Exercise 2.1:

Coin Flips. A fair coin is flipped until the first head occurs. Let $X$ denote the number of flips required.

(a) Find the entropy $H(X)$ in bits. The following expressions may be useful:
\[
\sum_{n=1}^{\infty} r^n = \frac{r}{1-r}, \quad \sum_{n=1}^{\infty} nr^n = \frac{r}{(1-r)^2}
\]

(b) A random variable $X$ is drawn according to this distribution. Find an “efficient” sequence of yes-no questions of the form, “Is $X$ contained in the set $S$?” Compare $H(X)$ to the expected number of questions required to determine $X$.

Solution:
The probability for the random variable is given by $P\{X = i\} = 0.5^i$. Hence,
\[
H(X) = -\sum_i p_i \log p_i \\
= -\sum_i 0.5^i \log(0.5^i) \\
= -\log(0.5) \sum_i i \cdot 0.5^i \\
= \frac{0.5}{(1-0.5)^2} \\
= 2
\]

Exercise 2.3:

Minimum entropy. What is the minimum value of $H(p_1, \ldots, p_n) = H(p)$ as $p$ ranges over the set of $n$-dimensional probability vectors? Find all $p$’s which achieve this minimum.

Solution:
Since $H(p) \geq 0$ and $\sum_i p_i = 1$, then the minimum value for $H(p)$ is 0 which is achieved when $p_i = 1$ and $p_j = 0$, $j \neq i$.

Exercise 2.11:

Average entropy. Let $H(p) = -p \log_2 p - (1-p) \log_2(1-p)$ be the binary entropy function.

(a) Evaluate $H(1/4)$.

(b) Calculate the average entropy $H(p)$ when the probability $p$ is chosen uniformly in the range $0 \leq p \leq 1$. 

Solution:

(a) \[
H(1/4) = -1/4 \log_2(1/4) - (1 - 1/4) \log_2(1 - 1/4) \\
= 0.8113
\] (3)

(b) \[
\bar{H}(p) = E[H(p)] \\
= \int_{-\infty}^{\infty} H(p)f(p)dp
\] (4)

Now, \[
f(p) = \begin{cases} 
1, & 0 \leq p \leq 1, \\
0, & \text{otherwise.}
\end{cases}
\] (5)

So, \[
\bar{H}(p) = \int_0^1 H(p)dp \\
= - \int_0^1 (p \log p + (1 - p) \log(1 - p)) dp \\
= - \left[ \int_0^1 p \log p dp - \int_1^0 q \log q dq \right] \\
= -2 \int_0^1 p \log p dp
\] (6)

Letting \( u = \ln p \) and \( v = p^2 \) and integrating by parts, we have:

\[
\bar{H}(p) = - \int u dv \\
= -\left[ uv - \int udv \right] \\
= -\left[ p^2 \ln p \ln 2 - \int p^2 \frac{1}{p \ln 2} dp \right] \\
= -\left[ p^2 \ln p \frac{1}{2 \ln 2} \right]_0^1 \\
= \frac{1}{2 \ln 2}
\] (7)

Exercise 2.16:

Example of joint entropy. Let \( p(x, y) \) be given by

\[
\begin{array}{c|cc}
X \setminus Y & 0 & 1 \\
0 & 1/3 & 1/3 \\
1 & 0 & 1/3 \\
\end{array}
\]

Find
(a) $H(X), H(Y)$.

(b) $H(X|Y), H(Y|X)$.

(c) $H(X, Y)$.

(d) $H(Y) - H(Y|X)$.

(e) $I(X; Y)$.

(f) Draw a Venn diagram for the quantities in (a) through (e).

