BEVEZETÉS AZ INF. ELM.-BE

1. előadás (02.12.)

Bevezetés, alapfogalmak

Legyen X egy diszkrét valószínűségi változó, amely értékét az $\mathcal{X}$ abc-ből veheti fel! A hozzá tartozó valószínűségi eloszlás: $p_X = Pr\{X = x\}$, $x \in \mathcal{X}$

Jelölés: Ha a valószínűségi eloszlás változóját ugyanazon a betű jelöli, mint a változót, akkor az indexet elhagyjuk: $p(x) = p_X(x)$

$p(y) = p_Y(y)$

Def: Egy diszkrét valószínűségi változó entrópiáján az alábbi:

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x).$$

rövid megjegyzés:
- A log alapja 2, ekkor a mértékegység bit.
- konvenció szerint $0 \cdot \log 0 = 0$. (mert $\lim_{x \to 0} x \log x = 0$)
- Előző miatt, egy 0 valségű esemény nem befolyásolja H értékét.

Egy X-től függő g változó várható értéke p eloszlásra:

$$E_p\, g(X) = \sum_{x \in \mathcal{X}} g(x) p(x).$$ (egyszerűbb jelölés: $E g(X)$)

Vegyük észre, hogy az entrópia nem más mint a $g(X) = \log \frac{1}{p(X)}$ változó értéke!

Ennek jelentősége akkor látszik, ha bevezetjük az alábbi megfontolásokat.

Def: Legyen $h(p)$ az az információmennyiség, amihez egy p valószínűségű esemény bekövetkezésével jutunk! Ennek 3 tulajdonságát kell teljesítenie:

- $h(p) \geq 0$, mert minden esemény infót hordoz.
- $h(pq) = h(p) + h(q)$, mert független események információjához külön-külön is hozzájuthatunk.
- $h(1/2) = 1$, ez csak egy konvenció, ami rögzíti az egységet.

Az egyetlen fv, ami ezeket teljesíti: $h(p) = -\log_2 p$

Mivel $E h(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) = H(X)$, ezért az entrópia nem más mint a változó által közvetített információ várható értéke.

Jelölések: • Ha a log alapja nem 2, akkor az entrópia:

$$H_b(X) = -\sum_{x\in X} p(x)\log_b p(x)$$

Ha $b = e$, akkor a mértékegység a nats

• Ha $X = \begin{cases} 1 & p \text{ valószínűséggel} \\ 0 & 1-p \text{ valószínűséggel} \end{cases}$ akkor

$$H(X) = -p\log p - (1-p)\log(1-p) \equiv H(p)$$

Néhány tulajdonság:
- $H(X) \geq 0$
- $H_b(X) = (\log_b a)\, H_a(X)$

  így pl. 1 nats $= \ln 2$ bit $\approx 0{,}69$ bit
- $H(p)$ konkáv függvény, szimmetrikus $\frac{1}{2}$-re $\Rightarrow$ $H(p)$ max, ha $p = \frac{1}{2}$.
- $H(X)$ jelentése, hogy várhatóan hány bináris jellel kérdezhető egy esemény, vagy hány kérdéssel található ki.

  pl.: lóverseny esélyek $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}, \frac{1}{64}\right)$
- $H(X)$ független X értékeitől, csak $p(x)$ eloszlástól függ.

Def: A X és Y változók együttes eloszlásának entrópiája az együttes entrópia:

$$H(X,Y) = -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(x,y).$$

Az egyváltozós esethez hasonlóan: $H(X,Y) = -E\log p(X,Y)$

Def: Feltételes entrópia $H(Y|X)$ az Y változó entrópiájának várható értéke, feltéve, hogy X értéke ismert:

$$H(Y|X) = \sum_{x\in X} p(x)\, H(Y|X=x) = -\sum_{x\in X} p(x)\sum_{y\in Y} p(y|x)\log p(y|x) =$$

$$= -\sum_{x\in X}\sum_{y\in Y} p(x,y)\log p(y|x) = -E\log p(Y|X)$$

Tétel (láncszabály): $H(X,Y) = H(X) + H(Y|X)$

Biz: $$H(X,Y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y) = -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x)p(y|x) =$$

$$= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y|x) =$$

$$= -\sum_{x\in\mathcal{X}} p(x)\log p(x) - \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(y|x) =$$

$$= H(X) + H(Y|X) \qquad \square$$

Következmény: $H(X,Y|Z) = H(X|Z) + H(Y|X,Z)$

Vegyük észre, hogy $H(X|Y) \neq H(Y|X)$, de $H(X) - H(X|Y) = H(Y) - H(Y|X)$!

Def: Egy $p(x)$ és $q(x)$ eloszlás Kullback-Leibler-távolsága (vagy a relatív entrópia) az alábbi:
$$D(p\|q) = \sum_{x\in\mathcal{X}} p(x)\log\frac{p(x)}{q(x)} = E_p \log\frac{p(X)}{q(X)}$$

ahol konvenciókat használtunk:
- $0 \log \frac{0}{0} = 0$
- $0 \log \frac{0}{q} = 0$ $(q>0\text{-ra})$
- $p \log \frac{p}{0} = \infty$ $(p>0\text{-ra})$.

Tehát $D(p\|q) \geq 0$ és $D(p\|q) = 0 \iff p = q$, azaz egyfajta távolságként értelmezhető, de valójában nem igazi metrika (pl. $D(p\|q) \neq D(q\|p)$).

Értelmezés: Ha X változó eloszlása ismert, akkor az átlagos hossza kb. $H(p)$. Ha viszont a kódunkban a betűk q eloszlást követnek, akkor a szükséges átlagos hossz $H(p) + D(p\|q)$.

Def: X és Y változó közötti kölcsönös információ az együttes eloszlásuk és a marginális eloszlásuk szorzatának relatív entrópiája:
$$I(X;Y) = D\big(p(x,y)\,\|\,p(x)p(y)\big) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} =$$
$$= E_{p(x,y)}\log\frac{p(X,Y)}{p(X)p(Y)}$$

Jelölés: Az I két változóját $;$ választja el, így tehát a $I(X;Y,Z)$ mennyiség a X és az (Y,Z) változók kölcsönös információja.

Vegyük észre, hogy $I(X;Y)$ kifejezhető az entrópiákkal:

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} = \sum_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)} =$$

$$= -\sum_{x,y} p(x,y) \log p(x) + \sum_{x,y} p(x,y) \log p(x|y) = -\sum_{x} p(x) \log p(x) - \left(-\sum_{x,y} p(x,y) \log p(x|y)\right) =$$

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

Tehát $I(X;Y)$ értelmezése: X bizonytalanságának csökkenése amiatt, hogy Y ismert.

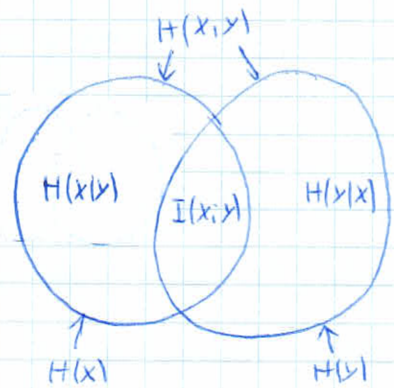
A láncszabály miatt: $I(X;Y) = H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y)$.

Továbbá: $I(X,X) = H(X) - H(X|X) = H(X)$.

Tehát a fontos összefüggések:

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

Grafikus szemléltetés:



BEV. INF. ELM.

2. előadás (02. 19.)

Eddig a láncszabályt és a H, I, D közötti kapcsolatokat 2 változóra írtuk fel, de általánosabb tételek is megfogalmazhatóak.

Tétel (általános láncszabály): Legyenek $X_1, X_2 \cdots X_n$ valószínűségi változók, a hozzájuk tartozó együttes eloszlás pedig $p(x_1, x_2, \cdots, x_n)$! Ekkor

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

Bizonyítás:

a) teljes indukcióval:

$n=2$-re láttuk: $H(X_1, X_2) = H(X_1) + H(X_2 \mid X_1)$

$n=3$-ra: $H(X_1, X_2, X_3) = H(X_1) + H(X_2, X_3 \mid X_1) = H(X_1) + H(X_2 \mid X_1) + H(X_3 \mid X_1, X_2)$

$\vdots$

ált n-re: $H(X_1, \cdots X_n) = H(X_{n-1}, X_n \mid X_1 \cdots X_{n-2}) + \sum_{i=1}^{n-2} H(X_i \mid X_{i-1}, \cdots, X_1) =$

$$= H(X_{n-1} \mid X_1, \cdots, X_{n-2}) + H(X_n \mid X_1, \cdots, X_{n-1}) + \sum \cdots =$$

$$= \sum_{i=1}^{n} H(X_i \mid X_1, \cdots X_{i-1}) \qquad \square$$

b) algebrailag:

Mivel $p(x_1, \cdots x_n) = \prod_{i=1}^{n} p(x_i \mid x_{i-1}, \cdots, x_1)$, ezért

$$H(X_1, \cdots X_n) = -\sum_{x_1 \cdots x_n} p(x_1, \cdots x_n) \log p(x_1 \cdots x_n) =$$

$$= -\sum_{x_1 \cdots x_n} p(x_1, \cdots, x_n) \log \prod_{i=1}^{n} p(x_i \mid x_{i-1}, \cdots x_1) =$$

$$= -\sum_{x_1 \cdots x_n} \sum_{i=1}^{n} p(x_1, \cdots x_n) \log p(x_i \mid x_{i-1}, \cdots x_1) = -\sum_{i=1}^{n} \sum_{x_1 \cdots x_i} p(x_1, \cdots, x_i) \log p(x_i \mid x_{i-1}, \cdots x_1)$$

$$= \sum_{i=1}^{n} H(X_i \mid X_{i-1}, \cdots X_1) \qquad \square$$

Def: A feltételes kölcsönös információ X valószínűségi változó között és adott Z melletti Y között:

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

$$= E_{p(x,y,z)} \log \frac{p(X, Y \mid Z)}{p(X \mid Z)\, p(Y \mid Z)}$$

Erre szintén felírható egy láncszabály

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

Biz: $I(X_1 \dots X_n ; Y) = H(X_1, \dots X_n) - H(X_1, \dots X_n | Y) =$

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

$$= \sum_{i=1}^{n} I(X_i ; Y | X_1 \dots X_{i-1}) \qquad \square$$

Def: A $p(x,y)$ és $q(x,y)$ eloszlások feltételes relatív entrópiája:

$$D(p(y|x) \| q(y|x)) = \sum_x p(x) \sum_y p(y|x) \log \frac{p(y|x)}{q(y|x)} =$$

$$= E_{p(x,y)} \log \frac{p(Y|X)}{q(Y|X)}$$

Tétel: $D(p(x,y) \| q(x,y)) = D(p(x) \| q(x)) + D(p(y|x) \| q(y|x))$

Biz: $D(p(x,y) \| q(x,y)) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{q(x,y)} = \sum_{x,y} p(x,y) \log \frac{p(x) p(y|x)}{q(x) q(y|x)} =$

$$= \sum_{x,y} p(x,y) \log \frac{p(x)}{q(x)} + \sum_{x,y} p(x,y) \log \frac{p(y|x)}{q(y|x)} \qquad \square$$

Egyenlőtlenségek

Emlékeztető: Egy fv konvex az (a,b) int-on, ha $\forall x_1, x_2 \in (a,b)$, $\lambda \in [0,1]$-re

$$f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

Konkáv, ha $-f$ konvex. Ha egy fv második deriváltja ≥ 0, akkor konvex, ha ≤ 0, akkor konkáv.

Jensen-egyenlőtlenség: ha f konvex, akkor $E f(X) \geq f(EX)$
ha f konkáv, akkor $E f(X) \leq f(EX)$.

Bizonyítás: 2 pontra: $E f(X) = p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 x_1 + p_2 x_2) = f(EX)$
n-re teljes indukció.

Tétel: Ha $p(x)$ és $q(x)$ két valószínűségi eloszlás, akkor $D(p \| q) \geq 0$, és egyenlőség akkor és csak akkor, ha $p = q$.

Biz: Legyen $A = \{x : p(x) > 0\}$ a tartó halmaz, ekkor:

$$-D(p \| q) = -\sum_{x \in A} p(x) \log \frac{p(x)}{q(x)} = \sum_x p(x) \log \frac{q(x)}{p(x)} \leq \log \sum_x p(x) \frac{q(x)}{p(x)} =$$

log konkávitása miatt

$$= \log \sum_x q(x) = \log 1 = 0.$$

egyenlőség akkor áll fenn, ha $\frac{q(x)}{p(x)} = \text{const} \iff q(x) = p(x)$. $\square$

Következmények: • $I(X;Y) \geq 0$ ("=" $\Leftrightarrow$ X és Y függetlenek)

• $D(p(y|x) \| q(y|x)) \geq 0$ ("=" $p(y|x) = q(y|x)$ $\forall x \in \{x \mid p(x) > 0\}$)

• $I(X;Y|Z) \geq 0$ ("=" $\Leftrightarrow$ X és Y függetlenek Z feltétel esetén)

Tétel: Ha $|\mathcal{X}|$ jelöli X értékkészlete elemeinek a számát, akkor az entrópia maximum

$$H(X) \leq \log |\mathcal{X}|.$$ Egyenlőség akkor és csak akkor, ha $p(X)$ egyenletes.

Biz: Legyen $u(x) = \frac{1}{|\mathcal{X}|}$, és legyen $p(x)$ egy tetszőleges eloszlás!

$$D(p\|u) = \sum_x p(x) \log \frac{p(x)}{u(x)} = \sum_x p(x) \log p(x) - \sum_x p(x) \log u(x) =$$

$$= \log |\mathcal{X}| - H(X).$$

Mivel $D(p\|u) \geq 0 \Rightarrow H(X) \leq \log |\mathcal{X}|$. Egyenlőség csak ha $u = p$. $\square$

Tétel: $H(X|Y) \leq H(X)$ „Az információ nem árthat."

Egyenlőség akkor és csak akkor, ha X és Y függetlenek.

Biz: $0 \leq I(X;Y) = H(X) - H(X|Y)$. $\square$

VIGYÁZAT! Y adott értékeire nem feltétlen csökken H, csak átlagban.

pl.:

$Y \backslash X$	1	2	
1	0	$\frac{3}{4}$	$\frac{3}{4}$
2	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$
	$\frac{1}{8}$	$\frac{7}{8}$	

$H(X) = 0{,}544$

$H(X|Y=1) = 0 < H(X)$

$H(X|Y=2) = 1 > H(X)$

$H(X|Y) = 0{,}25 < H(X)$

Tétel: $H(X_1, \dots X_n) \leq \sum_{i=1}^{n} H(X_i)$. Egyenlőség akkor és csak akkor, ha X_i-k függetlenek.

Biz: $$H(X_1 \dots X_n) = \sum_{i=1}^{n} H(X_i | X_{i-1}, \dots X_1) \leq \sum_{i=1}^{n} H(X_i)$$ $\square$

Tétel: Nemnegatív $a_1 \dots a_n, b_1 \dots b_n$ számokra $$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i\right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i}$$

Egyenlőség akkor és csak akkor, ha $\frac{a_i}{b_i} = \text{const}$.

Biz: Legyen $f(t) := t \log t$! Mivel $f''(t) = \frac{1}{t} \log e > 0$, ezért f konvex pozitív t-re.

Legyen $\alpha_i := \frac{b_i}{\sum_j b_j}$ és $t_i = \frac{a_i}{b_i}$! A vizsgált összefüggés:

$$\sum_i a_i \log \frac{a_i}{b_i} = \sum_j b_j \sum_i \alpha_i t_i \log t_i = \sum_j b_j \sum_i \alpha_i f(t_i) \overset{\text{Jensen miatt}}{\geq} \sum_j b_j f\left(\sum_i \alpha_i t_i\right) =$$

$$= \sum_i a_i \cdot \log \frac{\sum_i a_i}{\sum_i b_i}$$ $\square$

Tétel: $D(p\|q)$ konvex a (p,q) párra, azaz

$$D(\lambda p_1+(1-\lambda)p_2\|\lambda q_1+(1-\lambda)q_2)\le\lambda D(p_1\|q_1)+(1-\lambda)D(p_2\|q_2)\qquad\forall\lambda\in[0,1]$$

Biz: Alkalmazzuk a log-sum egyenlőtlenséget:

$$(\lambda p_1(x)+(1-\lambda)p_2(x))\log\frac{\lambda p_1(x)+(1-\lambda)p_2(x)}{\lambda q_1(x)+(1-\lambda)q_2(x)}\le\lambda p_1(x)\log\frac{\lambda p_1(x)}{\lambda q_1(x)}+(1-\lambda)p_2(x)\log\frac{(1-\lambda)p_2(x)}{(1-\lambda)q_2(x)}$$

x-re szummázva megkapjuk a kívánt egyenlőtlenséget. □

Tétel: $H(X)$ konkáv fv-e $p(x)$-nek.

Biz: Ha $u(x)=\frac{1}{|X|}$, akkor $H(X)=\underset{\text{konstans}}{\log|X|}-\underset{\text{konvex}}{D(p\|u)}\Rightarrow H$ konkáv □

Tétel: $I(X;Y)$ konkáv fv-e $p(x)$-nek fix $p(y|x)$-re és konvex fv-e $p(y|x)$-nek fix $p(x)$-re.

Egyelőre nem bizonyítjuk

Def: Az X,Y,Z változókat egy Markov láncból származónak hívjuk, és $X\to Y\to Z$-vel jelöljük, ha $p(x,y,z)=p(x)p(y|x)p(z|y)$.

(Ezzel ekvivalens, ha $p(z|x,y)=p(z|y)$.)

Tétel (Információ feldolgozási egyenlőtlenség): Ha $X\to Y\to Z$, akkor $I(X;Y)\ge I(X;Z)$.

Biz: Láncszabályok miatt: $I(X;Y,Z)=I(X;Z)+I(X;Y|Z)\ge I(X;Z)$

$$=I(X;Y)+I(X;Z|Y)=I(X;Y)$$

Mivel adott Y-ra X és Z független $I(X;Z|Y)=0$, továbbá $I(X;Y|Z)\ge 0$. □

A célunk általában, hogy egy üzenetet kódoljunk, és a kódból visszanyerjük az információt. A kérdés, hogy ezt milyen eséllyel tudjuk megtenni.

Tétel (Fano-egyenlőtlenség): Legyenek a $X\to Y\to\hat{X}$ egy markov-láncból származó változók, ahol $\hat{X}(Y)$ egy ismert fv (nem feltétlen determinisztikus). A hiba valószínűsége: $P_e=\Pr\{\hat{X}\ne X\}$. Erre igaz az alábbi állítás:

$$H(P_e)+P_e\log|X|\ge H(X|\hat{X})\ge H(X|Y)$$

Biz: Tekintsük az alábbi változót. $E=\begin{cases}1 & \text{ha } \hat{X}\ne X\\ 0 & \text{ha } \hat{X}=X\end{cases}$

Erre az entrópia: $H(E,X|\hat{X})=H(X|\hat{X})+\overbrace{H(E|X,\hat{X})}^{0}=H(X|\hat{X})$.

$$=H(E|\hat{X})+H(X|E,\hat{X})$$

Használjuk fel, hogy $H(E|\hat{X}) \leq H(E) = H(P_e)$.

