Elements of Information Theory - Chapter 2

.

December 25, 2024

[?]

1 Exercises

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:

   $$\sum_{n=1}^{\infty} r^n = \frac{r}{1-r}, \quad \sum_{n=1}^{\infty} nr^n = \frac{r}{(1-r)^2}$$

   Solution: we have the following pmf

   $$P(X=n) = \frac{1}{2^n}, \quad n \geq 1$$

   which gives use the entropy

   $$H(X) = \sum_{i=1}^{n} P(X=i) \times log(\frac{1}{P(X=i)}) = \sum_{i=1}^{n} \frac{1}{2^n} \times log(2^n) = \sum_{i=1}^{n} \frac{1}{2^n} \times n = \frac{\frac{1}{2}}{(1-\frac{1}{2})^2}$$

   (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: An efficient questioning strategy would be:

   - Q1: "Is $X = 1$?"
   - If no, Q2: "Is $X = 2$?"
   - If no, Q3: "Is $X = 3$?"
   - And so on...

The expected number of questions L needed is:

$$L = \sum_{n=1}^{\infty} n \cdot P(X = n) = \sum_{n=1}^{\infty} n \cdot \frac{1}{2^n} = 2$$

We found earlier that $H(X) = 2$ bits. Therefore, the expected number of questions equals the entropy: $L = H(X) = 2$. This achieves the theoretical lower bound, proving our questioning strategy is optimal.

2. Entropy of functions. Let X be a random variable taking on a finite number of values. What is the (general) inequality relationship of $H(X)$ and $H(Y)$ if

   (a) $Y = 2^X$?

   Solution: Since $Y = 2^X$ is a deterministic function of X, by the Data Processing Inequality:

   $$H(Y) \leq H(X)$$

   This is because a function of a random variable cannot increase entropy - any transformation can only preserve or lose information. The mapping from X to Y is one-to-one (injective) in this case, so actually:

   $$H(Y) = H(X)$$

   (b) $Y = \cos X$?

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

   Solution: The minimum value of the entropy is 0 bits. This occurs when one probability is 1 and all others are 0.

   Formally, the set of probability vectors that minimize $H(\mathbf{p})$ are:

   $$\mathbf{p} = (p_1, \ldots, p_n) \text{ where } p_i = 1 \text{ for some } i \text{ and } p_j = 0 \text{ for all } j \neq i$$

   This makes intuitive sense because when we know the outcome with certainty (probability 1), there is no uncertainty and thus zero entropy.

4. Axiomatic definition of entropy. If a sequence of symmetric functions $H_m(p_1, p_2, \ldots, p_m)$ satisfies the following properties:

   - Normalization: $H_2(\frac{1}{2}, \frac{1}{2}) = 1$
   - Continuity: $H_2(p, 1-p)$ is a continuous function of p
   - Grouping: $H_m(p_1, p_2, \ldots, p_m) = H_{m-1}(p_1 + p_2, p_3, \ldots, p_m) + (p_1 + p_2)H_2(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$

Prove that H_m must be of the form:

$$H_m(p_1, p_2, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log p_i, \quad m = 2, 3, \ldots$$

Solution: Let's prove this using functional equations.

(a) First, we'll show that for $m = 2$, $H_2(p, 1-p)$ must be of the form $-p \log p - (1-p)\log(1-p)$.

(b) Let $f(x) = H_2(x, 1-x)$. By the grouping property with $m = 3$:

$$H_3(x, y, 1-x-y) = f(x+y) + (x+y)f(\frac{x}{x+y})$$

(c) The same expression can be written differently by grouping y and $1-x-y$:

$$H_3(x, y, 1-x-y) = f(x) + (1-x)f(\frac{y}{1-x})$$

(d) Equating these expressions:

$$f(x+y) + (x+y)f(\frac{x}{x+y}) = f(x) + (1-x)f(\frac{y}{1-x})$$

(e) This functional equation, combined with continuity and $f(\frac{1}{2}) = 1$, has only one solution:

$$f(x) = -x \log x - (1-x)\log(1-x)$$

(f) Now for general m, repeated application of the grouping property shows:

$$H_m(p_1, \ldots, p_m) = -\sum_{i=1}^{m} p_i \log p_i$$

To verify: This solution satisfies all axioms:

- Normalization: $-\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$
- Continuity: The function is continuous for all $p \in (0, 1)$
- Grouping: Can be verified by direct substitution
- Symmetry: The sum is invariant under permutation of indices

The uniqueness follows from the fact that the functional equation has only one continuous solution satisfying the normalization condition.