**Solution:**

(a)

$$H(X) = -\frac{2}{3} \log\left(\frac{2}{3}\right) - \frac{1}{3} \log\left(\frac{1}{3}\right)$$

$$= \log 3 - \frac{2}{3}$$

$$= 0.9183$$

(b)

$$H(X|Y) = \sum_x \sum_y p(x, y) \log\left(\frac{p(y)}{p(x, y)}\right)$$

$$= \frac{1}{3} \log\left(\frac{1}{3}\right) + \frac{1}{3} \log\left(\frac{2}{3}\right) + 0 + \frac{1}{3} \log\left(\frac{2}{3}\right)$$

$$= \frac{2}{3} \log 2 + \frac{1}{3} \log 1$$

$$= \frac{2}{3}$$

$$H(Y|X) = \sum_x \sum_y p(x, y) \log\left(\frac{p(x)}{p(x, y)}\right)$$

$$= \frac{1}{3} \log\left(\frac{2}{3}\right) + \frac{1}{3} \log\left(\frac{2}{3}\right) + 0 + \frac{1}{3} \log\left(\frac{1}{3}\right)$$

$$= \frac{2}{3} \log 2 + \frac{1}{3} \log 1$$

$$= \frac{2}{3}$$
H(X, Y) = \sum_x \sum_y p(x, y) \log p(x, y)
\begin{align}
&= - \left[ \frac{1}{3} \log \left(\frac{1}{3}\right) + \frac{1}{3} \log \left(\frac{1}{3}\right) + 0 \log 0 + \frac{1}{3} \log \left(\frac{1}{3}\right) \right] \\
&= \log 3
\end{align}
(12)

(d)

H(Y) - H(Y|X) = \log 3 - \frac{2}{3} - \frac{2}{3}
= \log 3 - \frac{4}{3}
(13)

(e)

I(X:Y) = \sum_x \sum_y p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right)
\begin{align}
&= \frac{1}{3} \log \left(\frac{\frac{1}{3}}{\frac{1}{3}}\right) + \frac{1}{3} \log \left(\frac{\frac{1}{3}}{\frac{1}{3}}\right) + 0 + \frac{1}{3} \log \left(\frac{\frac{1}{3}}{\frac{1}{3}}\right) \\
&= \frac{2}{3} \log \frac{3}{2} + \frac{1}{3} \log \frac{3}{4} \\
&= \log 3 - \frac{4}{3}
\end{align}
(14)

Exercise 2.18:
Entropy of a sum. Let X and Y be random variables that take on values x_1, x_2, \ldots, x_r and y_1, y_2, \ldots, y_s. Let Z = X + Y.

(a) Show that H(Z|X) = H(Y|X). Argue that if X, Y are independent then H(Y) \leq H(Z) and H(X) \leq H(Z). Thus the addition of independent random variables adds uncertainty.

(b) Give an example (of necessarily dependent random variables) in which H(X) > H(Z) and H(Y) > H(Z).

(c) Under what conditions does H(Z) = H(X) + H(Y)?

Solution:

(a)

\begin{align}
H(Z|X) &= \sum_x \sum_z p(z, x) \log p(z|x) \\
&= \sum_x p(x) \sum_z p(z|x) \log p(z|x) \\
&= \sum_x p(x) \sum_y p(y|x) \log p(y|x) \\
&= H(Y|X)
\end{align}
(15)

If X, Y are independent, then H(Y|X) = H(Y). Now, H(Z) = H(Z|X) = H(Y|X) = H(Y). Similarly H(X) \leq H(Z)
(b) If $X, Y$ are dependent such that $Pr\{Y = -x|X = x\} = 1$, then $Pr\{Z = 0\} = 1$, so that $H(X) = H(Y) > 0$, but $H(Z) = 0$. Hence $H(X) > H(Z)$ and $H(Y) > H(Z)$. Another example is the sum of the two opposite faces on a dice, which always add to seven.

(c) The random variables $X, Y$ are independent and $x_i + y_j \neq x_m + y_n$ for all $i, m \in R$ and $j, n \in S$, i.e., the two random variables $X, Y$ never sum up to the same value. In other words, the alphabet of $Z$ is $r \times s$. The proof is as follows. Notice that $Z = X + Y = \phi(X, Y)$. Now

\[
H(Z) = H(\phi(X, Y)) \\
\leq H(X, Y) \\
= H(X) + H(Y|X) \\
\leq H(X) + H(Y)
\]  

(16)

Now if $\phi(\cdot)$ is a bijection (i.e., only one pair of $x, y$ maps to one value of $z$), then $H(\phi(X, Y)) = H(X, Y)$ and if $X, Y$ are independent then $H(Y|X) = H(Y)$. Hence, with these two conditions, $H(Z) = H(X) + H(Y)$.

