Homework 1 Solutions

- **Problem 1.** Let X be a discrete RV. Show that the entropy of a function of X is less than or equal to the entropy of X.

Solution

Let $y = g(x)$. Then

$$p(y) = \sum_{x:\, y=g(x)} p(x).$$

Consider any set of x's that map onto a single y. For this set

$$\sum_{x:\, y=g(x)} p(x) \log p(x) \leq \sum_{x:\, y=g(x)} p(x) \log p(y) = p(y) \log p(y),$$

since log is a monotone increasing function and $p(x) \leq \sum_{x:\, y=g(x)} p(x) = p(y)$. Extending this argument to the entire range of X (and Y), we obtain

$$\begin{aligned} H(X) &= -\sum_x p(x) \log p(x) \\ &= -\sum_y \sum_{x:\, y=g(x)} p(x) \log p(x) \\ &\geq -\sum_y p(y) \log p(y) \\ &= H(Y), \end{aligned}$$

with equality iff g is one-to-one with probability one.

- **Problem 2.** A function $f : 2^{\Omega} \to \mathbb{R}$ is said to be submodular if $\forall\, S \subseteq S' \subseteq \Omega$ and $\forall\, z \notin S'$, one has

$$f(S \cup \{z\}) - f(S) \geqslant f(S' \cup \{z\}) - f(S').$$

  1. Show that entropy and mutual information are submodular functions.

Solution

In this case Ω is a set of random variables. We use capital letters to denote sets of random variables (i.e. $X \subset \Omega$) and lower case letters to denote individual random variables $x \in \Omega$. Note that this contrasts with the usual notation of using capital letters for random variables and lower case letters for realizations of a random variable.

To show that $H : 2^{\Omega} \to [0, \infty)$ is submodular consider:

$$\underbrace{H(X, z) - H(X)}_{= \underbrace{H(z|X)}_{\text{cond.entropy}}} \geq \underbrace{H(Y, z) - H(Y)}_{= \underbrace{H(z|Y)}_{\text{cond.entropy}}} \tag{3.2}$$

where $H(z|Y) = H\left(z|X \cup (Y \setminus X)\right) \leq H(z|X)$, since conditioning cannot increase entropy.

$I(X, \Omega) = H(X) - H(X|\Omega) = H(X)$, which is therefore submodular (as function of X)

$I(X, \Omega \setminus X) = H(X) + H(\Omega \setminus X) - H(\Omega)$ is submodular (as function of X). In fact we have:

$$\begin{aligned} A_X &\equiv I(X \cup \{z\}, \Omega \setminus (X \cup \{z\})) - I(X, \Omega \setminus X) \\ &= H(X \cup \{z\}) + H(\Omega \setminus (X \cup \{z\})) - H(X) - H(\Omega \setminus X) \\ &= [H(X \cup \{z\}) - H(X)] + [H(\Omega \setminus (X \cup \{z\})) - H(\Omega \setminus X)] \end{aligned}$$

and similarly for A_Y. By submodularity of entropy, we have

$$H(X \cup \{z\}) - H(X) \geq H(Y \cup \{z\}) - H(Y)$$

and

$$H(\Omega \setminus Y) - H(\Omega \setminus (Y \cup \{z\})) \geq H(\Omega \setminus X) - H(\Omega \setminus (X \cup \{z\}))$$

since $\Omega \setminus (Y \cup \{z\}) \subseteq \Omega \setminus (X \cup \{z\})$. Therefore $A_X \geq A_Y$.

- **Problem 3.** Suppose that one has n coins, among which there may or may not be one counterfeit coin. If there is a counterfeit coin, it may be either heavier or lighter than the other coins. The coins are to be weighed by a balance. Find an upper bound on the number of coins n so that k weighings will find the counterfeit coin (if any) and correctly declare it to be heavier/lighter. Try to use information-theoretic arguments.

Solution

a) For n coins, there are $2n + 1$ possible situations or "states".

- One of the n coins is heavier.
- One of the n coins is lighter.
- They are all of equal weight.

Each weighing has three possible outcomes - equal, left pan heavier or right pan heavier. Hence with k weighings, there are 3^k possible outcomes and hence we can distinguish between at most 3^k different "states". Hence $2n + 1 \leq 3^k$ or $n \leq (3^k - 1)/2$.

Looking at it from an information theoretic viewpoint, each weighing gives at most $\log_2 3$ bits of information. There are $2n+1$ possible "states", with a maximum entropy of $\log_2(2n+1)$ bits. Hence in this situation, one would require at least $\log_2(2n + 1)/\log_2 3$ weighings to extract enough information for determination of the odd coin, which gives the same result as above.

