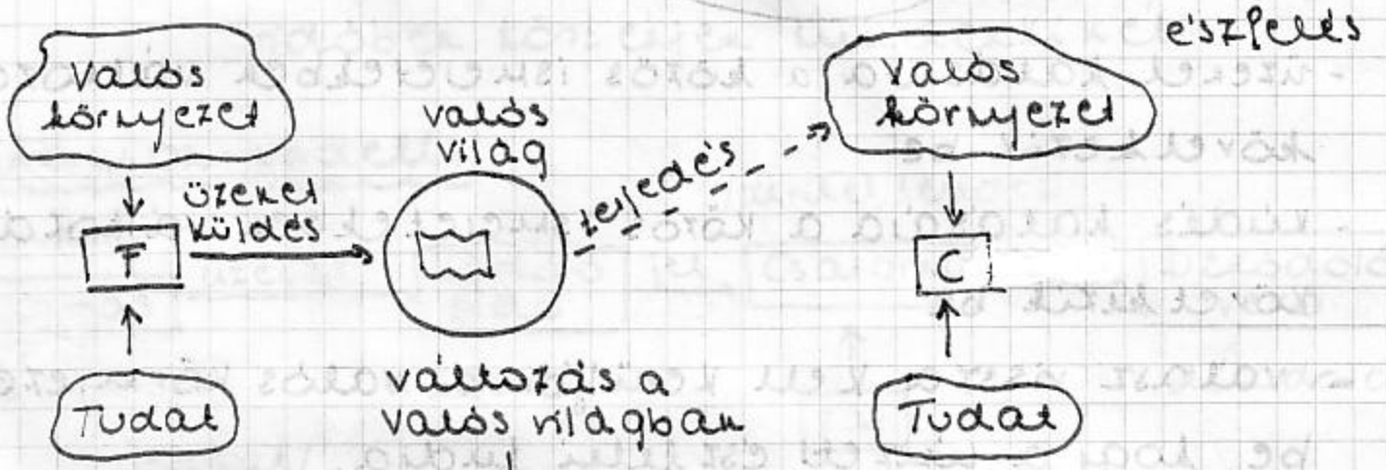
I. FA 2009.01.11

INFORMÁCIÓ ELMÉLET

Információelmélet alapjai:

kommunikáció: emberi tudatok kölcsönhatása
- észlelésből változásokra van szükség

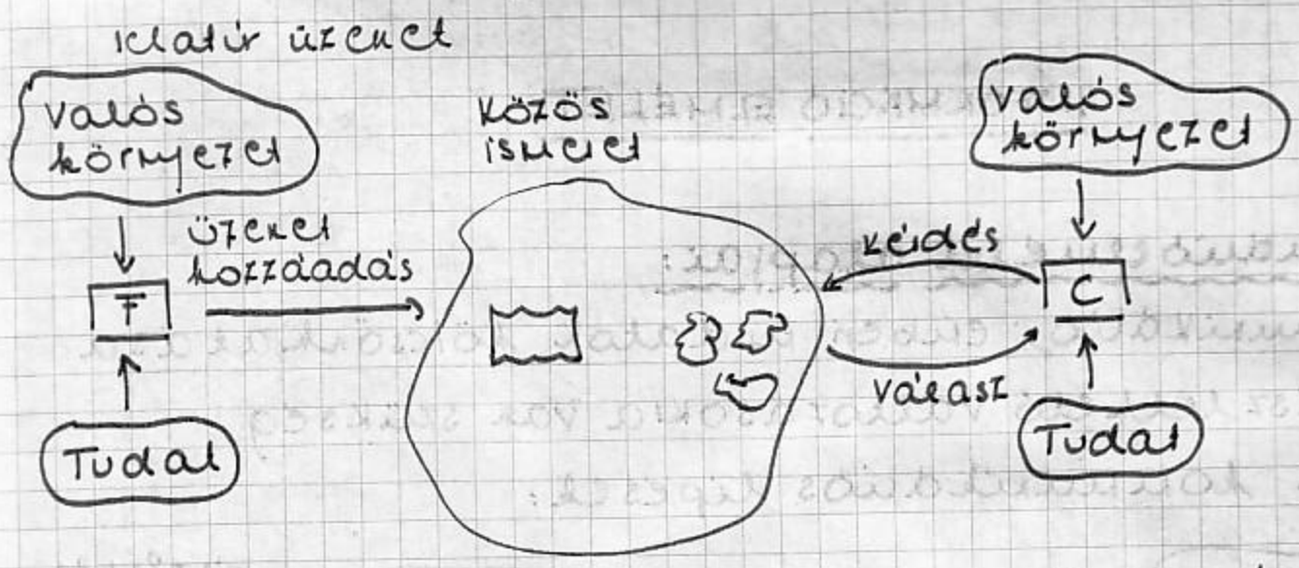
ebbeli kommunikációs lépések:



felismerkedő legyen

- valamit észlelünk a valós környezetben → kapcsolatba kerül a tudatunkkal → üzenet kibocsátása → változást vált ki a valós világban → → észleli a üzenet → kapcsolatba kerül a tudattal
- terjedés időben történik → zajok
- a kommunikáció működőképességének alapvető feltétele, hogy különböző üzenetekhez olyan különböző változás legyen létrehozható a valós világban, amit az észlelő is különbözőnek tud felismerni.
- hatékonyság: emlékek megőrzése

Információs-rendszer jellegű kommunikáció:



- üzenet hatására a közös ismeretekben változás következik be
- kérdés hatására a közös ismeretekben változás következik be

⇒ válasz vissza kell kerüljön a valós környezet-be, hogy a üzenet észlelni tudja.

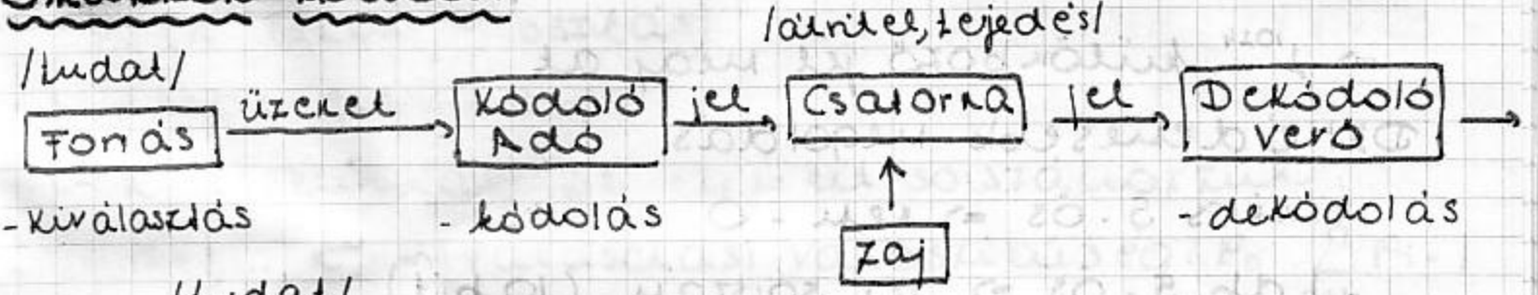
Történelmi állomások:

- beszéd kialakulása: hangképzés, hallás, gesztikulá-ció, testbeszéd, agy (a változások észleléséhez ⇒ megértsük, amit mondanak)
  - hangkultúrák terjedése
  - közös ismeret nagy része mindenkinél megtalálható ⇒ egy részét elérhetővé tesszük mások számára.
- feljegyzés: írás, rajz, kép, könyvtárak, kódexek, irattárak (középkor)
- nyomtatás: rögzített üzenet, azonos formában többszörösen elérhetővé válik
  - közös ismeret: könyvtárak, levéltárak, bürokrácia

- jelen: rádiófelvevés másolástórás ⇒ kikötés
  telefon ⇒ adatátvitel
  fényképgép ⇒ digitális műszerek
  videógép
  számítógép ⇒ adatbázolás, adatelérés,
  számítások (érzékszervek „jelentőssége" csökken)

- KÖZÖS ÉSZLELÉS: nem szükséges, hogy azonos időben történjen mindenkinél

SHANNON-MODELL:



elvárások:

(1) **mennyiségi szint**: mennyiségi követelmények, amelyek mellett a kívánt üzenet a csatornán

(2) **megértési szint**: (jelentés, szemantika): a címzett megérti-e, amit a forrás üzen
- gyenge megértési szint esetén erős mennyiségi szintre van szükség

(3) **hatékonysági szint**: kívánt hatás elérése

- **mennyiségi szint:**
  - forrás jellemzése: mennyi üzenetet bocsát ki
    ⇒ időegység - sebesség

- csatorna jellemzése: adó egység alatt átvihető, különböző csatornajelek száma

pl.: fonásra
- elemi üzenet: igen - nem
(i) pénzfeladás: fej · igen (1024x)
írás - nem
- átvitel: 0-1 sorozatok
⇒ 2^{1024} különböző jel megy át
(ii) 5. ös találat a heti LOTTÓN (1024x)
- átvitel: 0-1 sorozatok
⇒ 2^{1024} különböző jel megy át
DE! gyakorlatibb megoldás
- nincs 5. ös ⇒ nem - 0
- 1 db 5. ös ⇒ 10 sorszám (10 bit)
- több 5. ös ⇒ 11 sorszám (1024)
kód hosszának választható elve? /Poisson/
$P_0 = 0,999977$
$P_1 = 1,33 \cdot 10^{-5}$
P_2: elhanyagolható
- különbség (i) és (ii) között a bizonytalanság

2009.02.18. II. EA

• **mennyiségek:**
(1) csatorna: kapacitás
- bináris, szimmetrikus csatornával való ekvivalencia alapján
- egységnyi idő alatt átlagosan

ugyanannyi jelsorozat induló át, mint egy C bit/sec sebességű bináris csatornán (2^C)

(1) Forrás:

- véges sok, különböző szimbólumból választ (K db)
- a választás valamilyen véletlen körülmények közötti történik (valószínűségeloszlás-sal jellemezzük) ⇒ választás bizonytalansága
- N hosszú üzenet

⇒ eloszlás az N hosszú üzenetek halmazán (K^N elemű eloszlás)

- Shannon-entrópia:
  - K^N elemű halmaz elemeit sorszámozzuk: $i = 1, \dots, K^N$ ⇒ választási valószínűség: P_i, $\sum_{i=1}^{K^N} P_i = 1$

$$H_N = -\sum_{i=1}^{K^N} P_i \cdot \log_2 P_i \quad \text{/formula/}$$

  - H_N átlagát keressük szimbólumonként

$$H = \lim_{N \to +\infty} \frac{1}{N} \cdot H_N \quad \text{bit/szimbólum}$$

- egy szimbólumra jutó bizonytalanság

megválaszthatók: forrás sebessége ⇒ másodpercenként hány szimbólumot választ: V szimbólum/sec

C: bit/sec

H: bit/szimbólum

V: szimbólum/sec

- mennyiségi feladat megoldhatósága: veszteségmentesen átvihető minden üzenet

TÉTEL Zajmentes csatorna alaptétele /Shannon/

Nem lehet veszteségmentesen működtetni a kommunikációs rendszert, ha:

$$V > \frac{C}{H}$$

Tetszőleges $\varepsilon > 0$-hoz $\exists$ kódolás: $V < \frac{C}{H} - \varepsilon$

veszteségmentesen

sebességgel működtethető a rendszer.

(1) Csatorna kapacitás: (diszkrét, zajmentes)

- átvihető jelek: $s_1, \dots, s_M$
- átviteli idők: $t_1, \dots, t_M$ (jelenként)

szimmetrikus, bináris eset: $M = 2$, $t_1 = t_2 = t$

- T ideig működtetve hány különböző jelsorozat vihető át t idő alatt?

