Proof La.4.1.

a) $H(Y)$

Determine the distribution of Y:

$$P(Y=y_j) = \sum_{i=1}^{m} P(Y=y_j \mid X=x_i)\, P(X=x_i)$$

$$= \sum_{i=1}^{m} p_i w_{ij} = (pW)_j \,, \quad j=1,\dots,d$$

b) $H(Y \mid X=x_i) = H(w_i)$ by definition

c) $H(Y\mid X) = \sum_{i=1}^{m} p_i\, H(Y \mid X=x_i) = \sum_{i=1}^{m} p_i\, H(w_i)$ ☒

Slide 6.

$$I(X;Y) = I(p,W) = H(pW) - \sum_{i=1}^{m} p_i H(w_i)$$

$$= H\Big(\Big[\sum_{i=1}^{m} p_i w_{ij}\Big]_{j=1,\dots,d}\Big) + \sum_{i=1}^{m} p_i \sum_{j=1}^{d} w_{ij} \log w_{ij}$$

$$= -\sum_{j=1}^{d}\Big(\sum_{i=1}^{m} p_i w_{ij}\Big) \log\Big(\sum_{i=1}^{m} p_i w_{ij}\Big) + \sum_{i,j} p_i w_{ij} \log w_{ij}$$

$$= -\sum_{i,j} p_i w_{ij} \log\Big(\sum_{\ell=1}^{m} p_\ell w_{\ell j}\Big) + \sum_{i,j} p_i w_{ij} \log w_{ij}$$

$$= \sum_i p_i \Big[\sum_j w_{ij} \log \frac{w_{ij}}{\sum_\ell p_\ell w_{\ell j}}\Big]$$

$$= \sum_i p_i\, D(w_i \,\|\, pW)$$ ☒

RWTH AACHEN UNIVERSITY

Example 4.3. (BSC)

$$W = \begin{pmatrix} 1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon \end{pmatrix}$$

Mutual information $p = (p_0, p_1)$

$$I(X;Y) = \cancel{H}(pW) - \sum_{i=1}^{m} p_i H(w_i)$$

$$= H\big(p_0(1-\varepsilon) + p_1\varepsilon,\ \varepsilon p_0 + (1-\varepsilon)p_1\big) - p_0 H(1-\varepsilon, \varepsilon) - p_1 H(\varepsilon, 1-\varepsilon)$$

$$= H_2\big(p_0(1-\varepsilon) + p_1\varepsilon\big) - H_2(\varepsilon)$$

to be maximized over (p_0, p_1), $p_0, p_1 \geq 0$, $p_0 + p_1 = 1$:

$$H_2(q) = -q \log q - (1-q)\log(1-q), \quad 0 \leq q \leq 1$$

$$\leq \log 2$$

with equality if $q = \frac{1}{2}$ (Th. 2.1.8 a))

Hence $H_2(p_0(1-\varepsilon) + p_1\varepsilon)$ is maximal if

$$p_0(1-\varepsilon) + p_1\varepsilon = \frac{1}{2} .$$

This is achieved if $p_0 = p_1 = \frac{1}{2}$.

Capacity-achieving distribution is $p^* = \left(\frac{1}{2}, \frac{1}{2}\right)$

with capacity

$$C = \max_{(p_0, p_1)} I(X;Y) = \log 2 + (1-\varepsilon)\log(1-\varepsilon) + \varepsilon \log(\varepsilon)$$

$$= 1 + (1-\varepsilon)\log_2(1-\varepsilon) + \varepsilon \log_2 \varepsilon .$$

Thinking the Future
Zukunft denken

Proof of Th. 4.4.