Továbbá $H(X|E,\hat{X}) = (1-P_e)\cdot H(X|\hat{X}, E=0) + P_e \cdot H(X|\hat{X}, E=1) =$

$$= (1-P_e)\cdot 0 + P_e\, H(X|\hat{X}) \leq P_e \log|X|$$

Ebből tehát $H(P_e) + P_e \log|X| \geq H(X,\hat{X})$

Az információ feldolgozási egyenlőtlenségből: $I(X;\hat{X}) \leq I(X;Y) \Rightarrow H(X|\hat{X}) \geq H(X|Y)$

$\square$

Következmény: Mivel E bináris változó, $H(P_e)$ maximum 1, így a Fano-egyenlőtlenség alsó korlátot ad P_e-re:

$$1 + P_e \log|X| \geq H(X|Y) \Rightarrow P_e \geq \frac{H(X|Y)-1}{\log|X|}$$

BEV. INF. ELM.

3. előadás (02. 26.)

Asszimptotikus ekvipartíciós tulajdonság

Valószínűségből láttuk a nagy számok tv-ét, ami alapján sok kísérletben a gyakoriság a szemábyg valószínűségekhez tartott. Az inf elméletben ennek analógiáját, hogy egy hosszú sorozat hossza $n \cdot H$. Az alábbiakban ezt és néhány következményét vizsgáljuk.

jelölés: i.i.d. = független, azonos eloszlású változók.

Tétel (AEP): Legyenek $X_1, \dots, X_n$ i.i.d változók! Ekkor az együttes eloszlásukra:

$$-\frac{1}{n} \log p(X_1, \dots, X_n) \to H(X)$$

Biz: Mivel X_i-k függetlenek, az együttes eloszlás szorzatként írható:

$$-\frac{1}{n} \log p(X_1, \dots, X_n) = -\frac{1}{n} \sum_{i=1}^{n} \log p(X_i) \longrightarrow \text{ nagy számok tv miatt}$$

$$\longrightarrow -E \log p(X) = H(X)$$

$\square$

Def: $p(x)$ eloszlás esetén az X^n betűszekvenciák tipikus részhalmazának azon megvalósulások halmazát értjük, amikre $2^{-n(H(X)+\epsilon)} \leq p(x_1, \dots, x_n) \leq 2^{-n(H(X)-\epsilon)}$

Adott ϵ esetén a tipikus halmaz jelölése $A_\epsilon^{(n)}$.

Tétel: A tipikus halmaz az alábbi tulajdonságokkal rendelkezik:

1. Ha $(x_1, \dots, x_n) \in A_\epsilon^{(n)}$ akkor $H(X) - \epsilon \leq -\frac{1}{n} \log p(x_1, \dots x_n) \leq H(X) + \epsilon$.
2. $\Pr\{A_\epsilon^{(n)}\} > 1 - \epsilon$ elég nagy n-re.
3. $|A_\epsilon^{(n)}| \leq 2^{n(H(X)+\epsilon)}$
4. $|A_\epsilon^{(n)}| \geq (1-\epsilon) 2^{n(H(X)-\epsilon)}$ elég nagy n-re.

Biz:

1. definíció átrendezéséből triviális.
2. AEP miatt annak valség-e hogy $(X_1, \dots, X_n) \in A_\epsilon^{(n)}$ tart 1-hez $n \to \infty$ esetén, tehát $\forall \delta > 0$-ra $\exists n_0$ úgy, hogy $\forall n > n_0$-ra $\Pr\{A_\epsilon^{(n)}\} > 1 - \delta$.

   $\delta = \epsilon$-ra ez épp az állítás.
3. $$1 = \sum_{\underline{x} \in X^n} p(\underline{x}) \geq \sum_{\underline{x} \in A_\epsilon^{(n)}} p(\underline{x}) \geq \sum_{\underline{x} \in A_\epsilon^{(n)}} 2^{-n(H(X)+\epsilon)} = 2^{-n(H(X)+\epsilon)} \cdot |A_\epsilon^{(n)}|$$
4. Elég nagy n-re a 2.-ből: $1 - \epsilon < \Pr\{A_\epsilon^{(n)}\} \leq \sum_{\underline{x} \in A_\epsilon^{(n)}} 2^{-n(H(X)-\epsilon)} = 2^{-n(H(X)-\epsilon)} |A_\epsilon^{(n)}|$

$\square$

Tétel: Legyen X^n egy i.i.d.-ből álló szekvencia $p(x)$ valószínűséggel és $\epsilon > 0$!
Ekkor létezik olyan leképezés, ami az x^n szekvenciákhoz bináris kódszavakat rendel, úgy, hogy a szavak hosszának várható értéke:

$$E\left[\frac{1}{n}\ell(X^n)\right] \leq H(X) + \epsilon \qquad \text{minden } \epsilon\text{-ra elegendően nagy } n \text{ esetén.}$$

Biz: Bontsuk fel a lehetséges szekvenciák halmazát $A_\epsilon^{(n)}$-re és $A_\epsilon^{(n)C}$-re!
Az $A_\epsilon^{(n)}$-beli elemek leírásához legfeljebb $n(H+\epsilon)+1$ bit kell. ($+1$ azért mert $n(H+\epsilon)$ nem biztos, hogy egész.) Ezen szavak végére még tegyünk egy 0-t, hogy jelenítsük, ez az $A_\epsilon^{(n)}$-beli.
Az $A_\epsilon^{(n)}$-n kívüli szekvenciák kódolásához $n\log|X|+1$ kell. Ezekhez egy 1-t rakjunk hozzá. Így a kódolás szükséges hossz:
$$\ell(x^n) = \begin{cases} n(H+\epsilon)+2 & x^n \in A_\epsilon^{(n)} \\ n\log|X|+2 & x^n \notin A_\epsilon^{(n)} \end{cases}$$

A hossz várható értéke:

$$E(\ell(X^n)) = \sum_{x^n} p(x^n)\ell(x^n) = \sum_{x \in A_\epsilon^{(n)}} p(x^n)\ell(x^n) + \sum_{x \notin A_\epsilon^{(n)}} p(x^n)\ell(x^n) \leq$$

$$\leq \sum_{x \in A_\epsilon^{(n)}} p(x^n)(n(H+\epsilon)+2) + \sum_{x \notin A_\epsilon^{(n)}} p(x^n)(n\log|X|+2) =$$

$$= \Pr\{A_\epsilon^{(n)}\}(n(H+\epsilon)+2) + \Pr\{A_\epsilon^{(n)C}\}(n\log|X|+2) \leq$$

$$\leq n(H+\epsilon) + \epsilon n\log|X| + 2 = n(H+\epsilon')$$

ahol $\epsilon' = \epsilon + \epsilon\log|X| + \frac{2}{n}$, ami tetszőlegesen kicsi lehet megfelelő n esetén. $\square$

Def: Legyen $B_\delta^{(n)} \subset \mathcal{X}^n$ a legkisebb részhalmaza a szekvenciáknak amelynek valószínűsége:

$$\Pr\{B_\delta^{(n)}\} \geq 1-\delta \;!$$

Könnyen látható, hogy $B_\delta^{(n)}$-be a legnagyobb valószínű elemeket kell beválogatni addig, amíg el nem érik a $1-\delta$ határt. Belátható, hogy $B_\delta^{(n)}$ és $A_\epsilon^{(n)}$ nagyjából megegyeznek megfelelő δ és ϵ esetén.

Állítás: $\delta_n \to 0$ és $\epsilon_n \to$ konvergensek esetén $|B_{\delta_n}^{(n)}| \doteq |A_{\epsilon_n}^{(n)}| \doteq 2^{nH}$, ahol

$$a_n \doteq b_n \text{ jelentése } \lim_{n\to\infty} \frac{1}{n}\log\frac{a_n}{b_n} = 0.$$

Szemléletesen: $A_{\epsilon_n}^n$ és $B_{\delta_n}^n$ elemszáma ugyanúgy viselkedik az exponens első rendjében.

Sztochasztikus folyamatok entrópiarátája

A valóságban általában nem i.i.d. folyamatok vannak, így általánosabb esetet is vizsgálni kell. A következőkben látni fogjuk, hogy stacionárius folyamatok entrópiája is lineárisan nő, ahogy az i.i.d. is.

Def: Egy sztochasztikus folyamat stacionárius, ha az együttes eloszlás invariáns az indexek eltolására: $Pr\{X_1=x_1,\dots,X_n=x_n\} = Pr\{X_{\ell+1}=x_1,\dots,X_{\ell+n}=x_1\}$

$\forall n, \ell, x_1, x_2, \dots x_n$ -re.

Def: Egy diszkrét sztochasztikus folyamatot Markov-láncnak hívunk, ha egy adott elem feltételes valószínűsége csak az előző elemtől függ:

$$Pr(X_{n+1}=x_{n+1} \mid X_n=x_n, \dots X_1=x_1) = Pr(X_{n+1}=x_{n+1} \mid X_n=x_n)$$

$\forall x_1, x_2, \dots x_{n+1}$ -re.

Def: Egy Markov-lánc idő invariáns, ha $p(x_{n+1}|x_n)$ független n-től.

Egy időinvariáns Markov-lánc az átmeneti mátrixával jellemezhető: $P \in \mathbb{R}^{|X| \times |X|}$
$P_{ij} = Pr\{X_{n+1}=x_j \mid X_n=x_i\}$. Ekkor, ha az n. változó eloszlása $p(x_n)$, akkor az $n+1$.-é: $p(x_{n+1}) = \sum_{x_n} p(x_n) P_{x_n x_{n+1}}$. Ha $p(x_{n+1}) = p(x_n)$ $\forall x_n$-re akkor azt stacionárius eloszlásnak hívjuk. Ha a kezdeti állapot valószínűsége a stac. eloszlást követi, akkor a folyamat stacionárius.

Ha a Markov-láncban pozitív valószínűséggel el lehet jutni bármely állapotból bármelyikbe véges lépésen belül, a folyamat irreducibilis. Ha az állapotokból önmagukba vezető utak hosszainak legnagyobb közös tényezője 1, a folyamat aperiodikus. Ha egy véges állapottal rendelkező Markov-lánc irreducibilis és aperiodikus, akkor egyértelműen létezik egy stac. eloszlás és bármely kezdeti eloszlás esetén a lánc ehhez tart $n \to \infty$ esetén.

Def: Egy $\{X_i\}$ sztochasztikus folyamat entrópiarátája az alábbi limesz, ha létezik:

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

Definiálhatjuk az alábbi mennyiséget is: $H'(\mathcal{X}) = \lim_{n\to\infty} H(X_n \mid X_{n-1},\dots X_n)$

Tétel: Stac. stoch. folyamatra $H(X)$ és $H'(X)$ is létezik és egyenlőek.

Biz: Feltételes entrópia tulajdonságai miatt:

$$H(X_{n+1}|X_1,\dots,X_n) \le H(X_{n+1}|X_2,\dots,X_n) = H(X_n|X_1,\dots,X_{n-1}),$$

tehát $H(X_n|X_{n-1},\dots X_1)$ egy csökkenő pozitív sorozat, tehát konvergens.

Beláthatő, hogy ha $a_n \to a$ és $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$, akkor $b_n \to a$ (Cesàro-átlag).

Láncszabály miatt:

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

A jobb oldali tagoknak van határértéke a stac. miatt a bal oldal vele egyenlő ezzel a Cesàro-átlag miatt $\Rightarrow$ $H(X) = H'(X)$ $\square$

i.i.d szekvencia entrópiarátája: $H(X) = H(X_1)$

Markov-folyamat entrópiarátája: $H(X) = H(X_2|X_1) = -\sum_{i,j} \mu_i P_{ij} \log P_{ij}$.

Egyéb nem speciális esetben nem biztos, hogy létezik.

Termodinamika II. főtétele: Ezt nem írom le, de a könyv 4.4. fejezete részletesen leírja.

BEV INF ELM

4. előadás (03.05.)

Adattömörítés

Def: Egy X valószínűségi változóhoz tartozó kód egy leképezés $\mathcal{X}$-ről D^*-ra, ahol $\mathcal{X}$ az X értékkészlete, D^* pedig véges hosszúságú szavak a D-betűt tartalmazó ábécéből.

jelölések: kód: $C : \mathcal{X} \to D^*$

x-et kódoló szó: $C(x)$

$\ell(x)$: Az x-hez tartozó szó hossza.

Def: Egy kód várható hossza az $\ell(x)$ várható értéke:

$$L(C) = \sum_{x \in \mathcal{X}} p(x)\, \ell(x)$$

Def: Egy kód nemszinguláris, ha $\mathcal{X}$ tetszőleges értékeihez különböző szót rendel:

$$x \neq x' \Rightarrow C(x) \neq C(x')$$

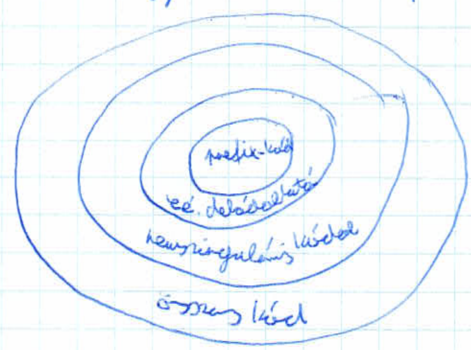
Def: Egy C kód kiterjesztése véges hosszúságú $\mathcal{X}$-stringekre az alábbi:

$$C(x_1 x_2 \cdots x_n) = C(x_1) C(x_2) \cdots C(x_n).$$

Def: Egy kód egyértelműen dekódolható, ha a kiterjesztése nemszinguláris.

Def: Egy kód prefix-kód vagy azonnali kód, ha egyik kódszó sem előtagja egy másiknak.

Szemléltetés:



pl.:

X	szinguláris	nemszinguláris, de nem eé. dekódolható	eé. dekódolható de nem prefix	prefix-kód
1	0	0	10	0
2	0	010	00	10
3	0	01	11	110
4	0	10	110	111

Egy hatékony kódoláshoz az egyértelmű dekódolhatóság a min, hogy ne kelljen külön elválasztó jel, de ha a kód nem prefix-kód, akkor meg kell várni az egészet.

Cél: A lehető legkisebb várható hosszúsággal rendelkező prefix-kód megtalálása.

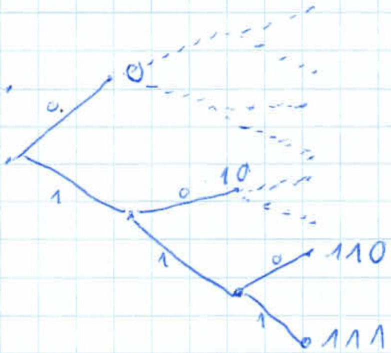
Tétel (Kraft-egyenlőtlenség): Minden prefix-kód egy D ábécé felett teljesíti az alábbit:

$$\sum_i D^{-l_i} \le 1$$

ahol l_i az i-ik kódszó.

Megfordítása: Ha adott a kódszóhosszúságoknak egy halmaza, ami teljesíti az egyenlőtlenséget, akkor létezik egy prefix-kód a megadott hosszúságú kódszavakkal.

Biz: Képzeljünk el egy D-áris fát, ahol az ágak reprezentálják a karaktereket és a csúcsok a szavakat! pl.: D=2-re:



A prefix-kód feltétele, hogy ha egy szó érvényes kódszó, akkor a hozzá tartozó csúcs leszármazottai nem azok.

Legyen a leghosszabb szó hossza l_{max}! Az l_{max} szinten a D-áris fának $D^{l_{max}}$ darab levele van, de ezek közül néhány kódszó, néhány nem. Egy l_i-ik szinten lévő kódszónak $D^{l_{max}-l_i}$ leszármazotta van az l_{max}-ik szinten, és ezek a halmazok diszjunktak, tehát az összes lehetséges kódszó leszármazott nem lehet többen, mint az összes levél:
$$\sum_i D^{l_{max}-l_i} \le D^{l_{max}}$$

$$\Rightarrow \sum_i D^{-l_i} \le 1$$

Megfordítás: Ha adott szóhosszak teljesítik az egyenlőtlenséget, akkor tudunk konstruálni egy megfelelő D-áris fát: Az 1. szó legyen az első l_1 hosszú lehetőség. Ez kizárja az összes leszármazottját és felmenőit. A 2. szó a maradék közül az első l_2-t, stb...

□

Tétel (kiterjesztett kraft-egyenlőtlenség): U. a. mint a kraft-egyenlőtlenség, csak megszámlálhatóan végtelen sok kódszóra:

$$\sum_{i=1}^{\infty} D^{-l_i} \leq 1 .$$

Biz: Legyen az ábécé a $\{0, 1, \dots, D-1\}$ és az i-ik kódszó $y_1 y_2 \cdots y_{l_i}$!

A $0,y_1 y_2 \cdots y_{l_i}$ egy valós szám a D-s számrendszerben:

$$0,y_1 y_2 \cdots y_{l_i} = \sum_{j=1}^{l_i} y_j D^{-j} .$$

Ez egyébként benne van a $\left[0,y_1 y_2 \cdots y_{l_i} ,\ 0,y_1 y_2 \cdots y_{l_i} + \frac{1}{D^{l_i}}\right)$ intervallumban, ahol a $0,y_1 y_2 \cdots y_{l_i}$-vel kezdődő számok vannak. A prefix-kód feltétele, hogy ezek az intervallumok diszjunktak, tehát az összegük nem lehet nagyobb, mint a $[0,1]$ intervallumé: $\sum_{i=1}^{\infty} D^{-l_i} \leq 1$.

Megfordítás: rendezzük át a sorrendet, hogy $l_1 \leq l_2 \leq \cdots$, és az adott hosszak sorrendjében kössük a $[0,1]$ intervallum egymás utáni $\frac{1}{D^{l_i}}$ hosszúságú intervallumait. pl. ha $l_1 = 1, l_2 = 2 \dots$ akkor $[0, \frac{1}{2})$, $[\frac{1}{2}, \frac{1}{4})$, $\dots$.

$\square$

Most van egy elégséges feltételünk a prefix kódokra, de jó lenne megtalálni a legjobb legrövidebb prefix kódot. Ez egyébként egy szimpla optimalizációs probléma:

optimalizálandó kifejezés: $L = \sum_i p_i l_i$

feltétel: $\sum_i D^{-l_i} \leq 1$.

Lagrange-multiplikátort használva: $J = \sum_i p_i l_i + \lambda \left(\sum_i D^{-l_i} \right)$

$$\frac{\partial J}{\partial l_i} = p_i - \lambda D^{-l_i} \ln D \overset{!}{=} 0 \quad \Rightarrow \quad D^{-l_i^*} = \frac{p_i}{\lambda \ln D}$$

A feltételbe visszahelyettesítve: $\lambda \geq (\ln D)^{-1} \quad \Rightarrow \quad D^{-l_i^*} \leq p_i \quad \Rightarrow$

$$\Rightarrow \quad l_i^* \geq -\log_D p_i \quad \Rightarrow \quad L^* = \sum_i p_i l_i \geq \sum_i p_i \log_D p_i = H_D(X)$$

Tehát az entrópia egy alsó korlát a várható hosszra, és egyenlőség csak akkor lehet, ha a megfelelő $\log_D p_i$ értékek egész számok.

Ez precízen is belátható.

Tétel: Egy kód L várható hossza bármely D-áris prefix-kódra és egy X random változóra:

$$L \geq H_D(X).$$

Egyenlőség akkor van, ha $D^{-l_i} = p_i$.

Biz:

$$L - H_D(X) = \sum_i p_i l_i - \sum_i p_i \log_D \frac{1}{p_i} = -\sum_i p_i \log_D D^{-l_i} + \sum_i p_i \log_D p_i = *$$

bevezetve a $r_i = \dfrac{D^{-l_i}}{\sum_j D^{-l_j}}$ és $c = \sum_i D^{-l_i}$ értékeket:

$$* = \sum_i p_i \log_D \frac{p_i}{r_i} - \log_D c = D_D(p \| r) + \log_D \frac{1}{c} \geq 0$$

Az első tag a relatív entrópia tulajdonságai miatt nemnegatív, a második pedig a Kraft-egyenlőtlenség miatt $c \leq 1$. □

Def: Egy eloszlás D-adikus, ha a valószínűségek D^{-n} alakban írhatók $n \in \mathbb{N}$-re.