  $\Rightarrow$ T/t jel vihető át

  $\Rightarrow$ összes lehetőség: $2^{T/t}$ $\quad N(T) = 2^{T/t}$

  $C = 1/t \Rightarrow N(T) = 2^{C \cdot T} \Rightarrow C = \frac{1}{T} \cdot \log_2 N(T)$

- csatornák ekvivalenciája: ha két csatornái a $N(T)$ megegyezik, akkor a két csatorna ekvivalens

$\Rightarrow$ a csatornák T idő alatt átvihető, különböző jelsorozatok száma $N(T) \rightarrow$ csatorna kapacitása: $C = \lim\limits_{T \to +\infty} \left(\frac{1}{T} \cdot \log_2 N(T) \right) = \frac{1}{t}$

$t^{-1} - \delta < \frac{1}{T} \cdot \log_2 N(T) < t^{-1} + \delta$

$2^{T(t^{-1} - \delta)} < N(T) < 2^{T(t^{-1} + \delta)}$ $\quad \delta > 0$, tetszőleges kicsi

általános eset: $s_1, \dots, s_M$; $t_1, \dots, t_M$ } szimmetrikus: $t_1 = \dots = t_M = t$

- $N(T) \cong M^{T/t}$, $\quad C = \lim\limits_{T \to +\infty} \left(\frac{1}{T} \cdot \log_2 M^{T/t} \right) = t^{-1} \log_2 M$

egyszerű, nem szimmetrikus eset:

$N(T) = N(T-t_1) + \dots + N(T-t_m)$ - differencia egyenlet

$1 - x^{-t_1} - \dots - x^{-t_m} = 0$

x_0 - legnagyobb pozitív gyök

$x_0^T = (x_0^{-t_1} + \dots + x_0^{-t_m}) \cdot x_0^T = x_0^{T-t_1} + \dots + x_0^{T-t_m}$

$N(T) = K \cdot x_0^T$

$C = \lim_{T \to +\infty} \left(\frac{1}{T} \cdot \log_2 N(T) \right) = \log_2 x_0$

(2) **Forrás jellemzése:**

- minden lehetséges üzenet választására működjön (sokszor ismétlődő)
- nem a szimbólumok a fontosak, hanem a választás bizonytalansága ⇒ Milyen eloszlás szerint
  - $(p_1, \dots, p_n) : p_i > 0 \wedge \sum_{i=1}^{n} p_i = 1$

⇒ Keressük a véges, diszkrét eloszlásokon egy $H(p_1, \dots, p_n)$ függvényt, melyre:

Shannon-elvárásai:

(0) H az eloszlástól van csak elemenként, azaz független $\{p_1, \dots, p_n\}$ permutációitól ⇒ eloszlástérünk

(1) H minden változójában folytonos

(2) $A(n) := H\left(\underbrace{\frac{1}{n}, \dots, \frac{1}{n}}_{n \text{ db}}\right)$ - egyenletes eloszlás bizonytalansága

$A(n) < A(m)$ $n < m$ monotonitás

(3) elágazási szabály (lépcsőzetesség):

$H(p_1, \dots, \underbrace{p_{n-1}, p_n}_{q := p_{n-1} + p_n}) = H(p_1, \dots, p_{n-2}, q) + q \cdot H\left(\frac{p_{n-1}}{q}, \frac{p_n}{q}\right)$

általános eset:

$$q_j := \sum_{i=1}^{n_j} p_{ji}\,, \quad 0 < p_{ji} < 1\,, \quad \sum_{j=1}^{\mu} q_j = 1$$

$$H\Big(\underbrace{p_{11}, \dots, p_{1n_1}}_{q_1}, \underbrace{p_{21}, \dots, p_{2n_2}}_{q_2}, \dots, \underbrace{p_{\mu 1}, \dots, p_{\mu n_\mu}}_{q_\mu}\Big) =$$

$$= H(q_1, \dots, q_\mu) + \sum_{j=1}^{\mu} q_j \cdot H\left(\frac{p_{j1}}{q_j}, \dots, \frac{p_{jn_j}}{q_j}\right) - \text{függvényegyenlet}$$

TÉTEL A (0)-(3) feltételeket kielégítő függvény csak $-K \cdot \sum_{i=1}^{n} p_i \cdot \log_2 p_i$ alakú lehet $(K > 0)$

Biz

$H\left(\frac{1}{2}, \frac{1}{2}\right) = 1$ - egységnyi entrópia $\Rightarrow -K \cdot (-1) = 1 \Rightarrow K = 1$

- racionális eloszlás esetén:

$$p_i = \frac{q_i}{\mu}\,, \quad \sum_{i=1}^{n} p_i = 1\,, \quad \sum_{i=1}^{n} q_i = \mu$$

- egyenletes eloszlásból csoportok képzése:

$$A(\mu) = H\underbrace{\left(\frac{1}{\mu}, \dots, \frac{1}{\mu}\right)}_{\mu \text{ db}} = H\left(\frac{q_1}{\mu}, \dots, \frac{q_n}{\mu}\right) + \sum_{i=1}^{n} \frac{q_i}{\mu} \cdot H\underbrace{\left(\frac{\frac{1}{\mu}}{\frac{q_i}{\mu}}, \dots, \frac{\frac{1}{\mu}}{\frac{q_i}{\mu}}\right)}_{q_i \text{ db}} =$$

$$= H\left(\frac{q_1}{\mu}, \dots, \frac{q_n}{\mu}\right) + \sum_{i=1}^{n} \frac{q_i}{\mu} \cdot A(q_i)$$

$$H\left(\frac{q_1}{\mu}, \dots, \frac{q_n}{\mu}\right) = H(p_1, \dots, p_n) = A(\mu) - \sum_{i=1}^{n} \frac{q_i}{\mu} \cdot A(q_i) =$$

$$= -\sum_{i=1}^{n} \left(\frac{q_i}{\mu} \cdot \Big(A(q_i) - A(\mu)\Big)\right) = -K \cdot \sum_{i=1}^{n} p_i \cdot \log_2 p_i$$

- $A(n) = k \cdot \log_2 n$ alakú:

$$A(n \cdot m) = H\left(\underbrace{\frac{1}{n \cdot m}, \dots, \frac{1}{n \cdot m}}_{n \cdot m \text{ db}}\right) = H\left(\frac{1}{n}, \dots, \frac{1}{n}\right) + n \cdot \frac{1}{n} \cdot H\left(\frac{\frac{1}{n \cdot m}}{\frac{1}{n}}, \dots, \frac{\frac{1}{n \cdot m}}{\frac{1}{n}}\right) =$$

(n db m-es csoport)

$$= A(n) + A(m)$$

$$A(n^m) = m \cdot A(n)$$

- nögzítsük: $s \in \mathbb{N}$ $(s > 0) \Rightarrow \forall t, n : \exists m :$

$$s^m \leq t^n < s^{m+1} \quad / \log_2 x \text{ (monotonitás)}$$

$$\log_2 s^m \leq \log_2 t^n < \log_2 s^{m+1}$$

$$m \cdot \log_2 s \leq n \cdot \log_2 t < (m+1) \cdot \log_2 s$$

$$\frac{m}{n} \cdot \log_2 s \leq \log_2 t < \frac{m}{n} \cdot \log_2 s + \frac{1}{n} \cdot \log_2 s$$

$$\frac{m}{n} \leq \frac{\log_2 t}{\log_2 s} < \frac{m}{n} + \frac{1}{n}$$

$$s^m \leq t^n < s^{m+1} \quad / A(x) \text{ (monotonitás)}$$

$$A(s^m) \leq A(t^n) < A(s^{m+1})$$

$$m \cdot (A(s)) \leq n \cdot (A(t)) < (m+1) \cdot (A(s))$$

$$\frac{m}{n} \cdot A(s) \leq A(t) < \frac{m}{n} \cdot A(s) + \frac{1}{n} \cdot A(s)$$

$$\frac{m}{n} \leq \frac{A(t)}{A(s)} < \frac{m}{n} + \frac{1}{n}$$

$n \to +\infty$: $\frac{m}{n} \to \frac{\log_2 t}{\log_2 s} = \frac{A(t)}{A(s)} \Rightarrow A(t) = \underbrace{\frac{A(s)}{\log_2 s}}_{\text{konstans} = K} \log_2 t \Rightarrow$

$\Rightarrow$ egyenletes eloszlás entrópiája: $A(n) = K \cdot \log_2 n$

- a polytonosság miatt a valós eset is teljesül ✓

2009.01.15. III. EA

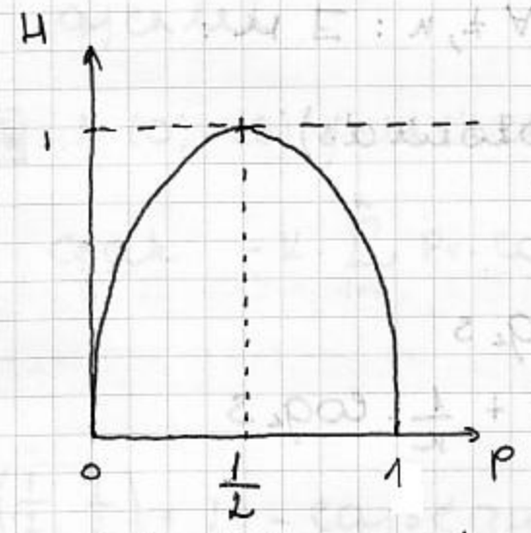
$$H(p_1, \dots, p_n) = -\sum_{i=1}^{n} p_i \cdot \log_2 p_i$$

Tulajdonságai:

(1) $H\left(\frac{1}{n}, \dots, \frac{1}{n}\right) = A(n) = \log_2 n$

(2) Kételemű eloszlás:

$$H(p, 1-p) = -p \cdot \log_2 p - (1-p) \cdot \log_2 (1-p)$$



$H(p, 1-p) \leq H\left(\frac{1}{2}, \frac{1}{2}\right) = 1$

$\Rightarrow H(p, 1-p) \uparrow \; [0, \frac{1}{2}]$

$\Rightarrow H(p, 1-p) \downarrow \; [\frac{1}{2}, 1]$

- minél kiegyenlítettebb az eloszlás, annál nagyobb a bizonytalanság

(3) $H(p_1, \dots, p_n) \leq \log_2 n = A(n)$

- egyenlőség felé közelítünk

$$H(p_1, \dots, p_{n-1}, q_1, q_2) = H(p_1, \dots, p_{n-1}, \underbrace{q_1, q_2}_{p}) + p \cdot H\left(\frac{q_1}{p}, \frac{q_2}{p}\right) \leq$$

$p_{n-1} + p_n = q_1 + q_2 = p$

(2)-es tulajdonság

$$\leq H(p_1, \dots, p_{n-2}, p) + p \cdot H\left(\frac{1}{2}, \frac{1}{2}\right)$$

- kiegyenlítés növeli az entrópiát

Valószínűségi változó entrópiája:

ξ v.v. elemei: $x_1, \dots, x_n$ (diszkrét, véges)

eloszlása: $P(\xi = x_i)$ $i \in [1 \dots n]$ $\sum_{i=1}^{n} P(\xi = x_i) = 1$

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

Kétdimenziós eloszlás entrópiája:

ξ, η v.v együttes eloszlása: $P(\xi = x_i, \eta = y_j)$

$i \in [1 \dots n], j \in [1 \dots m]$

$$H(\xi, \eta) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(\xi = x_i, \eta = y_j) \cdot \log_2 P(\xi = x_i, \eta = y_j)$$

$\Rightarrow \xi$ eloszlása:

- teljes eseményrendszer = diszjunkt események uniója $\Rightarrow$ A eseményt elmetszve ezekkel, szintén előállnak diszjunkt halmazokból (teljes valószínűség tétele)

$$P(\xi = x_i) = \sum_{j=1}^{m} P(\xi = x_i, \eta = y_j)$$

$$P(\eta = y_j) = \sum_{i=1}^{n} P(\xi = x_i, \eta = y_j)$$

Független eloszlás entrópiája,

- a függetlenség a legnagyobb bizonytalanságot jelenti

ξ, η v.v (függetlenek) együttes eloszlása

$P(\xi = x_i, \eta = y_j) = P(\xi = x_i) \cdot P(\eta = y_j)$

$i \in [1 \dots n], j \in [1 \dots m], \sum_{i=1}^{n} P(\xi = x_i) = 1, \sum_{j=1}^{m} P(\eta = y_j) = 1$

$$H(\xi, \eta) = -\sum_{i=1}^{n} \sum_{j=1}^{m} P(\xi = x_i) \cdot P(\eta = y_j) \cdot (\log_2 P(\xi = x_i) + \log_2 (\eta = y_j))$$

$$= -\sum_{i=1}^{n} \sum_{j=1}^{m} P(\xi = x_i) \cdot P(\eta = y_j) \cdot \log_2 P(\xi = x_i) -$$

$$- \sum_{i=1}^{n} \sum_{j=1}^{m} P(\xi = x_i) \cdot P(\eta = y_j) \cdot \log_2 P(\eta = y_j) = \dots = H(\xi) + H(\eta)$$

TÉTEL $H(\xi, \eta) \leq H(\xi) + H(\eta)$

$\Leftrightarrow \Rightarrow \xi, \eta$ függetlenek

Biz feltételes entrópiával

T.F.H. megfigyelünk: $\eta = y_j \quad j \in [1 \dots n]$

$P(\xi = x_i \mid \eta = y_j) \quad i \in [1 \dots n]$

$$P(\xi = x_i \mid \eta = y_j) = \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)} = \frac{P(\xi = x_i, \eta = y_j)}{\sum_{i=1}^{n} P(\xi = x_i, \eta = y_j)}$$

$$H(\xi = x_i \mid \eta = y_j) = -\sum_{i=1}^{n} P(\xi = x_i \mid \eta = y_j) \cdot \log_2 P(\xi = x_i \mid \eta = y_j)$$

Feltételes entrópia:

- Mennyi bizonytalanság maradt átlagosan ξ-ről, ha ismerjük η-t?

$H(\xi, \eta)$ - eddig sem figyeltük meg (teljes bizonytalanság)

$\Rightarrow \eta$ megfigyelése esetén eltűnik a megfigyelések bizonytalansága

$H(\xi \mid \eta) = H(\xi, \eta) - H(\eta)$ $\qquad H(\xi, \eta) = H(\eta) + \sum_{j=1}^{n} P(\eta = y_j) H(\xi = x_i \mid \eta = y_j)$

másképp gondolva:

$H(\eta)$: elágazásig (η megfigyeléséig véve az elágazást)

elágazás után maradó: $\sum_{j=1}^{n} P(\eta = y_j) \cdot H(\xi = x_i \mid \eta = y_j)$

$$\Rightarrow H(\xi = x_i \mid \eta = y_j) = -\sum_{j=1}^{n} P(\eta = y_j) \cdot \sum_{i=1}^{n} P(\xi = x_i \mid \eta = y_j) \cdot \log_2 P(\xi = x_i \mid \eta = y_j)$$

$$= -\sum_{j=1}^{n} P(\eta = y_j) \cdot \sum_{i=1}^{n} \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)} \cdot \log_2 \frac{P(\xi = x_i, \eta = y_j)}{P(\eta = y_j)}$$

$$= -\sum_{i=1}^{n} \sum_{j=1}^{n} P(\xi = x_i, \eta = y_j) \cdot \log_2 \frac{P(\eta = y_j)}{P(\xi = x_i, \eta = y_j)}$$

$H(\xi, \eta) \leq H(\xi) + H(\eta)$

$H(\xi, \eta) - H(\eta) \leq H(\xi)$

$H(\xi \mid \eta) \leq H(\xi)$

$\Leftrightarrow$ ξ, η függetlenek

T.F.H. megfigyelünk η-t v függvényét $f(\eta)$-t $\Rightarrow$

$\Rightarrow H(\xi|\eta) \leq H(\xi|f(\eta))$

- a bizonytalanság kevesbé érezhető közvetlen megfigyelés esetén

$\Leftrightarrow \forall z \; f(y)=z : P(\xi=x|\eta=y) = P(\xi=x|f(\eta)=z)$

$$P(\xi=x, f(\eta)=z) = \sum_{f(y)=z} P(\xi=x, \eta=y)$$

$$P(f(\eta)=z) = \sum_{f(y)=z} P(\eta=y)$$

$$H(\xi=x|f(\eta)=z) = \sum_x \sum_z \left(\sum_{f(y)=z} P(\xi=x, \eta=y) \right) \cdot \log_2 \frac{\sum_{f(y)=z} P(\eta=y)}{\sum_{f(y)=z} P(\xi=x, \eta=y)}$$

$$H(\xi=x|\eta=y) = \sum_x \underbrace{\sum_z \sum_{f(y)=z}}_{y} P(\xi=x, \eta=y) \cdot \log_2 \frac{P(\eta=y)}{P(\xi=x, \eta=y)}$$

$a_i \; i \in [1 \dots n] \quad a = \sum_{i=1}^{n} a_i$

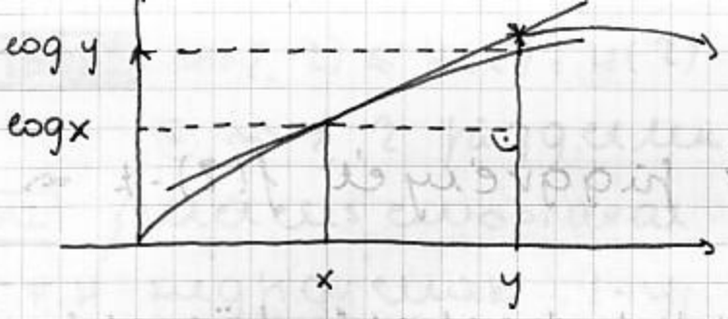
$b_i \; i \in [1 \dots n] \quad b = \sum_{i=1}^{n} b_i$

$$\sum_{i=1}^{n} a_i \cdot \log \frac{b_i}{a_i} \leq a \cdot \log \frac{b}{a}$$

$\Leftrightarrow \frac{b_i}{a_i} = \frac{b}{a} \quad i \in [1 \dots n]$

$\log y \leq \log x + C \cdot (y - x)$

$\Leftrightarrow x = y$



ezzel az ötlettel mindig felülről becsülhető

$$\log \frac{b_i}{a_i} \leq \log \frac{b}{a} + C \cdot \left(\frac{b_i}{a_i} - \frac{b}{a} \right)$$

$$(=) \iff \frac{b_i}{a_i} = \frac{b}{a}$$

$$\sum_{i=1}^{n} a_i \cdot \log \frac{b_i}{a_i} \leq \sum_{i=1}^{n} a_i \cdot \log \frac{b}{a} + C \cdot \sum_{i=1}^{n} a_i \left(\frac{b_i}{a_i} - \frac{b}{a} \right)$$

$(=) \iff \forall i \in [1 \dots n]$ - re teljesül

$$\sum_{i=1}^{n} a_i \cdot \log \frac{b_i}{a_i} \leq a \cdot \log \frac{b}{a} + C \cdot \left(\sum_{i=1}^{n} b_i - \frac{b}{a} \cdot \sum_{i=1}^{n} a_i \right) =$$

$$= a \cdot \log \frac{b}{a} + C \cdot \left(b - \frac{b}{a} \cdot a \right) =$$

$$= a \cdot \log \frac{b}{a}$$

$$P(\xi = x, \eta = y) \cdot \log_2 \frac{P(\eta = y)}{P(\xi = x, \eta = y)} \leq$$

$$\leq \sum_{f(y)=z} P(\xi = x, \eta = y) \cdot \log_2 \frac{\sum_{f(y)=z} P(\eta = y)}{\sum_{f(y)=z} P(\xi = x, \eta = y)}$$

$$(=) \iff \frac{P(\xi = x, \eta = y)}{P(\eta = y)} = \frac{\sum_{f(y)=z} P(\eta = y)}{\sum_{f(y)=z} P(\xi = x, \eta = y)} \iff$$

$$\iff P(\xi = x \mid \eta = y) = P(\xi = x \mid f(\eta) = z) = P(\xi = x)$$

$H(\xi|\eta) \leq H(\xi|f(\eta)) \Rightarrow H(\xi|\eta) \leq H(\xi)$ ⇒

$f(y) = c$ konstans függvényt választva

$\Rightarrow H(\xi,\eta) \leq H(\xi) + H(\eta)$ ✓

⇔ : $\forall y : f(y) = c$

$P(\xi = x | \eta = y) = P(\xi = x | f(\eta) = c) = P(\xi = x)$

$\frac{P(\xi = x, \eta = y)}{P(\eta = y)} = P(\xi = x)$

$P(\xi = x, \eta = y) = P(\xi = x) \cdot P(\eta = y) \Leftrightarrow \xi, \eta$ függetlenek ✓

ξ_1, ξ_2, ξ_3 → jelen, múlt, jövő

általánosan: Markov-lánc tulajdonság

$H(\xi_t = x_t | \xi_1 = x_1, \dots, \xi_{t-1} = x_{t-1}) \leq$

$\leq H(\xi_t = x_t | \xi_{t-1} = x_{t-1})$

$H(\xi_3 | \xi_1, \xi_2) \leq H(\xi_3 | \xi_2)$

⇔ $\Leftrightarrow P(\xi_3 = x | \xi_1 = y, \xi_2 = z) = P(\xi_3 = x | \xi_2 = z)$

IV. EA

2009.03.04

$a_i \; i \in [1 \dots n] \quad a = \prod_{i=1}^{n} a_i$

$b_i \; i \in [1 \dots n] \quad b = \prod_{i=1}^{n} b_i$

$\sum_{i=1}^{n} a_i \cdot \log \frac{b_i}{a_i} \leq a \cdot \log \frac{b}{a}$ következményei:

(1) $(p_1, \dots, p_n) \quad (q_1, \dots, q_n)$

$a_i \qquad b_i$

$\sum_{i=1}^{n} p_i \cdot \log_2 \frac{q_i}{p_i} \leq 0$

$$-\sum_{i=1}^{n} p_i \cdot \log_2 p_i \leq -\sum_{i=1}^{n} p_i \cdot \log_2 q_i$$

(2) $b_i := 1$

$(p_1, \dots, p_n)$
$\parallel$
a_i

$$\sum_{i=1}^{n} p_i \cdot \log_2 \frac{1}{p_i} \leq \log_2 n$$

Jensen egyenlőtlenség:

$h(x)$ - tetszőleges konvex függvény

ξ - tetszőleges V.V.

