1) Entropy

X descrete r.v. on $\mathcal{X}$ with $|\mathcal{X}| = K < \infty$

i.e. $\mathcal{X} = \{x_1, \dots, x_K\}$

a)

Let $p(X=x) = p(x)$ be the pmf of X, then the entropy is

$$H(X) = -\sum_{i=1}^{K} p(x_i) \log p(x_i)$$

Since $p(x_i)$ is a pmf, $p(x_i) \in [0,1]$.

Three cases:

i) $p(x_i) = 0$

We have that $p(x_i) \log p(x_i) = 0$ if $p(x_i) = 0$, by definition.

ii) $p(x_i) = 1$

Then $p(x_i) \log p(x_i) = 0$ since $\log 1 = 0$

iii) $p(x_i) \in (0,1)$,

then $p(x_i) \log p(x_i) < 0$

So

$H(X) = -\sum_{i=1}^{K} p(x_i) \log p(x_i)$ is a sum of non-negative terms (because of the minus sign) so it is true that

$$\boxed{H(X) \geq 0}$$

Because it is a sum of non-negative terms, the only way $H(X) = 0$ is when all terms in the sum are zero.

Because $P(x)$ is a pdf, $\sum_{i=1}^{K} p(x_i) = 1$, so the only way this is possible is that $p(x_j) = 1$ for some $j \in \{1, .., K\}$ & $p(x_i) = 0 \ \forall\ i \neq j$.

But assigning weight 1 to x_j & zero to all others means that sampling p returns x_j w/ 100% probability, and never the others. So X is a constant with value x_j

b) $p(x)$: pmf of X, $q(x)$: uniform on $\mathcal{X}$. i.e.

$$q(x) = \frac{1}{K} \quad \text{for } x \in \mathcal{X} = \{x_1, .., x_K\}$$

The KL divergence is

$$D(p\|q) = \sum_{i=1}^{K} p(x_i) \log \frac{p(x_i)}{q(x_i)}$$

$$= \underbrace{\sum_{i=1}^{K} p(x_i) \log p(x_i)}_{= -H(p)} - \sum_{i=1}^{K} p(x_i) \log q(x_i)$$

$$= -H(p) - \sum_{i=1}^{K} p(x_i) \log \frac{1}{K} \qquad \left(q(x_i) = \frac{1}{K} \right)$$

$$= -H(p) + \log K \underbrace{\sum_{i=1}^{K} p(x_i)}_{=1}$$

$$\Rightarrow \boxed{KL(p\|q) = -H(p) + \log K} \qquad q \text{ uniform.}$$

c) I found in b) that $KL(p\|q) = -H(p) + \log K$ when q is uniform

$$\Rightarrow H(p) = \log K - KL(p\|q)$$

Let me show that $KL(p\|p')$ is always non-negative, I will need the fact that $\log a \leq a-1$ from real analysis:

$$-KL(p\|p') = -\sum_x p(x) \log \frac{p(x)}{p'(x)}$$

$$= \sum_x p(x) \log \frac{p'(x)}{p(x)}$$

$$\leq \sum_x p(x) \left(\frac{p'(x)}{p(x)} - 1 \right)$$

$$= \sum_x p'(x) - \sum_x p(x) = 1 - 1 = 0.$$

$$\Rightarrow -KL(p\|p') \leq 0 \Rightarrow KL(p\|p') \quad \forall\, p, p'$$

Back to $H(p) = \log K - \underbrace{KL(p\|q)}_{\geq 0}$, q uniform.

So $\boxed{H(p) \leq \log K.}$ upper bound on $H(p)$

The equality is a satisfied when $KL(p\|q)$ is 0, i.e. when p is the same dist. as q, i.e. p is uniform!

So the distribution of maximum entropy on $\mathcal{X}$ is the uniform distribution.