BEV INF ELM

5. előadás (03.12.)

Láttuk: hogy a Kraft-egyenlőtlenség akkor egyenlő, ha az eloszlás D-adikus, és az átlagos kódszóhosszúság is akkor a minimális. De ennek megfelelően nem D-adikus esetekben is a legjobb esetet.

A minimális kódszóhossz van definiálva.

~~Def.: Egy $\{p_1, \dots p_n\}$ eloszláshoz tartozó C prefix kód optimális, ha a hozzá tartozó $\{l_1, \dots l_n\}$ kódszó hosszúságokból képzett $r_i = \frac{D^{-l_i}}{\sum_j D^{-l_j}}$ eloszlás relatív entrópiája minimális a p-eloszlásra nézve.~~

~~Azaz $\forall C'$ kódhoz tartozó r' eloszlásra $D_D(p\|r) \leq D_D(p\|r')$.~~

Tétel: Legyen a p eloszláshoz tartozó egy optimális kód szóhosszai rendre $\{l_1^*, \dots, l_m^*\}$ a várható hosszúság pedig $L^* = \sum_i p_i l_i^*$.
Ekkor

$$H_D(X) \leq L^* < H_D(X) + 1.$$

Bz: $H_D(X) \leq L^*$ előző órán látottak miatt.

Legyen egy példa kód olyan, hogy a kódszavak hossza $l_i = \lceil \log_D \frac{1}{p_i} \rceil$ alakú.

$\lceil a \rceil$ az a felső egész része! Ez kielégíti a Kraft egyenlőtlenséget, mert

$$\sum_i D^{-\lceil \log_D \frac{1}{p_i} \rceil} \leq \sum_i D^{-\log_D \frac{1}{p_i}} = \sum_i p_i = 1 \implies \text{Létezik prefix kód } \{l_i\} \text{ hosszakkal.}$$

$\lceil\,\rceil$ def-je szerint $\log_D \frac{1}{p_i} \leq l_i < \log_D \frac{1}{p_i} + 1$

Beszorozva p_i-vel és szummázva i-re: $H_D(X) \leq L < H_D(X) + 1$

Mivel L^* optimális, ezért $L^* \leq L < H_D + 1$. □

Sok honnan jöttek?

A hiba tovább csökkenthető, ha több karaktert rakunk egy blokkba, és ezeket kódoljuk, mint a $\mathcal{X}^n$-ből származó szavakat. Ha $l(x_1, \dots, x_n)$-nel jelöljük az $(x_1, \dots, x_n)$ szekvenciát kódoló szó hosszát, akkor a karakterenkénti várható hossz:

$$L_n = \frac{1}{n} \sum_{\underline{x} \in \mathcal{X}} p(x_1, \dots, x_n)\, l(x_1, \dots, x_n).$$

Erre alkalmazva az előbbi tételt: $H(X_1, \dots, X_n) \leq nL_n < H(X_1, \dots, X_n) + 1.$

iid változók esetén $H(X_1, \dots, X_n) = \sum H(X_i) = nH(X) \implies H(X) \leq L_n < H(X) + \frac{1}{n}$

vagyis nagy n-re $L_n \to H(X)$. Sztochasztikus esetben szintén $L_n \to H(\mathcal{X})$.

precízen kimondva:

Tétel: A minimális kavalteraenkénti kódszóhossz:

$$\frac{H(X_1,\dots,X_n)}{n} \le L_n^* < \frac{H(X_1,\dots,X_n)}{n} + \frac{1}{n}$$

Ha $X_1,\dots,X_n$ sztochasztikus folyamat, akkor $L_n^* \to H(X)$ ahol $H(X)$ az entrópiaráta.

Tétel (Hibás kód): Ha egy p eloszlású változót úgy kódolunk, mintha q eloszlású lenne, akkor a várható hossz

$$H(p) + D(p\|q) \le L < H(p) + D(p\|q) + 1.$$

Biz: $L = \sum_{x\in X} p(x) \ell(x) < \sum_{x\in X} p(x)\left(\log\frac{1}{q(x)} + 1\right) = \sum_{x\in X} p(x)\left[\log\left(\frac{p(x)}{q(x)}\frac{1}{p(x)}\right) + 1\right] =$

$$= \sum_{x\in X} p(x) \log\frac{p(x)}{q(x)} + \sum_{x\in X} p(x)\log\frac{1}{p(x)} + 1 = D(p\|q) + H(p) + 1$$

alsó korlát ugyanúgy $\square$

Tétel (McMillan): Minden egyértelműen dekódolható kód szóhosszai szintén teljesíti a kraft-egyenlőtlenséget: $\sum_i D^{-\ell_i} \le 1$

Megfordítva, ha kódszóhosszaknak egy halmaza teljesíti a kraft-egyenlőtlenséget, akkor létezik egy egyértelműen dekódolható kód.

Biz: A megfordítás triviális, hiszen a prefix kódra igaz az állítás, ami egyben dekódolható.

Legyen C^k a C kód k-ik kiterjesztése! Ekkor a kódhossz:

$\ell(x_1,x_2,\dots,x_n) = \sum_{i=1}^{k} \ell(x_i)$. Tekintsük az alábbit:

$$\left(\sum_{x\in X} D^{-\ell(x)}\right)^k = \sum_{x_1\in X}\cdots\sum_{x_k\in X} D^{-\ell(x_1)}\cdots D^{-\ell(x_k)} = \sum_{\underline{x}\in X^k} D^{-\ell(\underline{x})} = \sum_{m=1}^{k\ell_{max}} a(m) D^{-m}$$

ahol $a(m)$ az m hosszú kódszavak száma C^k-ban. Mivel egyértelműen dekódolható, ezért $a(m) \le D^m$, tehát

$$\left(\sum_{x\in X} D^{-\ell(x)}\right)^k \le \sum_{m=1}^{k\ell_{max}} D^m D^{-m} = k\ell_{max} \Rightarrow \sum_i D^{-\ell_i} \le (k\ell_{max})^{-1/k}$$

Mivel ez minden k-ra igaz, nézzük a $k\to\infty$ határesetet: $(k\ell_{max})^{-1/k} \to 1$

$$\Longrightarrow \sum_i D^{-\ell_i} \le 1$$

$\square$

Következmény: Az egyértelműen dekódolható kódok végtelen darab szó esetén is teljesítik a Kraft-egyenlőtlenséget.

Biz: A probléma az előző tétellel, hogy végtelen $|X|$ esetén l_{max} is lehet, hogy végtelen és a Σ nem biztos, hogy véges.

DE! A kód minden részhalmaza egyértelműen dekódolható, így vegyünk egy véges részhalmazt, és tartsunk a ∞-hez!

$$\sum_{i=1}^{\infty} D^{-l_i} = \lim_{N\to\infty} \sum_{i=1}^{N} D^{-l_i} \le 1$$

□

Huffman-kód

Jó lenne találni egy konkrét konstrukciót, ami megtalál egy optimális kódot. A Huffman kód pont ilyen.

Konstrukció: 1) rendezzük csökkenő sorrendbe a valószínűségek szerint a szavakat!

2) Adjuk össze a legkisebb két értéket! A hozzájuk tartozó szavakat tekintsük egynek ezentúl

3) Ismételjük a 1-2 lépéseket, amíg csak egy szó marad!

4) Egy adott szót kódoló kódszót visszafele haladva, minden elágazásnál tegyünk egy 0-t vagy 1-t!

+) D-áris kód esetén a 2. lépésben a legalsó D értéket kell összeadni. Ez azt jelenti, hogy minden lépésben D-1-gyel csökken a szavak száma, tehát kezdetben $1+k(D-1)$ db szónak kell lennie valamilyen k-ra. Ha ez nem teljesül, adjunk hozzá megfelelő mennyiségű szót, amelyek valószínűsége 0!

Példák: 1)

X	valószínűségek					kódszó
1	0,25	0,3	0,45	0,55	1	01
2	0,25	0,25	0,3	0,5		10
3	0,2	0,25	0,25			11
4	0,15	0,2				000
5	0,15					001

$H(X) = 2,29$

$L = 2,3$

2)

X	valószínűség			kódszó
1	0,25	0,5	1	1
2	0,25	0,25		2
3	0,2	0,25		00
4	0,15			01
5	0,15			02

$H_3(X) = 1{,}44$

$L = 1{,}75$

3)

X	valószínűség				kódszó
1	0,25	0,25	0,5	1	1
2	0,25	0,25	0,25		2
3	0,2	0,2	0,25		01
4	0,1	0,2			02
5	0,1	0,1			000
6	0,1				001
üres	0				002

$H_3(X) = 1{,}55$

$L = 1{,}7$

Vegyük észre, hogy a Huffman-kód nem egyértelmű még ekvivalencia szintjén sem!

4)

X	valószínűség				kódszó
1	1/3	1/3	2/3	1	1
2	1/3	1/3	1/3		00
3	1/4	1/3			010
4	1/12				011
1	1/3	1/3	2/3	1	00
2	1/3	1/3	1/3		10
3	1/4	1/3			01
4	1/12				11

$L_1 = 2$

$L_2 = 2$

$H(X) = 1{,}86$

A két kód még az átcserélgetés szintjén sem egyezik, de a relatív entrópia és a kódhossz igen.

Lemma: Minden eloszláshoz létezik egy optimális prefix kód a alábbi tulajdonságokkal:

1) Ha $p_j > p_k$ akkor $\ell_j < \ell_k$

2) A két leghosszabb szó hossza egyenlő

3) A két leghosszabb szó csak az utolsó bitjükben különbözik.

Biz: 1) Legyen C_m és C_m' két kód, amikben a j és k szó fel van cserélve! TFH $L(C_m') \geq L(C_m)$! Ekkor

$$0 \leq L(C_m') - L(C_m) = \sum_i p_i \ell_i' - \sum_i p_i \ell_i = p_j \ell_k + p_k \ell_j - p_j \ell_j - p_k \ell_k =$$

$$= (p_j - p_k)(\ell_k - \ell_j).$$

Mivel $p_j - p_k > 0 \Rightarrow \ell_k > \ell_j$.

2) Ha a két leghosszabb szó hossza nem egyezik, akkor kitörölve a leghosszabb szó utolsó karakterét továbbra is prefix kód marad.

3) Ha egy maximális hosszúságú szóban van testvér olyan, ami csak az utolsó karakterben különbözik, akkor kitörölve az utolsó karaktert továbbra is prefix marad. $\square$

Azon kódokat, amik teljesítik a lemma tulajdonságait kanonikus kódoknak hívjuk.

Tétel: A bináris Huffman-kód optimális.

Biz: Legyen $p = (p_1, \dots, p_m)$ egy eloszlás, ahol $p_1 \geq \dots \geq p_m$, a Huffman-redukciója pedig $p' = (p_1, p_2, \dots, p_{m-2}, p_{m-1} + p_m)$! Legyen $C^*_{m-1}(p')$ egy optimális kód p'-re, $C^*_m(p)$ pedig egy kanonikus kód p-re!

A C^*_{m-1} kiterjeszthető p-re az alábbi módon: A $p_1, \dots, p_{m-2}$ valószínűségekhez tartozó kódszavak egyeznek n. a., a $p_{m-1} + p_m$-hez tartozó szó végéhez pedig egy 0-t és egy 1-t téve rendeljük hozzá p_{m-1}-hez és p_m-hez.

Ekkor a várható kódhossz: $L(p) = L^*(p') + p_{m-1} + p_m$.

Mivel C^*_m kanonikus konstruálható belőle egy kód p'-re, egyszerűen úgy, hogy a p_{m-1} és p_m-hez tartozó szavak utolsó karakterét elhagyjuk. Erre a kódhossz:

$L(p') = L^*(p) - p_{m-1} - p_m$.

A kettőt összeadva: $L(p') + L(p) = L^*(p') + L^*(p) \Rightarrow (L(p) - L^*(p)) + (L(p') - L^*(p')) = 0$.

Mindkét tag nem negatív $\Rightarrow L(p) - L^*(p) = 0$.

Ha $m-1$-re a Huffman kód optimális, akkor m-re is, innen teljes indukció. $\square$

BEV INF ELM

6. előadás (03.19.)

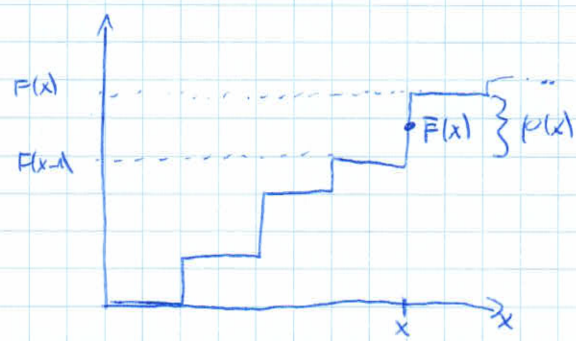
Shannon-Fano-Elias-kódolás

Láttuk, hogy a Huffman-kód optimális, de a valószínűség szerint kell rendezni hozzá a kódolandó szavakat. A SFE-kód nem lesz optimális, de a kódszavak és az eredeti szavak sorrendje megegyezik.

Legyen a kódolandó ábc: $X = \{1, 2, \dots, m\}$ adott sorrenddel, és a hozzájuk tartozó valószínűségek: $p(x) > 0$ $\forall x$-re. A kumulatív eloszlás fv:

$$F(x) = \sum_{a \leq x} p(a).$$

A módosított kumulatív eloszlásfv.: $\bar{F}(x) = \sum_{a < x} p(x) + \frac{1}{2} p(x)$.



Mivel a valószínűségek pozitívak, ezért $a \neq b \iff \bar{F}(a) \neq \bar{F}(b)$, tehát $\bar{F}(x)$-szel lehetne kódolni x-et. A gond, hogy $\bar{F}(x)$ valós szám, vagyis végtelen bitű.

Ötlet: Láttuk, hogy az $\ell(x) = \lceil \log \frac{1}{p(x)} \rceil$ kódhosszak kielégítik a Kraft-egyenlőtlenséget, ezért vágjuk le $\bar{F}(x)$-et $\ell(x)$ bitre és kódoljuk az x-et! Jelöljük ezt $\lfloor \bar{F}(x) \rfloor_{\ell(x)}$.

Tétel: $\bar{F}(x)$ becslése a két lépcső közé esik, vagyis $F(x-1) < \lfloor \bar{F}(x) \rfloor_{\ell(x)} < F(x)$

Biz: A második rész triviális, hiszen $\lfloor \bar{F}(x) \rfloor_{\ell(x)} \leq \bar{F}(x)$.

Az elsőhöz: $\bar{F}(x) - \lfloor \bar{F}(x) \rfloor_{\ell(x)} < 2^{-\ell(x)}$ (bináris számrendszerben).

Mivel $\ell(x) = \lceil \log \frac{1}{p(x)} \rceil + 1$, ezért

$$2^{-\ell(x)} < \frac{p(x)}{2} = \bar{F}(x) - F(x-1).$$

A kettő együtt: $\bar{F}(x) - \lfloor \bar{F}(x) \rfloor_{\ell(x)} < \bar{F}(x) - F(x-1) \Rightarrow F(x-1) < \lfloor \bar{F}(x) \rfloor_{\ell(x)}$ $\square$

Tétel: A SFE-kód prefix-kód.

Biz: Ennek bizonyításához rendeljük hozzá a $z_1 z_2 \cdots z_e$ kódszóhoz a $[0{,}z_1 z_2 \cdots z_e,\ 0{,}z_1 z_2 \cdots z_e + \frac{1}{2^e})$ intervallumot és lássuk be, hogy ezek diszjunktak!

Mivel $\frac{1}{2^{e(x)}} < \frac{p(x)}{2}$ ezért az $\frac{1}{2^e}$ rövidebb, mint a lépcső fele, tehát $\lfloor \bar{F}(x) \rfloor_{e(x)}$-ben korrigálva nem éri el a következő lépcsőt, így nem lóg bele a következő intervallumba. $\square$

Példa:

x	$p(x)$	$F(x)$	$\bar{F}(x)$	bináris	$e(x)$	kódszó
1	0,25	0,25	0,125	0,001	3	001
2	0,25	0,5	0,375	0,011	3	011
3	0,2	0,7	0,6	$0{,}1\overline{0011}$	4	1001
4	0,15	0,85	0,775	$0{,}110001\overline{1}$	4	1100
5	0,15	1,0	0,925	$0{,}111\overline{0110}$	4	1110

$L = 3{,}5 \qquad H(X) = 2{,}29$

Ahogy korábban is láttuk, az átlagos kódhossz nem optimális, de igaz rá, hogy $L < H(X) + 2$

Csatornakapacitás

A kommunikáció célja, hogy üzenetet küldjünk, és a vevő fél elég nagy valószínűséggel ugyanazt kapja meg. A kommunikációs rendszer vázlatos képe az ábrán, ezt szeretnénk matematikailag leírni.

$$W \rightarrow \boxed{\text{kódoló}} \xrightarrow{X^n} \boxed{\begin{matrix}\text{csatorna}\\ p(y|x)\end{matrix}} \xrightarrow{Y^n} \boxed{\text{dekoder}} \rightarrow \hat{W}$$

Def: Egy diszkrét csatornának nevezett rendszer tartalmaz egy $\mathcal{X}$ bemeneti ábécét, egy $\mathcal{Y}$ kimeneti ábécét és egy $p(y|x)$ átviteli mátrixot, ami megadja, hogy x elküldött jel esetén mekkora valséggel kapunk y-t.

A csatorna memóriamentes, ha $p(y|x)$ nem függ az előző karakterektől.

Def: Egy diszkrét memóriamentes csatorna információs csatornakapacitása
$$C := \max_{p(x)} I(X;Y)$$
ahol $p(x)$ a lehetséges bemeneti eloszlások.

[Példák]

Def: Egy csatorna szimmetrikus, ha az átmeneti mátrix sorai permutációi egymásnak, és ugyanígy az oszlopok.

pl.: $p(y|x) = \begin{pmatrix} 0,3 & 0,2 & 0,5 \\ 0,5 & 0,3 & 0,2 \\ 0,2 & 0,5 & 0,3 \end{pmatrix}$ ($p(y|x)$ az x-ik sor, y-ik oszlop)

Tétel: Szimmetrikus csatorna információs csatornakapacitása $C = \log|Y| - H(\underline{r})$, ahol $\underline{r}$ az átmeneti mátrix egyik sora.

Biz: $I(X;Y) = H(Y) - H(Y|X)$ = mivel a sorokban a valószínűségek ugyanazok
$= H(Y) - H(\underline{r}) \leq \log|Y| - H(\underline{r})$

ahol egyenlőség akkor van, ha Y egyenletes. Ez teljesül egyenletes X esetén hiszen

$$p(y) = \sum_{x \in X} p(y|x)\, p(x) = \frac{1}{|X|} \sum_{x \in X} p(y|x) = c\frac{1}{|X|} = \frac{1}{|Y|}$$

ahol c az oszlopok összege (ami fix). □

A fenti példa esetén $C = \log 3 - H(0,5;0,3;0,2) = 0,0995$

Def: Egy csatorna gyengén szimmetrikus, ha az átmeneti mátrix sorai permutációi egymásnak, és az oszlopok összege fix.

pl.: $p(y|x) = \begin{pmatrix} \frac{1}{3} & \frac{1}{6} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & \frac{1}{6} \end{pmatrix}$

Az előző tétel erre is igaz.