$E(h(\xi)) \leq h(E(\xi))$

- **forrás jellemzése/közelítések:**
- tipikus közelítések
- mindig a lehető legbizonytalanabbra készülünk => felülről közelítjük a forrás entrópiáját
- választható szimbólumok: $x_1, \dots, x_n$
- **0-adrendű közelítés:** nincs megfigyelés =>
  => n db különböző szimbólumot választunk
  - egyenletes eloszlást tételezünk fel:
    ξ eloszlása: $P(\xi_i = x_j) = \frac{1}{n}$ $j \in [1 \dots n]$ (1 szimbólum)
  - független választást tételezünk fel:
    $P(\xi_1 = x_{i_1}, \dots, \xi_N = x_{i_N}) = \left(\frac{1}{n}\right)^N$ (N szimbólum)

$H(\xi_i) = \log_2 n \Rightarrow H(\xi_1, \dots, \xi_N) = N \cdot H(\xi) = N \cdot \log_2 n$ (entrópia)

$H(\xi) = \log_2 n = \lim_{N \to +\infty} \left(\frac{1}{N} \cdot H(\xi_1, \dots, \xi_N)\right)$ (1 szimbólumra jutó)

- **Elsőrendű közelítés:** M hosszú üzenet megfigyelése
  - x_i előfordulása: M_i $i \in [1 \dots n]$

- eloszlás közelítése: $P(\xi_i = x_j) = \frac{\mu_j}{M} = p_j$ (1 szimbólumra)

- független választást feltételezünk fel:

$P(\xi_1 = x_{i1}, \dots, \xi_N = x_{iN}) \cdot \prod_{j=1}^{N} P(\xi_j = x_{ij}) = \prod_{i=1}^{n} p_i^{\mu_i}$ (N szimbólumra)

μ_i: statisztikai érték, hogy x_i hányszor fordul elő a kiválasztott szimbólumsorozatban. (véletlen mennyiség)

$H(\xi_i) = -\sum_{j=1}^{n} p_j \cdot \log_2 p_j \Rightarrow H(\xi_1, \dots, \xi_N) = N \cdot H(\xi) = -N \cdot \sum_{j=1}^{n} p_j \cdot \log_2 p_j$

(entrópia)

$H_i(P) = \lim_{N \to +\infty} \left(\frac{1}{N} \cdot H(\xi_1, \dots, \xi_N) \right)$ (1 szimbólumra jutó)

- <u>Másodrendű közelítés</u>: M hosszú üzenet megfigyelése
  - x_i előfordulása: $\mu_i \quad i \in [1 \dots n]$
  - x_i, x_j egymás utáni előfordulásainak száma: μ_{ij}
  - eloszlás közelítése:

$P(\xi = x_i) = \frac{\mu_i}{M}$ - önmagában való előfordulása

$P(\xi_t = x_i, \xi_{t+1} = x_j) = \frac{\mu_{ij}}{M}$ - két egymás utáni előfordulás

$P(\xi_{t+1} = x_j \mid \xi_t = x_i) = \frac{P(\xi_t = x_i, \xi_{t+1} = x_j)}{P(\xi_t = x_i)} = \frac{\frac{\mu_{ij}}{M}}{\frac{\mu_i}{M}} = \frac{\mu_{ij}}{\mu_i} = p_{j|i}$

$\Rightarrow P(\xi_1 = x_{i1}, \dots, \xi_N = x_{iN}) =$

$= P(\xi_N = x_{iN} \mid \xi_1 = x_{i1}, \dots, \xi_{N-1} = x_{iN-1}) \cdot P(\xi_1 = x_{i1}, \dots, \xi_{N-1} = x_{iN-1})$

⇒ Markov-láncok:

$P(\xi_n = x_{i_n} \mid \xi_1 = x_{i_1}, \dots, \xi_{n-1} = x_{i_{n-1}}) = P(\xi_n = x_{i_n} \mid \xi_{n-1} = x_{i_{n-1}}) =$

$= p_{i_n i_{n-1}}$ teljesüljön $\forall x_{i_1}, \dots, x_{i_n}$ értékre n-től függetlenül

- homogén egylépéses Markov-lánc:

$P(\xi_{t+1} = x_j \mid \xi_1 = x_{i_1}, \dots, \xi_{t-1} = x_{i_{t-1}}, \xi_t = x_i) =$

$= P(\xi_{t+1} = x_j \mid \xi_t = x_i) = p_{ji}$

⇒ átmeneti valószínűségek, együttes eloszlás:

$P(\xi_1 = x_{i_1}, \dots, \xi_n = x_{i_n}) =$

$= P(\xi_1 = x_{i_1}, \dots, \xi_{n-1} = x_{i_{n-1}}) \cdot P(\xi_n = x_{i_n} \mid \xi_{i_1} = x_{i_1}, \dots, \xi_{i_{n-1}} = x_{i_{n-1}}) =$

$= P(\xi_1 = x_{i_1}, \dots, \xi_{n-1} = x_{i_{n-1}}) \cdot P(\xi_n = x_{i_n} \mid \xi_{n-1} = x_{i_{n-1}}) = \dots =$

$= P(\xi_1 = x_{i_1}) \cdot P(\xi_2 = x_{i_2} \mid \xi_1 = x_{i_1}) \cdot \dots \cdot P(\xi_n = x_{i_n} \mid \xi_{n-1} = x_{i_{n-1}}) =$

$= P(\xi_1 = x_{i_1}) \cdot \prod_{i,j} p_{ji}^{\mu_{ji}}$ μ_{ji}: x_i, x_j átmenetek száma

$H(\xi_1, \dots, \xi_n) = H(\xi_n \mid \xi_1, \dots, \xi_{n-1}) + H(\xi_1, \dots, \xi_{n-1}) =$

$H(\xi, \eta) = H(\xi) + H(\eta)$ ∧
∧ $H(\xi) = H(\xi \mid \eta)$

$= H(\xi_n \mid \xi_{n-1}) + H(\xi_1, \dots, \xi_{n-1}) = \dots =$

$= H(\xi_1) + H(\xi_2 \mid \xi_1) + \dots + H(\xi_n \mid \xi_{n-1})$

$H(\xi_t \mid \xi_{t-1}) = \sum_{i=1}^{n} P(\xi_{t-1} = x_i) \cdot H(\xi_t \mid \xi_{t-1} = x_i)$

$H(\xi_t \mid \xi_{t-1} = x_i) = -\sum_{j=1}^{n} P(\xi_t = x_j \mid \xi_{t-1} = x_i) \cdot \log_2 \cdot P(\xi_t = x_j \mid \xi_{t-1} = x_i) =$

$= -\sum_{j=1}^{n} p_{ji} \cdot \log_2 p_{ji} \cdot H(\xi \mid x_i)$

$H(\xi_t \mid \xi_{t-1}) = \sum_{i=1}^{n} P(\xi_{t-1} = x_i) \cdot H(\xi \mid x_i)$

$P(\xi_t = x_i) \to \pi_i \quad (t \to +\infty)$

feltétel: minden érték rekurrens legyen, ne legyen

periódikus (ergodikus Markov. lánc)

⇒ Határeloszlás, stacionárius eloszlás:

$$P(\xi_t = x_i) = \sum_{j=1}^{n} P(\xi_{t-1} = x_j) \cdot p_{i|j} \quad \forall i \in [1 \dots k] \quad / \lim_{t \to +\infty}$$

$$\pi_i = \sum_{j=1}^{n} \pi_j \cdot p_{i|j}$$

$(\pi_1, \dots, \pi_k)$: átmenet valószínűségi mátrix sajátvektora a $\lambda = 1$ sajátértékkel

$$H(\xi_1, \dots, \xi_N) = H(\xi_1) + \sum_{j=1}^{n} H(\xi | x_j) \cdot \left(\sum_{t=2}^{n} P(\xi_{t-1} = x_j) \right) \quad \text{(entrópia)}$$

$$H(\xi) = \lim_{N \to +\infty} \left(\frac{1}{N} \cdot H(\xi_1, \dots, \xi_N) \right) = \lim_{N \to +\infty} \frac{1}{N} \cdot \sum_{j=1}^{n} H(\xi | x_j) \left(\sum_{t=2}^{n} P(\xi_{t-1} = x_j) \right) =$$

$$= \sum_{j=1}^{n} \pi_j \cdot H(\xi | x_j) \quad \text{(1 szimbólumra jutó)}$$

· Harmadrendű közelítés:

- x_i, x_j, x_k egymás utáni előfordulásainak a száma: μ_{ijk}
- eloszlás közelítése:

$$P(\xi_t = x_k \mid \xi_{t-1} = x_j, \xi_{t-2} = x_i) = \frac{\mu_{ijk}}{\mu_{ij}}$$

- kétlépéses Markov. lánc:

$$P(\xi_t = x_k \mid \xi_1 = x_{i1}, \dots, \xi_{t-3} = x_{i_{t-3}}, \xi_{t-2} = x_i, \xi_{t-1} = x_j) =$$

$$= P(\xi_t = x_k \mid \xi_{t-2} = x_i, \xi_{t-1} = x_j)$$

2009. 03. 11.

V. EA

Hosszú sorozatok jellemzése:

- forrás: független, azonos eloszlású
  - választható szimbólumok: $x_1, \dots, x_n$

$$P(\xi = x_i) = p_i$$

$$P(\xi_1 = x_{i_1}, \dots, \xi_N = x_{i_N}) = \prod_{i=1}^{n} p_i^{\mu_i}$$

μ_i: x_i előfordulásai

$P(\vec{\xi} = \vec{x})$ jelölésmód

  - nagyszámok törvénye: relatív gyakoriság tart az esemény valószínűségéhez

$$\mu_i \cong N \cdot p_i \quad i \in [1 \dots n]$$

$$P(\vec{\xi} = \vec{x}) = \prod_{i=1}^{n} p_i^{N \cdot p_i} \approx 2^{-N\left(-\sum_{i=1}^{n} p_i \log_2 p_i\right)} = 2^{-N \cdot H}$$

TÉTEL $\forall \varepsilon > 0 \; \forall \delta > 0 \; \exists N_0 : N > N_0$ esetén az N hosszú sorozatok két halmazba sorolhatók

(1) Lényegtelen halmaz és annak a valószínűsége, hogy ebből választunk $< \varepsilon$

(2) Tipikus halmaz és annak a valószínűsége, hogy ebből választunk $> 1 - \varepsilon$

$\vec{x}$ tipikus sorozat esetén:

$$\left| \frac{\log_2 P(\vec{\xi} = \vec{x})}{N} - H \right| < \delta$$

(tipikus halmazon, majdnem egyenlő tesse válik az eloszlás)

- **Ékek stratégia:**

- tipikus halmaz: H_N

CÉL: becsülni adni a halmaz elemszámára $\Rightarrow$

$$2^{-N(H+\delta)} < P(\vec{\xi} = \vec{x}) < 2^{-N(H-\delta)}$$

$$|M_0| < 2^{N(H+\delta)}$$

N elég nagy $\Rightarrow$ μ_i tipikus esetben: $(1-\delta) N \cdot P_i < \mu_i < (1+\delta) N \cdot P_i$

$$\Rightarrow |P(\vec{\xi} = \vec{x})| < 2^{\sum_{i=1}^{n} \mu_i \cdot \log_2 P_i}$$

(sorozat valószínűsége)

BIZ

$\vec{\xi}$, μ_i $i \in [1 \dots n]$, $\prod_{i=1}^{n} P_i^{\mu_i}$

μ_i előállítása és becslése:

- $S_j^{(i)}$ $j \in [1 \dots N]$

$$S_j^{(i)} = \begin{cases} 0 & \xi_j \neq x_i \\ 1 & \xi_j = x_i \end{cases}$$

- μ_i felírható N db, független, 0-1 értékű V.V. összegeként.

$$\mu_i = \sum_{j=1}^{N} S_j^{(i)} \qquad P(S_j^{(i)} = 1) = P_i$$

várható érték: $E(\mu_i) = N \cdot E(S^{(i)}) = N \cdot P_i$

szórásnégyzet: $D^2(\mu_i) = N \cdot D^2(S^{(i)}) = N \cdot (P_i - P_i^2) = N \cdot P_i \cdot (1 - P_i)$

Csebisev-egyenlőtlenség:

$$P((\xi - E(\xi))^2 \geq \gamma^2) \leq \frac{D^2(\xi)}{\gamma^2}$$

$\mu_i - E(\mu_i) = \mu_i - N \cdot P_i$ eltérést vizsgálva:

$$P((\mu_i - N \cdot P_i)^2 \geq (\delta' N)^2) \leq \frac{N \cdot P_i (1 - P_i)}{(\delta' \cdot N)^2} = \frac{K}{N} < \varepsilon'$$

