S-72.340 INFORMATION THEORY

Exercise 3, October 4, 2000 Solutions

1. For simplicity, we use the following notations:

$$p_k := p_X(k) \quad \text{and} \quad \mathbf{p} := \{p_0, p_1, p_2, \dots\}\,.$$

So, $H(X) = H(\mathbf{p})$. Besides the condition that $E(X) = A$, another constraint results from the fact that $\mathbf{p}$ must be a probability distribution. This means that

$$\sum_{k=0}^{\infty} p_k = 1\,.$$

Now we have two equality constraints:

$$g_1(\mathbf{p}) = \sum_{k=0}^{\infty} p_k - 1 = 0 \quad \text{and} \quad g_2(p) = \sum_{k=0}^{\infty} k p_k - A = 0\,.$$

We use the Lagrange multiplier method. The Lagrangian function for this problem is

$$\begin{aligned} L(\mathbf{p}, \lambda_1, \lambda_2) &= H(\mathbf{p}) + \lambda_1 g_1(\mathbf{p}) + \lambda_2 g_2(\mathbf{p}) \\ &= -\sum_{k=0}^{\infty} p_k \log p_k + \lambda_1 \cdot (\sum_{k=0}^{\infty} p_k - 1) + \lambda_2 \cdot (\sum_{k=0}^{\infty} k p_k - A)\,, \end{aligned}$$

where the Lagrange multipliers λ_1 and λ_2 are at this point unknown real numbers. The following conditions are necessary for the optimum distribution $\mathbf{p}$:

$$\frac{\partial L}{\partial p_k} = 0,\ k = 0, 1, 2, \dots \quad \text{and} \quad \frac{\partial L}{\partial \lambda_i} = 0,\ i = 1, 2.$$

The latter conditions are equivalent to the original equality constraints of the maximization problem. For simplicity, we use natural logarithms in solving the problem; the base number of the logarithm doesn't affect the solution, as long as the number is greater than one.

$$\begin{aligned} &\frac{\partial L}{\partial p_k} = -\ln p_k - 1 + \lambda_1 + k\lambda_2 = 0 \\ \Rightarrow\ &\ln p_k = -1 + \lambda_1 + k\lambda_2 \ \Rightarrow\ p_k = e^{-1+\lambda_1+k\lambda_2}\,. \end{aligned}$$

$$\frac{\partial L}{\partial \lambda_1} = 0 \Rightarrow \sum_{k=0}^{\infty} p_k = 1 \Rightarrow e^{-1+\lambda_1} \cdot \sum_{k=0}^{\infty} e^{k\lambda_2} = 1\,.$$

For the infinite sum in the last equality to converge, λ_2 must be negative. If it is, we get

$$\frac{e^{-1+\lambda_1}}{1-e^{\lambda_2}} = 1 \Rightarrow e^{-1+\lambda_1} = 1 - e^{\lambda_2} \Rightarrow p_k = (1-e^{\lambda_2}) \cdot (e^{\lambda_2})^k\,.$$

Since $\lambda_2 < 0$, $\exp(\lambda_2) \in (0,1)$ and is thus a probability. From this we conclude that $\mathbf{p}$ is the distribution of a geometrical random variable (the possible values of which begin from 0) with parameter $p = 1 - \exp(\lambda_2)$. That is,

$$p_k = pq^k\,,$$

where $q = 1 - p = \exp(\lambda_2)$.

$$\begin{aligned}
&\frac{\partial L}{\partial \lambda_2} = 0 \Rightarrow \sum_{k=0}^{\infty} k p_k = A \\
&\sum_{k=0}^{\infty} k p_k = p \cdot \sum_{k=1}^{\infty} k q^k = pq \cdot \sum_{k=1}^{\infty} k q^{k-1} = pq \cdot \frac{d}{dq}\left\{\sum_{k=1}^{\infty} q^k\right\} = pq \cdot \frac{d}{dq}\left\{\frac{q}{1-q}\right\} \\
&= \cdots = \frac{1-p}{p} = A \Rightarrow \cdots \Rightarrow p = \frac{1}{A+1}\,,\ q = \frac{A}{A+1} \\
&\Rightarrow p_k = \frac{1}{A+1} \cdot \left(\frac{A}{A+1}\right)^k,\ k = 0, 1, 2, \ldots
\end{aligned}$$

The resulting entropy is:

$$\begin{aligned}
&H(X) = H(\mathbf{p}) = -\sum_{k=0}^{\infty} p_k \log p_k = -\sum_{k=0}^{\infty} pq^k \log (pq^k) \\
&= -p \log p \cdot \sum_{k=0}^{\infty} q^k - p \log q \cdot \sum_{k=0}^{\infty} k q^k \\
&= \cdots = -\log p - \frac{1-p}{p} \log(1-p) = \log(A+1) + A \cdot \log\left(\frac{A+1}{A}\right).
\end{aligned}$$

2. a) The marginal distribution of X is given by

$$P(x) = \sum_y P(x,y)\,.$$

By definition, the entropy of X is

$$H(X) = -\sum_x P(x)\log P(x) = -\sum_{x,y} P(x,y)\log P(x)\,.$$

The expression for $H(Y)$ is derived in the same way.

b) The so-called *log sum inequality* will be used in the proof: for non-negative numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$,

$$\sum_{i=1}^n a_i \log\frac{a_i}{b_i} \geq \left(\sum_{i=1}^n a_i\right)\log\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}\,,$$

with equality if and only if a_i/b_i is a constant. By the definition of $H(X,Y)$ and using the results from a), we get

$$\begin{aligned}
H(X) + H(Y) - H(X,Y) &= -\sum_{x,y} P(x,y)[\log P(x) + \log P(y) - \log P(x,y)] \\
&= \sum_{x,y} P(x,y)\log\frac{P(x,y)}{P(x)P(y)} \geq \left(\sum_{x,y} P(x,y)\right)\log\frac{\sum_{x,y} P(x,y)}{\sum_{x,y} P(x)P(y)} \\
&= 1\cdot\log\frac{1}{1} = 0 \;\Rightarrow\; H(X) + H(Y) \geq H(X,Y)\,.
\end{aligned}$$