Mutual Information

$$I(X_1,X_2) \equiv \sum_{(x_1,x_2)\in \mathcal{X}_1\times\mathcal{X}_2} p_{12}(x_1,x_2)\log\frac{P_{12}(x_1,x_2)}{P_1(x_1)P_2(x_2)}$$

a) I'm going to prove that $I \geq 0$ by using the same trick: $\log a \leq a-1$

$$-I = -\sum_{x_1,x_2} P_{12}(x_1,x_2)\log\frac{P_{12}(x_1,x_2)}{P_1(x_1)P_2(x_2)}$$

$$= \sum_{x_1,x_2} P_{12}(x_1,x_2)\log\frac{P_1(x_1)P_2(x_2)}{P_{12}(x_1,x_2)}$$

$$\leq \sum_{x_1,x_2} P_{12}(x_1,x_2)\left(\frac{P_1(x_1)P_2(x_2)}{P_{12}(x_1,x_2)}-1\right) \quad (\log a \leq a-1)$$

$$= \underbrace{\sum_{x_1,x_2} P_{12}(x_1,x_2)}_{=1} - \sum_{x_1,x_2} P_1(x_1)P_2(x_2)$$

$$= 1 - \underbrace{\sum_{x_1} p(x_1)}_{=1}\underbrace{\sum_{x_2} p(x_2)}_{=1}$$

$$= 1-1 = 0$$

$$\Rightarrow -I(X_1,X_2) \leq 0 \Rightarrow \boxed{I(X_1,X_2) \geq 0} \quad \forall\, x_1,x_2 \in \mathcal{X}_1\times\mathcal{X}_2.$$

b) Entropy: $H(p) = -\sum_x p(x)\log p(x)$

$$I(X_1,X_2) = \sum_{x_1,x_2} P_{12}(x_1,x_2)\log\frac{P_{12}(x_1,x_2)}{P_1(x_1)P_2(x_2)}$$

$$= \underbrace{\sum_{x_1,x_2} P_{12}(x_1,x_2)\log P_{12}(x_1,x_2)}_{=H(Z)} - \sum_{x_1,x_2} P_{12}(x_{1,2})\log P_1(x_1)P_2(x_2)$$

$$= -H(Z) - \sum_{x_1,x_2} P_{12}(x_1,x_2)\log P_1(x_1) - \sum_{x_1,x_2} P_{12}(x_1,x_2)\log P_2(x_2)$$

$$= -H(Z) - \sum_{x_1}\Big(\big(\underbrace{\sum_{x_2} P_{12}(x_1,x_2)}_{=P_1(x_1)}\big)\log P_1(x_1)\Big) - \sum_{x_2}\Big(\big(\underbrace{\sum_{x_1} P_{12}(x_1,x_2)}_{=P_2(x_2)}\big)\log P_2(x_2)\Big)$$

$$= -H(Z) - \sum_{x_1} P_1(x_1)\log P_1(x_1) - \sum_{x_2} P_2(x_2)\log P_2(x_2)$$

$$\Rightarrow \boxed{I(X_1,X_2) = H(X_1)+H(X_2)-H(Z)}\ , \quad Z=(X_1,X_2)$$

c) I've shown that $I(X_1,X_2) = H(X_1)+H(X_2)-H(Z)$, $Z=(X_1,X_2)$

$$\Rightarrow H(Z) = H(X_1)+H(X_1)\underbrace{-I(X_1,X_2)}_{\le 0 \text{ since } I \geq 0 \text{ (shown in a))}}$$

So:

$$\boxed{H(Z) \le H(X_1)+H(X_2)}$$

$H(Z)$ is maximal when $I(X_1,X_2)$ is equal to zero. This happens when X_1 & X_2 are independent, because

$$I(X_1,X_2) = \sum_{x_1,x_2} P_{12}(x_1,x_2)\log \frac{P_{12}(x_1,x_2)}{P_1(x_1)P_2(x_2)}$$

$$= \sum_{x_1 x_2} P_{12}(x_1, x_2) \log \frac{P_1(x_1) P_2(x_2)}{P_1(x_1) P_2(x_2)}$$

$\left(\begin{array}{l} X_1 \,\&\, X_2 \text{ independent} \\ \Leftrightarrow P_{12}(x_1, x_2) = P_1(x_1) P_2(x_2) \end{array} \right)$

$$= \sum_{x_1, x_2} P_{12}(x_1, x_2) \log 1 = \underline{\underline{0}}$$

So for given marginals $P_1(x_1)$ & $P_2(x_2)$, the joint w/ maximal entropy is

$$\boxed{P_{12}(x_1, x_2) = P_1(x_1) P_2(x_2)}$$