$$\frac{\partial}{\partial p_k} H(pW)$$

$$= \frac{\partial}{\partial p_k}\left[-\sum_j \left(\sum_i p_i w_{ij}\right)\log\left(\sum_i p_i w_{ij}\right)\right]$$

$$= -\sum_j \left[w_{kj} \log\left(\sum_i p_i w_{ij}\right) + \cancel{\left(\sum_i p_i w_{ij}\right)} \frac{w_{kj}}{\cancel{\sum_i p_i w_{ij}}}\right]$$

$$= -\sum_j \left[w_{kj} \log\left(\sum_i p_i w_{ij}\right) + w_{kj}\right]$$

$$\frac{\partial}{\partial p_k} I(p,W) = \frac{\partial}{\partial p_k} H(pW) - \frac{\partial}{\partial p_k}\left(\sum_i p_i H(w_i)\right)$$

$$= -\sum_j w_{kj} \log\left(\sum_i p_i w_{ij}\right) + \sum_j w_{kj} \log w_{kj} \quad -1$$

$$= \sum_j w_{kj} \log \frac{w_{kj}}{\sum_i p_i w_{ij}} - 1$$

$$= D(w_k \| pW) - 1$$

Thinking the Future
Zukunft denken

RWTH AACHEN UNIVERSITY

Th. 4.5. (Proof)

p capacity-achieving iff $D(w_i \| pW) = \zeta \quad \forall i : p_i > 0$

Let $\rho(q) = -q \log q, \; q \geq 0$

T inverse of W, $T = W^{-1}$ so that

$$T 1_m = T W 1_m = I 1_m = 1_m$$

It holds:

$$\zeta = D(w_i \| pW)$$

$$= \sum_j w_{ij} \ln \frac{w_{ij}}{\sum_{\ell=1}^{m} p_\ell w_{\ell j}}$$

$$= -\sum_j \left[w_{ij} \ln \left(\sum_\ell p_\ell w_{\ell j} \right) + \rho(w_{ij}) \right], \quad i = 1, \dots, m$$

Hence $\forall k = 1, \dots, m$ by summation over i

$$\zeta \underbrace{\left(\sum_i t_{ki} \right)}_{=1} = -\sum_i t_{ki} \sum_j \left[w_{ij} \ln \left(\sum_\ell p_\ell w_{\ell j} \right) + \rho(w_{ij}) \right]$$

$$= -\sum_j \underbrace{\sum_i t_{ki} w_{ij}}_{\delta_{kj}} \ln \left(\sum_\ell p_\ell w_{\ell j} \right) - \sum_{i,j} t_{ki} \rho(w_{ij})$$

$$\overset{\uparrow}{=} -\ln \left(\sum_\ell p_\ell w_{\ell k} \right) - \sum_{i,j} t_{ki} \rho(w_{ij})$$

Resolve for $p = (p_1, \dots, p_m)$:

$$\sum_\ell p_\ell w_{\ell k} = \exp\left(-\zeta - \sum_{i,j} t_{ki} \rho(w_{ij}) \right) \quad \forall k = 1, \dots, m \quad (*)$$

Thinking the Future
Zukunft denken

Summation over k:

$$1 = \sum_k \exp\left(-\xi - \sum_{i,j} t_{ki}\, p(w_{ij})\right)$$

$$= \sum_k e^{-\xi}\, e^{-\sum_{i,j} t_{ki}\, p(w_{ij})}$$

It follows

capacity

$$\boxed{\xi = \ln\left(\sum_k e^{-\sum_{i,j} t_{ki}\, p(w_{ij})}\right) = C} \quad \text{(capacity)}$$

$\xi = C$ = capacity because of slide 6.

To determine p_e multiply (**) by t_{ks} and sum over k.

$$\sum_k t_{ks} \sum_e p_e w_{ek} = \sum_k t_{ks}\, e^{-C}\, e^{-\sum_{i,j} t_{ki}\, p(w_{ij})}$$

$$\sum_e p_e \underbrace{\sum_k w_{ek} t_{ks}}_{\delta_{es}}$$

$$\boxed{p_s = e^{-C} \sum_k t_{ks}\, e^{-\sum_{i,j} t_{ki}\, p(w_{ij})}, \quad s = 1, \dots, n.}$$

capacity-achieving distr.