1. **Entropy of functions of a random variable.**

Let $X$ be a discrete random variable. Show that the entropy of a function of $X$ is less than or equal to the entropy of $X$ by justifying the following steps:

\[
H(X, g(X)) \overset{(a)}{=} H(X) + H(g(X) | X) \overset{(b)}{=} H(X). \\
H(X, g(X)) \overset{(c)}{=} H(g(X)) + H(X | g(X)) \overset{(d)}{\geq} H(g(X)).
\]

Thus $H(g(X)) \leq H(X)$.

2. **Example of joint entropy.**

Let $p(x, y)$ be given by

\[
\begin{array}{ccc}
| & Y & 0 & 1 & \\
| X & 0 & \frac{1}{3} & \frac{1}{3} & \\
| & 1 & 0 & \frac{1}{3} & \\
\end{array}
\]

Find

(a) $H(X), H(Y)$.
(b) $H(X|Y), H(Y|X)$.
(c) $H(X,Y)$.
(d) $H(Y) - H(Y|X)$.
(e) $I(X;Y)$.

3. **Bytes.**

The entropy, $H_a(X) = -\sum p(x) \log_a p(x)$ is expressed in bits if the logarithm is to the base 2 and in bytes if the logarithm is to the base 256. What is the relationship of $H_2(X)$ to $H_{256}(X)$?
4. **Two looks.**

Here is a statement about pairwise independence and joint independence. Let $X, Y_1,$ and $Y_2$ be binary random variables. If $I(X; Y_1) = 0$ and $I(X; Y_2) = 0,$ does it follow that $I(X; Y_1, Y_2) = 0$?

(a) Yes or no?

(b) Prove or provide a counterexample.

(c) If $I(X; Y_1) = 0$ and $I(X; Y_2) = 0$ in the above problem, does it follow that $I(Y_1; Y_2) = 0$? In other words, if $Y_1$ is independent of $X,$ and if $Y_2$ is independent of $X,$ is it true that $Y_1$ and $Y_2$ are independent?

5. **A measure of correlation.**

Let $X_1$ and $X_2$ be identically distributed, but not necessarily independent. Let

$$\rho = 1 - \frac{H(X_2|X_1)}{H(X_1)}.$$ 

(a) Show $\rho = \frac{I(X_1;X_2)}{H(X_1)}$. (There is no typo in the definition of $\rho$)

(b) Show $0 \leq \rho \leq 1$.

(c) When is $\rho = 0$?

(d) When is $\rho = 1$?

6. **The value of a question.**

Let $X \sim p(x), \ x = 1, 2, \ldots, m.$

We are given a set $S \subseteq \{1, 2, \ldots, m\}.$ We ask whether $X \in S$ and receive the answer

$$Y = \begin{cases} 1, & \text{if } X \in S \\ 0, & \text{if } X \not\in S. \end{cases}$$

Suppose $\Pr\{X \in S\} = \alpha.$

(a) Find the decrease in uncertainty $H(X) - H(X|Y)$.

(b) Is the set $S$ with a given probability $\alpha$ is as good as any other $S' \neq S$ with $\Pr\{X \in S'\} = \alpha$?
7. **Relative entropy is not symmetric**

Let the random variable $X$ have three possible outcomes $\{a, b, c\}$. Consider two distributions on this random variable

<table>
<thead>
<tr>
<th>Symbol</th>
<th>$p(x)$</th>
<th>$q(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1/2</td>
<td>1/3</td>
</tr>
<tr>
<td>b</td>
<td>1/4</td>
<td>1/3</td>
</tr>
<tr>
<td>c</td>
<td>1/4</td>
<td>1/3</td>
</tr>
</tbody>
</table>

Calculate $H(p), H(q), D(p \parallel q)$ and $D(q \parallel p)$. Verify that in this case $D(p \parallel q) \neq D(q \parallel p)$.

8. **“True or False” questions**

Copy each relation and write true or false. Then, if it’s true, prove it. If it is false give a counterexample or prove that the opposite is true.

