Problem 1

**Example of Joint Entropy**

### a)

\[
H(X) = - \sum_x P_X(x) \log P_X(x) = - \frac{2}{3} \log \frac{2}{3} - \frac{1}{3} \log \frac{1}{3} = \log 3 - \frac{2}{3} = 0.918 \text{ bits},
\]

\[
H(Y) = - \sum_y P_Y(y) \log P_Y(y) = - \frac{1}{3} \log \frac{1}{3} - \frac{2}{3} \log \frac{2}{3} = \log 3 - \frac{2}{3} = 0.918 \text{ bits}.
\]

### b)

We need the conditional probabilities \(P_{X|Y}\) and \(P_{Y|X}\). With \(P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}\) we get

\[
\begin{array}{c|cc}
   P_{X|Y}(x|y) & x = 0 & x = 1 \\
   \hline
   y = 0 & 1 & 0 \\
   y = 1 & \frac{1}{2} & \frac{1}{2}
\end{array}
\]

\[
\begin{array}{c|cc}
   P_{Y|X}(y|x) & y = 0 & y = 1 \\
   \hline
   x = 0 & 1/2 & 1/2 \\
   x = 1 & 0 & 1
\end{array}
\]

Thus, we can calculate

\[
H(X|Y) = - \sum_y P_Y(y) \sum_{x \in \text{supp}(P_{X|Y}(|y))} P_{X|Y}(x|y) \log P_{X|Y}(x|y)
= - \frac{1}{3} (1 \log 1) - \frac{2}{3} \left( \frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2} \right) = \frac{2}{3} \text{ bits},
\]

\[
H(Y|X) = - \sum_x P_X(x) \sum_{y \in \text{supp}(P_{Y|X}(|x))} P_{Y|X}(y|x) \log P_{Y|X}(y|x)
= - \frac{2}{3} \left( \frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2} \right) - \frac{1}{3} (1 \log 1) = \frac{2}{3} \text{ bits}.
\]

### c)

\[
H(X,Y) = 3 \cdot \left( -\frac{1}{3} \log \frac{1}{3} \right) = \log 3 = 1.585 \text{ bits}.
\]

### d)

\[
H(Y) - H(Y|X) = \log 3 - \frac{4}{3} = 0.252 \text{ bits}.
\]

### e)

\[
I(X;Y) = H(Y) - H(Y|X) = \log 3 - \frac{4}{3} = 0.252 \text{ bits}.
\]
Problem 2

Zero Conditional Entropy

First, remember that $H(Y|X = x) \geq 0$ with equality if, and only if, for a given $X = x$, $Y$ is deterministic.

Next, note that $H(Y|X)$ can be written as

$$H(Y|X) = \sum_{x \in \text{supp}(P_X)} P_X(x) H(Y|X = x)$$

and thus is given by a sum of nonnegative terms. As $P_X(x) > 0$ for all $x \in \text{supp}(P_X)$, we can conclude that $H(Y|X) = 0$ if, and only if, $H(Y|X = x) = 0$ for all $x$ with $P_X(x) > 0$. But this is equivalent to saying that $Y$ is determined by $X$ for all $x$ with $P_X(x) > 0$, i.e., $Y$ is a function of $X$ with probability one.

Problem 3

Entropy of Functions of a Chance Variable

a) This follows directly from the chain rule.

b) This is a consequence of Problem 2.

c) This also holds because of the chain rule.

d) This inequality follows because the conditional entropy is nonnegative.

Thus, applying a function to a chance variable never increases the entropy. We have equality if and only if $H(X|g(X)) = 0$, which is satisfied if and only if $X$ is a function of $g(X)$ with probability one, i.e., if and only if the restriction of $g(\cdot)$ to the support of $P_X$ is injective. (The restriction of $g(\cdot)$ to the support of $P_X$ is the function $g|_{\text{supp}(P_X)}: \text{supp}(P_X) \rightarrow \mathcal{Y}: x \mapsto g(x)$.)

Problem 4

Entropy of a Sum

a) Observe that

$$H(X,Y,Z) \overset{(i)}{=} H(X) + H(Y|X) + H(Z|X,Y) = H(X) + H(Y|X),$$

$$H(X,Y,Z) \overset{(i)}{=} H(X) + H(Z|X) + H(Y|X,Z) = H(X) + H(Z|X),$$

where (i) follows from the chain rule and the underbraced terms are zero because $Z$ is a function of the pair $(X,Y)$ and $Y$ is a function of the pair $(X,Z)$. Therefore, we conclude that $H(Z|X) = H(Y|X)$.