Tétel: Az információs csatornakapacitás az alábbiakat teljesíti:

(i) $C \geq 0$ [Mivel $I(X;Y) \geq 0$]

(ii) $C \leq \log|X|$ [hiszen $C = \max I(X;Y) \leq \max H(X) = \log|X|$]

(iii) $C \leq \log|Y|$ [ugyanezért]

(iv) Mivel $I(X;Y)$ folytonos konkáv fv-e $p(x)$-nek, ezért a lokális maximum globális maximum is. Továbbá (ii) és (iii) miatt ez véges érték, amit $I(X;Y)$ fel is vesz, tehát C értéke a szokásos optimalizálási eljárásokkal (pl.: gradiens módszer) megtalálható.

Azt szeretnénk bizonyítani, hogy C az a szám, ahány különböző n hosszúságú üzenetet küldhetünk hiba nélkül, előtte azonban ezt formalizálni kell néhány definícióval.

Def: Az $(\mathcal{X}, p(y|x), \mathcal{Y})$ diszkrét csatorna n-ik kiterjesztése a $(\mathcal{X}^n, p(y^n|x^n), \mathcal{Y})$

ahol $p(y_k|x^k, y^{k-1}) = p(y_k|x_k) \quad \forall k = 1, 2, \dots n$-re.

megjegyzés: Ha a csatornában nincs visszacsatolás (és általában ilyeneket nézünk), akkor $p(y^n|x^n) = \prod_{i=1}^{n} p(y_i|x_i)$

Def: Az $(\mathcal{X}, p(y|x), \mathcal{Y})$ csatornához tartozó (M, n) kódhoz az alábbiak szükségesek:

1. Egy bemeneti index-halmaz: $\{1, 2, \dots, M\}$ (ebből veszünk W-t)
2. Egy kódoló fv: $X^n: \{1, 2, \dots, M\} \Rightarrow \mathcal{X}^n$ „kódkönyv"
3. Egy dekódoló fv.: $g: \mathcal{Y}^n \to \{1, 2, \dots, M\}$

Def: Adott i indexhez tartozó feltételes hiba valószínűség alatt az alábbit értjük:

$$\lambda_i = \Pr\left(g(Y^n) \neq i \mid X^n = x^n(i)\right)$$

Def: Az (M, n) kódhoz tartozó maximális hiba valószínűség: $\lambda^{(n)} = \max_i \lambda_i$

Def: Az (M, n) kódhoz tartozó átlagos hiba valószínűség $P_e^{(n)} = \frac{1}{M} \sum_{i=1}^{M} \lambda_i$

Def: Az (M, n) kód rátája $R = \frac{\log M}{n}$ bit per átvitel.

Def: Az R rátát elérhetőnek nevezzük, ha létezik olyan kód $(2^{\lceil nR \rceil}, n)$ kóddal, hogy a $\lambda^{(n)} \to 0$ ha $n \to \infty$.

Def: Egy csatorna kapacitása az elérhető ráták szuprémuma.

A csatorna kapacitás tételével belátjuk, hogy az megegyezik az információs kapacitással

BEVINFELM

7. előadás (03.26.)

Def: Az $A_\varepsilon^{(n)}$ az együttes tipikus szekvenciák halmaza. Elemei olyan $\{(x^n, y^n)\}$ szekvenciapárok, amelyek külön külön és együtt is tipikusak, azaz

$$A_\epsilon^{(n)} = \Big\{ (x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n : \left| -\frac{1}{n} \log p(x^n) - H(X) \right| < \epsilon \text{ és }$$
$$\left| -\frac{1}{n} \log p(y^n) - H(X) \right| < \epsilon \text{ és }$$
$$\left| -\frac{1}{n} \log p(x^n, y^n) - H(X,Y) \right| < \epsilon \Big\}$$

Tétel (együttes AEP): Legyen (X^n, Y^n) n db i.i.d-ből álló pár szekvenciája, azaz $p(x^n, y^n) = \prod_{i=1}^{n} p(x_i, y_i)$. Ekkor teljesülnek az alábbiak:

1. $\Pr\{(X^n, Y^n) \in A_\epsilon^{(n)}\} \to 1$ ha $n \to \infty$

2. $|A_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+\epsilon)}$

át kell írni, mert így
$I(X;Y)=0$

3. Ha $(\tilde{X}^n, \tilde{Y}^n) \sim p(x^n)p(y^n)$ (azaz, ha $\tilde{X}^n$ és $\tilde{Y}^n$ függetlenek)

$$\Pr\{(\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\} \leq 2^{-n(I(X;Y)-3\epsilon)}$$

$$\Pr\{(\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\} \geq (1-\epsilon)\, 2^{-n(I(X;Y)+3\epsilon)}$$ elég nagy n-re.

Biz: 1. A nagy számok tv.-e miatt elég nagy n-re külön-külön teljesülnek az alábbiak:

$$\Pr\left\{ \left| -\frac{1}{n} \log p(X^n) - H(X) \right| \geq \epsilon \right\} < \frac{\epsilon}{3}$$

$$\Pr\left\{ \left| -\frac{1}{n} \log p(Y^n) - H(Y) \right| \geq \epsilon \right\} < \frac{\epsilon}{3}$$

$$\Pr\left\{ \left| -\frac{1}{n} \log p(X^n, Y^n) - H(X,Y) \right| \geq \epsilon \right\} < \frac{\epsilon}{3}$$

Válasszunk olyan nagy n-et, hogy mind teljesüljön, ekkor a, hogy az unió teljesüljön $< \epsilon$ valószínű.

2. $$1 = \sum_{(x^n, y^n)} p(x^n, y^n) \geq \sum_{A_\epsilon^{(n)}} p(x^n, y^n) \geq |A_\epsilon^{(n)}| \cdot 2^{-n(H(X,Y)+\epsilon)}$$

$$\Rightarrow |A_\epsilon^{(n)}| \leq 2^{n(H(X,Y)+\epsilon)}$$

3. $$\Pr\{(\tilde{X}^n, \tilde{Y}^n) \in A_\epsilon^{(n)}\} = \sum_{(x^n, y^n) \in A_\epsilon^{(n)}} p(x^n)\, p(y^n) \leq 2^{n(H(X,Y)+\epsilon)} \cdot 2^{-n(H(X)-\epsilon)} \cdot 2^{-n(H(Y)-\epsilon)} =$$
$$= 2^{-n(I(X;Y)-3\epsilon)}$$

Mivel elég nagy n-re: $P(A_\epsilon^{(n)}) \geq 1-\epsilon$, ezért

$$1-\epsilon \le \sum_{(x^n,y^n)\in A_\epsilon^{(n)}} p(x^n, y^n) \le |A_\epsilon^{(n)}|\, 2^{-n(H(X,Y)-\epsilon)}$$

$$\Rightarrow |A_\epsilon^{(n)}| \geq (1-\epsilon)\, 2^{n(H(X,Y)-\epsilon)}$$

Ebből: $\Pr\{(\tilde{X}^n,\tilde{Y}^n)\in A_\epsilon^{(n)}\} = \sum_{A_\epsilon^n} p(x^n)p(y^n) \geq$

$$\geq (1-\epsilon)2^{n(H(X,Y)-\epsilon)} \cdot 2^{-n(H(X)+\epsilon)} \cdot 2^{-n(H(Y)+\epsilon)} =$$

$$= (1-\epsilon)\cdot 2^{-n(I(X;Y)+3\epsilon)}$$

□

Tétel (Csatorna kódolás tétele):

Egy diszkrét memóriamentes csatornára minden C-nél kisebb ráta elérhető. Azaz, minden $R < C$ esetén létezik egy $(2^{nR}, n)$ kódsorozat úgy, hogy a maximális hiba $\lambda^{(n)} \to 0$.

Megfordítva: Minden $(2^{nR}, n)$ kódsorozatra, amelyre $\lambda^{(n)} \to 0$ a ráta $R \le C$.

Biz: Adjunk egy konstrukciót, ami $\forall R$-re generál egy megfelelő kódot!

Ha van egy adott $p(x)$ eloszlásunk legeneráltatunk 2^{nR} darab kódszót az alábbi eloszlással: $p(x^n) = \prod_{i=1}^{n} p(x_i)$. A kódot a C mátrixszal reprezentálhatjuk, amelynek a sorai a kódszavak:

$$C = \begin{pmatrix} x_1(1) & x_2(1) & \cdots & x_n(1) \\ \vdots & \vdots & \ddots & \vdots \\ x_1(2^{nR}) & x_2(2^{nR}) & \cdots & x_n(2^{nR}) \end{pmatrix}$$

Adott C generálásának a valsége: $\Pr(C) = \prod_{w=1}^{2^{nR}} \prod_{i=1}^{n} p(x_i(w))$

TFH az üzenet egyenletes eloszlású, tehát $\Pr(W=w) = 2^{-nR}$ és a kódolás során a w-ik üzenethez a w-ik sor tartozik! A dekódoló az alábbi szerint dolgozzon:

- $\hat{W}$-t ad vissza, ha $(X^n(\hat{W}), Y^n)$ tipikus, és nincs más $W' \neq \hat{W}$, amire $(X^n(W'), Y^n)$ tipikus lenne.
- Hibát dob, ha nincs megfelelő $\hat{W}$ vagy ha több is van.

Lássuk be, hogy a tévedés valószínűsége $\to 0$!

A tévedés-esemény: $\mathcal{E} = \{\hat{W}(Y^n) \neq W\}$. Ennek valószínűsége:

$$\Pr(\mathcal{E}) = \sum_C \Pr(C) P_e^{(n)}(C) = \sum_C \Pr(C) \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \lambda_w(C) =$$

$$= \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \sum_C \Pr(C) \lambda_w(C) =$$

mivel C konstrukciója szimmetrikus volt W-n, ezért az összeg w-től független. TFH $w=1$, és számoljunk ezzel:

$$= \sum_C \Pr(C) \lambda_1(C) = P(\mathcal{E} \mid W=1)$$

Tekintsük az alábbi eseményt: $E_i = \{(X^n(i), Y^n) \in A_\epsilon^{(n)}\}$ $i \in \{1, \dots 2^{nR}\}$!

Tévedés két okból lehet:
- a $(X^n(1), Y^n)$ nem tipikus, ekkor E_1^c teljesül
- más $(X^n(i), Y^n)$ is tipikus $i \neq 1$-re, ekkor $E_2 \cup E_3 \cup \dots \cup E_{2^{nR}}$ teljesül.

Ebből:

$$\Pr(\mathcal{E} \mid W=1) = P(E_1^c \cup E_2 \cup E_3 \cup \dots \cup E_{2^{nR}} \mid W=1) \leq$$

$$\leq P(E_1^c \mid W=1) + \sum_{i=2}^{2^{nR}} P(E_i \mid W=1).$$

AEP miatt $P(E_1^c \mid W=1) \leq \epsilon$ és $P(E_i \mid W=1) \leq 2^{-n(I(X,Y)-3\epsilon)}$ elég nagy n-re.

$$P(\mathcal{E} \mid W=1) \leq \epsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X,Y)-3\epsilon)}$$ = egyenletes eloszlása miatt

$$= \epsilon + (2^{nR}-1) \cdot 2^{-n(I(X,Y)-3\epsilon)} \leq$$

$$\leq \epsilon + 2^{3n\epsilon} \cdot 2^{-n(I(X,Y)-R)}$$ Ha $R < I(X;Y)$ akkor megfelelő ϵ és n-re

$$\leq 2\epsilon.$$

Korlát élesítések:

1. $p(x)$ legyen az, amivel a $I(X;Y)$ maximális, így a bizonyítás minden $R < C$-re érvényes lesz.
2. $P(\mathcal{E} \mid W=1) \leq 2\epsilon$ az átlag C kódra igaz, de ha azt választjuk amire a hiba legkisebb, annál 2ϵ-nál biztos kisebb lesz, tehát létezik egy ilyen kód.
3. Mivel W egyenletes, ezért $P(\mathcal{E} \mid W=1) = \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} \lambda_i(C^*) \leq 2\epsilon$, tehát a szavak legalább felére $\lambda_i \leq 4\epsilon$. Eldobva a másik felét, a maximális hibát is korlátoztuk $\to 0$-ra. Ugyan a szavak száma így 2^{nR-1} és a ráta $R \to R - \frac{1}{n}$, de nagy n-re ez u.a. □

A megfordítás bizonyításához kell pár lemma:

Lemma: Egy diszkrét memóriamentes csatornán használt egyenletes eloszlású W üzenet esetén teljesül a $$H(W|\hat{W}) \le 1 + P_e^{(n)} nR$$

Biz.: A Fano-egyenlőtlenségből
$1 + P_e \log|W| \ge H(W|\hat{W})$, ha $|W| = 2^{nR}$ éppen ezt kapjuk. □

Lemma: Legyen Y^n egy X^n bemenet eredménye egy C kapacitású csatornán! Ekkor
$I(X^n, Y^n) \le nC$ minden $p(x^n)$-re.

Biz.:
$$I(X^n, Y^n) = H(Y^n) - H(Y^n|X^n) = H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1, \dots Y_{i-1}, X^n) = \text{memóriamentesség miatt}$$
$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) \le \sum_{i=1}^{n} (Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) = \sum_{i=1}^{n} I(X_i; Y_i) \le nC$$

Csatorna kódolási tétel megfordításának bizonyítása:

Van egy $(2^{nR}, n)$ kódunk $\lambda^{(n)} \to 0$-val, mutassuk meg, hogy $R \le C$!

Adott kódoló és dekódoló fv esetén a $W \to X^n(W) \to Y^n \to \hat{W}$ egy Markov-lánc, ezért alkalmazható a Fanó-egyenlőtlenség. Ha W eloszlása egyenletes, akkor

$nR = H(W) = H(W|\hat{W}) + I(W;\hat{W}) \le$ lemma miatt

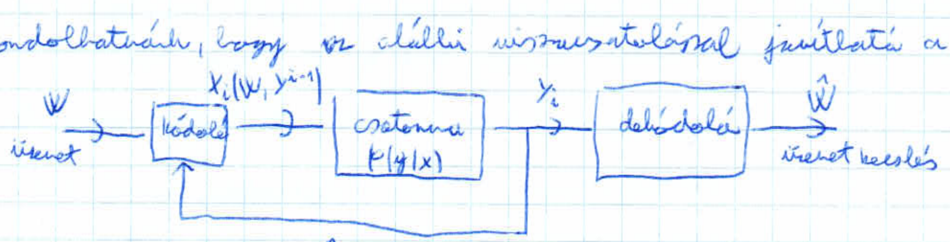
$\le 1 + P_e^{(n)} nR + I(W;\hat{W}) \le$ adatfeldolgozási egyenlőtlenség miatt

$\le 1 + P_e^{(n)} nR + I(X^n, Y^n) \le$ lemma miatt

$\le 1 + P_e^{(n)} nR + nC$.

Leosztva n-nel: $R \le \frac{1}{n} + P_e^{(n)} R + C$ $\quad n \to \infty$ esetén: $R \le C$. □

Azt gondolhatnánk, hogy az előbbi visszacsatolással javítható a kapacitás:



Itt az Y_i üzenetet hiba nélkül visszaküldi a kódolóba, és az i-ik kódszó az összes előbbi Y_i-től függ.

A valóságban azonban ez nem lehetséges, csak a dekódolás egyszerűsítésében segíthet.

Def: A visszacsatolási kapacitás egy diszkrét memória nélküli csatornán visszacsatolásos kóddal elérhető ráták szuprémuma. jele: C_{FB}.

Tétel: $$C_{FB} = C = \max_{p(x)} I(X;Y)$$

Biz. A csatorna kódolási tétel megfordításánál használt lemma itt nem teljesül, mert $H(Y_i \mid Y_1, \dots, Y_{i-1}, X^n) \neq H(Y_i \mid X_i)$, ezért mást kell felhasználni.

Mivel a visszacsatolás nélküli kód a visszacsatolásos kód spec. esete, ezért $C_{FB} \geq C$. A másik irányhoz az alábbi kell:

Egyenletes W esetén $nR = H(W) = H(W \mid \hat{W}) + I(W;\hat{W}) \leq$ Fano miatt

$\leq 1 + P_e^{(n)} nR + I(W;\hat{W}) \leq$ adatfeld. egyenl. miatt

$\leq 1 + P_e^{(n)} nR + I(W;Y^n)$.

Ebben az információ:

$$I(W;Y^n) = H(Y^n) - H(Y^n \mid W) = H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid Y_1, \dots, Y_{i-1}, W) =$$

mivel X_i az Y-k és W fv-e, ezért a feltételen nem ront

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid Y_1, \dots, Y_{i-1}, W, X_i) =$$

mivel Y_i csak X_i-től függ

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i \mid X_i) \leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i \mid X_i) = \sum_{i=1}^{n} I(X_i;Y_i) \leq$$

$\leq nC$

Visszaírva: $nR \leq 1 + P_e^{(n)} nR + nC$ leosztva n-nel és $n \to \infty$-ben

$R \leq C \Rightarrow C_{FB} \leq C$. □

Hamming - kód

A Shannon-féle kódolási tétel csak azt mondja ki, hogy minden $R < C$ rátához létezik olyan kód, ami tetszőlegesen kis hibával továbbítható, de egy könnyen dekódolható hatékony kódot találni nehéz.

A hiba csökkentésének legegyszerűbb módja a redundancia növelése, azaz felesleges bitek küldése. Pl.: 0 helyett küldjük a 00000 üzenetet, így ha ebből néhány el is romlik, még az üzenet kitalálható.

paritásbit: A kódszó végére egy plusz 1, ha a kódszó páratlan és 0, ha a kódszó páros. Ezzel észlelhető, ha páratlan db hiba történik.

Még hatékonyabb, ha a kódnak több részére is adunk paritásbitet. A Hamming-kód erre egy példa.

Konstrukció: Írjuk le az összes lehetséges 3 hosszú sorozatot oszlopvektorként:

$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Ennek a rangja 3, ezért dimenzió miatt a magtere $7-3=4$ dimenziós. A kódszavak legyenek azon szavak, amik H-val megszorozva $(0\,0\,0)$-t adnak! Ezek:

0000000	0100101	1000011	1100110
0001111	0101010	1001100	1101001
0010110	0110011	1010101	1110000
0011001	0111100	1011010	1111111

Ez $2^4 = 16$ db kódszó, amik mivel rang $H = 3$ legalább 3 db 1-est tartalmaznak (a 0-n kívül). Mivel ezek egy 4 dimes alteret alkotnak, ezért bármely kettő legalább 3 bitben különbözik $\Rightarrow$ Ha az egyik bit romlik, még mindig kitalálható, hogy melyik kódszóhoz tartozik. Hogyan?

Legyen $\underline{c}$ egy kódszó, $\underline{e}_i$ pedig az a vektor, amelynek az i-ik bitje 1, a többi 0! Ha $\underline{c}$-nek az i-ik bitje változott, akkor $\underline{v} = \underline{c} + \underline{e}_i$-t kapjuk. Megszorozva H-val:

$H\underline{v} = H(\underline{c}+\underline{e}_i) = H\underline{c} + H\underline{e}_i = H\underline{e}_i$ ami a H i-ik oszlopa.

$\Rightarrow$ Minden lehetséges $\underline{v}$-hez megtalálható a hozzá legközelebbi.

