

Solutions to Homework Set #6

1. One bit quantization of a single Gaussian random variable

Let $X \sim \mathcal{N}(0, \sigma^2)$ and let the distortion measure be squared error. Here we do not allow block descriptions. Show that the optimum reproduction points for 1 bit quantization are $\pm\sqrt{\frac{2}{\pi}}\sigma$, and that the expected distortion for 1 bit quantization is $\frac{\pi-2}{\pi}\sigma^2$.

Compare this with the distortion rate bound $D = \sigma^2 2^{-2R}$ for $R = 1$.

Solution: One bit quantization of a single Gaussian random variable

With one bit quantization, the obvious reconstruction regions are the positive and negative real axes. Proving that such is optimal needs a bit of care, which we will skip here. The reconstruction point is the centroid of each region. For example, for the positive real line, the centroid a is given as

$$\begin{aligned} a &= \int_0^{\infty} x \frac{2}{\sqrt{2\pi}\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \int_0^{\infty} \sigma \sqrt{\frac{2}{\pi}} e^{-y} dy \\ &= \sigma \sqrt{\frac{2}{\pi}}, \end{aligned}$$

using the substitution $y = x^2/2\sigma^2$. The expected distortion for one bit quantization is

$$\begin{aligned} D &= \int_{-\infty}^0 \left(x + \sigma \sqrt{\frac{2}{\pi}}\right)^2 \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx \\ &\quad + \int_0^{\infty} \left(x - \sigma \sqrt{\frac{2}{\pi}}\right)^2 \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= 2 \int_{-\infty}^{\infty} \left(x^2 + \sigma^2 \frac{2}{\pi}\right) \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx \\ &\quad - 2 \int_0^{\infty} \left(-2x\sigma \sqrt{\frac{2}{\pi}}\right) \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \sigma^2 + \frac{2}{\pi}\sigma^2 - 4 \frac{1}{\sqrt{2\pi}} \sigma^2 \sqrt{\frac{2}{\pi}} \end{aligned}$$

$$\begin{aligned}
&= \sigma^2 \frac{\pi - 2}{\pi} \\
&\approx .3634 \sigma^2,
\end{aligned}$$

which is significantly larger than the distortion rate bound $D(1) = \sigma^2/4$.

2. Rate distortion function with infinite distortion

Find the rate distortion function $R(D) = \min I(X; \hat{X})$ for $X \sim \text{Bernoulli}(\frac{1}{2})$ and distortion

$$d(x, \hat{x}) = \begin{cases} 0, & x = \hat{x}, \\ 1, & x = 1, \hat{x} = 0, \\ \infty, & x = 0, \hat{x} = 1, \end{cases}$$

where $x, \hat{x} \in \{0, 1\}$. Thus reconstructing a 0 by a 1 is never allowed.

Solution: Rate distortion function with infinite distortion

We wish to evaluate the rate distortion function

$$R(D) = \min_{p(\hat{x}|x): \sum_{(x, \hat{x})} p(x)p(\hat{x}|x)d(x, \hat{x}) \leq D} I(X; \hat{X}).$$

Since $d(0, 1) = \infty$, we must have $p(0, 1) = 0$ for any finite distortion D . Thus, the distortion $D = p(1, 0)$, and hence we have the following joint distribution for (X, \hat{X}) (assuming $D \leq \frac{1}{2}$):

$$p(x, \hat{x}) = \begin{bmatrix} \frac{1}{2} & 0 \\ D & \frac{1}{2} - D \end{bmatrix}.$$

The mutual information for this joint distribution is

$$\begin{aligned}
R(D) &= I(X; \hat{X}) = H(X) - H(X|\hat{X}) \\
&= H\left(\frac{1}{2}, \frac{1}{2}\right) - \left(\frac{1}{2} + D\right)H\left(\frac{\frac{1}{2}}{\frac{1}{2} + D}, \frac{D}{\frac{1}{2} + D}\right) \\
&= 1 + \frac{1}{2} \log \frac{\frac{1}{2}}{\frac{1}{2} + D} + D \log \frac{D}{\frac{1}{2} + D},
\end{aligned}$$

which is the rate distortion function for this binary source if $0 \leq D \leq \frac{1}{2}$. Since we can achieve $D = \frac{1}{2}$ with zero rate (use $p(\hat{x} = 0) = 1$), we have $R(D) = 0$ for $D \geq \frac{1}{2}$.