$\downarrow$

megválasztható

Megválaszthatók N-el olyan nagynak, hogy ε'-től függően ez teljesüljön $\forall i \in [1 \dots n]$-re

$$\Rightarrow P(|\mu_i - N \cdot P_i| < \delta' \cdot N) > 1 - \varepsilon'$$

$\Rightarrow$ ene az N. ne: lényegtelen halmaz

$P(\text{legalább 1 } i\text{-ne}: |\mu_i - N \cdot P_i| > \delta' \cdot N) <$

$< \sum_{i=1}^{k} P(|\mu_i - N \cdot P_i| > \delta' N) \leq k \cdot \varepsilon' \Rightarrow \varepsilon' := \frac{\varepsilon}{k}$ esetén kap-

pk a lényegtelen halmaz ✓

komplementer: tipikus halmaz

$\forall i \in [1 \dots k]: |\mu_i - N \cdot P_i| < \delta' \cdot N$

$N(P_i - \delta') < \mu_i < N \cdot (P_i + \delta')$

$$\prod_{i=1}^{k} P_i^{N(P_i+\delta')} < \prod_{i=1}^{k} P_i^{\mu_i} < \prod_{i=1}^{k} P_i^{N(P_i-\delta')} \qquad / \log_2$$

$$N \cdot \sum_{i=1}^{k} \log_2 P_i \cdot (P_i + \delta') < \log_2 P(\vec{\xi} = \vec{x}) < N \cdot \sum_{i=1}^{k} \log_2 P_i (P_i - \delta')$$

$$H + \delta' \cdot \sum_{i=1}^{k} \log_2 P_i < \frac{\log_2 P(\vec{\xi} = \vec{x})}{N} < H - \delta' \cdot \sum_{i=1}^{k} \log_2 P_i$$

$$\Rightarrow \left| \frac{\log_2 P(\vec{\xi} = \vec{x})}{N} - H \right| < \delta \qquad \delta := -\delta' \cdot \sum_{i=1}^{k} \log_2 P_i \quad ✓$$

- **tipikus halmaz:**

- legvalószínűbb válaszadások:

$P_1 \geq \dots \geq P_k$ legvalószínűbb válaszadás: N ab x_i

$(\mu_1 \cdot N)$

DE! ez nem lesz tipikus

$\vec{x}_1, \vec{x}_2, \dots$ esetén: $P(\vec{\xi} = \vec{x}_i) \leq P(\vec{\xi} = \vec{x}_{i-1})$

$$\sum_{i=1}^{L} P(\vec{\xi} = \vec{x}_i) \geq 1 - \lambda$$

$$\sum_{i=1}^{L-1} P(\vec{\xi} = \vec{x}_i) < 1 - \lambda$$

L: λ-ból függő $\wedge$ $2^{N(H-\delta)} < L(\lambda) < 2^{N(H+\delta)}$

kódolás és csatornakapacitás:

- forrás: $\xi_1, \dots, \xi_N$
- csatornák: $\eta_1, \dots, \eta_{M(\xi_1,\dots,\xi_N)}$ (kódolás eredménye)
- kódolás: legegyszerűbb

$f(\xi_1, \dots, \xi_N) \cdot (\eta_1, \dots, \eta_{M(\xi_1,\dots,\xi_N)})$

$H(f(\xi)) \leq H(\xi)$

$\ominus \Leftrightarrow f$ invertálható

$H(\xi_1, \dots, \xi_N) \geq H(\eta_1, \dots, \eta_{M(\xi_1,\dots,\xi_N)}) = H(f(\xi_1, \dots, \xi_N))$

nincs veszteség $\Leftrightarrow$ invertálható a kódolás $\Leftrightarrow$

$\Leftrightarrow \exists f^{-1}: f^{-1}(\eta_1, \dots, \eta_M) = (\xi_1, \dots, \xi_N)$

$H(\eta_1, \dots, \eta_M) \geq H(\xi_1, \dots, \xi_N)$

$\Rightarrow H(\xi_1, \dots, \xi_N) = H(\eta_1, \dots, \eta_{M(\xi_1,\dots,\xi_N)})$

- T ideig működtetve a csatornát

$\Rightarrow T$ idő alatt átvihető különböző jelsorozatok száma: $N(T) \Rightarrow$ csatorna kapacitása:

$$C = \lim_{T \to +\infty} \left(\frac{1}{T} \cdot \log_2 N(T) \right) \Rightarrow \text{egyenletes eloszlás}$$

esetén a maximális: $\log_2 N(T) = T(C \pm \delta)$ entrópia

$2^{T(C-\delta)} < N(T) < 2^{T(C+\delta)}$ $\quad \delta > 0$, tetszőleges kicsi

TÉTEL Zajmentes csatorna alaptétele /Shannon/

(a) Nem lehet $\frac{C}{H}$-nál nagyobb sebességgel működtetni a csatornát úgy, hogy minden üzenet a csatorna kimenetéből visszaállítható legyen.

(Nem lehet veszteségmentes)

(b) Tetszőleges $\delta > 0$-hoz $\exists$ kódolás: $\frac{C}{H} \cdot \delta$ sebességgel veszteségmentesen lehet minden üzenetet átvinni.

Biz

(a) T idő, V sebesség, H entrópia

- forrásentrópia T idő alatt:

$$V \cdot T \cdot (H - \delta) < H(T) < V \cdot T \cdot (H + \delta)$$

- csatornakimenet max. entrópiája: $(T(C-\delta), T(C+\delta))$

$V \cdot T \cdot (H - \delta) > T(C+\delta)$ → biztos veszteség ⇒

$$\Rightarrow V \cdot T \cdot (H - \delta) \leq T \cdot (C + \delta)$$

$$V \leq \frac{C+\delta}{H-\delta} \to \frac{C}{H} \quad \delta, \delta > 0, \text{ tetszőleges kicsik} \checkmark$$

2009.03.18. **VI. EA**

(b) $\delta > 0$

$\frac{C}{H} - \delta$ sebességgel veszteségmentes legyen az üzenetátvitel

görög betű: tetszőleges kicsi lehet, és ha elég nagy a hosszúság (N), akkor az egyenlőtlenség teljesül.

N hosszú üzenet

kérdés: Milyen hosszú csatornajel tartozik hozzá? → tipikusan rövidebb / éppen eltéréssel hosszabb

- forrás: független, azonos eloszlású:
  - választható szimbólumok: $x_1, \ldots, x_m$

N: üzenetek hossza

(i) tipikus üzenetek: $1-\varepsilon$-nál nagyobb valószínűség.

gel

$\Rightarrow$ elemszáma $\leq 2^{N\cdot(H+\delta)}$

(ii) nem tipikus üzenetek: ε-nál kisebb valószínűséggel

$\Rightarrow$ elemszáma $\leq K^N = 2^{N\cdot\log_2 K}$

(i) kódolás: T_1

$$2^{T_1\cdot(C-\delta)} \geq 2^{N(H+\delta)}$$

$$T_1 \geq N\cdot\frac{H+\delta}{C-\delta} = N\cdot\left(\frac{H}{C}+\alpha\right) \qquad \alpha: \text{tetszőleges kicsi}$$

(ii) marad felhasználatlan T hosszú csatornajel $\Rightarrow$ vegyünk egy ilyet $\Rightarrow$ T_2 hosszú jellel kódoljuk a nem tipikus üzeneteket $2^{T_2\cdot(C-\delta)} \geq 2^{N\cdot\log_2 K}$

$$T_2 \geq N\cdot\frac{\log_2 K}{C-\delta} = N\cdot\left(\frac{\log_2 K}{C}+\beta\right) \qquad \beta: \text{tetszőleges kicsi}$$

- kódhossz várható értéke: átviteli idő

$$(1-\varepsilon)\cdot T_1 + \varepsilon\cdot(T_1+T_2) = N\cdot\left(\frac{H}{C}+\alpha\right) + \varepsilon\cdot\left(N\cdot\frac{\log_2 1}{C}+\beta\right)$$

- 1 szimbólumra jutó kódhossz:

$$\frac{H}{C}+\alpha+\varepsilon\cdot\frac{\log_2 K}{C}+\varepsilon\beta = \frac{H}{C}+\varepsilon \Rightarrow V = \frac{C}{H}-\delta \checkmark$$

$\downarrow$ 1 sec-re jutó

Nevezetes kódolások:

Előre meg kell mondani, hogy mivel történik a kódolás - a jövőben véletlen eloszlás szerint érkeznek a szimbólumok

-

• Bináris kódolás:

- valaszthato szimbólumok: $x_1, \dots, x_n$

  x_i - w_i bináris kód

  $\ell(w_i)$ kódhossz

- feladat: visszafejthetőség = eredeti sorozat visszaállítható legyen

  elégséges feltétel:

  pl.: prefix kódok: $\forall i, j \in [1 \dots n]$ w_i nem kezdő-szelete w_j-nek

$$\sum_{i=1}^{n} 2^{-\ell(w_i)} = 1$$ (véges, teljes kódrendszer)

$$\sum_{i=1}^{\infty} 2^{-\ell(w_i)} \leq 1$$ (Kraft-egyenlőtlenség: végtelen kódrendszer)

- szimbólumok eloszlása:

  $x_1, \dots, x_n$ - $p_1, \dots, p_n$ valószínűségekkel

  $\Rightarrow$ kódhossz várható értéke: $\sum_{i=1}^{n} p_i \cdot \ell(w_i)$

TÉTEL $H(p_1, \dots, p_n) \leq \sum_{i=1}^{n} p_i \cdot \ell(w_i)$ $\forall$ prefix kódra

$\exists$ prefix kód

$\sum_{i} p_i \cdot \ell(w_i) < H(p_1, \dots, p_n) + 1$

Biz:

a_i $i \in [1 \dots n]$ $\quad a = \sum_{i=1}^{n} a_i$

b_i $i \in [1 \dots n]$ $\quad b = \sum_{i=1}^{n} b_i$

$$\sum_{i=1}^{n} a_i \cdot \log \frac{b_i}{a_i} \leq a \cdot \log \frac{b}{a}$$

$$\sum_{i=1}^{n} p_i \cdot \log_2 \frac{q_i}{p_i} \leq 1 \cdot 0$$

$$H(P_1, \dots, P_n) \leq -\sum_{i=1}^{n} P_i \cdot \log_2 q_i$$

$$\sum_{i=1}^{n} 2^{-\ell(w_i)} = 1, \quad q_i = 2^{-\ell(w_i)}$$

$$H(P_1, \dots, P_n) \leq -\sum_{i=1}^{n} P_i \cdot (-\ell(w_i)) = \sum_{i=1}^{n} P_i \cdot \ell(w_i) \checkmark$$

• Nevezetes kódolások

(1) Shannon-Fano kód

- $P_1, \dots, P_n : P_1 \geq \dots \geq P_n$ (rendezett)

$$Q_1 := 0$$

$$Q_i := \sum_{j=1}^{i-1} P_j \quad i \in [1 \dots n]$$

- bináris kódoljuk a Q_i számokat ($= w_i$)

