Physics of Information Technology

Jiaming Liu

February 25 2025

Solutions to Problems 4.1–4.6

Problem 4.1: Properties of the Entropy Function

The Shannon (discrete) entropy is defined by

$$H(p) = -\sum_{i=1}^{X} p_i \ln p_i. \tag{4.1}$$

We now verify that this function satisfies four important properties.

Continuity

Each term $-p_i \ln p_i$ is continuous for $p_i > 0$; by convention, we set $0 \ln 0 = 0$. Thus, small changes in the probabilities p_i lead to small changes in $H(p)$.

Non-negativity

For $0 \le p_i \le 1$, the product $p_i \ln p_i$ is less than or equal to zero. Hence,

$$H(p) \ge 0,$$

with equality if and only if one of the p_i equals 1 (i.e. the outcome is certain).

Boundedness

When all outcomes are equally likely, i.e. $p_i = \frac{1}{X}$ for all i, we have

$$H(p) = -\sum_{i=1}^{X} \frac{1}{X} \ln \frac{1}{X} = \ln X.$$

Thus, the entropy is maximized by the uniform distribution and is bounded above by $\ln X$.

Additivity for Independent Variables

If we have two independent random variables x and y with joint probability $p(x, y) = p(x)p(y)$, then the joint entropy is:

$$\begin{aligned}
H(x, y) &= -\sum_{x,y} p(x, y) \ln[p(x)p(y)] \\
&= -\sum_{x,y} p(x)p(y) \ln p(x) - \sum_{x,y} p(x)p(y) \ln p(y) \\
&= -\sum_{x} p(x) \ln p(x) - \sum_{y} p(y) \ln p(y) \\
&= H(x) + H(y).
\end{aligned}$$

Thus, the entropy of independent events is additive.

Problem 4.2: Equivalent Forms of Mutual Information

The mutual information between two random variables x and y is given by several equivalent expressions:

$$I(x,y) = H(x) + H(y) - H(x,y) = H(y) - H(y|x) = H(x) - H(x|y) = \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{p(x)p(y)}. \tag{4.10}$$

Derivation

1. **Using the Chain Rule:** The chain rule for entropy tells us that
$$H(x,y) = H(x) + H(y|x).$$
Rearranging gives
$$H(y|x) = H(x,y) - H(x),$$
so that
$$H(y) - H(y|x) = H(x) + H(y) - H(x,y).$$

2. Similarly, by writing
$$H(x,y) = H(y) + H(x|y),$$
we find
$$H(x) - H(x|y) = H(x) + H(y) - H(x,y).$$

3. **Logarithmic Form:** By definition, the mutual information is also given by
$$I(x,y) = \sum_{x,y} p(x,y) \ln \frac{p(x,y)}{p(x)p(y)}.$$
A detailed manipulation (using properties of logarithms and the definitions of conditional and joint entropies) shows that this expression is equivalent to $H(x) + H(y) - H(x,y)$.

Each of these forms emphasizes a different aspect of the information shared by x and y.

Problem 4.3: Error Reduction in a Binary Channel with Error Probability ϵ

Assume a binary channel where each transmitted bit has an error probability ϵ. We use repetition coding and majority voting to reduce the error probability.

(a) Three Transmissions (Single Majority Vote)

When the same bit is transmitted three times, the probability of exactly k errors is given by the binomial formula:
$$P(k \text{ errors}) = \binom{3}{k} \epsilon^k (1-\epsilon)^{3-k}.$$
An error occurs if the majority of bits are wrong; that is, if there are 2 or 3 errors. Thus, the overall error probability is
$$P_{\text{error}} = \binom{3}{2} \epsilon^2 (1-\epsilon) + \binom{3}{3} \epsilon^3.$$
For small ϵ, the dominant term is
$$P_{\text{error}} \approx 3\epsilon^2.$$

(b) Two-Level Majority Voting (3 Groups of 3 Transmissions)

Suppose now that we send 9 bits organized as 3 groups of 3. First, each group of 3 is decoded using majority voting (with error probability approximately $p \approx 3\epsilon^2$ as above). Next, we take a majority vote of the three group decisions. The error probability at the second stage is then

$$P_{\text{final}} = \binom{3}{2} p^2(1-p) + p^3.$$

For small p, we approximate

$$P_{\text{final}} \approx 3p^2 \approx 3\,(3\epsilon^2)^2 = 27\,\epsilon^4.$$

(c) General N-Level Majority Voting

If we repeat the majority voting process N times (each level using groups of 3 bits, for a total of 3^N base bits), the error probability at each level is approximately given by

$$p_{k+1} \approx 3\,(p_k)^2,$$

with $p_0 = \epsilon$. Hence, after N levels the error probability behaves roughly as

$$p_N \sim C\,\epsilon^{2^N},$$

where C is a constant resulting from the repeated binomial coefficients. This shows a rapid (roughly double-exponential) reduction in error probability as N increases.

