Quantum Information Theory: Sheet 1

Otto Pyper

Exercise 1. a) By definition, if $\underline{u} \in J^n$ then

$$\begin{aligned} 2^{-n(H(u)+\varepsilon)} &\leq p(u_1, \dots, u_n) \leq 2^{-n(H(U)-\varepsilon)} \\ \implies -n(H(U)+\varepsilon) &\leq \log p(u_1, \dots, u_n) \leq -n(H(U)-\varepsilon) \\ \implies H(U)-\varepsilon &\leq -\frac{1}{n} p(u_1, \dots, u_n) \leq H(U)+\varepsilon \end{aligned}$$

c) We have that $\mathbb{P}(T_\varepsilon^{(n)}) = \sum_{u \in T_\varepsilon^{(n)}} p(u)$. Therefore:

$$(1-\delta) < \mathbb{P}(T_\varepsilon^{(n)}) \leq |T_\varepsilon^{(n)}| p_{\max} \leq |T_\varepsilon^{(n)}| 2^{-n(H(U)-\varepsilon)}$$

and the result follows. Similarly:

$$2^{-n(H(U)+\varepsilon)} |T_\varepsilon^{(n)}| \leq |T_\varepsilon^{(n)}| p_{\min} \leq \mathbb{P}(T_\varepsilon^{(n)}) \leq 1$$

and again the result follows.

Exercise 2. $p(0) = 0.4$, $p(1) = 0.6$, binary source described by U_1, U_2, U_3.

1. The most probable sequence in $\{0,1\}^3$ is 111, which occurs with probability 0.216
2. We first calulate the entropy, which is given by $H(U) = -0.4 \log 0.4 - 0.6 \log 0.6 \approx 0.971$. For $\varepsilon = 0.2$, the typical sequences are then those that occur with probability p, where $0.0876 \leq p \leq 0.201$. So the typical set is $\{001, 010, 100, 011, 101, 110\}$.
3. The total probability of these sequences is 0.72.
4. A smallest set of probability at least 0.72 is $\{111, 011, 101, 110\} \cup \{x\}$, for any $x \in \{001, 010, 100\}$.
5. This set of higher probability thus has its benefits in that it will yield a lower error rate in the compression scheme. However, it is in general impractical to use a 'high probability set' where the criteria for determining whether something is in the set or not is unclear; we had to made an arbitrary choice to create such a set. In proofs it is more convenient to have a more general, simpler definition of a typical set.

Exercise 3.

1. We have that $H(X) = -\sum_{x \in J_X} p(x) \log p(x) = -\sum_x \sum_y p(x,y) \log p(x)$, and hence:

$$-H(X,Y) + H(X) + H(Y) = \sum_x \sum_y p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

which we recognise as the relative entropy of the two distributions $\{p(x,y)\}_{x,y}$ and $\{p(x)p(y)\}_{x,y}$, noting that the first is absolutely continuous with respect to the second since if $p(x)p(y) = 0$ then either $p(x) = 0$ or $p(y) = 0$, and in either case $p(x,y) = 0$ since not both of x, y can occur.

The relative entropy of two probability distributions is always non-negative, and equals zero if and only if the two probaility distributions are identical, *i.e.* for each x, y we have $p(x,y) = p(x)p(y)$; so X, Y are independent.

2. Define $f(\lambda)$ as:

$$\begin{aligned} f(\lambda) &= H(\lambda p + (1-\lambda)q) - \lambda H(p) - (1-\lambda)H(q) \\ &= -\sum_x (\lambda p(x) + (1-\lambda)q(x)) \log[\lambda p(x) + (1-\lambda)q(x)] + \lambda \sum_x p(x)\log p(x) + (1-\lambda)\sum_x q(x)\log q(x) \\ \therefore f'(\lambda) &= H(q) - H(p) - \sum_x (p(x) - q(x)) \log[\lambda p + (1-\lambda)q] - \sum_x (p(x) - q(x)) \\ \therefore f''(\lambda) &= -\sum_x ((p(x) - q(x))^2 \cdot \frac{1}{\lambda p(x) + (1-\lambda)q(x)} \leq 0 \end{aligned}$$

with equality iff $p(x) = q(x)$ for all x. So f is concave, and $f(0) = 0$, $f(1) = 0$ hence $f(\lambda) \geq 0$ for all $0 < \lambda < 1$.

Exercise 4. The inequality (1) was derived using Jensen's inequality, for which equality holds iff the function φ in question is linear or the inputs are all equal; log is not linear hence equality holds in (1) iff $q(x) = p(x)$ for all x.