Gondolhatunk a Hamming-kódra az alábbi módon is:

16 kódszó van, ami 2^4, tehát elég lenne 4 bit is, de mi használtunk 3 ellenőrző bitet. Vegyük észre, hogy a kódszavak első 4 karaktere egyedi, vagyis ezek az információs bitek, a másik 3 pedig 3 bithármas paritásbitjei.

pl.: 0110|011

Általános Hamming kód esetén:

Ha H-ban ℓ sor van, a blokkhosszúság $n = 2^{\ell} - 1$, az információs bitek száma pedig $k = 2^{\ell} - \ell - 1$. A minimális távolság ekkor is 3, feltéve hogy minden lehetséges kódszót felhasználunk.

BEV. INF. ELM.

8. előadás (04.09.)

Differenciális entrópia

Def: Ha az X valószínűségi változóhoz tartozó $F(x) = Pr(X \leq x)$ kummulatív eloszlásfüggvény folytonos, akkor az X változót folytonosnak nevezzük.
$f(x) = F'(x)$ függvényt hívjuk a változó valószínűségi sűrűségfv.-ének,
a $S = \{x \in X \mid f(x) > 0\}$ pedig az értékkészlete.

Def: Egy f sűrűségfv.-el rendelkező folytonos valószínűségi változó differenciális entrópiája az alábbi:

$$h(X) = -\int_S f(x) \log f(x)\, dx$$

(Mivel h csak f-től függ, ezért a jelölés néha $h(f)$.)

Példák: 1) egyenletes eloszlás.

A $[0, a]$ intervallumon egyenletes eloszlás sűrűségfv.-e: $f(x) = \frac{1}{a}$

$$h(X) = -\int_0^a \frac{1}{a} \log \frac{1}{a}\, dx = \log a$$

NÉH $a < 1 \Rightarrow h(X) < 0$, tehát a diff. entrópia lehet negatív.

2) Normál eloszlás. (itt érdemes nat-ban számolni)

A sűrűségfv.: $\phi = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}$

$$h(\phi) = -\int_{-\infty}^{\infty} \phi(x) \left[-\frac{x^2}{2\sigma^2} - \ln\sqrt{2\pi\sigma^2}\right] = E_X\left(\frac{X^2}{2\sigma^2}\right) + \frac{1}{2}\ln(2\pi\sigma^2) =$$

$$= \frac{1}{2} + \frac{1}{2}\ln(2\pi\sigma^2) = \frac{1}{2}\ln(2\pi e\sigma^2)$$

A diszkrét esethez hasonlóan itt is bizonyítható az AEP tulajdonság.

Tétel: Legyen $X_1, \dots, X_n$ néhány folytonos i.i.d. változó sokasága azonos f sűrűséggel! Ekkor

$$-\frac{1}{n} \log f(X_1, \dots, X_n) \to E(-\log f(X)) = h(X)$$

Biz: Triviális a nagy számok tv.-éből.

Def: Adott ϵ és n értékekre az $A_\epsilon^{(n)}$ tipikus halmaz az f sűrűségfv-re:

$$A_\epsilon^{(n)} = \left\{ (x_1, \dots, x_n) \in S^n : \left| -\frac{1}{n} \log f(x_1, \dots, x_n) - h(X) \right| \le \epsilon \right\}$$

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

Def: Egy $A \subset \mathbb{R}^n$ halmaz térfogata $Vol(A) = \int_A dx_1 \cdots dx_n$.

Tétel: Az $A_\epsilon^{(n)}$-re teljesülnek az alábbiak:

1. $Pr(A_\epsilon^{(n)}) > 1-\epsilon$ elég nagy n-re.
2. $Vol(A_\epsilon^{(n)}) \le 2^{n(h(X)+\epsilon)}$ minden n-re.
3. $Vol(A_\epsilon^{(n)}) \ge (1-\epsilon) 2^{n(h(X)-\epsilon)}$ elég nagy n-re.

Bz: 1. Az előző tételből következik, hogy a def-ben lévő $|\cdot| \to 0$, így a valószínűség $\to 1$.

2.
$$1 = \int_{S^n} f(x_1, \dots, x_n)\, dx_1 \cdots dx_n \ge \int_{A_\epsilon^{(n)}} f(x_1, \dots, x_n)\, dx_1 \cdots dx_n \ge$$
$$\ge \int_{A_\epsilon^{(n)}} 2^{-n(h(X)+\epsilon)}\, dx_1 \cdots dx_n = 2^{-n(h(X)+\epsilon)} \int_{A_\epsilon^{(n)}} dx_1 \cdots dx_n =$$
$$= 2^{-n(h(X)+\epsilon)} Vol(A_\epsilon^{(n)}) \implies Vol(A_\epsilon^{(n)}) \le 2^{n(h(X)+\epsilon)}$$

3. Ha 1. teljesül, akkor
$$1-\epsilon \le \int_{A_\epsilon^{(n)}} f(x_1, \dots, x_n)\, dx_1 \cdots dx_n \le \int_{A_\epsilon^{(n)}} 2^{-n(h(X)-\epsilon)}\, dx_1 \cdots dx_n =$$
$$= 2^{-n(h(X)-\epsilon)} \int_{A_\epsilon^{(n)}} dx_1 \cdots dx_n = 2^{-n(h(X)-\epsilon)} Vol(A_\epsilon^{(n)})$$
$\square$

Tétel: $A_\epsilon^{(n)}$ a legkisebb térfogatú halmaz, amely $1-\epsilon$-nál nagyobb valséggel fordul elő.

Bz: u.a. mint diszkrét esetben.

Nézzük milyen viszonyban áll h H-val, ha diszkretizáljuk a folytonos változót! Felosztva X értelmezési tartományát Δ hosszú intervallumokra, akkor az átlagérték tétel alapján $f(x_i)\Delta = \int_{i\Delta}^{(i+1)\Delta} f(x)\, dx$.

Ez alapján definiálható az alábbi diszkrét valószínűségi változó:

$$X^\Delta = x_i \quad \text{ha} \quad i\Delta \le X \le (i+1)\Delta$$

$$p_i = f(x_i)\Delta \implies \sum_i p_i = 1.$$

Tétel: Ha az f sűrűségfv Riemann-integrálható, akkor

$$H(X^{\Delta}) + \log \Delta \to h(X) \quad \text{amikor } \Delta \to 0.$$

Biz:
$$H(X^{\Delta}) = -\sum_{-\infty}^{\infty} p_i \log p_i = -\sum_{-\infty}^{\infty} f(x_i)\Delta \log\left(f(x_i)\Delta\right) =$$
$$= -\sum \Delta f(x_i) \log f(x_i) - \sum f(x_i)\Delta \log \Delta =$$
$$= \underbrace{-\sum \Delta f(x_i) \log(x_i)}_{\Delta \to 0 \text{ esetén a def szerint } = -\int f(x) \log f(x)\,dx = h(X)} - \log \Delta$$
□

Következmény: Egy folytonos változó n-bit kvantálásának entrópiája
$$H(X^{2^n}) \approx h(X) + n.$$

Def: Az $X_1, \dots, X_n$ változók együttes eloszlásához tartozó entrópia:
$$h(X_1, \dots, X_n) = -\int f(x^n) \log f(x^n)\, dx^n$$

Def: Ha az X, Y változók együttes sűrűségfv-e $f(x,y)$, a feltételes entrópiájuk:
$$h(X|Y) = -\int f(x,y) \log f(x|y)\, dx\, dy.$$

Mivel $f(x|y) = f(x,y)/f(y)$, ezért $h(X|Y) = h(X,Y) - h(Y)$.

Tétel: Ha $X_1, \dots, X_n$ eloszlása egy többváltozós normáleloszlás μ átlaggal és K kovarianciamátrixszal, akkor együttes entrópiájuk:
$$h(X_1, \dots, X_n) = h(\mathcal{N}(\mu, K)) = \frac{1}{2} \log\left[(2\pi e)^n |K|\right]$$

Biz: Mivel az együttes eloszlás sűrűségfv-e: $f(\underline{x}) = \dfrac{e^{-\frac{1}{2}(\underline{x}-\mu)^T K^{-1} (\underline{x}-\mu)}}{(\sqrt{2\pi})^n |K|^{1/2}}$

ezért
$$h(f) = -\int f(\underline{x}) \left[-\frac{1}{2}(\underline{x}-\mu)^T K^{-1} (\underline{x}-\mu) - \ln\left[(\sqrt{2\pi})^n |K|\right]\right] d\underline{x} =$$
$$= \frac{1}{2} E_{\underline{x}}\left[\sum_{i,j} (x_i - \mu_i)(K^{-1})_{ij}(x_j - \mu_j)\right] + \ln\left[(\sqrt{2\pi})^n |K|\right] =$$
$$= \frac{1}{2} \sum_{i,j} E_{\underline{x}}\left((x_i-\mu_i)(x_j-\mu_j)\right)(K^{-1})_{ij} + \ln\left[(\sqrt{2\pi})^n |K|\right] =$$
$$= \frac{1}{2} \sum_{i,j} K_{ij} \cdot (K^{-1})_{ij} + \ln\left[(\sqrt{2\pi})^n |K|\right] = \frac{1}{2} \sum_j I_{jj} + \ln\left[(\sqrt{2\pi})^n |K|\right] =$$
$$= \frac{n}{2} + \frac{1}{2} \ln(2\pi \cdot |K|) = \frac{1}{2} \ln(2\pi e |K|)$$
□

Def: Az f és g sűrűségfv-ek által definiált folytonos változók relatív entrópiája

$$D(f\|g)=\int_{S_f} f(x)\log\frac{f(x)}{g(x)}\,dx$$

(Az értelmesség feltétele, hogy $S_f \subseteq S_g$.)

Def: Az $f(x,y)$ együttes sűrűségfv-nel jellemzett változók kölcsönös információja

$$I(X;Y)=\int f(x,y)\log\frac{f(x,y)}{f(x)\,f(y)}\,dx\,dy.$$

Algebrai átalakítások alapján

$$\begin{aligned} I(X;Y) &= h(X)-h(X|Y)\\ &= h(Y)-h(Y|X)\\ &= h(X)+h(Y)-h(X,Y)\\ &= D\big(f(x,y)\,\|\,f(x)f(y)\big)\end{aligned}$$

VÉH a kölcsönös információ a kvantálásra invariáns:

$$I(X^\Delta;Y^\Delta)=H(X^\Delta)-H(X^\Delta|Y^\Delta)\approx h(X)-\log\Delta-\big(h(X|Y)-\log\Delta\big)=I(X;Y)$$

Ezáltal, ha kvantáljuk a változók partícióját, és azzal definiálnánk a kölcsönös információt, ugyanazt kapnánk, mint az előbb.

Tétel: $D(f\|g)\geq 0$. Egyenlőség, ha $f=g$ majdnem mindenütt.

Biz: $$-D(f\|g)=\int_S f\log\frac{g}{f}\underset{\text{Jensen miatt}}{\leq}\log\int_{S_f} f\,\frac{g}{f}=\log\int_{S_f} g\underset{S_f\subseteq S_g\text{ miatt}}{\leq}\log 1=0$$ □

Következmények:
- $I(X;Y)\geq 0$ (egyenlőség ha X és Y független)
- $h(X|Y)\leq h(X)$
- $h(X_1,\dots,X_n)=\sum_{i=1}^n h(X_i|X_1,\dots,X_{i-1})$
- $h(X_1,\dots,X_n)\leq\sum_{i=1}^n h(X_i)$ ha X_i-k iid.

Példa az alkalmazásra: Hadamard-egyenlőtlenség.
Legyen $\underline{X}\sim\mathcal{N}(0,K)$! Ekkor $|K|\leq\prod_{i=1}^n K_{ii}$

Biz: alkalmazzuk az utolsó következményt és emeljük e hatványra

Tétel: (i) $h(X+c)=h(X)$ (ii) $h(aX)=h(X)+\log a$

Biz: (i) triviá

(ii) Ha $Y=aX$, akkor $f_Y(y)=\frac{1}{|a|}f_X\left(\frac{y}{a}\right)$.

$$h(aX)=-\int f_Y(y)\log f_Y(y)\,dy=-\int\frac{1}{|a|}f_X\left(\frac{y}{a}\right)\log\left(\frac{1}{|a|}f_X\left(\frac{y}{a}\right)\right)dy=$$

$$=-\int f_X(x)\log f_X(x)\,dx+\log|a|=h(X)+\log|a| \qquad \square$$

Következmény: Vektorváltozóra: $h(A\underline{X})=h(X)+\log|\det(A)|$

Tétel: Legyen $\underline{X}\in\mathbb{R}^n$, 0 átlaggal és $K=E_{\underline{X}}(\underline{X}\underline{X}^T)$ kovarianciamátrixszal! Ekkor $h(\underline{X})\leq\frac{1}{2}\log(2\pi e)^n|K|$. Egyenlőség, ha $\underline{X}\sim\mathcal{N}(0,K)$.

Biz: Legyen $g(\underline{x})$ a sűrűségfv! Feltétel miatt $\int g(\underline{x})x_ix_j\,d\underline{x}=K_{ij}$.

Legyen $\phi_K(\underline{x})$ a $\mathcal{N}(0,K)$ sűrűségfv-e! Ekkor

$$0\leq D(g\|\phi_K)=\int g\log\left(\frac{g}{\phi_K}\right)=-h(g)-\int g\log\phi_K=$$

mivel g és ϕ_K első két momentuma megegyezik

$$=-h(g)-\int\phi_K\log\phi_K=-h(g)+h(\phi_K) \qquad \square$$

Tétel: Legyen X egy véletlen változó $h(X)$ entrópiával, $\hat{X}$ pedig egy becslése! A négyzetes hiba várható értéke:

$$E(X-\hat{X})^2\geq\frac{1}{2\pi e}e^{2h(X)}.$$ Egyenlőség, ha X gauss-eloszlás.

Biz: $$E(X-\hat{X})^2\geq\min_{\hat{X}}E(X-\hat{X})^2=E(X-E(X))^2=\mathrm{var}(X)\geq\frac{1}{2\pi e}e^{2h(X)}$$

abból következően, hogy adott varianciájú eloszlások közül a gauss-eloszlásnak a legnagyobb az entrópiája.

Következmény: Ha van Y plusz infónk, amitől függ $\hat{X}$, akkor

$$E(X-\hat{X}(Y))^2\geq\frac{1}{2\pi e}e^{2h(X|Y)}$$

BEV INF ELM

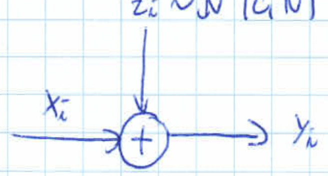
9. előadás (04.16.)

Gauss-i csatorna

Def: Gauss-i csatorna egy olyan folytonos ábécét továbbító csatorna, ahol egy Gauss-eloszlásból adódó zaj adódik hozzá a jelhez.

Az i. időpillanatban: $Y_i = X_i + Z_i$

ahol $Z_i \sim N(0, N)$



Mivel a bemenet és a kimenet végtelen halmazok, a különböző kódokat tetszőlegesen messze választhatjuk egymástól, így a hiba 0-ra csökkenthető, és a kapacitás végtelen.

Mivel ezt általában valós rendszerek modellezésére használják, ezért szokott lenni valamilyen kikötés az energiára. Sőt általában az alábbi teljesítmény-kikötéssel korlátozzák:

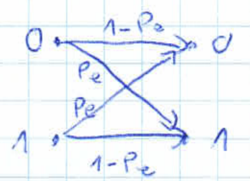
$$\frac{1}{n} \sum_{i=1}^{n} x_i^2 \le P.$$

A legegyszerűbb (szuboptimális eset), ha egy bináris ábécét akarunk továbbítani, és a kódszavak $\pm\sqrt{P}$. A dekódolásnak egyszerűen el kell döntenie, hogy Y_i pozitív vagy negatív. A hiba valószínűsége:

$$P_e = \frac{1}{2} \Pr\left(Y<0 \mid X=\sqrt{P}\right) + \frac{1}{2} \Pr\left(Y>0 \mid X=-\sqrt{P}\right) =$$

$$= \frac{1}{2} \Pr\left(Z<-\sqrt{P} \mid X=\sqrt{P}\right) + \frac{1}{2} \Pr\left(Z>\sqrt{P} \mid X=-\sqrt{P}\right) = \Pr\left(Z>\sqrt{P}\right) =$$

$$= 1 - \Phi\left(\sqrt{P/N}\right) \qquad \text{ahol} \quad \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt$$



A korábban, diszkrét csatornákra látott definíciók és tételek itt is használhatóak, de a kikötéssel együtt kell kezelni.

Def: Az információs kapacitás egy Gauss-i csatornára a P teljesítményfeltétellel:

$$C = \max_{EX^2 \le P} I(X;Y)$$

A kapacitás analitikusan kiszámolható, hiszen

$$I(X;Y) = h(Y) - h(Y|X) = h(Y) - h(X+Z|X) = h(Y) - h(Z|X) = \text{mivel } X \text{ és } Z \text{ független}$$
$$= h(Y) - h(Z)$$

Mivel Z Gauss-eloszlású, ezért $h(Z) = \frac{1}{2}\log(2\pi e N)$.

Továbbá $EY^2 = E(X+Z)^2 = \underbrace{EX^2}_{=P \text{ a megkötés miatt}} + 2\underbrace{EXEZ}_{0} + \underbrace{EZ^2}_{=N} = P + N$

Ez alapján tehát $h(Y) \le \frac{1}{2}\log(2\pi e(P+N))$.

$$I(X;Y) = h(Y) - h(Z) \le \frac{1}{2}\log(2\pi e(P+N)) - \frac{1}{2}\log(2\pi e N) = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$$

$$C = \max I(X;Y) = \frac{1}{2}\log\left(1+\frac{P}{N}\right).$$

ezt akkor éri el, ha $X \sim N(0,P)$

Def: Egy (M,n) kód egy Gaussi csatornán P-hatvány kikötéssel az alábbiakból áll:

1. Indexhalmaz $\{1,\dots,M\}$
2. Kódoló fv, amely eleget tesz a kikötésnek: $x: \{1,\dots,M\} \to \mathcal{X}^n$ úgy, hogy $\frac{1}{n}\sum_{i=1}^{n} x_i^2(w) \le P \quad \forall w$-re.
3. Dekódoló függvény: $g: \mathcal{Y}^n \to \{1,\dots,M\}$.

Def: Egy R ráta elérhető egy P-hatvány kikötésű Gaussi csatornára, ha létezik egy $(2^{nR}, n)$ kód, aminek a kódszavai teljesítik a kikötést és a maximális hiba valószínűsége tart a 0-hoz.

Def: A csatornakapacitás az elérhető ráták maximuma.

Tétel: Egy P-hatvány kikötésű gaussi csatorna csatornakapacitása $C = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$.

Szemlélet: Egy n hosszú kódszó esetén a hiba varianciája nN, vagyis az $\mathcal{Y}^n$-térben a küldött üzenet körüli $\sqrt{nN}$ sugarú gömbön belül lesz a továbbított szó nagy valószínűséggel.

Mivel az Y varianciája $n(P+N)$, ezért az összes szó valahol egy $\sqrt{n(P+N)}$ sugarú gömbben lesz.

Mivel egy n-dimenziós gömb térfogata $C_n r^n$, ezért a dekódolható szavak száma:
$$\frac{C_n[n(P+N)]^{n/2}}{C_n(nN)^{n/2}} = \left(1+\frac{P}{N}\right)^{n/2} = 2^{\frac{n}{2}\log\left(1+\frac{P}{N}\right)}$$

Tétel első fele: Minden $R < C = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$ ráta elérhető

Biz: 1. Először generáljunk egy kódkönyvet! Minden szó minden elemét generáljuk egy $P-\epsilon$ varianciájú normáleloszlásból, ezáltal nagy n-re $\frac{1}{n}\sum_i X_i^2 \to P-\epsilon < P$.