- **Problem 4.** Three squares have average area $\bar{a} = 100\mathrm{m}^2$. The average of the lengths of their sides is $\bar{l} = 10\mathrm{m}$. What can be said about the area of the larges square?

Solution

Solution: Let x be the length of the side of a square, and let the probability of x be $(1/3, 1/3, 1/3)$ over the three lengths (l_1, l_2, l_3). Then the information that we have is that $E[x] = 10$ and $E[f(x)] = 100$, where $f(x) = x^2$ is the function mapping lengths to areas. This is a strictly convex function. We notice that the equality $E[f(x)] = f(E[x])$ holds, therefore x is a constant, and the three lengths must all be equal. The area of the largest square is $100m^2$.

- **Problem 5.** The Rényi entropy of order $\alpha \geqslant 0$, $\alpha \neq 1$ of a discrete RV X supported on a set of cardinality M is defined as

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\sum_i^M p_i^\alpha \right).$$

1. Show that $H_0 \geqslant H_1 \geqslant H_2 \geqslant H_\infty$. Observe that the subscripts 1 and ∞ are to be taken in the sense of a limit (i.e., $\alpha \to 1$, $\alpha \to \infty$, respectively).
2. Show that Rényi entropy is non-negative, and that it is concave for $\alpha \leqslant 1$.

Solution

Proof. For $0 \leq \alpha < \beta$ with $\alpha \neq 1$ and $\beta \neq 1$,

$$\begin{aligned}
H_\alpha(X) &= \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} P_X(x)^\alpha \\
&= -\log \mathrm{E}\Big[P_X(X)^{\alpha-1}\Big]^{\frac{1}{\alpha-1}} \\
&= -\log \mathrm{E}\Big[P_X(X)^{\alpha-1}\Big]^{\frac{\beta-1}{\alpha-1}\frac{1}{\beta-1}} \\
&\geq -\log \mathrm{E}\Big[P_X(X)^{(\alpha-1)\frac{\beta-1}{\alpha-1}}\Big]^{\frac{1}{\beta-1}} \\
&= -\log \mathrm{E}\Big[P_X(X)^{\beta-1}\Big]^{\frac{1}{\beta-1}} \\
&= \frac{1}{1-\beta} \log \sum_{x \in \mathcal{X}} P_X(x)^\beta \\
&= H_\beta(X).
\end{aligned}$$

We observe that x^c is convex-$\cup$ for $c \geq 1$ or $c \leq 0$ and convex-$\cap$ for $0 \leq c \leq 1$. The inequality in the above derivation follows from the Jensen inequality in the following cases:

$\beta > \alpha > 1:\ c = \frac{\beta-1}{\alpha-1} > 1$, x^c is convex-$\cup$ and $\frac{1}{\beta-1} > 0$;

$\beta > 1 > \alpha \geq 0:\ c = \frac{\beta-1}{\alpha-1} < 0$, x^c is convex-$\cup$ and $\frac{1}{\beta-1} > 0$;

$1 > \beta > \alpha \geq 0:\ 1 > c = \frac{\beta-1}{\alpha-1} > 0$, x^c is convex-$\cap$ and $\frac{1}{\beta-1} < 0$.

For $\alpha = 1$ or $\beta = 1$, the Jensen inequality can be applied directly. The conditions for equality in (2.14) follow directly from the Jensen inequality. $\square$

$$H_\alpha(\lambda P+(1-\lambda)Q)=\frac{1}{1-\alpha}\log\left[\Sigma(\lambda p_i+(1-\lambda)q_i)^\alpha\right] \quad (14)$$

$$>\frac{1}{1-\alpha}\log\left[\Sigma(\lambda p_i^\alpha+(1-\lambda)q_i^\alpha)\right] \quad (15)$$

$$=\frac{1}{1-\alpha}\log\left[\lambda\Sigma p_i^\alpha+(1-\lambda)\Sigma q_i^\alpha\right]. \quad (16)$$

Since the log function is concave, (16) is strictly greater than

$$\frac{1}{1-\alpha}\left[\lambda\log(\Sigma p_i^\alpha)+(1-\lambda)\log(\Sigma q_i^\alpha)\right]$$

$$=\lambda H_\alpha(P)+(1-\lambda)H_\alpha(Q) \quad (17)$$

(Solution Sources: TC Solution Manual, Aarti Singh Lecture notes, ECE 534 UIC notes, Renyi's entropy and probability of error, Ben-Bassat, Raviv, Cachin PhD dissertation.)