Clearly, equality holds if and only if $P(X,Y) = P(X)P(Y)$; that is, when X and Y are independent.

c) From the lectures, we know that

$$\begin{aligned}
&H(X,Y) = H(Y) + H(X|Y) \;\Rightarrow\; H(X|Y) = H(X,Y) - H(Y) \\
&\Rightarrow\; H(X) - H(X|Y) = H(X) + H(Y) - H(X,Y)\,.
\end{aligned}$$

From the result in b), it follows that $H(X) \geq H(X|Y)$ with equality if and only if X and Y are independent.

3. The marginal distributions of X and Y are:

$$P_X(0) = \frac{2}{3}\,,\; P_X(1) = \frac{1}{3} \quad \text{and} \quad P_Y(0) = \frac{1}{3}\,,\; P_Y(1) = \frac{2}{3}\,.$$

$$\begin{aligned}
&H(X) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = \frac{2}{3}\log\frac{3}{2} + \frac{1}{3}\log 3\,. \\
&H(Y) = H(X) \\
&H(X,Y) = -3\cdot\frac{1}{3}\log\frac{1}{3} = \log 3\,. \\
&H(X|Y) = H(X,Y) - H(Y) = \frac{2}{3}(\log 3 - \log\frac{3}{2}) = \frac{2}{3}\log 2\,. \\
&H(Y|X) = H(X|Y)\,. \quad \text{(This is not generally true.)}
\end{aligned}$$

4. If $Y = g(X)$ defines a one-to-one transformation of a random variable X, then

$$f_Y(y) = f_X(x)\cdot|dx/dy| = f_X(g^{-1}(y))\cdot|dx/dy|\,.$$

In this case,

$$\begin{aligned}
&y = ax + b \Rightarrow x = \frac{y-b}{a} \Rightarrow \frac{dx}{dy} = \frac{1}{a} \\
&\Rightarrow f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).
\end{aligned}$$

$$\begin{aligned}
H(Y) &= -\int_{-\infty}^{\infty} f_Y(y)\log f_Y(y)\,dy = -\int_{-\infty}^{\infty} \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)\log\left[\frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)\right] dy \\
&= -\int_{-\infty}^{\infty} f_X(x)\,\log\left(\frac{1}{|a|} f_X(x)\right) dx\,,
\end{aligned}$$

where we have made the change of variable $x = (y-b)/a$. The logarithm can be split, and we get:

$$\begin{aligned}
H(Y) &= -\int_{-\infty}^{\infty} f_X(x)\log f_X(x)\,dx - \log\frac{1}{|a|}\cdot\int_{-\infty}^{\infty} f_X(x)\,dx \\
&= H(X) + \log|a|\,.
\end{aligned}$$

5. a) The distribution of X :

$$p(x) = \begin{cases} \alpha \cdot p_1(x) & x \in [1, m] \\ (1-\alpha) \cdot p_2(x) & x \in [m+1, n] \,. \end{cases}$$

The entropy of X is

$$\begin{aligned} H(X) &= -\sum_{x=1}^{n} p(x) \log p(x) \\ &= -\sum_{x=1}^{m} \alpha p_1(x) \log [\alpha p_1(x)] - \sum_{x=m+1}^{n} (1-\alpha) p_2(x) \log [(1-\alpha) p_2(x)] \\ &= -\alpha \log \alpha \sum_{x=1}^{m} p_1(x) - \alpha \sum_{x=1}^{m} p_1(x) \log p_1(x) \\ &- (1-\alpha) \log (1-\alpha) \sum_{x=m+1}^{n} p_2(x) - (1-\alpha) \sum_{x=m+1}^{n} p_2(x) \log p_2(x) \\ &= \alpha H(X_1) + (1-\alpha) H(X_2) - \alpha \log \alpha - (1-\alpha) \log (1-\alpha) \,. \end{aligned}$$

b) This is not a rigorous solution. At first, we make a wild guess that $\alpha = 1/2$ maximizes $H(X)$. It maximizes the part that does not depend on the distributions of X_1 and X_2 ; that is, the expression

$$-\alpha \log \alpha - (1-\alpha) \log (1-\alpha) \,.$$

If logarithms to the base 2 are used and $\alpha = 1/2$, this expression is equal to $\log 2$. So,

$$\begin{aligned} 2^{H(X)} &= 2^{-\alpha \log \alpha} \cdot 2^{-(1-\alpha)\log(1-\alpha)} \cdot 2^{\alpha H(X_1)} \cdot 2^{(1-\alpha) H(X_2)} \\ &\leq 2^{\log 2} \cdot 2^{[H(X_1)+H(X_2)]/2} \leq 2\left(\frac{1}{2} \cdot 2^{H(X_1)} + \frac{1}{2} \cdot 2^{H(X_2)}\right) \\ &= 2^{H(X_1)} + 2^{H(X_2)} \,. \end{aligned}$$

The last inequality was due to the convexity of the function 2^x .

Interpretation: Entropy $H(X)$ represents the average number of bits needed to describe the random variable X . Thus, $2^{H(X)}$ can be interpreted as the effective size of the alphabet needed. The result just 'derived' says that if two disjoint alphabets are combined randomly like here, the effective size of the new alphabet is at most the sum of the sizes of the component alphabets.