2. A küldő elküldi a w szóhoz tartozó $X^n(w)$ kódszót.

3. A fogadó azt a w szót választja dekódoltnak, amivel a $(X^n(w), Y^n)$ pár benne van a tipikus halmazban.

Ha nincs benne egyik w-vel sem, vagy ha $X^n(w)$ nem teljesíti a kikötést, akkor hibát dob.

4. A hiba valószínűségét az alábbi módon számolhatjuk:

Legyen $E_0 = \left\{ \frac{1}{n}\sum_{j=1}^{n} X_j^2(1) > P \right\}$

$$E_i = \left\{ (X^n(i), Y^n) \in A_\epsilon^{(n)} \right\}$$

Hiba történik, ha E_0 vagy E_1^c fordul elő, és hibás dekódolás, ha $E_2 \cup E_3 \cup \dots \cup E_{2^{nR}}$.

$$P(\mathcal{E}) = P\left(E_0 \cup E_1^c \cup E_2 \cup E_3 \cup \dots \cup E_{2^{nR}}\right) \le$$

$$\le P(E_0) + P(E_1) + \sum_{i=2}^{2^{nR}} P(E_i)$$

$P(E_0)$ és $P(E_1)$ a nagy számok tv-e és a AEP miatt $\le \epsilon$

Mivel $X^n(1)$ és $X^n(i)$ a generálás miatt függetlenek, ezért Y^n és $X^n(i)$ is, ezért a AEP miatt $P(E_i) \le 2^{-n(I(X;Y)-3\epsilon)}$

Ebből $P(\mathcal{E}) \le 2\epsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X;Y)-3\epsilon)} = 2\epsilon + (2^{nR}-1)\,2^{-n(I(X;Y)-3\epsilon)} \le$

$$\le 2\epsilon + 2^{3n\epsilon}\,2^{-n(I(X;Y)-R)} \le 3\epsilon \qquad \square$$

Tétel második fele: Az $R > C = \frac{1}{2}\log\left(1+\frac{P}{N}\right)$ ráták nem érhetőek el.

Biz: Azt kell belátni, hogy ha $(2^{nR}, n)$ kód hibájának valsége $P_e^{(n)} \to 0$, akkor $R \le C$.

Vegyünk egy tetszőleges kódot, amire teljesül, hogy $\frac{1}{n}\sum_{i=1}^{n} x_i^2(w) \le P$ $\forall w$-re, és legyen W egyenletes eloszlású!

A diszkrét esetnél bizonyított lemma alapján

$W \to X^n(W) \to Y^n \to \hat{W}$ lánc esetén

$$H(W \mid \hat{W}) \le 1 + nRP_e^{(n)} = n\epsilon_n \qquad \text{ahol } \epsilon_n \to 0,\ P_e^{(n)} \to 0.$$

Ebből:

$$nR = H(W) = I(W;\hat{W}) + H(W|\hat{W}) \le I(W;\hat{W}) + n\epsilon_n \le I(X^n;Y^n) + n\epsilon_n =$$

$$= h(Y^n) - h(Y^n|X^n) + n\epsilon_n = h(Y^n) - h(Z^n) + n\epsilon_n \le \sum_{i=1}^{n} h(Y_i) - h(Z^n) + n\epsilon_n$$

$$= \sum_{i=1}^{n} h(Y_i) - \sum_{i=1}^{n} h(Z_i) + n\epsilon_n = \sum_{i=1}^{n} I(X_i;Y_i) + n\epsilon_n$$

Legyen $P_i = \frac{1}{2^{nR}} \sum_{w} x_i^2(w)$. Mivel a kódszavak mind kielégítik a határfeltételt, ezért az átlaguk is: $\frac{1}{n}\sum_i P_i \le P$.

Mivel $Y_i = X_i + Z_i$ ahol X_i és Z_i függetlenek, ezért $EY_i^2 = P_i + N$, így $h(Y_i) \le \frac{1}{2}\log(2\pi e(P_i+N))$. Ebből:

$$nR \le \sum_{i=1}^{n} \left(h(Y_i) - h(Z_i)\right) + n\epsilon_n \le \sum_{i=1}^{n} \left(\frac{1}{2}\log(2\pi e(P_i+N)) - \frac{1}{2}\log(2\pi e N)\right) + n\epsilon_n =$$

$$= \sum_{i=1}^{n} \frac{1}{2}\log\left(1+\frac{P_i}{N}\right) + n\epsilon_n$$

Jensen-egyenlőtlenség miatt

$$R \le \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2}\log\left(1+\frac{P_i}{N}\right) \le \frac{1}{2}\log\left(1+\frac{1}{n}\sum_{i=1}^{n}\frac{P_i}{N}\right) \le \frac{1}{2}\log\left(1+\frac{P}{N}\right) \qquad \square.$$

Párhuzamos Gauss-i csatornák

Ha a jeltovábbítás több frekvencián zajlik, és a zaj nem fehérzaj, akkor az úgy modellezhető, mint k párhuzamos, egymástól független Gauss-i csatorna:

$$Y_i = X_i + Z_i \qquad i = 1, 2, \dots k \qquad \text{ahol} \qquad Z_i \sim \mathcal{N}(0, N_i)$$

A határfeltételt a párhuzamos csatornákra közösen értelmezhetjük: $E\sum_{i=1}^{k} X_i^2 \le P$

A csatornakapacitás:

$$C = \max_{\sum EX_i^2 \le P} I(X_1 \dots X_k; Y_1 \dots Y_k)$$

Mivel Z_i-k függetlenek X_i-ktől:

$$I(X_1,\dots,X_k;Y_1,\dots,Y_k) = h(Y_1,\dots,Y_k) - h(Y_1,\dots,Y_k|X_1,\dots,X_k) =$$

$$= h(Y_1,\dots,Y_k) - h(Z_1,\dots,Z_k|X_1,\dots,X_k) =$$

$$= h(Y_1,\dots,Y_k) - h(Z_1,\dots,Z_k) = h(Y_1,\dots,Y_k) - \sum_i h(Z_i) \le$$

$$\le \sum_i \left(h(Y_i) - h(Z_i)\right) \le \sum_i \left(\frac{1}{2}\log\left(1+\frac{P_i}{N_i}\right)\right)$$

ahol $P_i = EX_i^2$ és $\sum_i P_i \le P$.

Egyenlőség akkor van, ha $(X_1,\dots,X_k) \sim \mathcal{N}\left(0, \begin{pmatrix} P_1 & & & \\ & P_2 & & \\ & & \ddots & \\ & & & P_k \end{pmatrix}\right)$

a kérdés, hogy milyen $(P_1, P_2, \dots P_k)$ kombinációra:

Ez egy egyszerű optimalizáció, ami Lagrange-multiplikátorral elvégezhető:

$$J(P_1,\dots,P_k) = \sum_{i=1}^{k} \frac{1}{2} \log\left(1+\frac{P_i}{N_i}\right) + \lambda\left(\sum_{i=1}^{k} P_i\right)$$

Variálva J-t, az alábbit kapjuk: $\frac{1}{2}\frac{1}{P_i+N_i} + \lambda = 0 \quad \Rightarrow \quad P_i = \nu - N_i$.

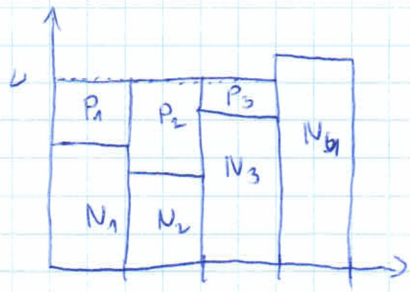
Bár $P_i \geq 0$ kell lennie, de az alábbinál ez nem feltétlen teljesül.

Ehelyett: Kuhn-Tucker-feltétel

$$P_i = (\nu - N_i)^+ \quad \text{úgy, hogy teljesüljön a} \quad \sum_{i=1}^{k} (\nu - N_i)^+ = P$$

$$\text{ahol} \quad (x)^+ = \begin{cases} x & \text{ha } x \geq 0 \\ 0 & \text{else} \end{cases}$$

Ábrázolva:



„víz feltöltős metódus"

BEV INF ELM

10. előadás (04.23.)

Maximum entrópia

Fizikában általában a rendszer adott feltételei mellett megvalósulható makroállapotok közül az valósul meg, amelyhez tartozó mikroállapotok száma a legnagyobb. Ezzel ekvivalens feltétel, ha azt mondjuk, hogy a maximális entrópiájú makroállapot lesz a megvalósuló.

Ennek az elvnek az analógiája matematikában is alkalmazható: gyakran lehet szükségünk arra az eloszlásra, amely bizonyos feltételek mellett a maximális entrópiával rendelkezik.

Feladat: Keressük meg azt az f valószínűségi sűrűségfv-t, amelyre $h(f)$ maximális és teljesíti a alábbi feltételeket:

1. $f(x) \geq 0$ a teljes S értelmezési tartományon
2. $\int_S f(x)\,dx = 1$
3. $\int_S f(x)\, r_i(x)\,dx = \alpha_i \qquad 1 \leq i \leq m$ adott r_i korlátozások és α_i kikötésekre.

1. Megoldási szemlélet (Kalkulus):

Mivel $h(f)$ konkáv fv, az optimalizációs módszerek használhatóak:

$$J(f) = -\int f \ln f + \lambda_0 \int f + \sum_{i=1}^{m} \lambda_i \int f r_i$$

Ezt variálva: $$\frac{\partial J}{\partial f(x)} = -\ln f(x) - 1 + \lambda_0 + \sum_{i=1}^{m} \lambda_i r_i(x) = 0$$

$$\Longrightarrow f(x) = e^{\lambda_0 - 1 + \sum_i \lambda_i r_i(x)}$$

ahol $\lambda_0, \lambda_1, \dots \lambda_m$ olyanok, hogy f teljesítse a kikötéseket.

Ez azonban csak a fv alakját határozza meg. A maximum bizonyításához 2. variált kellene, de egyszerűbb infelmmel bizonyítani.

2. Megoldási szemlélet (Információs egyenlőtlenség):

Ha g teljesíti a 1-3 feltételeket, és f^* az előbbi exp alak, akkor

$$0 \leq D(g\|f^*) = -h(g) + h(f)$$

(hiszen egy g-t f^*-gal súlyozva ugyanúgy viselkedik, mint g-vel súlyozva.)

Ebből $h(g) \leq h(f)$.

Formális bizonyítás az alábbi.

Tétel: Az $f^*(x) = e^{\lambda_0 + \sum_i \lambda_i v_i(x)}$ függvény, ahol $\lambda_0, \lambda_1, \dots, \lambda_m$ olyanok, hogy a kikötések teljesüljenek, egyértelműen maximalizálja az entrópiát.

Biz: Legyen g olyan fv, ami teljesíti a feltételeket! Ekkor

$$h(g) = -\int_S g \ln g = -\int_S g \ln\left(\frac{g}{f^*} f^*\right) = -D(g\|f^*) - \int g \ln f^* \leq$$

$$\leq -\int_S g \ln f^* = -\int_S g\left(\lambda_0 + \sum_i \lambda_i v_i\right) = \text{ mivel } g \text{ és } f^* \text{ is teljesíti a kikötést}$$

$$= -\int_S f^* \left(\lambda_0 + \sum_i \lambda_i v_i\right) = -\int_S f^* \ln f^* = h(f^*)$$

Egyenlőség akkor, ha $D(g\|f^*) = 0 \iff g = f^*$. $\square$

Példák:

1) 1D-s gáz, hőmérséklet kikötéssel.

A kikötések: $EX = 0$ és $EX^2 = \sigma^2$. Mivel az eloszlás alakja $e^{\lambda_0 + \lambda_1 x + \lambda_2 x^2}$, ezért a megoldás normáleloszlás lesz. Ennek momentumait ismerjük, tehát a megoldás:

$X \sim \mathcal{N}(0, \sigma^2)$, $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$

2) Dobókocka, kikötés nélkül.

Legyen $S = \{1,2,3,4,5,6\}$! A diszkrét entrópia maximális, ha az eloszlás egyenletes.

$\Rightarrow p(x) = \frac{1}{6} \quad \forall x \in S$

3) Dobókocka megadott átlaggal: $EX = \sum_i i p_i = \alpha$

TFH n dobásból az i-t n_i-szer dobtuk! A mikroállapotok száma $\binom{n}{n_1 \dots n_6}$, és mindegyik valószínűsége $\frac{1}{6^n}$. Hogy lássuk melyik a legvalószínűbb, azt kell maximalizálnunk a $\sum_i i n_i = n\alpha$ feltétellel.

A Stirling-formulát alkalmazva: $n! \approx \left(\frac{n}{e}\right)^n$

$$\binom{n}{n_1, \dots, n_6} \approx \frac{\left(\frac{n}{e}\right)^n}{\prod_i \left(\frac{n_i}{e}\right)^{n_i}} = \prod_i \left(\frac{n}{n_i}\right)^{n_i} = e^{\sum_i n_i \ln\left(\frac{n}{n_i}\right)} = e^{n H\left(\frac{n_1}{n}, \dots, \frac{n_6}{n}\right)}$$

Tehát a mikroállapotszám maximalizálása az entrópia maximalizálásával ekvivalens. Mivel a feltétel az első momentumra vonatkozott, ezért $p_i = \frac{e^{\lambda i}}{\sum_i e^{\lambda i}}$

Ez épp a Boltzmann-eloszlás.

4) Véges intervallum nincs kikötés. A tétel alapján a megoldás az egyenletes eloszlás.

5) $S=[0,\infty)$, $EX=\mu$. $\Rightarrow$ $f(x)=\frac{1}{\mu}e^{-\frac{x}{\mu}}$.

Fizikai interpretáció: Legyen X egy adott molekula magassága a légkörben!
Mivel az átlagos potenciális energia adott, ezért $E(mgX)$ rögzített.
Így kapjuk a barometrikus magasságformulát, ami volt ez.

Univerziális kódolás

Láttuk, hogy ha ismerjük a kódolandó szavak eloszlását, akkor található egy olyan kód, ami közel optimális az átlagos szóhossz tekintetében.
Sokszor azonban nem ismerjük előre az eloszlást, hanem folyamatában kell kódolni. Egyszerű valamelyik feladat, hogy így is az ideálishoz közel eső kódszóhosszat kapjunk.

3 példa:

(1) Aritmetikai kódolás

Lemma: Legyen Y egy véletlen változó, és $F(y)$ az eloszlásfv-e!
Legyen $U=F(Y)$ egy másik véletlen változó! Ekkor U egyenletes a $[0,1]$-en.

Biz: Mivel $F(y)\in[0,1]$, ezért U itt van definiálva. Adott $u\in[0,1]$-re:

$$F_U(u)=Pr(U\le u)=Pr(F(Y)\le u)=Pr(Y\le F^{-1}(u))=F(F^{-1}(u))=u \qquad \square$$

Most képzeljünk el egy $X_1, X_2, \dots$ végtelen sorozatot a $\mathcal{X}=0,\dots,m$ ábécéből! Beírva elé egy "$0,$"-t egy valós számot kapunk az $m+1$ számrendszerben felírva. Az eredeti sorozat eloszlásfv-e: $F_X(x)=Pr(X\le x=0,x_1x_2\dots)=$

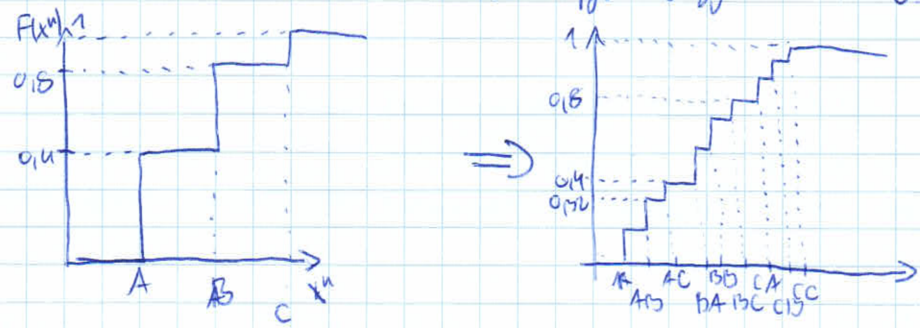
$$=Pr(0,X_1X_2\dots\le 0,x_1x_2\dots)=$$
$$=Pr(X_1<x_1)+Pr(X_1=x_1, X_2<x_2)+\dots$$

Ekkor az $U=F_X(X)=0,F_1F_2\dots$ változó a lemma értelmében egyenletes az $[0,1]$-en, tehát a számjegyei független, egyenletes eloszlásúak lesznek.
Ez a szekvencia már tömöríthetetlen, tehát U egy ideális tömörítése X-nek.

~~Metódus: Legyen $X_1,\dots,X_n$ egy véges bináris szekvencia! Ekkor hozzárendelhetjük az összes olyan valós számot, amely $0,X_1X_2\dots X_n$-nel kezdődik. Ez a $[0,X_1X_2\dots X_n;\ 0,X_1X_2\dots X_n+2^{-n})$ intervallum. Pontosabban~~

9

metódus: Válasszuk ki a [0,1) intervallumból azt, amelyikbe az első karakter tartozik a SFE-kód alapján! Ezt továbbá bontjuk fel a második karakter alapján és így tovább. Pl.: $P(A,B,C) = \{0,4; 0,4; 0,2\}$



Ahhoz, hogy univerzális kód legyen, adaptív módon kell alkalmazni, azaz a valószínűségi eloszlást minden lépés után frissíteni kell a beolvasott karakterrel.

② Csúszóablakos Lempel-Ziv-kód (LZ77)

Van egy W ablakméret, és minden "mondat" után megnézzük, hogy az elmúlt W karakterben van-e olyan szekvencia, amelyre válik a következő szakasz, és hogy mi az ilyen leghosszabb szekvencia. Ha van, akkor elküldjük a P pozíciót és a L hosszat. Ha nincs, akkor csak magát a következő karaktert.

Pl.: W = 4, üzenet: ABBABBABBBAABABA

felbontva mondatokra: A, B, B, ABBABB, BA, A, BA, BA

kódolva: (0,A), (0,B) (1,1,1), (1,3,6), (1,4,2) (1,1,1), (1,3,2), (1,2,2)

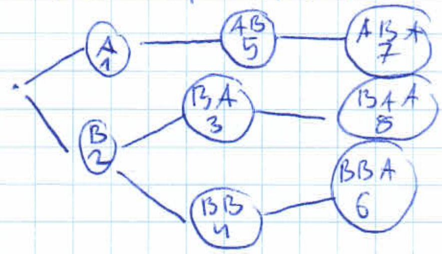
③ Fa-strukturált Lempel-Ziv-kód (LZ78)

Minden mondatot úgy kódolunk, mint az eddig nem szerepelt legrövidebb mondat. Ezáltal, csak a már korábban szerepelt mondatot kell megadni, és azt, hogy milyen karaktert fűzünk hozzá. Így egy fa gráffal reprezentálhatjuk a mondatokat

üzenet: ABBABBABBBAABABAA

felbontva: A, B, BA, BB, AB, BBA, ABA, BAA

fa gráf:



kódolva: (0,A), (0,B), (2,A), (2,B), (4,A), (5,A) (3,A)

BEV INFELM

11. előadás (04.30)

Hálózati információelmélet

A kommunikáció általában több résztvevő között zajlik, akik bármely másikkal kommunikálnak. Ez egy bonyolult hálózatot eredményez, ami az információátadás elemzését bonyolítja. Csak a legegyszerűbb definíciókat és a Slepian-Wolf-tételt nézzük meg.