(a) $H(X) \geq H(X|Y)$

(b) $H(X) + H(Y) \leq H(X, Y)$

(c) Let $X, Y$ be two independent random variables. Then

$$H(X - Y) \geq H(X).$$

(d) Let $X, Y, Z$ be three random variables that satisfies $H(X, Y) = H(X) + H(Y)$ and $H(Y, Z) = H(Z) + H(Y')$. Then the following holds

$$H(X, Y, Z) = H(X) + H(Y) + H(Z).$$

(e) For any $X, Y, Z$ and the deterministic function $f, g$ $I(X; Y|Z) = I(X, f(X, Y); Y, g(Y, Z)|Z)$.

9. **Joint Entropy** Consider $n$ different discrete random variables, named $X_1, X_2, ..., X_n$. Each random variable separately has an entropy $H(X_i)$, for $1 \leq i \leq n$.

(a) What is the upper bound on the joint entropy $H(X_1, X_2, ..., X_n)$ of all these random variables $X_1, X_2, ..., X_n$ given that $H(X_i)$, for $1 \leq i \leq n$ are fixed?

(b) Under what conditions will this upper bound be reached?
(c) What is the lower bound on the joint entropy \( H(X_1, X_2, \ldots, X_n) \) of all these random variables?

(d) Under what condition will this upper bound be reached?

10. **More question of True or False**

Let \( X, Y, Z \) be discrete random variable. Copy each relation and write true or false. If it’s true, prove it. If it is false give a counterexample or prove that the opposite is true.

For instance:

- **H(X) \geq H(X|Y)** is true. Proof: In the class we showed that \( I(X; Y) > 0 \), hence \( H(X) - H(X|Y) > 0 \).
- **H(X) + H(Y) \leq H(X, Y)** is false. Actually the opposite is true, i.e., \( H(X) + H(Y) \geq H(X, Y) \) since \( I(X; Y) = H(X) + H(Y) - H(X, Y) \geq 0 \).

(a) If \( H(X|Y) = H(X) \) then \( X \) and \( Y \) are independent.

(b) For any two probability mass functions (pmf) \( P, Q \),

\[
D\left(\frac{P + Q}{2} || Q\right) \leq \frac{1}{2} D(P || Q),
\]

where \( D(||) \) is a divergence between two pmfs.

(c) Let \( X \) and \( Y \) be two independent random variables. Then

\[
H(X + Y) \geq H(X).
\]

(d) \( I(X; Y) - I(X; Y|Z) \leq H(Z) \)

(e) If \( f(x, y) \) is a convex function in the pair \((x, y)\), then for a fixed \( y \), \( f(x, y) \) is convex in \( x \), and for a fixed \( x \), \( f(x, y) \) is convex in \( y \).

(f) If for a fixed \( y \) the function \( f(x, y) \) is a convex function in \( x \), and for a fixed \( x \), \( f(x, y) \) is convex function in \( y \), then \( f(x, y) \) is convex in the pair \((x, y)\). (Examples of such functions are \( f(x, y) = f_1(x) + f_2(y) \) or \( f(x, y) = f_1(x)f_2(y) \) where \( f_1(x) \) and \( f_2(y) \) are convex.)
(g) Let \(X, Y, Z, W\) satisfy the Markov chain \(X \rightarrow Y \rightarrow Z\) and \(Y \rightarrow Z \rightarrow W\). Does the Markov \(X \rightarrow Y \rightarrow Z \rightarrow W\) hold? (The Markov \(X \rightarrow Y \rightarrow Z \rightarrow W\) means that \(P(x|y, z, w) = P(x|y)\) and \(P(x, y|z, w) = P(x, y|z)\).)

(h) \(H(X|Z)\) is concave in \(P_{X|Z}\) for fixed \(P_Z\).

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 the object \(X\) and the question \(Q\) 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 the value of a single question in the sense that \(I(X; Q_1, A_1, Q_2, A_2) \leq 2I(X; Q_1, A_1)\).