(2) is proved similarly using Jensen; let P denote the r.v. that takes values $p(x)$ each with probability $p(x)$. Then we have:

$$\begin{aligned} H(X) &= -\sum_{x \in J_X} p(x) \log p(x) \\ &= \sum_{x \in J_X} \log \frac{1}{p(x)} \\ &= \mathbb{E}[\log \frac{1}{P}] \\ &\leq \log \mathbb{E}\frac{1}{P} = \log |J_X| \end{aligned}$$

so again by Jensen we have equality iff the values that P takes are constant, *i.e.* each $x \in J_X$ occurs with equal probability. Hence we have equality in (2) iff X is uniform.

Exercise 5. We have already seen that

$$I(X:Y) \coloneqq H(X) + H(Y) - H(X,Y) = D(\{p_{X,Y}(x,y)\} || \{p_X(x)p_Y(y)\})$$

Moreover, it can be seen that:

$$\begin{aligned}
H(Y|X) &\coloneqq \sum_{x\in J} p_X(x) H(Y|X=x) \\
&= -\sum_{x\in J} p_X(x) \sum_{y\in J} p_{Y|X}(y|x) \log p_{Y|X}(y|x) \\
&= -\sum_{x,y\in J} p(x,y) \log p(y|x) \\
&= -\sum_{x,y\in J} p(x)p(y|x) \log \frac{p(y|x)p(x)}{p(x)} \\
&= -D(\{p(x,y)\}_{x,y\in J} || \{p(x)/|J|\}_{x,y\in J}) + \sum_{x,y} p(x)p(y|x) \log |J| \\
&= \log |J| - D(\{p(x,y)\}_{x,y\in J} || \{p(x)/|J|\}_{x,y\in J}) \\
&= -D(\{p(x,y)\}_{x,y\in J} || \{p(x)\}_{x,y\in J})
\end{aligned}$$

where we remark that the latter function on x, y in the relative entropy is not a probability distribution.

Exercise 6.

1. We know that $H(X|Y) = H(X,Y) - H(Y)$, and $I(X:Y) = H(X) + H(Y) - H(X,Y)$. It is then easy to see that $I(X:Y) = H(X) - H(X|Y)$.

2. If X, Y are independent then $H(X|Y) = H(X)$, so $I(X:Y) = H(X) - H(X|Y) = H(X) - H(X) = 0$.

Exercise 7.

1. I believe that by 'equal' here it is mean that $P(X = x|Y = x) = 1$ for all x, but this isn't generally how I would interpret equal; I would say they are equal if they are i.i.d, for instance, or if they have the same distribution but are not independent (and this could split into a variety of cases).

   In this case we have $I(X:Y) = H(X) - H(X|Y) = -\sum_x p(x)\log p(x) - \sum_x p(x)H(X|Y = x)$. $H(X|Y = x) = \sum_{x'} p(x'|x)\log p(x'|x) = 0$. So $I(X:Y) = H(X)$.

2. $I(X:Y) = H(X) - H(X|Y)$. Therefore:

   $$\begin{aligned}
   I(X:Y) &= -\frac{1}{2}\log 2^{-1} - \frac{1}{2}\log 2^{-1} - H(X|Y) \\
   &= 1 - p(Y=0)H(X|Y=0) - p(Y=1)H(X|Y=1)
   \end{aligned}$$

   Note that $p(Y=0) = p(Y=0|X=1)p(X=1) + p(Y=0|X=0)p(X=0) = \frac{1}{2}(1-p) + \frac{1}{2}p = \frac{1}{2}$. In particular, $p(x|y) = p(y|x)$.

   So $H(X|Y=1) = -p(1|1)\log p(1|1) - p(0|1)\log p(0|1) = -p\log p - (1-p)\log(1-p) = h(p)$. Similarly $H(X|Y=0) = h(p)$. So $I(X:Y) = 1 - \frac{1}{2}h(p) - \frac{1}{2}h(p) = 1 - h(p)$.

Exercise 8. WLOG say $p(0) = 1 - \varepsilon$. Then we have:

$$\begin{aligned}
H(X) &= -\sum_{x\in J} p(x)\log p(x) \\
&= -(1-\varepsilon)\log(1-\varepsilon) - \sum_{x\neq 0} p(x)\log p(x)
\end{aligned}$$

Now consider the function $f(x) = x\log(x)$. This function is convex:

$$\begin{aligned} f(x) &= x\log(x) \\ \implies f'(x) &= \log(x) + \frac{1}{\log_e(2)} \\ \implies f''(x) &= \frac{1}{x\log_e(2)} \end{aligned}$$

so f is convex for $0 < x < 1$. So given t_i and x_i such that $\sum t_i = 1$, we have that $f(\sum t_i x_i) \leq \sum t_i f(x_i)$. Setting $t_i = \frac{1}{m-1}$ and $x_i = p(x)$ then gives:

$$\begin{aligned} f(\sum p(x)/(m-1)) &\leq \frac{1}{m-1}\sum p(x)\log p(x) \\ \implies (m-1)f(\varepsilon/(m-1)) &\leq \sum p(x)\log p(x) \\ \implies \varepsilon\log(\varepsilon/(m-1)) &\leq \sum p(x)\log p(x) \\ \therefore H(X) &\leq -(1-\varepsilon)\log(1-\varepsilon) - \varepsilon\log(\varepsilon/(m-1)) \\ &= h(\varepsilon) + \varepsilon\log(m-1) \end{aligned}$$

which is the desired inequality.

