**Thm.:** (bounds on Shannon entropy)

Let $p \in \mathbb{R}^n$ be a probability distribution and define $H_2(p) := -\log \|p\|_2^2$. Then

$$0 \overset{(i)}{\leq} H_2(p) \overset{(ii)}{\leq} H(p) \overset{(iii)}{\leq} H_0(p) \overset{(iv)}{\leq} \log n$$

where equality holds in

(i) iff $\exists x: p(x) = 1$

(ii) iff $\exists m \in \{1, \dots, n\}\ \forall x: p(x) \in \{0, \frac{1}{m}\}$

(iii) iff $-"-$

(iv) iff $\forall x: p(x) > 0$

**proof:** (i) $-\log \sum_x p(x)^2 \geq -\log \sum_x p_x = 0$

$\uparrow$ '=' iff $\forall x: p(x)^2 = p(x) \Leftrightarrow \forall x: p(x) \in \{0,1\}$

(ii) $H(p) = -\sum_x p(x) \log p(x) \geq -\log \sum_x p(x)^2$

$\uparrow$

$-\log$ is strictly convex

(iii) $H(p) = \sum_{x \in X_0} p(x) \log \frac{1}{p(x)} \leq \log \sum_{x \in X_0} \frac{p(x)}{p(x)} = H_0(p)$

$\uparrow$

Log is strictly concave

(iv) ✓ □

## II.2. Conditional entropy & mutual information

**Def.:**

- "conditional entropy" $H(X|Y) := H(X,Y) - H(Y)$
- "mutual information" $I(X:Y) := H(X) + H(Y) - H(X,Y)$
- "cond. mutual info." $I(X:Y|Z) := H(X,Z) + H(Z,Y) - H(X,Y,Z) - H(Z)$

**Interpretation:** • $H(X|Y) = \sum_y p(y) \underbrace{\left( \sum_x p(x|y) \log p(x|y)^{-1} \right)}$

entropy of $X$ if $Y=y$ is known

= average uncertainty about $X$ if $Y$ is known

• $I(X:Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = I(Y:X)$

= reduction of uncertainty (increase of information) about $X$ after learning $Y$ (and vice versa)

= information of $X$ about $Y$ and v.v.

• $I(X:Y|Z) = H(X|Z) - H(X|Z,Y)$

= reduction of uncertainty about $X$ when we learn $Y$ and already know $Z$.

**Thm.:** a) $H(X|Y) \geq 0$ with '=' iff $\forall y \, \exists x: p(x,y) = p(y)$

b) $I(X:Y) \geq 0$ with '=' iff $\forall x,y: p(x,y) = p(x)p(y)$

c) $I(X:Y|Z) \geq 0$

**proof:** a) $H(X,Y) - H(X) = \sum_{x,y} p(x,y) \log \underbrace{\left( \frac{\sum_{y'} p(x,y')}{p(x,y)} \right)}_{\geq 1} \geq 0$

"=" $\Leftrightarrow \forall x,y: \; p(x,y) = 0 \vee p(x) = p(x,y)$

b) $I(X:Y) = -\sum_{x,y} p(x,y) \log \left( \frac{p(x)p(y)}{p(x,y)} \right) \geq -\log \sum_{x,y} p(x)p(y) = 0$

$\uparrow$ $-\log$ strictly convex

'=' iff $\frac{p(x)p(y)}{p(x,y)} = \text{const.} = 1$

$\uparrow$ normalization

c) $\rightarrow$ exercise ...

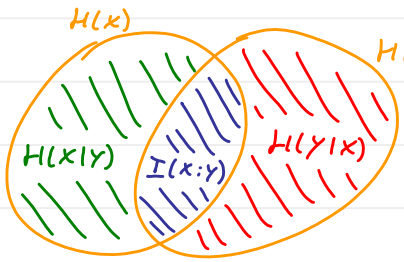
□

**remark:** a) $\Leftrightarrow H(X,Y) \geq H(Y)$ can be seen as:

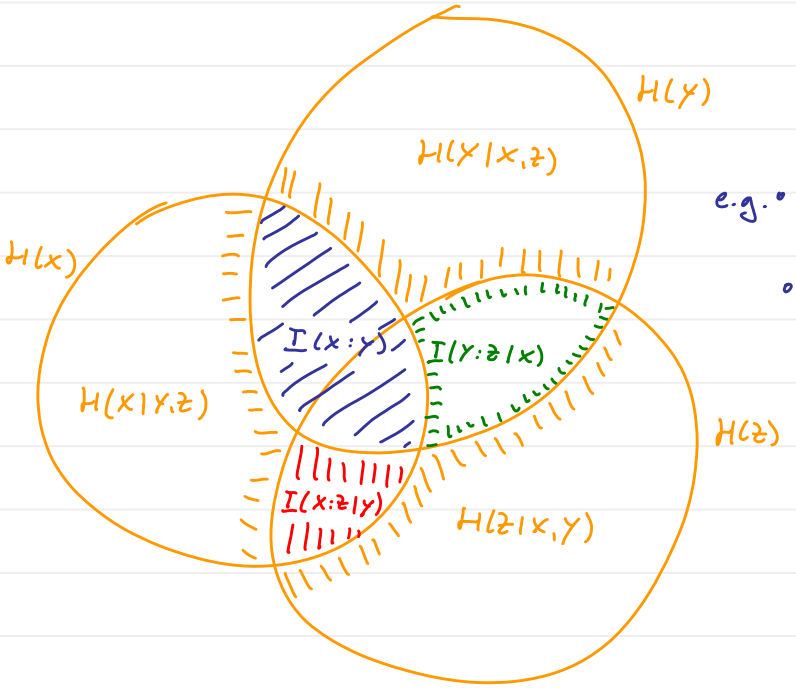
the entropy of a subsystem never exceeds the entropy of the whole system.