12. Entropy bounds.
Let \(X \sim p(x)\), where \(x\) takes values in an alphabet \(X\) of size \(m\). The entropy \(H(X)\) is given by

\[
H(X) \equiv -\sum_{x \in X} p(x) \log p(x) = E_p \log \frac{1}{p(X)}.
\]

Use Jensen’s inequality \((Ef(X) \leq f(EX), \text{if } f \text{ is concave})\) to show

(a) \(H(X) \leq \log E_p \frac{1}{p(X)} = \log m\).
(b) \(-H(X) \leq \log(\sum_{x \in X} p^2(x))\), thus establishing a lower bound on \(H(X)\).
(c) Evaluate the upper and lower bounds on \(H(X)\) when \(p(x)\) is uniform.
(d) Let \(X_1, X_2\) be two independent drawings of \(X\). Find \(\Pr\{X_1 = X_2\}\) and show \(\Pr\{X_1 = X_2\} \geq 2^{-H}\).

Suppose a (non-stationary) Markov chain starts in one of \(n\) states, necks
down to $k < n$ states, and then fans back to $m > k$ states. Thus $X_1 \rightarrow X_2 \rightarrow X_3$, $X_1 \in \{1, 2, \ldots, n\}$, $X_2 \in \{1, 2, \ldots, k\}$, $X_3 \in \{1, 2, \ldots, m\}$, and $p(x_1, x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_2)$.

(a) Show that the dependence of $X_1$ and $X_3$ is limited by the bottleneck by proving that $I(X_1; X_3) \leq \log k$.
(b) Evaluate $I(X_1; X_3)$ for $k = 1$, and conclude that no dependence can survive such a bottleneck.

14. Convexity of Halfspaces, hyperplanes and polyhedron

Let $x$ be a real vector of finite dimension $n$, i.e., $x \in \mathbb{R}^n$. A halfspace is the set of all $x \in \mathbb{R}^n$ that satisfies $a^T x \leq b$, where $a \neq 0$. In other words a halfspace is the set
$$\{x \in \mathbb{R}^n : a^T x \leq b\}.$$ 
A hyperplan is the set of the form
$$\{x \in \mathbb{R}^n : a^T x = b\}.$$

(a) Show that a halfspace and a hyperplan are convex sets.
(b) show that for any two sets $A$ and $B$ that are convex the intersection $A \cap B$ is also convex.
(c) A polyhedron is an intersection of halfspaces and a hyperplans. Deduce that a polyhedron is a convex set.
(d) A probability vector $x$ is such that each element is positive and it sums to 1. Is the set of all vector probabilities of dimension $n$ (called the probability simplex) a halfspace, hyperplan or polyhedron?

15. Some sets of probability distributions.

Let $X$ be a real-valued random variable with $\Pr(X = a_i) = p_i$, $i = 1, \ldots, n$, where $a_1 < a_2 < \ldots < a_n$. Let $p$ denote the vector $p_1, p_2, \ldots, p_n$. Of course $p \in \mathbb{R}^n$ lies in the standard probability simplex. Which of the following conditions are convex in $p$? (That is, for which of the following conditions is the set of $p \in \mathbb{P}$ that satisfy the condition convex?)
(a) \( \alpha \leq E[f(X)] \leq \beta \), where \( E[f(X)] \) is the expected value of \( f(X) \), i.e. \( E[f(x)] = \sum_{i=1}^{n} p_i f(a_i) \) (The function \( f : \mathbb{R} \rightarrow \mathbb{R} \) is given.)

(b) \( \Pr(X > \alpha) \leq \beta \)

(c) \( E[|X^2|] \leq \alpha E[|X|] \).

(d) \( \text{var}(X) \leq \alpha \), where \( \text{var}(X) = E(X - EX)^2 \) is the variance of \( X \).

(e) \( E[X^2] \leq \alpha \)

(f) \( E[X^2] \geq \alpha \)

16. **Perspective transformation preserve convexity** Let \( f(x), f : \mathbb{R} \rightarrow \mathbb{R} \), be a convex function.

