TUTORIAL VII

1 Sum of random variables

(From [1]). Let X, Y be integer-valued random variables and let $Z = X + Y$.

1. Prove that $H(Z|X) = H(Y|X)$. (Hint: Expand $H(Z|X)$ using the definition of conditional entropy.)

   A: *By definition, we have*

$$\begin{aligned} H(Z|X) &= \sum_x p_X(x) H(Z|X=x) \\ &= -\sum_x p_X(x) \sum_z p_{Z|X}(z|x) \log p_{Z|X}(z|x) \\ &= -\sum_x p_X(x) \sum_z p_Y(z-x|x) \log p_Y(z-x|x) \\ &= -\sum_x p_X(x) \sum_y p_Y(y|x) \log p_Y(y|x) \\ &= \sum_x p_X(x) H(Y|X=x) \\ &= H(Y|X). \end{aligned}$$

   Similarly, it can be shown that $H(Z|Y) = H(X|Y)$.

2. Prove that if X, Y are independent, then $H(Z) \geq \max\{H(X), H(Y)\}$. That is, addition of independent random variables increases entropy.

   A: *The mutual information between* Z *and* Y *is non-negative, i.e.,* $I(Z;Y) = H(Z) - H(Z|Y) \geq 0$*. It follows from the previous question that* $H(Z) \geq H(Z|Y) = H(X|Y)$*. Since* X, Y *are independent,* $H(X|Y) = H(X)$ *and so* $H(Z) \geq H(X)$*. Similar arguments imply that* $H(Z) \geq H(Y)$ *and so* $H(Z) \geq \max\{H(X), H(Y)\}$*.*

3. Give an example of random variables X, Y for which $H(Z) < \min\{H(X), H(Y)\}$.

   A: *Suppose that* X *is uniformly distributed over a fixed finite set of integers (say of size* m*) with equal probabilities and let* $Y = -X$*. Then both* $H(X)$ *and* $H(Y)$ *are* $\log_2 m$ *while* $H(Z)$ *is* 0 *since* Z *always takes the value* 0*.*

4. State and prove a necessary and sufficient condition for when the entropy of the sum equals the sum of the entropies, i.e., $H(Z) = H(X) + H(Y)$.

   A: $H(X,Y) \leq H(X) + H(Y)$ *since* $I(X;Y) = H(X) + H(Y) - H(X,Y) \geq 0$*. We know that* Z *is a function of* X, Y *and hence* $H(Z) \leq H(X,Y)$ *from which the required inequality follows. Equality holds when* X*,* Y *are independent and when the function is a bijection.*

2 Huffman codes with costs.

Words such as "Run!", "Help!", and "Fire!" are short, not because they are used frequently, but perhaps because time is precious in the situations in which these words are required. Suppose that $X = i$ with probability p_i, for $i = 1, 2, ..., m$. Let l_i be the number of binary symbols in the codeword associated with $X = i$, and let c_i denote the cost per letter of the codeword when $X = i$. Thus, the average cost C of the description of X is $C = \sum_{i=1}^{m} p_i c_i l_i$.

1. Minimise C over all $l_1, l_2, \ldots, l_m$ such that $\sum_{i=1}^m 2^{-l_i} \leq 1$. Ignore any implied integer constraints on l_i. Exhibit the minimising $l_1^*, l_2^*, \ldots, l_m^*$ and the associated minimum value C^*.

   A: *Assume that $\sum_{i=1}^m 2^{-l_i} = 1$. Let $q_i = 2^{-l_i}$ for $i = 1, \ldots, m$ and $S = \sum_{i=1}^m p_i c_i$. Then $\mathbf{r} = (\frac{p_1 c_1}{S}, \ldots, \frac{p_m c_m}{S})$ is a probability distribution. We have*

$$\begin{aligned} C &= \sum_{i=1}^m p_i c_i l_i \\ &= -\sum_{i=1}^m S r_i \log q_i \\ &= -S \sum_{i=1}^m r_i \log \frac{q_i r_i}{r_i} \\ &= S \left(\sum_{i=1}^m r_i \log \frac{r_i}{q_i} - \sum_{i=1}^m r_i \log r_i \right) \\ &= S \left(D(\mathbf{r}||\mathbf{q}) + H(\mathbf{r}) \right). \end{aligned}$$

   C can be minimised by choosing $\mathbf{q} = \mathbf{r}$. The minimising l_i^* is given by $l_i^* = \log \frac{C}{p_i c_i}$ and $C^* = S \cdot H(\mathbf{r})$.

2. How would you use the Huffman code procedure to minimise C over all uniquely decodable codes? Let C_{Huffman} denote this minimum.

   A: *Run Huffman algorithm on input distribution $\mathbf{r}$. The resulting code minimises the expected cost.*

3. Show that $C^* \leq C_{\text{Huffman}} \leq C^* + \sum_{i=1}^m p_i c_i$.

   A: *Since we run the Huffman algorithm on the distribution $\mathbf{r}$ the code lengths are $l_i = \lceil -\log r_i \rceil$ for $i = 1, \ldots, m$. Then we have*

$$\begin{gathered} -\log r_i < l_i < -\log r_i + 1 \\ -\sum_{i=1}^m p_i c_i \log r_i < \sum_{i=1}^m p_i c_i l_i < -\sum_{i=1}^m p_i c_i (\log r_i + 1) \\ -\sum_{i=1}^m p_i c_i \log r_i < C_{\text{Huffman}} < -\sum_{i=1}^m p_i c_i \log r_i + S \\ C^* < C_{\text{Huffman}} < C^* + S. \end{gathered}$$

3 Maximum likelihood decoding

(From [1]). Let W be a symmetric binary input channel with output alphabet $\mathcal{Y}$ and defined by the transition probabilities $p(y|0)$ and $p(y|1)$ for all $y \in \mathcal{Y}$.