Exercise 9. Let q_j be the probability distribution given by $\{p(x_{i+j-1}|y_j)\}_i$, and let $Q = \sum_{j=1}^m p(y_j)q_j$ be the distribution given by their weighted sum.

Then $\mathbb{P}(Q = 1) = \sum_{j=1}^m p(y_j)p(x_j|y_j) = \sum_{j=1}^m p(x_j, y_j) = 1 - \varepsilon$. Hence we can apply (8) to the random variable Q to see that $H(Q) \leq h(\varepsilon) + \varepsilon\log(m-1)$.

However, since H is itself concave, we have that:

$$\begin{aligned} H(Q) &= H(\sum_{j=1}^m p(y_j)q_j) \\ &\geq \sum_{j=1}^m p(y_j)H(q_j) \end{aligned}$$

Note that q_j has identical entropy to $X|Y = y_j$; the probabilities are the same, they just apply to different values that the variable can take; this has no impact on entropy.

Hence $H(X|Y) = \sum_{j=1}^m p(y_j)H(q_j) \leq H(Q) \leq h(\varepsilon) + \varepsilon\log(m-1)$, as required.

Exercise 10. We can express $H(Y, Z, X) - H(X, Y, Z) = 0$ as:

$$\begin{aligned} 0 =& H(Y) + H(Z|Y) + H(X|Y, Z) \\ &- (H(X) + H(Y|X) + H(Z|X, Y)) \end{aligned}$$

But $H(Z|X, Y) = \sum_{x,y} p(x, y)H(Z|X = x, Y = y) = \sum_{x,y} p(x, y)H(Z|Y = y) = \sum_y p(y)H(Z|Y = y) = H(Z|Y)$. So the above simplifies to:

$$H(Y) - H(Y|X) + H(X|Y, Z) - H(X) = 0$$

and $I(X : Y) = H(Y) - H(Y|X)$, $I(X : Z) = H(X) - H(X|Z)$, so we have that

$$\begin{aligned} I(X : Y) - I(X : Z) &= H(X|Z) - H(X|Y, Z) \\ &= I(X : Y|Z) \geq 0 \end{aligned}$$

since the mutual information between any two r.v.s is non-negative, as can be seen here:

$$\begin{aligned} H(X|Z) - H(X|Y,Z) &= -\sum_{x,y,z} p(x,y,z) \log \frac{p(x,z)p(y,z)}{p(z)p(x,y,z)} \\ &= \mathbb{E}\left[-\log \frac{p(x,z)p(y,z)}{p(x,y,z)p(z)} \right] \end{aligned}$$

i.e. is the expectation of the negative logarithm of the random variable that takes the value $p(x,z)p(y,z)/(p(z)p(x,y,z))$ with probability $p(x,y,z)$. Then, by Jensen:

$$\begin{aligned} H(X|Z) - h(X|Y,Z) &\geq -\log \mathbb{E}\left[\frac{p(x,z)p(y,z)}{p(z)p(x,y,z)} \right] \\ &= -\log \left(\sum_{x,y,z} \frac{p(x,z)p(y,z)}{p(z)} \right) \\ &= -\log \left(\sum_{y,z} p(y,z) \sum_{x} p(x|z) \right) \\ &= -\log \left(\sum_{y,z} p(y,z) \right) \\ &= 0 \end{aligned}$$

Exercise 11. Let $p(X=0) = q$, and $p(X=1) = 1-q$. We then calculate $I(X:Y) = H(X) - H(X|Y)$.

Note that $H(X|Y=0) = H(X|Y=1) = 0$, since the outputs 0 and 1 can only arise from inputs 0 and 1 respectively.

So $H(X|Y) = p(Y=e)H(X|Y=e) = ph(q)$. Moreover, $H(X)$ is the binary entropy $h(q)$.

Hence $I(X:Y) = (1-p)h(q)$, which is maximised at $q = 1/2$, giving $\mathcal{C} = (1-p) = 2/3$ for $p = 1/3$.

Exercise 12. If $a \neq -1, 1$ then the output uniquely identifies the input; $H(X|Y) = 0$, so the capacity is the max of $H(X) = h(q)$, which is achieved at $q = 1/2$, giving capacity 1.

If $a = 1$, then we have exactly the same situation as the above, with $p = 1/2$, $e = 1$, and 2 is now recognised as 1. So the capacity of this channel is given by the same formula, which is $\max(1-p)h(q)$. $p = 1/2$, so the capacity is 1/2. Ditto $a = -1$.

Exercise 13. This is trivial.

Exercise 14.