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

(a) Show that $H(Z|X) = H(Y|X)$. Argue that if X, Y are independent, then $H(Y) \le H(Z)$ and $H(X) \le H(Z)$. 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: *Entropy of a sum.*

(a) $Z = X + Y$. Hence $p(Z = z|X = x) = p(Y = z - x|X = x)$.

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

If X and Y are independent, then $H(Y|X) = H(Y)$. Since $I(X;Z) \ge 0$, we have $H(Z) \ge H(Z|X) = H(Y|X) = H(Y)$. Similarly we can show that $H(Z) \ge H(X)$.

(b) Consider the following joint distribution for X and Y Let

$$X = -Y = \begin{cases} 1 & \text{with probability } 1/2 \\ 0 & \text{with probability } 1/2 \end{cases}$$

Then $H(X) = H(Y) = 1$, but $Z = 0$ with prob. 1 and hence $H(Z) = 0$.

(c) We have

$$H(Z) \le H(X,Y) \le H(X) + H(Y)$$

because Z is a function of (X,Y) and $H(X,Y) = H(X) + H(Y|X) \le H(X) + H(Y)$. We have equality iff (X,Y) is a function of Z and $H(Y) = H(Y|X)$, i.e., X and Y are independent.

41. **Random questions**

One wishes to identify a random object $X \sim p(x)$. A question $Q \sim r(q)$ is asked at random according to $r(q)$. This results in a deterministic answer $A = A(x,q) \in \{a_1, a_2, \ldots\}$. Suppose X and Q are independent. Then $I(X;Q,A)$ is the uncertainty in X removed by the question-answer (Q,A).

(a) Show $I(X;Q,A) = H(A|Q)$. Interpret.

(b) Now suppose that two i.i.d. questions $Q_1, Q_2, \sim r(q)$ are asked, eliciting answers A_1 and A_2. Show that two questions are less valuable than twice a single question in the sense that $I(X;Q_1,A_1,Q_2,A_2) \leq 2I(X;Q_1,A_1)$.

Solution: *Random questions.*

(a)

$$\begin{aligned}
I(X;Q,A) &= H(Q,A) - H(Q,A,|X) \\
&= H(Q) + H(A|Q) - H(Q|X) - H(A|Q,X) \\
&= H(Q) + H(A|Q) - H(Q) \\
&= H(A|Q)
\end{aligned}$$

The interpretation is as follows. The uncertainty removed in X by (Q,A) is the same as the uncertainty in the answer given the question.

(b) Using the result from part a and the fact that questions are independent, we can easily obtain the desired relationship.

$$\begin{aligned}
I(X;Q_1,A_1,Q_2,A_2) &\stackrel{(a)}{=} I(X;Q_1) + I(X;A_1|Q_1) + I(X;Q_2|A_1,Q_1) + I(X;A_2|A_1, \\
&\stackrel{(b)}{=} I(X;A_1|Q_1) + H(Q_2|A_1,Q_1) - H(Q_2|X,A_1,Q_1) + I(X;A \\
&\stackrel{(c)}{=} I(X;A_1|Q_1) + I(X;A_2|A_1,Q_1,Q_2) \\
&= I(X;A_1|Q_1) + H(A_2|A_1,Q_1,Q_2) - H(A_2|X,A_1,Q_1,Q_2) \\
&\stackrel{(d)}{=} I(X;A_1|Q_1) + H(A_2|A_1,Q_1,Q_2) \\
&\stackrel{(e)}{\leq} I(X;A_1|Q_1) + H(A_2|Q_2) \\
&\stackrel{(f)}{=} 2I(X;A_1|Q_1)
\end{aligned}$$

(a) Chain rule.
(b) X and Q_1 are independent.

(c) Q_2 are independent of X, Q_1, and A_1.
(d) A_2 is completely determined given Q_2 and X.
(e) Conditioning decreases entropy.
(f) Result from part a.

27. **Grouping rule for entropy:** Let $\mathbf{p} = (p_1, p_2, \ldots, p_m)$ be a probability distribution on m elements, i.e, $p_i \geq 0$, and $\sum_{i=1}^m p_i = 1$. Define a new distribution $\mathbf{q}$ on $m-1$ elements as $q_1 = p_1$, $q_2 = p_2, \ldots, q_{m-2} = p_{m-2}$, and $q_{m-1} = p_{m-1} + p_m$, i.e., the distribution $\mathbf{q}$ is the same as $\mathbf{p}$ on $\{1, 2, \ldots, m-2\}$, and the probability of the last element in $\mathbf{q}$ is the sum of the last two probabilities of $\mathbf{p}$. Show that

$$H(\mathbf{p}) = H(\mathbf{q}) + (p_{m-1} + p_m) H\left(\frac{p_{m-1}}{p_{m-1}+p_m}, \frac{p_m}{p_{m-1}+p_m}\right). \tag{2.54}$$