If $X$ and $Y$ are independent, we have $H(Y) = H(Y|X)$, so

$$H(Y) = H(Y|X) = H(Z|X) \leq H(Z),$$

where the last inequality follows because conditioning does not increase entropy. Likewise, one can show that $H(X) \leq H(Z)$ if $X$ and $Y$ are independent.

b) Let $X$ and $Y$ be fair coin flips that are influenced by each other in such way that whenever $X$ equals one, $Y$ equals zero and the other way round, i.e., $P_{Y|X}(0|0) = 0$, $P_{Y|X}(1|0) = 1$, $P_{Y|X}(0|1) = 1$, and $P_{Y|X}(1|1) = 0$. In this case, $Z = 1$ with probability 1. Thus, $H(Z) = 0$, however, $H(X) = H(Y) = 1$ bit. Note that $Y$ is a function of $X$. 

© Amos Lapidoth, 2016 2
c) Note that $Z$ is a function of the pair $(X,Y)$, so $H(Z) \leq H(X,Y)$ and

$$H(Z) \leq H(X,Y)$$

$$\overset{(i)}{=} H(X) + H(Y) - I(X;Y)$$

$$\overset{(ii)}{\leq} H(X) + H(Y),$$

where (i) follows from the definition of the mutual information and (ii) follows because mutual information is nonnegative. The first inequality holds with equality if and only if the pair $(X,Y)$ can be recovered from $Z$ with probability one. The second inequality holds with equality if and only if $I(X;Y) = 0$, i.e., if and only if $X$ and $Y$ are independent. Therefore, $H(Z) = H(X) + H(Y)$ if and only if $X$ and $Y$ are independent and the pair $(X,Y)$ can be recovered from $Z$ with probability one.

Note that in our case where $Z = X + Y$, the pair $(X,Y)$ can be recovered from $Z$ if the alphabets $\mathcal{X}$ and $\mathcal{Y}$ are such that for each possible pair $(x_i, y_j) \in \mathcal{X} \times \mathcal{Y}$, the sum $x_i + y_j$ is unique. An example of such a situation is $\mathcal{X} = \{1, \ldots, 10\}$ and $\mathcal{Y} = \{0, 100\}$.

Problem 5

Jensen’s Inequality

Remember what Jensen’s inequality states:

Lemma 1. If $f$ is a concave function and $X$ is a random variable, then

$$E[f(X)] \leq f(E[X]). \quad (1)$$

Moreover, if $f$ is strictly concave, then equality in (1) implies $X = \text{const}$.

Let $A$ be a uniformly distributed random variable over the set $\mathcal{A} = \{a_1, a_2, \ldots, a_n\}$. Then,

$$E[A] = \sum_{k=1}^{n} \frac{1}{n} \cdot a_k = \frac{1}{n} \sum_{k=1}^{n} a_k.$$

a) Let $f : \mathbb{R}^+ \rightarrow \mathbb{R}$, $x \mapsto f(x) = \log x$. The function $f$ is strictly concave, so

$$\log \left( \prod_{k=1}^{n} a_k \right)^{\frac{1}{n}} = \frac{1}{n} \log \left( \prod_{k=1}^{n} a_k \right)$$

$$= \frac{1}{n} \sum_{k=1}^{n} \log a_k$$

$$= E[\log A]$$

$$\leq \log E[A]$$

$$= \log \left( \frac{1}{n} \sum_{k=1}^{n} a_k \right).$$

Because the function $f$ is monotonically increasing, we have

$$\left( \prod_{k=1}^{n} a_k \right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{k=1}^{n} a_k.$$

Since $f$ is strictly concave, equality holds if and only if $A$ is a deterministic random variable, i.e., $a_k = a = \text{constant}$, $\forall k$. 

© Amos Lapidoth, 2016 3
b) If $\beta \geq 1$, then $f: \mathbb{R}^+ \to \mathbb{R}$, $x \mapsto f(x) = x^\beta$ is a convex function. Appropriately modified, the lemma above can be used for convex functions. Using the random variable $A$ again, we get

$$\sum_{k=1}^{n} \frac{1}{n} a_k^\beta = \mathbb{E}[A^\beta] \geq (\mathbb{E}[A])^\beta = \left(\sum_{k=1}^{n} \frac{1}{n} a_k\right)^\beta,$$

which proves the claim.

For $0 < \beta \leq 1$, the function $f$ is concave. Therefore, $\frac{1}{n} \sum_{k=1}^{n} a_k \leq \left(\frac{1}{n} \sum_{k=1}^{n} a_k\right)^{\beta}$.

c) Considering Part b) for $\beta = 2$, we see that $\sqrt{\frac{1}{n} \sum_{k=1}^{n} a_k^2}$ is always at least as large as $\frac{1}{n} \sum_{k=1}^{n} a_k$. For example, if your scores in six exams are 1, 2, 3, 4, 5 and 6, respectively, then $\frac{1}{n} \sum_{k=1}^{n} a_k = 3.5$, while $\sqrt{\frac{1}{n} \sum_{k=1}^{n} a_k^2} = 3.89$. 

© Amos Lapidoth, 2016