## Venn-diagrams:

graphical depiction of relations between entropic quantities in terms of relations between sets:



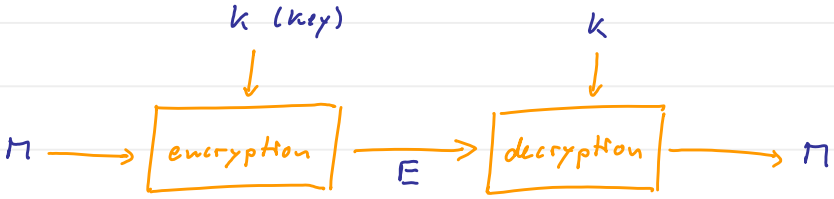

e.g. $I(X:Y) + H(Y|X) = H(Y)$




e.g. • $H(Y|X,Z) + I(Y:Z|X) = H(Y|X)$

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

# II.3. Application for crypto systems




**Def.:** We say that random variables $M, E, K$ describe a „perfectly secure crypto system" iff

(i) $I(M:E) = 0$ (E contains no info about M without K)

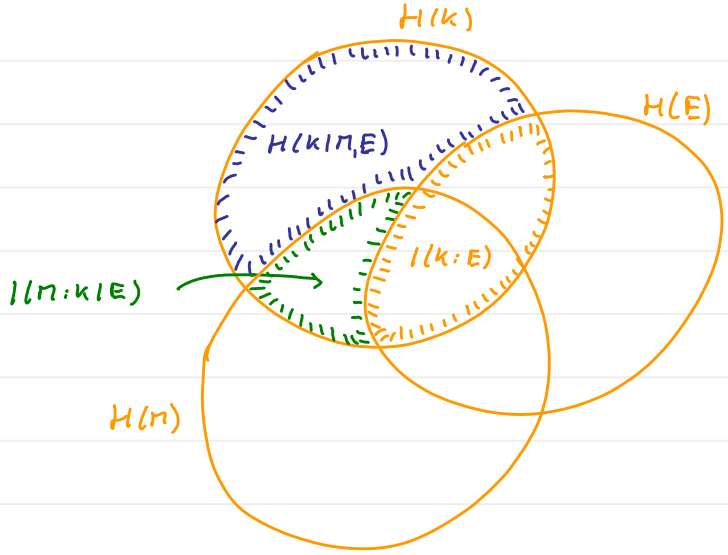
(ii) $H(M|KE) = 0$ (once E and K are known, M can be perfectly recovered)

**Thm.:** (Shannon '49)

A perfectly secure crypto system requires $H(K) \geq H(M)$.

**proof:** $H(K) = I(M:K|E) + \underbrace{I(K:E)}_{\geq 0} + \underbrace{H(K|M,E)}_{\geq 0}$

$\geq I(M:K|E) = H(M) - \underbrace{H(M|K,E)}_{=0} - \underbrace{I(M:E)}_{=0} = H(M)$ □




**remark:** • loosely speaking, this means that the key must not be shorter than the message.

• in practice, of course, weaker requirements are imposed.

## II.4. Chain rules

**Thm.:** (a) $H(X_1,\dots,X_n) = \sum_{i=1}^{n} \underbrace{H(X_i \mid X_{i-1},\dots,X_1)}_{:= H(X_1) \text{ for } i=1}$

(b) $H(X_1,\dots,X_n \mid Y) = \sum_{i=1}^{n} \underbrace{H(X_i \mid X_{i-1},\dots,X_1, Y)}_{:= H(X_1 \mid Y) \text{ for } i=1}$

(c) $I(X_1,\dots,X_n : Y) = \sum_{i=1}^{n} \underbrace{I(X_i : Y \mid X_{i-1},\dots,X_1)}_{:= I(X_1 : Y) \text{ for } i=1}$

**proof** (sketch): for (a) use $p(x_1,\dots,x_n) = \prod_{i=1}^{n} \underbrace{\frac{p(x_i,\dots,x_1)}{p(x_{i-1},\dots,x_1)}}_{\leftarrow 1 \text{ for } i=1}$

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

and similarly for (b).

(c): $I(X_1,\dots,X_n : Y) = H(X_1,\dots,X_n) - H(X_1,\dots,X_n \mid Y)$

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

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

□