egyszerre csak akkor igazak ha egyenlők!

$$\lim_{n \to \infty} H_n(X) \overset{(B)}{\geq} \lim_{n \to \infty} H(X_n | X_1 \dots X_{n-1})$$

ez akkor van ha erősen stac. a folyamat

ha van forrás kiterjesztés:

$$H(X_1 \dots X_n) \le L_{X_1, \dots X_n} < H(X_1 \dots X_n) + 1$$

$$H_n(X) \le L'_X < H_n(X) + \frac{1}{n}$$

ha $n \to \infty \to L'_X = H_n X$ (ha teljesen ismerem a folyamatot!)

az átl. kódhossz elmehet az entrópiáig.

$L_X = H_\infty(X)$

optimum

apriori ismernem kell a forrást!

L-Z kódolás (Lempel-Ziv)

- apriori ismeretek nélkül is működik és közelíti az optimumot
- UNIX-ban ez működik

elve: könyvet át akarok vinni, ehhez felépít egy könyvtárat és csak a címet kell megadnom.

$L_X = 5$ fix hosszú kóddal

a könyvtár:	tárhely	tartalom	kód | tartalom
1	0001	0	0000 | 0
2	0010	1	0000 | 1
			itt van az ‚1'
3	0011	11	0010 | 1
4	0100	00	0001 | 0
⋮	0101	10	0010 | 0
	0110	100	0101 | 0

és így tovább

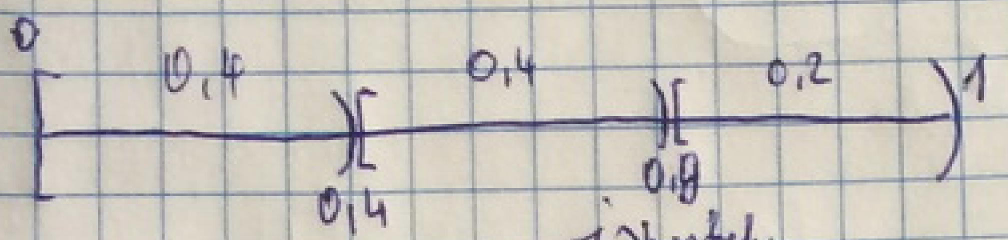
0, 1, 11, 00, 10

[0 1 1 1 0 0 1 0 1 0 0 1 1 1 1 0 0 1] bitsorozat

→ Nem kell közzé ismernem apriori az eloszlásokat.

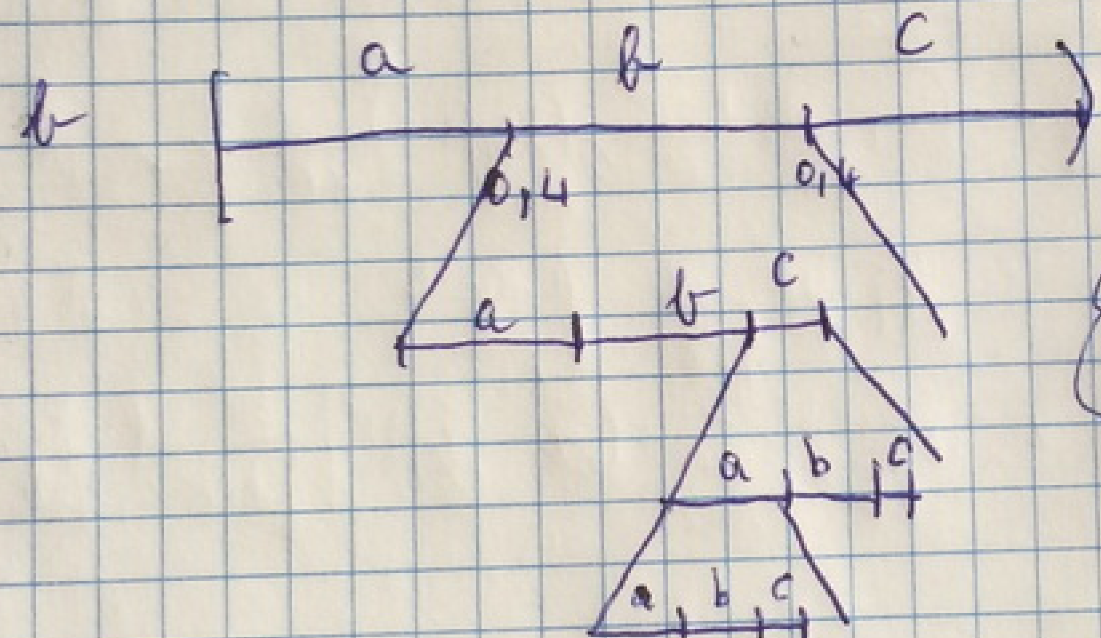
Aritmetikai kódolás:

- kell a $P(x)$
- tart az optimális kompr. kódhoz
- ritka sorozat → rövid kód, gyakori sorozat → rövid kód



$X = \{ a, b, c, \underbrace{p(a)=0{,}4,\ p(b)=0{,}4,\ p(c)=0{,}2}_{\text{ismertek}} \}$

$[b, c, a]$ átviendő sorozat



elég 1. rendben ismerni az eloszlást!
(még nem feltétlen kell formáiskiterjesztés)

$\boxed{0{,}75 = 0.\overset{2^{-1}}{1}\overset{2^{-2}}{1}}$ bca

ezt küldöm át
mert ez egyértelműen meghatároz egy részintervallumot!

de változó hosszú a kód

kell mindenképp egy stop szimbólum p(!)

~~nem prefix nem tud~~

1. ZH eddig