Problem 4.4: Differential Entropy of a Gaussian Process

Consider a continuous random variable x that is normally distributed:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\,\exp\!\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

The differential entropy is defined by

$$H = -\int_{-\infty}^{\infty} p(x)\ln p(x)\,dx. \tag{4.13}$$

Derivation

1. **Compute $\ln p(x)$:**

$$\begin{aligned}\ln p(x) &= \ln\left[\frac{1}{\sqrt{2\pi\sigma^2}}\right] + \ln\left[\exp\!\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right]\\ &= -\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}.\end{aligned}$$

2. **Substitute into the Entropy Integral:**

$$H = -\int_{-\infty}^{\infty} p(x)\left[-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2}\right]dx.$$

   This simplifies to

$$H = \int_{-\infty}^{\infty} p(x)\left[\frac{1}{2}\ln(2\pi\sigma^2) + \frac{(x-\mu)^2}{2\sigma^2}\right]dx.$$

3. **Separate the Integral:**

$$H = \frac{1}{2}\ln(2\pi\sigma^2)\int_{-\infty}^{\infty} p(x)\,dx + \frac{1}{2\sigma^2}\int_{-\infty}^{\infty} p(x)(x-\mu)^2\,dx.$$

   The first integral is 1 (normalization) and the second is the variance σ^2:

$$H = \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\cdot\sigma^2.$$

4. **Combine the Results:**

$$H = \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2}.$$

5. **Write in Compact Form:**
   Recognize that $\frac{1}{2} = \frac{1}{2}\ln e$ (since $\ln e = 1$); therefore,

$$H = \frac{1}{2}\left[\ln(2\pi\sigma^2) + \ln e\right] = \frac{1}{2}\ln(2\pi e\,\sigma^2).$$

This is the result stated as Equation (4.25).

Problem 4.5: Capacity of a Telephone Line

A standard telephone line is specified to have a bandwidth of

$$\Delta f = 3300\,\text{Hz},$$

and a signal-to-noise ratio (SNR) of 20 dB.

The Shannon capacity for a band-limited Gaussian channel is given by:

$$C = \Delta f\,\log_2\left(1 + \frac{S}{N}\right) \quad \text{(bits per second).}$$

(a) Calculate the Capacity

1. **Convert SNR from dB to Linear:**
   20 dB corresponds to

$$\frac{S}{N} = 10^{20/10} = 100.$$

2. **Apply the Capacity Formula:**

$$C = 3300\,\log_2(1+100) = 3300\,\log_2(101).$$

3. **Approximation:**
   Since $\log_2(101)$ is approximately 6.66, we have

$$C \approx 3300 \times 6.66 \approx 21\,978\,\text{bits per second},$$

   which is roughly 22 kbps.

(b) SNR Required for 1 Gbit/s Capacity

Set the capacity equal to 10^9 bits per second:

$$10^9 = 3300 \log_2\left(1 + \frac{S}{N}\right).$$

1. **Solve for the Logarithm Term:**

$$\log_2\left(1 + \frac{S}{N}\right) = \frac{10^9}{3300} \approx 303030.3.$$

2. **Solve for S/N:**

$$1 + \frac{S}{N} = 2^{303030.3} \quad \Longrightarrow \quad \frac{S}{N} \approx 2^{303030.3} - 1.$$

3. **Express in Decibels:**
   The SNR in decibels is

$$\text{SNR (dB)} = 10 \log_{10}\left(\frac{S}{N}\right).$$

   Noting that $\log_{10}(2) \approx 0.301$, we have

$$\text{SNR (dB)} \approx 10 \times 303030.3 \times 0.301 \approx 912\,000\,\text{dB}.$$

   This value is astronomically high and clearly impractical.

Problem 4.6: The Sample Mean and the Cramér–Rao Bound

Let $x_1, x_2, \ldots, x_n$ be independent and identically distributed samples from the Gaussian distribution

$$N(x_0, \sigma^2).$$

We consider the sample mean estimator:

$$\hat{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Unbiasedness

Since $E[x_i] = x_0$ for all i,

$$E[\hat{x}] = \frac{1}{n} \sum_{i=1}^{n} E[x_i] = \frac{n\, x_0}{n} = x_0.$$

Thus, $\hat{x}$ is an unbiased estimator of x_0.

Variance of the Sample Mean

Because the x_i are independent,

$$\text{Var}(\hat{x}) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(x_i) = \frac{n\, \sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Fisher Information and the Cramér–Rao Lower Bound

For a single observation from $N(x_0, \sigma^2)$, the Fisher information with respect to x_0 is

$$J(x_0) = \frac{1}{\sigma^2}.$$

For n independent observations, the total Fisher information is

$$J_n(x_0) = \frac{n}{\sigma^2}.$$

The Cramér–Rao Lower Bound (CRLB) states that the variance of any unbiased estimator is at least

$$\mathrm{Var}(\hat{x}) \geq \frac{1}{J_n(x_0)} = \frac{\sigma^2}{n}.$$

Since the variance of the sample mean $\hat{x}$ is exactly $\frac{\sigma^2}{n}$, it achieves the CRLB.