5. 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:

$$
\begin{aligned}
H(X, g(X)) &\overset{(a)}{=} H(X) + H(g(X)|X) && (1)\\
&\overset{(b)}{=} H(X); && (2)\\
H(X, g(X)) &\overset{(c)}{=} H(g(X)) + H(X|g(X)) && (3)\\
&\overset{(d)}{\geq} H(g(X)) && (4)
\end{aligned}
$$

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

Solution:

(a) by the entropy chain rule

(b) $H(g(X)|X) = 0$ as $g(X)$ depends on X

(c) by symmetry of the entropy chain rule

(d) $H(X|g(X)) \geq 0$ since conditional entropy is always non-negative (this is because entropy measures uncertainty, and uncertainty cannot be negative)

6. Zero conditional entropy. Show that if $H(Y|X) = 0$, then Y is a function of X, i.e., for all x with $p(x) > 0$, there is only one possible value of y with $p(x, y) > 0$.

   Solution: Let's prove this by contradiction.

   Recall that $H(Y|X) = \sum_x p(x)H(Y|X = x)$, where $H(Y|X = x) = -\sum_y p(y|x)\log p(y|x)$.

   Suppose there exists some x_0 with $p(x_0) > 0$ for which there are at least two values y_1 and y_2 with $p(y_1|x_0) > 0$ and $p(y_2|x_0) > 0$.

   Then:

$$
\begin{aligned}
H(Y|X) &= \sum_x p(x)H(Y|X = x)\\
&\geq p(x_0)H(Y|X = x_0)\\
&= p(x_0)(-\sum_y p(y|x_0)\log p(y|x_0))\\
&> 0
\end{aligned}
$$

   The last inequality follows because:

   - $p(x_0) > 0$ by assumption
   - $H(Y|X = x_0) > 0$ since it has at least two non-zero probabilities (and entropy of a non-deterministic distribution is always positive)

This contradicts our assumption that $H(Y|X) = 0$. Therefore, for each x with $p(x) > 0$, there must be exactly one value of y with $p(y|x) > 0$ (which must equal 1). This means Y is a function of X.

7. Infinite entropy. This problem shows that the entropy of a discrete random variable can be infinite. Let $A = \sum_{n=2}^{\infty}(n \log^2 n)^{-1}$. (It is easy to show that A is finite by bounding the infinite sum by the integral of $(x \log^2 x)^{-1}$.) Show that the integer-valued random variable X defined by $\Pr(X = n) = (An \log^2 n)^{-1}$ for $n = 2, 3, \ldots$ has $H(X) = +\infty$.

   Solution: Let's calculate the entropy directly:

$$
\begin{aligned}
H(X) &= -\sum_{n=2}^{\infty} \Pr(X = n) \log \Pr(X = n) \\
&= -\sum_{n=2}^{\infty} \frac{1}{An \log^2 n} \log\left(\frac{1}{An \log^2 n}\right) \\
&= \sum_{n=2}^{\infty} \frac{1}{An \log^2 n} [\log A + \log n + 2 \log \log n] \\
&= \frac{1}{A} \sum_{n=2}^{\infty} \frac{\log A}{n \log^2 n} + \frac{1}{A} \sum_{n=2}^{\infty} \frac{1}{n \log n} + \frac{2}{A} \sum_{n=2}^{\infty} \frac{1}{n \log^2 n}
\end{aligned}
$$

   The first and third terms are finite, but the middle sum $\sum_{n=2}^{\infty} \frac{1}{n \log n}$ diverges (this can be shown using the integral test). Therefore, $H(X) = +\infty$.

References