$\ell_1, \dots, \ell_n$ (ℓ_i pozitív)

$$2^{-\ell_i} \leq P_i < 2^{-\ell_i + 1}$$

$$\ell_i \geq -\log_2 P_i > \ell_i - 1$$

$$-\sum_{i=1}^{n} P_i \cdot \log_2 P_i > \sum_{i=1}^{n} P_i \cdot \ell_i - 1$$

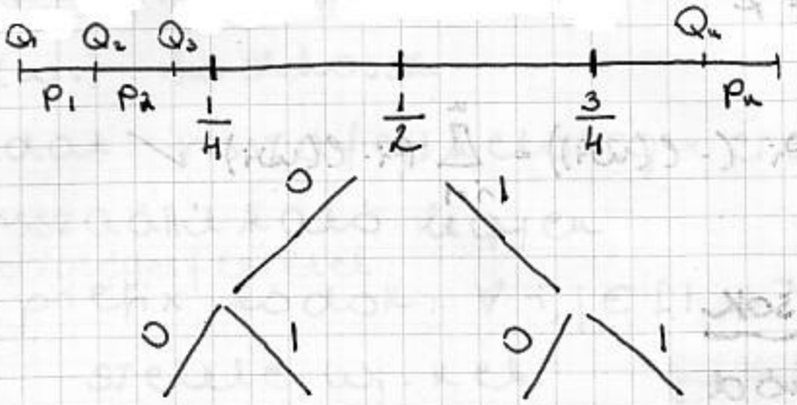
$$\sum_{i=1}^{n} P_i \cdot \ell_i < H + 1$$

prefix kód:
- egy kódszó csak a nagyobb indexű prefixe lehet ⇒ már a következők prefixe
- w_i nem prefixe w_{i+1}-nek:
  w_i hossza ℓ_i : P_i első értékes jegye az ℓ_i-edik
  $Q_{i+1} = Q_i + P_i \Rightarrow w_i$ első ℓ_i bitben változik

- szimbólumok gyakoriság szerinti csökkenő sorendbe való rendezése
- lista két részre osztása úgy, hogy a két részben a szimbólumok gyakoriságainak összege közel egyenlő legyen (0. / 1 kód hozzárendelése)
- a felezés addig tart, míg 1 szimbólum ma-

rad az intervallumban

- $Q_1, \dots, Q_n$ az intervallumok kezdőpontja



⇒ kódfa építhető (prefix kód)

- gyakrabban szereplő elemek kódja rövidebb

(2) Gilbert-Moore kód

- $P_1, \dots, P_n$ (nincs rendezés)
- $Q_1, \dots, Q_n$ az intervallumok felezőpontja
- a kódját ugyanúgy, mint a S-F kódnál, felezésekkel építik fel

$$\sum_{i=1}^{n} p_i \cdot \ell_i < H + 2$$

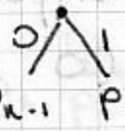
(3) Huffman kód:

- $(P_1, \dots, P_n)$
- kiválasztjuk mindig a két legkisebbet és így építünk kódfát (rekurzív)

W0 - P_{n-1}

W1 - P_n

$W = P_{n-1} + P_n = q$



- $(P_1, \dots, P_{n-2}, q)$: ismét végrehajtjuk az előző lépést

⇒ optimális kód:

- bármely más prefix kódra: w_i'

$$\sum_{i=1}^{n} p_i \cdot \ell(w_i') \geq \sum_{i=1}^{n} p_i \cdot \ell(w_i)$$

$$\sum_{i=1}^{n} p_i \cdot \ell(w_i) < H + 1$$

⇒ gyakoriság:

- x_i : n_i előfordulással

$$\sum_{i=1}^{n} n_i = N$$

minimalizálás: $\sum_{i=1}^{n} n_i \cdot \ell(w_i)$

$$\sum_{i=1}^{n} \frac{n_i}{N} \cdot \ell(w_i) \Rightarrow \text{már eloszlás} \Rightarrow$$

⇒ mindegy, hogy eloszlással vagy gyakorisággal számolunk.

- $(\xi_1, \dots, \xi_N)$ együttes eloszlás esetén: $H(\xi_1, \dots, \xi_N)$ nevezetes kód:

⇒ kódhossz várható értéke $< H(\xi_1, \dots, \xi_N) + 1$

⇒ 1 szimbólumra jutó kódhossz:

$$\frac{1}{N} \cdot H(\xi_1, \dots, \xi_N) + \frac{1}{N} \leq H + \delta + \frac{1}{N}$$

2009.03.25. VII. EA

Veszteséges tömörítések:

- **Lempel-Ziv kódok**

(1) **LZ77 kód**

- ha hosszúságú csúszóablak

  $h_a = h_k + h_e$

  h_k: mái megnézett (kereső)

  h_e: következő betű (előretekintő)

- keresőpufferben megkeresi az előretekintő puffer első karakterének előfordulásait
- a megtalált pozíciókkal kezdődően megnézi, hogy hol a leghosszabb az egyezés → kiválasztja
- alkülödött kód: $< t, h, c >$

  t: távolság (kereső - előretekintő)

  h: egyezés hossza

  c: a már nem egyező, első szimbólum kódja
- n szimbólum: $\lceil \log_2 n \rceil$

- $h(t) = \lceil \log_2 h_k \rceil$
- $h(h) = \lceil \log_2 h_e \rceil$
- stacionárius, ergodikus

(2) **LZ78 kód**

- szótár használata (folyamatosan bővül)
- megnézi, hogy az új szövegrészben milyen az előzőekkel a leghosszabb egyezés

  → amivel egyezik, az mái a szótárban van:

  kódja: i

- aktuális kód: <i, c>
  i: egyező rész szótárkódja
  c: a mai nem egyező, első szimbólum kódja
- új szótárbejegyzés

(3) **LZW kód** (T, Welch)
- 1978-ban megfelelően működik
  előírás: aktuális kód: i
    i: egyező rész szótárkódja
    - következő kódolandó az előző végével kezdődik
    új szótárbejegyzés

pl.: UNIX
GIF

- szótár mérete:
  - korlátozzuk
  - újraépítjük
- tömörített szöveg már nem tömöríthető

Kolmogorov bonyolultság:
- algoritmikus információ elmélet
Shannon: előre megmondjuk, hogyan fogunk kódolni (várható érték)
Kolmogorov: hogyan tudunk tömörebben kódolni (rosszabbul ez a probléma)

• **tömöríthetőség:**
- tömörítendő szöveg: x
- kód: p
- f(p) = x kiszámítható

x alakja	$f(p)$	x kódja	kódhossz
$x \in \mathbb{N}$	$f_0(p) = p$	x	$h(x) \cdot \log_2 x$
$x = 2^k$	$f_1(p) = 2^p$	k	$\log_2 h(x)$
$x = 2^{2^k}$	$f_2(p) = 2^{2^p}$	k	$\log_2(\log_2 h(x))$
$x = 2^{2^{\cdot^{\cdot^{2^k}}}}$ (k)	$f_3(p) = 2^{2^{\cdot^{\cdot^{2^p}}}}$ (p)	k	$\underbrace{\log_2 \dots \log_2}_{k+1}(h(x))$

- x kódolása: $\langle i, p \rangle$ (kiszámítható függvények)
  - i: melyik függvényt használjuk
  - p: függvények száma

pl.: $x = 1025$

$x = 2^{10} + 1$

$f_{11}(p) = 2^p + 1$

- általában: véges sok döntést választva, ezek a módszerek tudunk egy további választani, ami spec. számokra kömörebb

⇒ rekurzívan felsorolható halmaz

- **parciális rekurzív függvények:**

$f: \mathbb{N}^k \to \mathbb{N}$

alapfüggvények:

zérófüggvény: $Z: \mathbb{N} \to \mathbb{N}$ $Z(x) = 0$

rákövetkező függvény: $S: \mathbb{N} \to \mathbb{N}$ $S(x) = x + 1$

vetítő függvény: $P_{n,i}: \mathbb{N}^n \to \mathbb{N}$ $P(x_1, \dots, x_i, \dots x_n) = x_i$

függvényképzések:

kompozíció: $f: \mathbb{N}^n \to \mathbb{N}$ képződik

$g_1, \dots, g_k: \mathbb{N}^n \to \mathbb{N}$, $h: \mathbb{N}^k \to \mathbb{N}$

$f(x_1, \dots, x_n) = h(g_1(x_1, \dots, x_n), \dots, g_k(x_1, \dots, x_n))$

rekurzió (primitív): $f: \mathbb{N}^{n+1} \to \mathbb{N}$ képződik

$g: \mathbb{N}^n \to \mathbb{N}$, $h: \mathbb{N}^{n+2} \to \mathbb{N}$

$f(x_1, \dots, x_n, 0) = g(x_1, \dots, x_n)$

$f(x_1, \dots, x_n, y+1) = h(x_1, \dots, x_n, y, f(x_1, \dots, x_n, y))$

⇒ primitív rekurzív függvények osztálya:

- zárt a kompozícióra
- zárt a primitív rekurzióra
- $Z, S, P_{n,i} \in$ osztály

⇒ generáló sorozat: $g_1, \ldots, g_n = f$

g_i : alapfüggvény ∨

való kisebb indexűekből kompozícióval ∨

való kisebb indexűekből primitív rekurzióval

Tétel Primitív rekurzív függvények osztálya = primitív rekurzív függvényosztályok metszete

pl.: Nem minden kiszámítható függvény primitív rekurzív

(1) univerzális függvény

$\exists\ U: \mathbb{N}^2 \to \mathbb{N}$ univerzális kiszámítható:

$\forall\ f: \mathbb{N} \to \mathbb{N}$ primitív rekurzív függvénye

$\exists\ K: U(x, K) = f(x)$ $(\forall x$-re$)$ $\wedge\ \forall K: U(x, K)$ primitív rekurzív függvény

- $U(x, K)$-t úgy kapjuk, hogy kikeressük a K. generáló sorozatot és kiszámítjuk a generálás által adott számítást x-re.

$h(x) := U(x, x) + 1$

⇒ $h(x)$ kiszámítható

DE: nem primitív rekurzív

Ui: indirekt

T.F.H. $h(x)$ primitív rekurzív (számára $a = k$) $\Rightarrow$

$\Rightarrow h(k) = \begin{cases} = U(k,k)+1 \Rightarrow \text{def. miatt} \\ = U(k,k) \Rightarrow \text{diadás miatt} \end{cases}$ ↯

(2) Ackermann-függvények:

$A: \mathbb{N}^2 \to \mathbb{N}$

$A(0,y) = y+1$

$A(x+1, 0) = A(x,1)$

$A(x+1, y+1) = A(x, A(x+1, y))$

$a_n: \mathbb{N} \to \mathbb{N}$

$a_n(x) := A(n,x)$

$a_{n+1}(x+1) = A(n+1, x+1) = A(n, A(n+1, x)) =$

$= a_n(A(n+1, x)) = a_n(a_{n+1}(x))$

$\Rightarrow$ kapok apriori növekszenek

$f: \mathbb{N}^k \to \mathbb{N}$ r A-szinten belül van, ha

$\forall x = \max(x_1, \dots, x_k): f(x_1, \dots, x_n) \leq a_r(x)$

f primitív rekurzív $\Rightarrow \exists r: f$ r A-szinten belül van (függvénykifejezések csak véges szinttel emelik a függvény A-szintjét

$\Rightarrow f(x) := a_x(x)$ nem primitív rekurzív

minimalizálás függvény: $f: \mathbb{N}^n \to \mathbb{N}$ képződik

$g: \mathbb{N}^{n+1} \to \mathbb{N}$

$f(\vec{x}) = \mu y([g(\vec{x}, y) = 0])$

$f(x_1, \dots, x_n) = k$, ha

(1) $g(x_1, \dots, x_n, k) = 0$

(2) $g(x_1, \dots, x_n, y) \neq 0 \quad (y < k)$

(3) $g(x_1, \dots, x_n, y)$ értelmezve van $y \leq k$-ra

$\Rightarrow$ parciálisan rekurzív függvények osztálya
(= Turing-Kisztámítható függvények)

Szigorú (regulális) minimalizálás:
- ua., mint a minimalizálás +
(4) $\forall (x_1, \dots, x_n): \exists\, g(x_1, \dots, x_n) = 0$
(5) g totális