(a) Show that the function \( tf\left(\frac{x}{t}\right) \),

\[ tf\left(\frac{x}{t}\right), \quad (1) \]

is a convex function in the pair \((x, t)\) for \( t > 0 \). (The function \( tf\left(\frac{x}{t}\right) \) is called perspective transformation of \( f(x) \).)

(b) Is the preservation true for concave functions too?

(c) Use this property to prove that \( D(P||Q) \) is a convex function in \((P, Q)\).

17. **Coin Tosses**

Consider the next joint distribution: \( X \) is the number of coin tosses until the first head appears and \( Y \) is the number of coin tosses until the second head appears. The probability for a head is \( q \), and the tosses are independent.

a. Compute the distribution of \( X \), \( p(x) \), the distribution of \( Y \), \( p(y) \), and the conditional distributions \( p(y|x) \) and \( p(x|y) \).

b. Compute \( H(X), H(Y|X), H(X,Y) \). Each term should not include a series. Hint: Is \( H(Y|X) = H(Y - X|X) \)?

c. Compute \( H(Y), H(X|Y), \) and \( I(X;Y) \). If necessary, answers may include a series.

18. **Inequalities** Copy each relation to your notebook and write \( \leq, \geq \) or \( = \), prove it.

(a) Let \( X \) be a discrete random variable. Compare \( \frac{1}{2p(x)} \) vs. \( \max_x p(x) \).
(b) Let $H_b(a)$ denote the binary entropy for $a \in [0, 1]$ and $H_{\text{ter}}$ is the ternary entropy i.e. $H_{\text{ter}}(a, b, c) = -a \log a - b \log b - c \log c$, where $p_1, p_2, p_3 \in [0, 1]$, and $p_1 + p_2 + p_3 = 1$. Compare $H_{\text{ter}}(a b, a \bar{b}, \bar{a})$ vs $H_b(a) + \bar{a}H_b(b)$.

19. **True or False of a constrained inequality:**

Given are three discrete random variables $X, Y, Z$ that satisfy $H(Y|X, Z) = 0$.

(a) Copy the next relation to your notebook and write **true** or **false**.

\[ I(X; Y) \geq H(Y) - H(Z) \]

(b) What are the conditions for which the equality $I(X; Y) = H(Y) - H(Z)$ holds.

(c) Assume that the conditions for $I(X; Y) = H(Y) - H(Z)$ are satisfied. Is it true that there exists a function such that $Z = g(Y)$?

20. **True or False of**: Copy each relation to your notebook and write **true** or **false**. If true, prove the statement, and if not provide a counterexample.

(a) Let $X - Y - Z - W$ be a Markov chain, then the following holds:

\[ I(X; W) \leq I(Y; Z). \]

(b) For two probability distributions, $p_{XY}$ and $q_{XY}$, that are defined on $\mathcal{X} \times \mathcal{Y}$, the following holds:

\[ D(p_{XY}||q_{XY}) \geq D(p_X||q_X). \]

(c) If $X$ and $Y$ are dependent and also $Y$ and $Z$ are dependent, then $X$ and $Z$ are dependent.

21. **Cross entropy**: Often in Machine learning, cross entropy is used to measure performance of a classifier model such as neural network. Cross entropy is defined for two PMFs $P_X$ and $Q_X$ as

\[ H(P_X, Q_X) \triangleq - \sum_{x \in \mathcal{X}} P_X(x) \log Q_X(x). \]
In a shorter notation we write as

\[ H(P, Q) \triangleq - \sum_{x \in \mathcal{X}} P(x) \log Q(x). \]

Copy each of the following relations to your notebook and write \textbf{true} or \textbf{false} and provide a proof/disproof.

(a) \( 0 \leq H(P, Q) \leq \log |\mathcal{X}| \) for all \( P, Q \).
(b) \( \min_Q H(P, Q) = H(P, P) \) for all \( P \).
(c) \( H(P, Q) \) is concave in the pair \( (P, Q) \).
(d) \( H(P, Q) \) is convex in the pair \( (P, Q) \).