Jelölés: Legyen $(X_1, X_2, \dots, X_k)$ véges számú véletlen változó gyűjteménye, adott valószínűségi eloszlással: $p(x_1, x_2, \dots, x_k)$!

Legyen S ezeknek egy részhalmaza (pl.: $S = (X_j, X_e)$), és képezzük el belőle n darabot! Ekkor a valószínűségi eloszlás:

$$\Pr(\mathbf{S} = \mathbf{s}) = \prod_{i=1}^{n} \Pr(S_i = s_i) \qquad \mathbf{s} \in \mathcal{S}^n$$

Nagy n-ekre:

$$-\frac{1}{n} \log p(S_1, S_2, \dots, S_n) = -\frac{1}{n} \sum_{i=1}^{n} \log p(S_i) \to H(S)$$

Def: Legyen $A_\epsilon^{(n)}$ az n-szekvenciák ϵ-tipikus halmaza, azaz

$$A_\epsilon^{(n)} = \left\{ (\underline{x}_1, \dots, \underline{x}_k) : \left| -\frac{1}{n} \log p(\underline{s}) - H(S) \right| < \epsilon \quad \forall S \subseteq \{X_1, \dots, X_k\} \right\}$$

Def: Legyen $A_\epsilon^{(n)}(S)$ ennek a megszorítása pusztán S részhalmazain!

Pl. ha $S = (X_1, X_2)$, akkor

$$A_\epsilon^{(n)}(X_1, X_2) = \Big\{ (\underline{x}_1, \underline{x}_2) : \left| -\frac{1}{n} \log p(\underline{x}_1, \underline{x}_2) - H(X_1, X_2) \right| < \epsilon,$$
$$\left| -\frac{1}{n} \log p(\underline{x}_1) - H(X_1) \right| < \epsilon,$$
$$\left| -\frac{1}{n} \log p(\underline{x}_2) - H(X_2) \right| < \epsilon \Big\}$$

Def: Jelölje $a_n \doteq 2^{n(b \pm \epsilon)}$ azt, ha $\left| \frac{1}{n} \log a_n - b \right| < \epsilon$ elég nagy n-re!

Tétel: Bármely $\epsilon > 0$-ra, létezik elég nagy n, melyre

1. $P(A_\epsilon^{(n)}(S)) \geq 1 - \epsilon \quad \forall S \subseteq \{X_1; X_2; \dots X_n\}$
2. $\underline{s} \in A_\epsilon^{(n)}(S) \Rightarrow p(\underline{s}) \doteq 2^{-n(H(S) \pm \epsilon)}$
3. $|A_\epsilon^{(n)}(S)| \doteq 2^{n(H(S) \pm 2\epsilon)}$
4. Legyenek $S_1, S_2 \subseteq \{X_1, \dots X_k\}$! Ha $(\underline{s}_1, \underline{s}_2) \in A_\epsilon^{(n)}(S_1, S_2)$, akkor
$$p(\underline{s}_1 | \underline{s}_2) \doteq 2^{-n(H(S_1|S_2) \pm 2\epsilon)}$$

Biz: 1. A nagy számok tv-éből és $A_\epsilon^{(n)}(S)$ defjéből következik.

2. $A_\epsilon^{(n)}(S)$ defjét átrendezve, azt kapjuk.

3. $$1 \geq \sum_{\underline{s} \in A_\epsilon^{(n)}(S)} p(\underline{s}) \geq \sum_{\underline{s} \in A_\epsilon^{(n)}(S)} 2^{-n(H(S)+\epsilon)} = |A_\epsilon^{(n)}(S)|\, 2^{-n(H(S)+\epsilon)}$$

$$1-\epsilon \leq \sum_{\underline{s} \in A_\epsilon^{(n)}(S)} p(\underline{s}) \leq \sum_{\underline{s} \in A_\epsilon^{(n)}(S)} 2^{-n(H(S)-\epsilon)} = |A_\epsilon^{(n)}(S)|\, 2^{-n(H(S)-\epsilon)}$$

A kettőből következik az állítás.

4. Ha $(\underline{s}_1, \underline{s}_2) \in A_\epsilon^{(n)}(S_1, S_2)$, akkor $p(\underline{s}_1) \doteq 2^{-n(H(S_1) \pm \epsilon)}$ és $p(\underline{s}_2) = 2^{-n(H(S_2) \pm \epsilon)}$.

Ebből $p(\underline{s}_2 | \underline{s}_1) = \dfrac{p(\underline{s}_1, \underline{s}_2)}{p(\underline{s}_1)} \doteq 2^{-n(H(S_2|S_1) \pm 2\epsilon)}$ □

Tétel: Legyen $S_1, S_2 \subseteq \{X_1, \dots, X_k\}$, és jelölje $A_\epsilon^{(n)}(S_1 | \underline{s}_2)$ azon $\underline{s}_1 \in S_1$-ek halmazát, amik együttesen tipikusak $\underline{s}_2$-vel!
Ha $\underline{s}_2 \in A_\epsilon^{(n)}(S_2)$, akkor elég nagy n-re

$$|A_\epsilon^{(n)}(S_1|\underline{s}_2)| \leq 2^{n(H(S_1|S_2)+2\epsilon)} \quad \text{és} \quad (1-\epsilon)\, 2^{n(H(S_1|S_2)-2\epsilon)} \leq \sum_{\underline{s}_2} p(\underline{s}_2) |A_\epsilon^{(n)}(S_1|\underline{s}_2)|.$$

Biz: $$1 \geq \sum_{\underline{s}_1 \in A_\epsilon^{(n)}(S_1|\underline{s}_2)} p(\underline{s}_1|\underline{s}_2) \geq \sum_{\underline{s}_1 \in A_\epsilon^{(n)}(S_1|\underline{s}_2)} 2^{-n(H(S_1|S_2)+2\epsilon)} =$$

$$= |A_\epsilon^{(n)}(S_1|\underline{s}_2)|\, 2^{-n(H(S_1|S_2)+2\epsilon)}$$ ebből az első következik.

$$1-\epsilon \leq \sum_{\underline{s}_2} p(\underline{s}_2) \sum_{\underline{s}_1 \in A_\epsilon^{(n)}(S_1|\underline{s}_2)} p(\underline{s}_1|\underline{s}_2) \leq \sum_{\underline{s}_2} p(\underline{s}_2) \sum_{\underline{s}_1 \in A_\epsilon^{(n)}(S_1|\underline{s}_2)} 2^{-n(H(S_1|S_2)-2\epsilon)} =$$

$$= \sum_{\underline{s}_2} p(\underline{s}_2)\, |A_\epsilon^{(n)}(S_1|\underline{s}_2)|\, 2^{-n(H(S_1|S_2)-2\epsilon)}$$ ebből a második következik □

Tétel: Legyen $S_1, S_2, S_3 \subseteq \{X_1, \dots, X_k\}$ és tegyük fel, hogy S_1 és S_2 csak S_3-tól függ, azaz $\Pr(S_1 = \underline{s}_1, S_2 = \underline{s}_2, S_3 = \underline{s}_3) = \prod_{i=1}^{n} p(s_{1i}|s_{3i})\, p(s_{2i}|s_{3i})\, p(s_{3i})$!

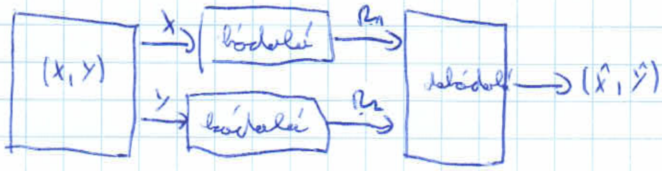
Ekkor $\Pr\left((S_1, S_2, S_3) \in A_\epsilon^{(n)}\right) \doteq 2^{n(I(S_1;S_2|S_3) \pm 6\epsilon)}$

Biz: $$P\left((S_1, S_2, S_3) \in A_\epsilon^{(n)}\right) = \sum_{(\underline{s}_1, \underline{s}_2, \underline{s}_3) \in A_\epsilon^{(n)}} p(\underline{s}_3)\, p(\underline{s}_1|\underline{s}_3)\, p(\underline{s}_2|\underline{s}_3) \doteq$$

$$\doteq \underbrace{|A_\epsilon^{(n)}(S_1, S_2, S_3)|}_{2^{n(H(S_1,S_2,S_3) \pm \epsilon)}}\, 2^{-n(H(S_3) \pm \epsilon)} \cdot 2^{-n(H(S_1|S_3) \pm 2\epsilon)} \cdot 2^{-n(H(S_2|S_3) \pm 2\epsilon)} \doteq$$

$$\doteq 2^{-n(I(S_1;S_2|S_3) \pm 6\epsilon)}$$ □

Tekintsük az alábbi szituációt: Két helyről akarunk információt eljuttatni egy harmadikba, de ezek nem korrelálatlanok. Mivel a kódolás külön történik, az elérhető rátájuk külön-külön $R_1 > H(X)$, $R_2 > H(Y)$. Ha viszont az információt együtt értelmeznénk, akkor a teljes ráta $R > H(X,Y)$. Az alábbiakban ezt vizsgáljuk meg formálisan.



Def: Egy $((2^{nR_1}, 2^{nR_2}), n)$ elosztott kód egy (X,Y) együttes eloszláshoz két kódoló fv-ből és egy együttes dekódoló fv-ből áll:

$$f_1 : \mathcal{X}^n \to \{1, 2, \dots, 2^{nR_1}\}$$

$$f_2 : \mathcal{Y}^n \to \{1, 2, \dots, 2^{nR_2}\}$$

$$g : \{1, 2, \dots, 2^{nR_1}\} \times \{1, 2, \dots, 2^{nR_2}\} \to \mathcal{X}^n \times \mathcal{Y}^n.$$

(R_1, R_2) a kódhoz tartozó rátapár.

Def: A hiba valószínűsége: $P_e^{(n)} = Pr\left(g(f_1(X^n), f_2(Y^n)) \neq (X^n, Y^n)\right)$

Def: Egy (R_1, R_2) rátapár elérhető, ha létezik olyan $((2^{nR_1}, 2^{nR_2}), n)$ kód, amire $P_e^{(n)} \to 0$.

VIGYÁZAT!!! Ez az R nem kompatibilis a csatorna kódolásnál látott R-rel! Bár meg lehetett volna oldani úgy is, de a könyv szerzői gyötörtek valahol.

Tétel (Slepian-Wolf): Egy elosztott kódra a (X,Y) forrásnál és és $p(x,y)$ valószínűségi eloszlásnál az elérhető ráták szükséges és elégséges feltétele:

$$R_1 \geq H(X|Y)$$

$$R_2 \geq H(Y|X)$$

$$R_1 + R_2 \geq H(X,Y).$$

A bizonyítás előtt vezessünk be egy kódot, amely nagy n-re tetszőleges $R > H(X)$ rátát elér!

Minden X^n szekvenciához rendeljünk hozzá véletlenszerűen egy indexet a $\{1, \dots, 2^{nR}\}$ halmazból, amelyeket hívjunk rekesznek! A dekódoláshoz a kapott index alapján csak azt nézzük meg, hogy adott rekeszben valóban egyetlen tipikus elem! Ha nincs ilyen, vagy ha több van, akkor dobjunk hibát!

Hiba akkor lehet ha
- a küldött szekvencia nem tipikus vagy
- másik tipikus szekvencia is került ugyanabba a vödörbe.

Nagy n esetén az első valószínűsége elhanyagolható, így ha sokkal több vödör van, mint tipikus elem, akkor a második is. Formálisan:

$$\Pr\left(g(f(\underline{x})) \neq \underline{x}\right) \leq \Pr\left(\underline{x} \notin A_\epsilon^{(n)}\right) + \sum_{\underline{x}} \Pr\left(\exists \underline{x}' \neq \underline{x} : \underline{x}' \in A_\epsilon^{(n)},\ f(\underline{x}') = f(\underline{x})\right) p(\underline{x})$$

$$\leq \epsilon + \sum_{\underline{x}} \sum_{\substack{\underline{x}' \in A_\epsilon^{(n)} \\ \underline{x}' \neq \underline{x}}} \Pr\left(f(\underline{x}') = f(\underline{x})\right) p(\underline{x})$$

$$\leq \epsilon + \sum_{\underline{x}} \sum_{\underline{x}' \in A_\epsilon^{(n)}} 2^{-nR} p(\underline{x}) = \epsilon + \sum_{\underline{x}' \in A_\epsilon^{(n)}} 2^{-nR} \underbrace{\sum_{\underline{x}} p(\underline{x})}_{1} \leq$$

$$\leq \epsilon + 2^{n(H(X)+\epsilon)} 2^{-nR} \leq 2\epsilon \quad \text{ha } R > H(X) + \epsilon.$$

Tehát minden $R > H(X)$ elérhető ily módon.

Slepian - Wolf - tétel bizonyítása:

- Elérhetőség: Az (R_1, R_2) rátapár elérhető az alábbi módon:

Minden X^n szekvenciához rendeljünk hozzá egy indexet a $\{1, \dots, 2^{nR_1}\}$ halmazból, és a Y^n szekvenciákhoz a $\{1, \dots, 2^{nR_2}\}$ halmazból!

Az (i_c, j_c) indexpárt dekódoljuk $(\underline{x}, \underline{y})$-ként, ha ez az egyetlen tipikus szekvenciapár a neki megfelelő vödörben! Ha a vödörben nincs tipikus pár, vagy több ilyen van, akkor dobjunk hibát!

Az alábbi hibaeseményeket tekintsük:

$$E_0 = \{(\underline{x}, \underline{y}) \notin A_\epsilon^{(n)}\}$$

$$E_1 = \{\exists \underline{x}' \neq \underline{x} : f_1(\underline{x}') = f_1(\underline{x}),\ (\underline{x}', \underline{y}) \in A_\epsilon^{(n)}\}$$

$$E_2 = \{\exists \underline{y}' \neq \underline{y} : f_2(\underline{y}') = f_2(\underline{y}),\ (\underline{x}, \underline{y}') \in A_\epsilon^{(n)}\}$$

$$E_{12} = \{\exists (\underline{x}', \underline{y}') : \underline{x}' \neq \underline{x},\ \underline{y}' \neq \underline{y},\ f(\underline{x}') = f(\underline{x}),\ f(\underline{y}) = f(\underline{y}'),\ (\underline{x}', \underline{y}') \in A_\epsilon^{(n)}\}$$

A hiba valószínűsége: $P_e = \Pr(E_0 \cup E_1 \cup E_2 \cup E_{12}) \leq \Pr(E_0) + \Pr(E_1) + \Pr(E_2) + \Pr(E_{12})$

$\Pr(E_0) < \varepsilon$ az AEP miatt.

$$\Pr(E_1) = \sum_{(\underline{x}, \underline{y})} p(\underline{x}, \underline{y}) \Pr\left(\exists \underline{x}' \neq \underline{x} : f_1(\underline{x}') = f(\underline{x}),\ (\underline{x}', \underline{y}) \in A_\epsilon^{(n)}\right) \leq$$

$$\leq \sum_{(\underline{x}, \underline{y})} p(\underline{x}, \underline{y}) \sum_{\substack{\underline{x} \neq \underline{x}' \\ (\underline{x}', \underline{y}) \in A_\epsilon^{(n)}}} \Pr\left(f_1(\underline{x}') = f(\underline{x})\right) = \sum_{(\underline{x}, \underline{y})} p(\underline{x}, \underline{y}) 2^{-nR_1} |A_\epsilon(X|\underline{y})| \leq$$

$$\leq 2^{-nR_1} 2^{n(H(X|Y)+\epsilon)} \to 0 \quad \text{ha } R_1 > H(X|Y).$$

$\Pr(E_2) \to 0$ ha $R_2 > H(Y|X)$ ugyanígy.

$\Pr(E_{12}) \to 0$ ha $R_1 + R_2 > H(X, Y)$ ugyanígy. $\Rightarrow P_e \to 0$

• szükségesség

Ha (R_1, R_2) elérhető, akkor $R_1 + R_2 \geq H(X,Y)$ egyszerűen önálló kódolási gyakorlatból gondolva rá.

Ha (R_1, R_2) elérhető, akkor $(R_1, H(Y))$ is, hiszen csak Y redundanciáját növelné.

Emiatt $R_1 + H(Y) \geq H(X,Y) \Rightarrow R_1 \geq H(X|Y)$

$(H(X), R_2)$-re ugyanígy $\Rightarrow R_2 \geq H(Y|X)$. □

A tétel kiterjeszthető tetszőleges sok változóra:

Tétel: Legyen $(X_{1i}, X_{2i}, \dots, X_{mi})$ iid-változók a $p(x_1, \dots, x_m)$ eloszlásból! Erre a valószínűségi változókra az elérhető rátára a szükséges és elégséges feltétel:

$$R(S) > H\left(X(S) \mid X(S^c)\right) \qquad \forall S \subseteq \{1, \dots, m\}\text{-re.}$$

ahol $R(S) = \sum_{i \in S} R_i$ és $X(S) = \{X_j : j \in S\}$.

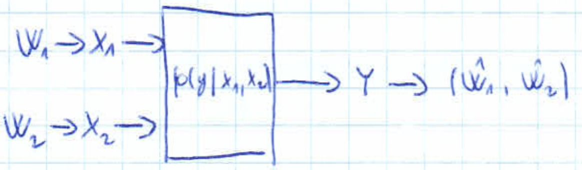
[Lineáris kódok]

BEV INFELM

12. előadás (05.14.)

Sok rendszerben találkozunk olyannal, hogy több forrás jelét kell értelmezni, és ezek egymásra hatnak. Pl.: műhold különböző adói, különböző telefoncsatornák, amelyek interferálnak, stb.. Ezek elvezetnek minket a többszörös hozzáférésű csatornához.

Def.: Egy diszkrét memóriamentes többszörös hozzáférésű csatorna 3 ábécéből (két küldő és egy fogadó) és egy átmenetvalószínűségből áll.



Def.: Egy $((2^{nR_1}, 2^{nR_2}), n)$ kód egy többszörös hozzáférésű csatornához a két indexhalmazból: $W_1 = \{1, 2, \dots, 2^{nR_1}\}$, $W_2 = \{1, 2, \dots, 2^{nR_2}\}$, két kódolóból:

$X_1 : W_1 \to X_1^n$, $X_2 : W_2 \to X_2^n$ és egy dekódolóból:

$g : Y^n \to W_1 \times W_2$ áll.

Tegyük fel, hogy az üzenetek küldése a $W_1 \times W_2$ egyenletes valószínűséggel történik!
Ekkor a hiba valószínűsége: $P_e^{(n)} = \frac{1}{2^{n(R_1+R_2)}} \sum_{(w_1,w_2) \in W_1 \times W_2} \Pr\left(g(Y^n) \neq (w_1, w_2) \mid (w_1, w_2)\right)$

Def: A (R_1, R_2) rátapár elérhető, ha létezik olyan $((2^{nR_1}, 2^{nR_2}), n)$ kód, amire $P_e^{(n)} \to 0$.

Def: A kapacitásrégió az elérhető (R_1, R_2) párok halmazának lezártja.

Tétel: Egy többszörös hozzáférésű csatorna kapacitásrégiója azon (R_1, R_2) párok konvex burkának a lezártja, amelyek teljesítik a

$$R_1 < I(X_1; Y | X_2)$$
$$R_2 < I(X_2; Y | X_1)$$
$$R_1 + R_2 < I(X_1, X_2; Y)$$

feltételeket valamilyen $p(x_1, x_2) = p(x_1)p(x_2)$ szorzateloszlásra.