$\Rightarrow$ rekurzív függvények osztálya

- Zártság és a generáló sorozatok ugyanúgy definiálhatók, mint a primitív esetben
  különbségek
- parc. rek. fv.-ek generáló sorozata: szintaktikusan jellemezhető
- rek. fv-ek generáló sorozata: szemantikus (nem ellenőrizhető)

Univerzális függvény:
$\exists\, U: \mathbb{N}^2 \to \mathbb{N}$: univerzális:
$\forall f: \mathbb{N} \to \mathbb{N}$ parciális rekurzív függvénype
$\exists\, K$: $U(x, K) = f(x)$ ($\forall x$-re) $\wedge\ \forall K: U(x, K)$ parciális rekurzív függvény
$h(x) = U(x, x) + 1$ nem ellenpélda

$h(n) = \begin{cases} U(n, n) + 1 \\ U(n) \end{cases}$ csak annyit jelent, hogy nincs értelmezve n-ben

- U konstruálása ugyanaz, mint a primitív esetben

2009.04.01 VIII. EA

• Kleen-féle normálalak:

$\exists U: \mathbb{N} \to \mathbb{N}$ ∧ $V: \mathbb{N}^3 \to \mathbb{N}$ primitív rekurzív:

$\forall f: \mathbb{N} \to \mathbb{N}$ parciális rekurzív függvénye

$\exists K: f(x) = U(\mu t\,(V(K, x, t) = 0))$ $(\forall x \cdot rc)$

$A \subseteq \mathbb{N}$ rekurzív, ha

$$k_A(x) = \begin{cases} 0 & x \notin A \\ 1 & x \in A \end{cases} \quad \text{rekurzív (eldönthetőség)}$$

$A \subseteq \mathbb{N}$ rekurzívan felsorolható, ha

$\exists f: \mathbb{N} \to \mathbb{N}$ rekurzív: $A = f(\mathbb{N})$ $\quad x \in A \Leftrightarrow \exists n: f(n) = x$

• Rice-tétel:

TÉTEL A részleges részhalmaza a parciális rekurzív függvényeknek: $A = \{n \mid f_n \in A\}$

A rekurzív $\Leftrightarrow$ $A = \emptyset$ v A az összes parciális rekurzív függvény halmaza

• Kolmogorov-entrópia:

- önkorlátozó kódolás:

$x = a_1 \dots a_n$

$\overline{x} = \underbrace{1 \dots 1}_{\text{x hossza}} \underbrace{0}_{\text{záró}} \underbrace{a_1 \dots a_n}_{x}$

$c(\overline{x}) = 2 \cdot c(x) + 1$

$x' = \overline{n}\,x$

$c(x') = 2 \cdot \log_2 n + c(x) + 1$

(x,y) esetén:

$\gamma_0(x,y) = \bar{x}y = z_0 - \pi_{0,1}(z_0) = x \quad \pi_{0,2}(z_0) = y$

$\gamma_1(x,y) = x'y = z_1 - \pi_{1,1}(z_1) = x \quad \pi_{1,2}(z_1) = y$

Feladat: x adott

Mennyire tömöríthető? ⇒ Legrövidebb kód, amely-ne a dekódolást végző függvény kiszámítható ⇒ ⇒ kód + dekódoló kódjának hosszainak mini-muma a tömöríthetőség alsó határa

- $\mathbb{N} = \{0,1,\dots\}$, $\Lambda = \{0,1\}^*$
- egy természetes számnak kölcsönösen egyértel-műen megfelelt elhető
- x bonyolultsága: $f: \mathbb{N} \to \mathbb{N}$ p.r. fv.
  Az $x \in \mathbb{N}/\Lambda$ bonyolultsága az f függvény szerint: $C_f(x) = \min_p \{ \ell(p) \mid f(p) = x \}$
  (Ha nincs ilyen p, akkor végtelen)

TÉTEL Alaptétel / Invariancia tétel

$\exists$ optimális $f_0: \mathbb{N} \to \mathbb{N}$ p.r. fv.:

$\forall f: \mathbb{N} \to \mathbb{N}$ p.r. fv.-re: $\exists K_f$ konstans:

$C_{f_0}(x) \leq C_f(x) + K_f \quad (\forall x \in \mathbb{N}\text{-re})$

⇒ optimális függvények szerinti bonyolultság csak konstansban tér el egymástól

f_0, g_0 optimális: $|C_{g_0}(x) - C_{f_0}(x)| \leq K_{f_0 g_0}$

- rögzítjük az f_0 optimális függvényt

Az $x \in \mathbb{N}$ Kolmogorov bonyolultsága $C(x) = C_{f_0}(x)$

Biz:

x, $f: \mathbb{N} \to \mathbb{N}$ tetszőleges

$C_f(x) := k \Rightarrow \exists p : f(p) = x \wedge \ell(p) = k$

$U(y, n)$ univerzális p.r. fv. $\wedge$ f sorszáma $= n \Rightarrow$

$\Rightarrow U(p, n) = f(p)$

$z = \overline{n}\,p$ kód: $f_0(z) = U(\Pi_{1,2}(z), \Pi_{1,1}(z)) = U(p, n) = x$

$C_{f_0}(x) \leq \ell(z) = \ell(n') + \ell(p) = \ell(p) + k_f = C_f(x) + k_f$ ✓

Tulajdonságai:

(1) Alaptétel: $C(x)$ konkét $C_f(x)$ ismeretében, csak felülről becsülhető

(2) $C(x) > k$ jelentése: $\exists p : f_0(p) = x \wedge \ell(p) > k$ (2^{k+1} kód)

(3) $C(x)$ korlátlan függvény: $C(x) \leq \ell(x) + k$

$f(p) = p \Rightarrow C(x) \leq C_f(x) + k_f \leq \ell(x) + k$

(4) $C(x)$ nem kiszámítható

T.F.H. $C(x)$ kiszámítható

x_0: első x : $C(x) > 1 \cdot 2^0$

x_1 : — " — : $C(x) \geq 2^1 \wedge x_1 > x_0$

x_k : — " — : $C(x) \geq 2^k \wedge x_k > x_{k-1}$

$f(p) := x_p$

$C_f(x_p) \leq \ell(p) = \log_2 p \leq \log_2 (\log C(x_p))$

$C(x_p) \geq 2^p$

$C(x_p) \leq C_f(x_p) + k_f \leq \ell(p) + k_f \leq \log_2 (\log C(x_p)) + k_f \Rightarrow$

$\Rightarrow$ csak véges sok p-re teljesülhet ↯

(5) $\lim_{x \to +\infty} C(x) = +\infty$

Ha ∞ sokszor n alatt maradna $\Rightarrow$ 2^n kód hasz nálata ↯

(6) $\mu(x) := \min_{y>x} C(y)$ (alsó burkoló)

$\Rightarrow \lim_{x\to+\infty} \mu(x) = +\infty \wedge \mu(x)$ ∀ kiszámítható,

monoton, ∞-hez tartó függvénynél lassabban nő és nem kiszámítható

(7) $x+h$ bonyolultsága:

$C(x+h) \leq C(x) + C(h) + 2\cdot\log_2 C(h) + K$

z = $\bar{h}p$ kód, $f_0(p) = x$, $C(x) = \ell(p)$

$f(z) = f_0(\Pi_{1,2}(z)) + \Pi_{11}(z) \cdot f_0(p) + h = x + h$

$C_1(x+h) \leq \ell(p) + \ell(h) + 2\cdot\log_2 \ell(h) + K_1$

$C(x+h) \leq C(x) + C(h) + 2\cdot\log_2 C(h) + K$

- a $g(x,h) = x+h$ helyett tetszőleges két változós kiszámítható függvényt vehetünk

IX. EA 2009.04.15.

- kivételek ⇒ konvarok (XML)

X. EA 2009.04.22.

Óra eleje: XSLT

(7) $x+h$ bonyolultsága

$g(x,h) = y$

$C(g(x,h)) \leq C(x) + C(h) + 2\cdot\log_2 C(h) + K_g$

$C(x) = K$, $C(h) = \mu$, $f_0(p) = x$, $f_0(q) = h$, $\ell(p) = K$, $\ell(h) = \mu$

$r = \bar{q}p$ kód

$f(r) = g(f_0(\underbrace{\Pi_{1,2}(r)}_{p}), f_0(\underbrace{\Pi_{11}(r)}_{q})) = y$

$C_f(q(x,k)) \leq \ell(p) + \ell(q) + 2\cdot\log_2 \ell(q) + K_f$

$C(q(x,k)) \leq C(x) + C(k) + 2\cdot\log_2 C(k)$

• Feltételes Kolmogorov-entrópia:

- x kiszámításához mennyit segít y ismerete

Mekkora a legrövidebb program kódjának a hossza, ami y-ból x-et kiszámítja?

- x feltételes bonyolultsága y ismeretében:

$f: \mathbb{N}^2 \to \mathbb{N}$ p.r. fv. $f(y,p) = x$

Az $x \in \mathbb{N}$ / $\mathcal{A}$ bonyolultsága (feltételes-y ismeretében) az f függvény szerint:

$$C_f(x|y) = \min_p \{\ell(p) \mid f(y,p) = x\}$$

TÉTEL Alaptétel / Invariancia tétel

$\exists$ optimális $f_0^{(2)}: \mathbb{N}^2 \to \mathbb{N}$ p.r. fv.:

$\forall f: \mathbb{N}^2 \to \mathbb{N}$ p.r. fv-re $\exists K_f$ konstans:

$C_{f_0^{(2)}}(x|y) \leq C_f(x|y) + K_f \quad (\forall x,y\text{-ra})$

$\Rightarrow$ optimális függvények szerinti feltételes bonyolultság csak konstansban tér el egymástól

$f_0^{(2)}, g_0^{(2)}$ optimális: $|C_{f_0^{(2)}}(x|y) - C_{g_0^{(2)}}(x|y)| \leq K_{f_0^{(2)}, g_0^{(2)}}$

- rögzítjük az $f_0^{(2)}$ optimális függvényt

Az $x \in \mathbb{N}$ feltételes Kolmogorov bonyolultsága

$C(x|y) = C_{f_0^{(2)}}(x|y)$

Biz

- ugyanúgy, mint az egyszerű esetben

$U(x,k,p)$ univerzális függvénnyel

XI. EA 2009.04.29.

Tulajdonságai:

(1) Alaptétel: $C(x|y)$ konkét $C_f(x|y)$ ismeretében csak felülről becsülhető

(2) $C(x|y) = K$ jelentése: $\exists p: f_0(p,y) = x \wedge \ell(p) = K$

(3) $C(x) \leq C(x|y) + C(y) + 2 \cdot \log_2 C(y) + K$

(4) halmaz szerinti feltételes bonyolultság

$A = \{a_1, \ldots, a_N\}$

$C(x|A) = ?\quad x \in A$

$\langle A \rangle = \bar{a}_1 \bar{a}_2 \ldots \bar{a}_N$

$\Rightarrow C(x|A) = C(x|\langle A \rangle)$

$\Downarrow$

véges halmaz elemeinek a halmaz szerinti feltételes bonyolultsága

általános eset:

adott: A, $|A| = N$ kódja: p

$|\{x \mid x \in A \wedge C(x|A) < \log_2 N - \delta\}|$?