1. Prove that the capacity of W is equal to

$$\frac{1}{2} \sum_{b \in \{0,1\}} \sum_{y \in \mathcal{Y}} p(y|b) \log_2 \frac{2p(y|b)}{p(y|0) + p(y|1)}.$$

A: *The capacity of the channel is given by $C(W) = \max_{p_X} I(X;Y)$. We have*

$$\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) \\
&= \sum_{y\in\mathcal{Y}} p(y)\log\frac{1}{p(y)} - \sum_{b\in\{0,1\}} p(b)H(Y|X=b) \\
&= \sum_{y\in\mathcal{Y}}\left(\sum_{b\in\{0,1\}} p(b)p(y|b)\right)\log\frac{1}{p(y)} - \sum_{b\in\{0,1\}} p(b)\sum_{y\in\mathcal{Y}} p(y|b)\log\frac{1}{p(y|b)} \\
&= \sum_{b\in\{0,1\}}\sum_{y\in\mathcal{Y}} p(b)p(y|b)\log\frac{1}{p(y)} - \sum_{b\in\{0,1\}}\sum_{y\in\mathcal{Y}} p(b)p(y|b)\log\frac{1}{p(y|b)} \\
&= \sum_{b\in\{0,1\}}\sum_{y\in\mathcal{Y}} p(b)p(y|b)\log\frac{p(y|b)}{p(y)} \\
&= \sum_{b\in\{0,1\}}\sum_{y\in\mathcal{Y}} p(b)p(y|b)\log\frac{p(y|b)}{p(y|0)p(0)+p(y|1)p(1)},
\end{aligned}$$

which is maximised when X follows the uniform distribution since W is symmetric. Setting $p(0) = p(1) = \frac{1}{2}$ gives the required expression for capacity.

2. Suppose a bit c was transmitted and we receive $y \in \mathcal{Y}$. We decode y into the bit 0 if $p(y|0) > p(y|1)$, and 1 otherwise. Show that the probability that we make a decoding error is at most $Z(W) = \sum_{y\in\mathcal{Y}} \sqrt{p(y|0)\cdot p(y|1)}$.

A: *Denote the estimate for c as $\hat{c}$. We make a decoding error whenever $c \neq \hat{c}$ which happens if $p(y|\hat{c}) > p(y|c)$. Let δ_y be an indicator function for this event ($\delta_y = 1$ if $p(y|\hat{c}) > p(y|c)$ and 0 otherwise). The probability of error is given by*

$$\Pr[c \neq \hat{c}] = \sum_{y\in\mathcal{Y}} p(y|c)\delta_y.$$

We can bound δ_y as follows: $\delta_y \leq \left(\frac{p(y|\hat{c})}{p(y|c)}\right)^{1/2}$. If $\delta_y = 1$ then $\left(\frac{p(y|\hat{c})}{p(y|c)}\right)^{1/2} > 1$ and otherwise < 1. Therefore, we have,

$$\begin{aligned}
\Pr[c \neq \hat{c}] &= \sum_{y\in\mathcal{Y}} p(y|c)\delta_y \\
&\leq \sum_{y\in\mathcal{Y}} p(y|c)\left(\frac{p(y|\hat{c})}{p(y|c)}\right)^{1/2} \\
&\leq \sum_{y\in\mathcal{Y}} \sqrt{p(y|c)p(y|\hat{c})} \\
&= \sum_{y\in\mathcal{Y}} \sqrt{p(y|0)p(y|1)}.
\end{aligned}$$

3. Suppose a code $C \subseteq \{0,1\}^n$ is used for transmitting a sequence of n bits on the discrete memoryless channel W. Consider the following maximum likelihood decoding rule at the receiver: if $\mathbf{y} \in \mathcal{Y}^n$ is received, output a codeword $\mathbf{c} \in C$ for which $p(\mathbf{y}|\mathbf{c}) = \prod_{i=1}^n p(y_i|c_i)$ is maximum, ties are broken arbitrarily.

   Prove that if a codeword $\mathbf{c}_0 \in C$ was transmitted on the channel, and $\mathbf{y}_0$ was received, the probability that the above decoding rule outputs a codeword different from $\mathbf{c}_0$ is at most $\sum_{j=1}^n d_j Z(W)^j$, where d_j is the number of codewords in C at Hamming distance j from $\mathbf{c}_0$ (i.e., the number of codewords that are different from $\mathbf{c}$ at exactly j positions).

A: $\Pr[\mathbf{c} \neq \mathbf{c}_0] = \sum_{\mathbf{x} \neq \mathbf{c}_0} \Pr[\mathbf{c} = \mathbf{x}]$. *For a fixed* $\mathbf{x} \neq \mathbf{c}_0$*, we have* $\Pr[\mathbf{c} = \mathbf{x}] = \prod_i \Pr[c_i = x_i] \leq \prod_{c_i \neq c_{0,j}} Z(W) = Z(W)^{h(\mathbf{c}, \mathbf{c}_0)}$*, where* $h(\mathbf{c}, \mathbf{c}_0)$ *is the hamming distance between* $\mathbf{c}$ *and* $\mathbf{c}_0$*. The required expression now follows easily.*

References

[1] Course 15-859: Information theory and its applications in theory of computation, by Venkatesan Guruswami and Mahdi Cheraghchi at CMU.