Solution:

$$\begin{aligned}
H(\mathbf{p}) &= -\sum_{i=1}^m p_i \log p_i && (2.55)\\
&= -\sum_{i=1}^{m-2} p_i \log p_i - p_{m-1}\log p_{m-1} - p_m \log p_m && (2.56)\\
&= -\sum_{i=1}^{m-2} p_i \log p_i - p_{m-1}\log \frac{p_{m-1}}{p_{m-1}+p_m} - p_m \log \frac{p_m}{p_{m-1}+p_m} && (2.57)\\
&\quad -(p_{m-1}+p_m)\log(p_{m-1}+p_m) && (2.58)\\
&= H(\mathbf{q}) - p_{m-1}\log \frac{p_{m-1}}{p_{m-1}+p_m} - p_m \log \frac{p_m}{p_{m-1}+p_m} && (2.59)\\
&= H(\mathbf{q}) - (p_{m-1}+p_m)\left(\frac{p_{m-1}}{p_{m-1}+p_m}\log \frac{p_{m-1}}{p_{m-1}+p_m} - \frac{p_m}{p_{m-1}+p_m}\log \frac{p_m}{p_{m-1}+p_m}\right) && (2.60)\\
&= H(\mathbf{q}) + (p_{m-1}+p_m) H_2\left(\frac{p_{m-1}}{p_{m-1}+p_m}, \frac{p_m}{p_{m-1}+p_m}\right), && (2.61)
\end{aligned}$$

where $H_2(a,b) = -a\log a - b \log b$.

28. **Mixing increases entropy.** Show that the entropy of the probability distribution, $(p_1, \ldots, p_i, \ldots, p_j, \ldots, p_m)$, is less than the entropy of the distribution $(p_1, \ldots, \frac{p_i+p_j}{2}, \ldots, \frac{p_i+p_j}{2}, \ldots, p_m)$. Show that in general any transfer of probability that makes the distribution more uniform increases the entropy.

Solution:

Mixing increases entropy.

This problem depends on the convexity of the log function. Let

$$\begin{aligned}
P_1 &= (p_1, \ldots, p_i, \ldots, p_j, \ldots, p_m)\\
P_2 &= (p_1, \ldots, \frac{p_i+p_j}{2}, \ldots, \frac{p_j+p_i}{2}, \ldots, p_m)
\end{aligned}$$

Then, by the log sum inequality,

$$\begin{aligned}
H(P_2) - H(P_1) &= -2\left(\frac{p_i+p_j}{2}\right)\log\left(\frac{p_i+p_j}{2}\right) + p_i \log p_i + p_j \log p_j\\
&= -(p_i+p_j)\log\left(\frac{p_i+p_j}{2}\right) + p_i \log p_i + p_j \log p_j\\
&\geq 0.
\end{aligned}$$

Thus,

$$H(P_2) \geq H(P_1).$$

18. **World Series.** The World Series is a seven-game series that terminates as soon as either team wins four games. Let X be the random variable that represents the outcome of a World Series between teams A and B; possible values of X are AAAA, BABABAB, and BBBAAAA. Let Y be the number of games played, which ranges from 4 to 7. Assuming that A and B are equally matched and that the games are independent, calculate $H(X)$, $H(Y)$, $H(Y|X)$, and $H(X|Y)$.

Solution:

World Series. Two teams play until one of them has won 4 games.

There are 2 (AAAA, BBBB) World Series with 4 games. Each happens with probability $(1/2)^4$.

There are $8 = 2\binom{4}{3}$ World Series with 5 games. Each happens with probability $(1/2)^5$.

There are $20 = 2\binom{5}{3}$ World Series with 6 games. Each happens with probability $(1/2)^6$.

There are $40 = 2\binom{6}{3}$ World Series with 7 games. Each happens with probability $(1/2)^7$.

The probability of a 4 game series ($Y = 4$) is $2(1/2)^4 = 1/8$.
The probability of a 5 game series ($Y = 5$) is $8(1/2)^5 = 1/4$.
The probability of a 6 game series ($Y = 6$) is $20(1/2)^6 = 5/16$.
The probability of a 7 game series ($Y = 7$) is $40(1/2)^7 = 5/16$.

$$\begin{aligned} H(X) &= \sum p(x) log \frac{1}{p(x)} \\ &= 2(1/16)\log 16 + 8(1/32)\log 32 + 20(1/64)\log 64 + 40(1/128)\log 128 \\ &= 5.8125 \end{aligned}$$

$$\begin{aligned} H(Y) &= \sum p(y) log \frac{1}{p(y)} \\ &= 1/8\log 8 + 1/4\log 4 + 5/16\log(16/5) + 5/16\log(16/5) \\ &= 1.924 \end{aligned}$$

Y is a deterministic function of X, so if you know X there is no randomness in Y. Or, $H(Y|X) = 0$.

Since $H(X) + H(Y|X) = H(X,Y) = H(Y) + H(X|Y)$, it is easy to determine $H(X|Y) = H(X) + H(Y|X) - H(Y) = 3.889$