**Exercise 2.21:**

*Data processing.* Let $X_1 \rightarrow X_2 \rightarrow X_3 \rightarrow \cdots \rightarrow X_n$ form a Markov chain in this order; i.e., let

\[
p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2|x_1)\cdots p(x_n|x_{n-1})
\]

(17)

Reduce $I(X_1; X_2, \ldots, X_n)$ to its simplest form.

**Solution:**

\[
I(X_1; X_2, \ldots, X_n) = H(X_1) - H(X_1|X_2, \ldots, X_n) \\
= H(X_1) - [H(X_1, X_2, \ldots, X_n) - H(X_2, \ldots, X_n)] \\
= H(X_1) - \left[\sum_{i=1}^{n} H(X_i|X_{i-1}, \ldots, X_1) - \sum_{i=2}^{n} H(X_i|X_{i-1}, \ldots, X_2)\right] \\
= H(X_1) - \left[H(X_1) + \sum_{i=2}^{n} H(X_i|X_{i-1}) - \left(H(X_2) + \sum_{i=3}^{n} H(X_i|X_{i-1})\right)\right] \\
= H(X_2) - H(X_2|X_1) \\
= I(X_2; X_1) \\
= I(X_1; X_2)
\]

(18)

**Exercise 2.33:**

*Fano’s inequality.* Let $Pr(X = i) = p_i, i = 1, 2, \ldots, m$ and let $p_1 \geq p_2 \geq p_3 \geq \cdots \geq p_m$. The minimal probability of error predictor of $X$ is $\hat{X} = 1$, with resulting probability of error $P_e = 1 - p_1$. Maximise $H(p)$ subject to the constraint $1 - p_1 = P_e$ to find a bound on $P_e$ in terms of $H$. This is Fano’s inequality in the absence of conditioning.

**Solution:**

We want to maximise $H(p) = \sum_{i=1}^{m} p_i \log p_i$ subject to the constraints $1 - p_1 = P_e$ and $\sum p_i = 1$. Form the Lagrangian:

\[
\mathcal{L} = H(p) + \lambda(P_e - 1 + p_1) + \mu\left(\sum p_i - 1\right)
\]

(19)
and take the partial derivatives for each $p_i$ and the Lagrangian multipliers:

$$\begin{align*}
\frac{\partial L}{\partial p_i} &= -(\log p_i + 1) + \mu, \ i \neq 1 \\
\frac{\partial L}{\partial p_1} &= -(\log p_1 + 1) + \lambda + \mu \\
\frac{\partial L}{\partial \lambda} &= p_e - 1 + p_1 \\
\frac{\partial L}{\partial \mu} &= \sum p_i - 1
\end{align*}$$

(20)

Setting these equations to zero, we have

$$\begin{align*}
p_i &= 2^{\mu - 1} \\
p_1 &= 2^{\lambda + \mu - 1} \\
p_1 &= 1 - P_e \\
\sum p_i &= 1
\end{align*}$$

(21)

We proceed by eliminating $\mu$. Since the probabilities must sum to one,

$$1 - P_e + \sum_{i \neq 1} 2^{\mu - 1} = 1$$

$$\Rightarrow \mu = 1 + \log \left( \frac{P_e}{m - 1} \right)$$

(22)

Hence, we have

$$\begin{align*}
p_1 &= 1 - P_e \\
p_i &= \frac{P_e}{m - 1}, \ \forall i \neq 1
\end{align*}$$

(23)

Since we know that for these probabilities the entropy is maximised,

$$\begin{align*}
H(p) &\leq - \left[ (1 - P_e) \log(1 - P_e) + \sum_{i \neq 1} \frac{P_e}{m - 1} \log \left( \frac{P_e}{m - 1} \right) \right] \\
&= - \left[ (1 - P_e) \log(1 - P_e) + P_e \log P_e + P_e \sum_{i \neq 1} \frac{1}{m - 1} \log \left( \frac{1}{m - 1} \right) \right] \\
&= H(P_e) + P_e \log (|\mathcal{H}| - 1)
\end{align*}$$

(24)

from which we get Fano’s inequality in the absence of conditioning.