Keressük: p kód: $f_0(p, \langle A \rangle) = x \in A$

$$1 + 2 + 2^2 + \ldots + 2^{\log_2 N - \delta - 1} = 2^{\log_2 N - \delta} - 1$$

$$\frac{|\{x \mid x \in A \wedge C(x|A) < \log_2 N - \delta\}|}{|\{x \mid x \in A\}|} \leq 2^{-\delta}$$

$\rightarrow$ A elemeinek többségére az A szerinti feltételes bonyolultság nem lehet δ-nál kisebb, mint $\log_2 N$

OK! $\log_2 N - K$-nél δ-val rövidebb kódok száma, amit f_0 használhat $\leq N \cdot 2^{-\delta}$

$\Rightarrow$ egyenletes kódhossz választása az indokolt

$\Rightarrow$ szükséges a megfelelő halmaz megtalálása

pl.: személyek adatait relációs AB-t építünk

R(csnev, unev)

csnev: 10 byte = 160 bit

unev: 20 byte = 160 bit

A: összes lehetséges reláció előfordulás

- lehetséges különböző sorok száma: 2^{320}
- lehetséges előfordulások száma: $2^{2^{320}}$

[lehetséges különböző sorok számának a részhalmaza)

$\Rightarrow |A| = 2^{2^{320}}$

$\Rightarrow \{I \mid I \in A\}\ C(I|A) = 2^{320}$

- sorok száma < 10 milliárd

csnev: első 256 leggyakoribb (80%)

unev: első 256 leggyakoribb (80%)

- H megvan készítése

1. Minden nap gyakori: 60% - 1 byte
2. Családnév gyakori: 16% - 22 byte
3. Utónév gyakori 16% - 22 byte
4. Egyik nap sem gyakori 4% - 4 byte

• paraméteres halmazseregek:

- a: paraméter (kód)
- B_a: véges halmaz
- $h(a) = B_a$ h totális, rekurzív függvény

$A \subseteq \mathbb{N} \times \mathbb{N}$ rekurzíván rekurzíván felsorolható

$B_a = \{x \mid (x, a) \in A\}$

megszorítás: $|B_a| = ha$ véges

- a alapján B_a elemei felsorolhatóak algoritmikusan

TÉTEL $\exists K_A$ konstans: $C(x|B_a) = C(x|a) \leq$
$\leq \log_2 \mu_a + K_A$ $(\forall x \in B_a, \forall a \cdot \text{na})$

Biz:

- B_a elemei: A-hoz $\exists$ felsorolós függvény:
  $f(t,a)$ konstruálása
- futtatjuk A felsorolásait
  ⇒ első $x \in B_a$ ⇒ $f(0,a) = x_1$
  ⇒ megyünk tovább ⇒ $f(1,a) = x_2, \ldots, f(\mu-1, a) = x_\mu$
  ⇒ B_a $\forall$ eleme felsorolásra került ⇒ $f(t,a)$ ez után nincs értelmezve
⇒ $C(x|a) \leq C_f(x|a) + K_f$ $\in (0, \mu_{a-1})$ közötti érték
⇒ $C(x|a) \leq \log_2 \mu_a + K_A$ ✓

t értel jelentése: séma + paraméter jó megválasztása

- **érdekes halmazsereg:**

$D_K = \{x \mid C(x) \leq K\}$

- f_0 optimális függvény

hely / lépés	0	1	2	3
1	1	1	2	3
1	1	2	3	
2	2	3		
3	3			

- ha $x \in D_K$: valamelyik K hosszú értéke kapunk $f_0(p) = x$ eredményt

$D^*_K = \{x \mid C(x) = K\}$

$D'_K = \{x \mid C(x) < K\}$

Kolmogorov- és Shannon entrópia összevetés

- minél nagyobb a jelleség, annál szorosabb a kapcsolat közöttük

- N hosszú sorozat:

  ⇒ tipikus halmaz: 1-ε valószínűség

  $$2^{N\cdot(H-\sigma)} < M(N) < 2^{N\cdot(H+\sigma)}$$

  $M(N)$ ↓ tipikus halmaz elemszáma

  Kolmogorov: x üzenet tipikus ⇒ statisztika jellemzi

  $$x \in B_N : C(x|B_N) = C(x|N) \leq \log_2 M(N) + K_{fonds}$$

- $$P_{TIP}(x) = \begin{cases} 0 & x \notin \text{tipikus} \\ 1 & x \in \text{tipikus} \end{cases}$$

  $$A := [\, x \mid \ell(x) = N \wedge P_{TIP}(x) = 1 \,]$$

  B_N = { tipikus N hosszú sorozatok }

- $\log_2 M(N) = \log_2 |B_N|$

  $$\log_2 |B_N| - \varepsilon = N\cdot(H-\sigma) - \varepsilon \leq C(x|B_N) \leq N\cdot(H+\sigma) + K_{fonds}$$

- 1 szimbólumra:

  $$H - \sigma - \frac{\varepsilon}{N} \leq \frac{C(x|B_N)}{N} < H + \sigma + \frac{K_{fonds}}{N}$$

Egy szimbólumra jutó entrópia a tipikus halmaz elemeit megegyezik a feltételes Kolmogorov entrópia egy szimbólumra jutó részével.

XII. EA 2009.05.06

· Információ-növekedés törvénye

- Mennyi információ nyerhető ki az adatbázisból?

⇒ a kinyert információ nem lehet több, mint amennyit beírtunk

- adatbázis tartalma: x
- kérdés: y
- válasz: a

$f(y, x) = a$

$$C(a|x) \leq C_f(a|x) + K_f \leq C(y) + K_f$$

$$\| \\ C(y)$$

· Prefix Kolmogorov bonyolultság:

- ua., mint a Kolmogorov bonyolultság

DE! prefixmentes függvényeket használunk

$f: \mathbb{N} \to \mathbb{N}$ prefixmentes, ha f értelmezve van p-n és q-n, akkor egyik sem prefixe a másiknak

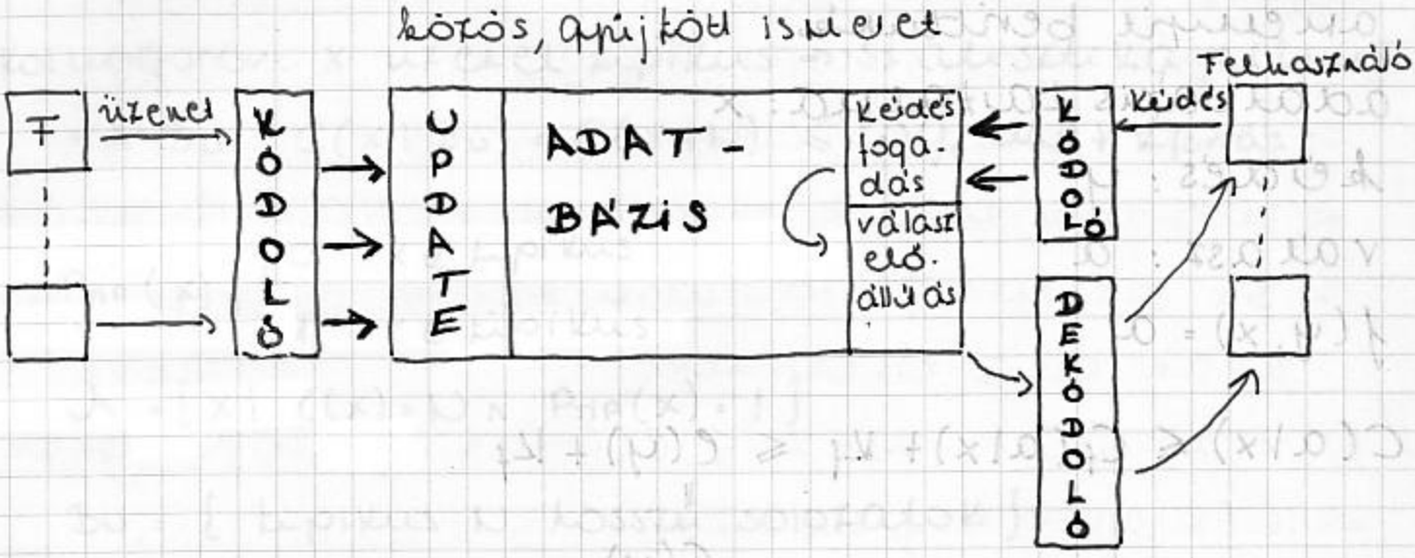
- minden tulajdonságot és az Alaptétel ugyanúgy értelmezünk

$$K(x, y) \leq K(x) + K(y|x) + 2 \cdot \log_2 K(x)$$

2009.05.13 XIII. EA

ABKR: elvi modell

Diagram: információs rendszer



megszorítások:
- Mi lehet az adatbázis tartalma?
- Milyen új üzenetek lehetségesek?
- Hogyan épülnek be a régi ismeretekbe?
- Milyen kérdések fogalmazhatóak meg?
- Mik a válaszfüggvények?

ABKR: formális modell

(1) logikai fogalmi modell:

- valóság modellezése → absztrakt adattípusok → adatmodellek
- kódrendszer → kiszámíthatósági, bonyolultságelméleti elemzések

- kódrendszer = bináris szavak rendszere

(a) Mi lehet az adatbázis lehetséges állapota?

$\Sigma \subseteq \Lambda$ Σ: rekurzív v. rekurzívan felsorolható

- sémák használata: séma előírás = adat

$S \subseteq \Lambda$

$s \in S$: Σ_s séma szerinti lehetséges adatok (bináris kódok)

$\langle s, y \rangle = x = s'y$

$s_1, \ldots, s_k$
$y_1, \ldots, y_k$ } $\langle s_i, y_i \rangle$ felsorolós kód

(b) Hogyan lehet módosítani?

Y: módosítandó adatok halmaza

$M: Y \times \Sigma \to \Sigma$ módosító függvény

$u \in M, y \in Y$: $u(y, x) = x'$

- Zárt sági elvárás: $M(Y \times \Sigma) \subseteq \Sigma$

- szemantika az adatmodellben dől el ⇒ generikus

- ekvivalencia-osztályok Σ-ban

pl.: üres módosítás: $x \equiv x'$, ha $M(\Lambda, x) = M(\Lambda, x')$

- sémakerülő módosítások ⇒ sémán belüli módosítás, sémán belül kell maradjon

$s \in S$: Σ_s ∧ $M_s \subseteq M$ sémán belüli módosító függvény

- adatmennyiségek:
  - hossz (redundancia)
  - $K(x)$, $C(x)$
  - sémaváltozás: $|\Sigma_s| = N$

$x \in \Sigma_s$: $K(x|s) \sim \log_2 N$

• relációs-séma:

$S = R(A_1, \dots, A_k)$ $A_i - |D_i| = \mu_i$

I: az S előfordulása

$\Rightarrow$ lehetséges előfordulások száma: $2^{\prod_{i=1}^{k} \mu_i} = |\Sigma_S|$ (sorok száma)

- legfeljebb K sor: $N - \binom{\mu}{K} \sim \mu^K$

- kód: $K \cdot \log_2 \mu$

- minden lehetséges függőség korlátozza a sorok számát $\Rightarrow$ tömörebb ábrázolási lehetőség

(c) <u>Lekérdezések megfogalmazása:</u>

Q: kérdések halmaza

$q \in Q$: Σ_q: válaszok halmaza

$A: Q \times \Sigma \to \bigcup_{q \in Q} \Sigma_q$: válasz függvény

$q \in Q$; $a \in A$: $a(q, x) = v \in \Sigma_q$ (értelmes)

- kérdések bonyolultsága $\Rightarrow$ generikus

(1) <u>Fizikai modell</u>

- bedarabolódik egy bináris kód címezhető, véges kapacitású rekeszekbe

fizikai kód: $(c_1, l_1, y_1), \dots, (c_k, l_k, y_k)$

cím hossz adat

ϕ: fizikai kódok rendszere:

$f: \Sigma \to \phi$: $x \in \Sigma$: $\phi_x \subseteq \phi$

módosítások: $y \in Y$: ϕ_y

$w \in \phi_y$, $z \in \phi_x$: $\phi_{xy} = \phi_y \phi_x = wz \xrightarrow{f_\mu} v \in \phi_{\mu(y,x)}$

Levezetések: $q \in Q : Q_q$

$\phi_q, \phi_x, \phi_{ABKR} \rightarrow \phi_{A(q,x)}$