Példák:
- független bináris szimmetrikus csatornák
- bináris szorzó csatorna
- bináris törlő többszörös hozzáférésű csatorna

Bizonyítás: Két részben bizonyítjuk, először azt, hogy a feltételeknek megfelelő (R_1, R_2)-k elérhetőek, aztán, hogy az elérhető (R_1, R_2)-ek szükséges feltétele a zártság.

1) Elérhetőség.

Az $\underline{X}_1(i)$ kódszót generáljuk n db i.i.d. változóval $p_1(x_{1i})$-ből $\forall i \in \{1, \dots, 2^{nR_1}\}$-re. Az $\underline{X}_2(j)$ kódszót ugyanígy $p_2(x_{2j})$-ből.

Legyen $A_\epsilon^{(n)}$ a tipikus $(\underline{x}_1, \underline{x}_2, \underline{y})$ szekvenciák halmaza! ~~Adott~~ ~~(i,j) kapott kódszóra a dekódoló adja vissza azt a y~~ vett. Adott $\underline{y}$ kapott kódra a dekódoló adja vissza azt a (i, j) indexpárt, amire $(\underline{x}_1(i), \underline{x}_2(j), \underline{y}) \in A_\epsilon^{(n)}$! Ha nincs ilyen, vagy több ilyen van, akkor dobjon hibát!

TFH az $(1,1)$ volt elküldve (a random generálás miatt mindegy)! Legyen
$$E_{ij} = \{(\underline{X}_1(i), \underline{X}_2(j), \underline{Y}) \in A_\epsilon^{(n)}\}$$
A hiba valószínűsége:
$$P_e^{(n)} = \Pr\left[E_{11}^c \cup \left(\bigcup_{(i,j)\neq(1,1)} E_{ij}\right)\right] \leq \Pr(E_{11}^c) + \sum_{\substack{i\neq 1\\ j=1}} \Pr(E_{ij}) + \sum_{\substack{i=1\\ j\neq 1}} \Pr(E_{ij}) + \sum_{\substack{i\neq 1\\ j\neq 1}} \Pr(E_{ij})$$

AEP miatt $\Pr(E_{11}^c) \to 0$. Előző órán belátott általános tételek miatt $i \neq 1$-re:
$$\Pr(E_{i1}) = \Pr\left((\underline{X}_1(i), \underline{X}_2(1), \underline{Y}) \in A_\epsilon^{(n)}\right) =$$
$$= \sum_{(\underline{x}_1, \underline{x}_2, \underline{y}) \in A_\epsilon^{(n)}} p(\underline{x}_1, \underline{x}_2, \underline{y}) = \sum_{(x_1, x_2, y) \in A_\epsilon^{(n)}} p(x_1)\, p(x_2, y)$$
mivel a random kódolás miatt i-től függetlenek a dolgok

$$\leq |A_\epsilon^{(n)}|\, 2^{-n(H(X_1)-\epsilon)}\, 2^{-n(H(X_2,Y)-\epsilon)} \leq$$
$$\leq 2^{-n(H(X_1)+H(X_2,Y)-H(X_1,X_2,Y)-3\epsilon)} = 2^{-n(I(X_1;X_2,Y)-3\epsilon)} = 2^{-n(I(X_1;Y|X_2)-3\epsilon)}$$

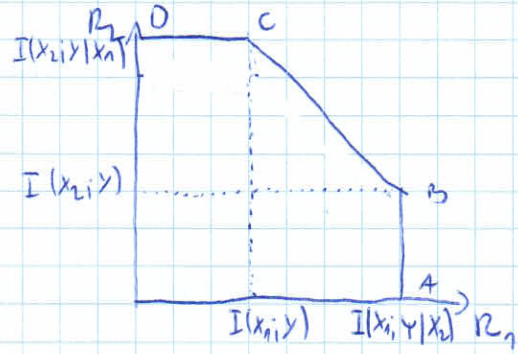
ugyanígy $\Pr(E_{1j}) \leq 2^{-n(I(X_2;Y|X_1)-3\epsilon)}$ és $\Pr(E_{ij}) \leq 2^{-n(I(X_1,X_2;Y)-4\epsilon)}$

Ebből
$$P_e^{(n)} \leq \Pr(E_{11}^c) + 2^{nR_1}\, 2^{-n(I(X_1;Y|X_2)-3\epsilon)} + 2^{nR_2}\, 2^{-n(I(X_2;Y|X_1)-3\epsilon)} +$$
$$+ 2^{n(R_1+R_2)}\, 2^{-n(I(X_1,X_2;Y)-4\epsilon)}$$

Ha $\epsilon \to 0$ és a (R_1, R_2) a feltételeknek megfelelnek, akkor $P_e^{(n)} \to 0$ $n \to \infty$-re.

[A $\Pr(\cdot)$-ke mindenhol képzeljük oda, hogy a feltétel, hogy az $(1,1)$ lett elküldve.]

megjegyzés: ez általában így néz ki



A pont ahol X_1 rátája maximális, és X_2 nem érdekel minket. Ez meg kijelölhető, hiszen

$$\max R_1 = \max_{p(x_1)p(x_2)} I(X_1;Y|X_2)$$

ahol $$I(X_1;Y|X_2) = \sum_{x_2} p_2(x_2)\, I(X_1;Y|X_2=x_2) \leq \max_{x_2} I(X_1;Y|X_2=x_2)$$

Tehát ha X_2 nem érdekel, akkor járunk a legjobban, ha az optimális x_2-t küldi állandóan.

B pont az a pont, ahol X_1 rátája a maximális, és a X_2-t is a lehető legjobban kiszűrjük. Ez olyan, mintha X_1 a zaj része lenne és X_2 érdekel.

C és D pont ugyanezek csak fordítva.

2) Megfordított: minden elérhető (R_1,R_2) ráta benne van a kérdéses halmazban.

A tétel megfordításának bizonyításához alakítsuk át a kapacitási régió alakját! Ehhez szükség van néhány lemmára.

Lemma 1.: A C kapacitási régió konvex, azaz ha $(R_1,R_2),(R_1',R_2') \in C$ akkor $(\lambda R_1+(1-\lambda)R_1', \lambda R_2+(1-\lambda)R_2') \in C$ $\forall \lambda \in [0,1]$.

Biz: Használjuk a vektor jelölést azaz $\underline{R} = (R_1,R_2)$. Ha $\underline{R}$ és $\underline{R}'$ elérhető, akkor $\lambda\underline{R}+(1-\lambda)\underline{R}'$ is az, ha az első kódkönyv λn betűjét és a második kódkönyv $(1-\lambda)n$ betűjét használjuk. Ekkor a lekódolható üzenetek száma az új kódban $2^{n\lambda R_1} \cdot 2^{n(1-\lambda)R_1'} = 2^{n(\lambda R_1+(1-\lambda)R_1)}$

A hiba valószínűsége az egyes részek összege, de az is $\to 0$. □

Ezután tisztázzuk a konvex kombináció használatát! Képzeljük el az alábbi 2 halmazt:

$C_1 = \{(x,y) : 0 \leq x \leq 10,\ 0 \leq y \leq 10,\ x+y \leq 100\}$

$C_2 = \{(x,y) : 0 \leq x \leq 20,\ 0 \leq y \leq 20,\ x+y \leq 20\}$.

Azt hihetnénk, hogy az $(\frac{1}{2},\frac{1}{2})$ eh.-kal vett kombinációjuk az alábbi:

$C = \{(x,y) : 0 \leq x \leq 15,\ 0 \leq y \leq 15,\ x+y \leq 60\}$

de könnyen látható, hogy $(15,15) \in C$ míg $(15,15) \notin C_1 \cup C_2$.
A tévedés oka, hogy C_1-ben a $x+y \leq 100$ feltétel nem aktív. Ennek elkerülésére szorítkozzunk kizárólag az ötszög alakú halmazokra!

Legyenek $I_1 := I(X_1;Y|X_2)$ $\quad I_2 := I(X_2;Y|X_1)$ $\quad I_3 := I(X_1,X_2;Y)$

és $\underline{I} = (I_1, I_2, I_3)$ $\quad$ adott $p(x_1,x_2) = p(x_1)p(x_2)$ és $p(y|x_1,x_2)$ eloszlásra

Az $\underline{I}$ által definiált régió:

$$C_{\underline{I}} = \{(R_1,R_2) : 0 \le R_1 \le I_1,\ 0 \le R_2 \le I_2,\ R_1+R_2 \le I_3\}$$

Mivel az (X_1, X_2) eloszlás szorzatalakú, ezért

$$I(X_2;Y|X_1) = H(X_2|X_1) - H(X_2|Y,X_1) = H(X_2) - H(X_2|Y,X_1) = I(X_2;Y,X_1) =$$

$$= I(X_2;Y) + I(X_2;X_1|Y) \ge I(X_2;Y)$$

Ebből $I(X_1;Y|X_2) + I(X_2;Y|X_1) \ge I(X_1;Y|X_2) + I(X_2;Y) = I(X_1,X_2;Y)$

azaz $I_1 + I_2 \ge I_3$ $\quad$ tehát a C terület valóban ötszög.

Lemma 2: Legyenek $\underline{I}_1$ és $\underline{I}_2$ ilyen vektorok és $C_{\underline{I}_1}$ és $C_{\underline{I}_2}$ a hozzájuk tartozó régiók! Legyen $\underline{I}_\lambda = \lambda \underline{I}_1 + (1-\lambda)\underline{I}_2$ és $C_{\underline{I}_\lambda}$ a hozzá tartozó régió!

Ekkor $\quad C_{\underline{I}_\lambda} = \lambda C_{\underline{I}_1} + (1-\lambda) C_{\underline{I}_2}$.

Biz: • $C_{\underline{I}_1}$ minden pontja teljesíti a def relációkat $\underline{I}_1$-gyel, és ugyanígy $C_{\underline{I}_2}$ minden pontja $\underline{I}_2$-vel. Ez alapján a $(\lambda, 1-\lambda)$-val kombinált pontok is teljesítik a $(\lambda, 1-\lambda)$-val kombinált relációkat, vagyis

$$\lambda C_{\underline{I}_1} + (1-\lambda) C_{\underline{I}_2} \subseteq C_{\underline{I}_\lambda}$$

• A $C_{\underline{I}}$ régiót az alábbi 5 pont konvex burka definiálja:

$$(0,0),\ (I_1,0),\ (I_1, I_3-I_1),\ (I_3-I_2, I_2),\ (0, I_2).$$

Ha $I_3 = I_1 + I_2$, akkor a 3. és a 4. is azonos, és végig $\underline{I}_\lambda$-ra ugyanígy felírható ez az 5 pont, $\underline{I}_1$ és $\underline{I}_2$ kombinációjaként. Mivel $C_{\underline{I}_\lambda}$ minden sarokpontja a $C_{\underline{I}_1}$ és $C_{\underline{I}_2}$ sarokpontjainak kombinációja, ezért mindet tartalmazza a kombináció, azaz

$$\lambda C_{\underline{I}_1} + (1-\lambda) C_{\underline{I}_2} \supseteq C_{\underline{I}_\lambda} \qquad \square$$

(ha $I_3 > I_1 + I_2$ lenne, akkor a $C_{\underline{I}_\lambda}$ valamelyik sarokpontja nem lenne benne a két mint halmaz kombinációjában.)

Lemma 3: Tekintsük azon (R_1,R_2) rátapárok halmazának lezártját, amelyek kielégítik a

$$R_1 < I(X_1;Y|X_2,Q)$$
$$R_2 < I(X_2;Y|X_1,Q)$$
$$R_1 + R_2 < I(X_1,X_2;Y|Q)$$

feltételeket valamilyen $p(x_1,x_2,y,q) = p(q)p(x_1|q)p(x_2|q)p(y|x_1,x_2)$ eloszlásra, ahol $|Q| \le 4$! ~~Ez megegyezik a kapacitás régióval.~~

Ekkor minden rátapár elérhető

Biz.: Legyen $\underline{R}$ egy olyan pár, amely benne van a Szántban!

$$I(X_1;Y|X_2,Q) = \sum_{q=1}^{m} p(q) I(X_1;Y|X_2,Q=q) =$$

$$= \sum_{q=1}^{m} p(q) I(X_1;Y|X_2)_{p_{1q},p_{2q}}.$$

ahol $m = |Q|$ és $p_{1q}(x_1) = p_1(x_1|q)$, $p_{2q}(x_2) = p_2(x_2|q)$.

Ugyanígy $I(X_2;Y|X_1,Q)$-ra.

Legyen $\underline{R}_q = (R_{1q}, R_{2q})$ olyan rátapár, amely a p_{1q} és p_{2q} eloszlással teljesíti az eredeti tétel feltételeit, ezen $\underline{R}_q$ elérhető az 54. o. miatt. Mivel $\underline{R}$ teljesíti a lemma 3 feltételit, azt felbonthatni ilyenekre:

$$\underline{R} = \sum_{q=1}^{m} p(q) \underline{R}_q$$

Lemma 1 miatt, az $\underline{R}_q$-k halmazai konvexek, és Lemma 2 miatt ezek konvex kombinációja is az információk konvex kombinációjuk elérhető rátái, tehát ha $\underline{R}_q$-k elérhetők, akkor $\underline{R}$ is. □

Megjegyzés: m felső korlátja a Carathéodory-tétel következménye. Eszerint, egy d-dimenziós Euklideszi téren vett kompakt halmaz bármely pontja reprezentálható, mint maximum $d+1$ pont konvex kombinációja. Mivel a $\underline{I}$ vektor 3D-s, ezért 4 db tagból lehet összerakni bármelyiket. Ez azért fontos, mert ha a Q változó bármilyen nagy ábécéből vehetne fel értékeket, akkor nem tudnánk hogy eggyel több q bevétele növelné-e a kapacitási régiót.

Most még rátérhetünk a tétel megfordítottjának bizonyítására. Azt kell belátni, hogy minden elérhető (R_1, R_2) pár benne van a Lemma 3-ban definiált halmazban.

Nézzük adott n-re a $W_1 \times W_2 \times X_1^n \times X_2^n \times Y^n$ halmaz elemeinek az eloszlását:

$$p(w_1,w_2,x_1^n,x_2^n,y^n) = 2^{-nR_1}\cdot 2^{-nR_2}\ p(x_1^n|w_1)\,p(x_2^n|w_2)\prod_{i=1}^{n} p(y_i|x_{1i},x_{2i})$$

A következőben a különböző infókat ezzel számoljuk ki.

Ha (R_1,R_2) elérhető, akkor Y^n-ből jól becsülhető (W_1,W_2), így a feltételes entrópiája kicsi:

$$H(W_1,W_2|Y^n) \le n(R_1+R_2)P_e^{(n)} + H(P_e^{(n)}) \triangleq n\epsilon_n$$

Ebből $H(W_1|Y^n) \le H(W_1,W_2|Y^n) \le n\epsilon_n$

$$H(W_2|Y^n) \le H(W_1,W_2|Y^n) \le n\epsilon_n$$

Ezzel a R_1 és R_2 ráták felülbecsülhetők:

$$nR_1 = H(W_1) = I(W_1;Y^n) + H(W_1|Y^n) \leq \text{ előzőből } \leq I(W_1;Y^n) + n\epsilon_n \leq$$

$$\leq I(X_1^n(W_1);Y^n) + n\epsilon_n = H(X_1^n(W_1)) - H(X_1^n(W_1)|Y^n) + n\epsilon_n \leq$$

$$\leq H(X_1^n(W_1)|X_2^n(W_2)) - H(X_1^n(W_1)|Y^n, X_2^n(W_2)) + n\epsilon_n =$$

$$= I(X_1^n(W_1);Y^n|X_2^n(W_2)) + n\epsilon_n =$$

$$= H(Y^n|X_2^n(W_2)) - H(Y^n|X_1^n(W_1), X_2^n(W_2)) + n\epsilon_n =$$

$$= H(Y^n|X_2^n(W_2)) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, X_1^n(W_1), X_2^n(W_2)) + n\epsilon_n =$$

$$= H(Y^n|X_2^n(W_2)) - \sum_{i=1}^{n} H(Y_i|X_{1i}(W_1), X_{2i}(W_2)) + n\epsilon_n \leq$$

$$\leq \sum_{i=1}^{n} H(Y_i|X_2^n(W_2)) - \sum_{i=1}^{n} H(Y_i|X_{1i}(W_1), X_{2i}(W_2)) + n\epsilon_n \leq$$

$$\leq \sum_{i=1}^{n} H(Y_i|X_{2i}(W_2)) - \sum_{i=1}^{n} H(Y_i|X_{1i}(W_1), X_{2i}(W_2)) + n\epsilon_n =$$

$$= \sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i}) + n\epsilon_n$$

$$\Rightarrow R_1 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i}) + \epsilon_n$$

ugyanígy $R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{2i};Y_i|X_{1i}) + \epsilon_n$

Az összegük pedig:

$$n(R_1+R_2) = H(W_1,W_2) = I(W_1,W_2;Y^n) + H(W_1,W_2|Y^n) \leq I(W_1,W_2;Y^n) + n\epsilon_n \leq$$

$$\leq I(X_1^n(W_1), X_2^n(W_2);Y^n) + n\epsilon_n = H(Y^n) - H(Y^n|X_1^n(W_1), X_2^n(W_2)) + n\epsilon_n =$$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y^{i-1}, X_1^n(W_1), X_2^n(W_2)) + n\epsilon_n =$$

$$= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_{1i}, X_{2i}) + n\epsilon \leq$$

$$\leq \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_{1i}, X_{2i}) + n\epsilon_n = \sum_{i=1}^{n} I(X_{1i}, X_{2i};Y_i) + n\epsilon$$

$$\Rightarrow R_1 + R_2 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i}, X_{2i};Y_i) + \epsilon_n$$

Az így kapott felső korlátok a kódkönyv i oszlopaiból vett eloszlások átlaga.
Bevezetve a $Q = i \in \{1, \dots, n\}$ változót

$$R_1 \leq \frac{1}{n}\sum_{i=1}^{n} I(X_{1i};Y_i|X_{2i}) + \epsilon_n = \frac{1}{n}\sum_{i=1}^{n} I(X_{1q};Y_q|X_{2q}, Q=i) + \epsilon_n =$$

$$= I(X_{1Q};Y_Q|X_{2Q}, Q) + \epsilon_n = I(X_1;Y|X_2,Q) + \epsilon$$

ahol $X_1 \triangleq X_{1Q}$, $X_2 \triangleq X_{2Q}$, $Y \triangleq Y_Q$

Mivel $X_{1i}(W_1)$ és $X_{2i}(W_2)$ függetlenek

$$\Pr\left(X_{1i}(W_1)=x_1,\ X_{2i}(W_2)=x_2\right) \stackrel{\Delta}{=} \Pr\left(X_{1Q}=x_1 \mid Q=i\right) \Pr\left(X_{2Q}=x_2 \mid Q=i\right)$$

Ez alapján tehát, ha $n\to\infty$-re $P_e^{(n)}\to 0$, akkor tudunk olyan

$p(q, x_1, x_2, y) = p(q)\,p(x_1|q)\,p(x_2|q)\,p(y|x_1,x_2)$ eloszlást választani, hogy

$$R_1 \subseteq I(X_1; Y \mid X_2, Q)$$

$$R_2 \subseteq I(X_2; Y \mid X_1, Q)$$

$$R_1 + R_2 \subseteq I(X_1, X_2; Y \mid Q)$$ teljesüljön.

Ha $|Q|$ lecsonkítjuk 4-re, a terület nem változik. $\square$