3. Properties of $R(D)$

Here we wish to show that we can consider the minimum distortion to be zero. Consider a discrete source $X \in \mathcal{X} = \{1, 2, \dots, m\}$ with distribution p_1, p_2, \dots, p_m and a distortion measure $d(x, \hat{x}), x, \hat{x} \in \{1, 2, \dots, m\}$. Let $R(D)$ be the rate distortion function for this source and distortion measure. Let $d'(i, j) = d(i, j) - w_i$ be a new distortion measure and let $R'(D)$ be the corresponding rate distortion function. Show that $R'(D) = R(D + \bar{w})$, where $\bar{w} = \sum p_i w_i$, and use this to show that there is no essential loss of generality in assuming that $\min_{\hat{x}} d(i, \hat{x}) = 0$, i.e., for each $x \in \mathcal{X}$, there is one symbol \hat{x} which reproduces the source with zero distortion.

Solution: Properties of $R(D)$

By definition,

$$R'(D') = \min_{p(\hat{x}|x): \sum p(\hat{x}|x)p(x)d'(x, \hat{x}) \leq D'} I(X; \hat{X}). \quad (1)$$

For any conditional distribution $p(\hat{x}|x)$, we have

$$D' = \sum_{x, \hat{x}} p(x)p(\hat{x}|x)d'(x, \hat{x}) \quad (2)$$

$$= \sum_{x, \hat{x}} p(x)p(\hat{x}|x)(d(x, \hat{x}) - w_x) \quad (3)$$

$$= \sum_{x, \hat{x}} p(x)p(\hat{x}|x)d(x, \hat{x}) - \sum_x p(x)w_x \sum_{\hat{x}} p(\hat{x}|x) \quad (4)$$

$$= D - \sum_x p(x)w_x \quad (5)$$

$$= D - \bar{w}, \quad (6)$$

or $D = D' + \bar{w}$. Hence

$$R'(D') = \min_{p(\hat{x}|x): \sum p(\hat{x}|x)p(x)d'(x, \hat{x}) \leq D'} I(X; \hat{X}) \quad (7)$$

$$= \min_{p(\hat{x}|x): \sum p(\hat{x}|x)p(x)d(x, \hat{x}) \leq D' + \bar{w}} I(X; \hat{X}) \quad (8)$$

$$= R(D' + \bar{w}). \quad (9)$$

For any distortion matrix, we can set $w_i = \min_{\hat{x}} d(i, \hat{x})$, hence ensuring that $\min_{\hat{x}} d'(x, \hat{x}) = 0$ for every x . This produces only a shift in the rate distortion function and does not change the essential theory. Hence, there is no essential loss of generality in assuming that for each $x \in \mathcal{X}$, there is one symbol \hat{x} which reproduces it with zero distortion.

4. Erasure distortion

Consider $X \sim \text{Bernoulli}(\frac{1}{2})$, and let the distortion measure be given by the matrix

$$d(x, \hat{x}) = \begin{bmatrix} 0 & 1 & \infty \\ \infty & 1 & 0 \end{bmatrix},$$

where $x \in \{0, 1\}$ and $\hat{x} \in \{0, e, 1\}$. Calculate the rate distortion function for this source. Can you suggest a simple scheme to achieve any value of the rate distortion function for this source?

Solution: Erasure distortion

The infinite distortion constrains $p(0, 1) = p(1, 0) = 0$. Hence by symmetry (and convexity of mutual information) the joint distribution of (X, \hat{X}) is of the form shown in Figure 1.

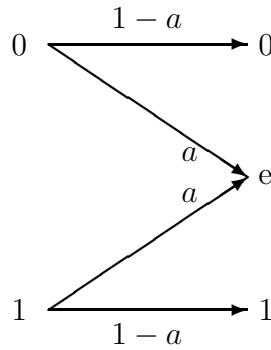


Figure 1: Joint distribution for erasure rate distortion of a binary source.

For this joint distribution, it is easy to calculate the distortion $D = a$ and that $I(X; \hat{X}) = H(X) - H(X|\hat{X}) = 1 - a$. Hence we have $R(D) = 1 - D$ for $0 \leq D \leq 1$. For $D > 1$, $R(D) = 0$.

We can achieve this rate distortion with a very simple encoding scheme. If D is rational, say k/n , then we send only the first $n-k$ of any block of n bits. We reproduce these bits exactly and reproduce the remaining bits as erasures. Hence we can send information at rate $1 - D$ and achieve a distortion D . If D is irrational, we can get arbitrarily close to D by using longer and longer block lengths.

5. **Rate distortion with two constraints**

Let X_i be iid $\sim p(x)$. We are given two distortion functions $d_1(x, \hat{x})$ and $d_2(x, \hat{x})$. We wish to describe X^n at rate R and reconstruct it with distortions $Ed_1(X^n, \hat{X}_1^n) \leq D_1$, and $Ed_2(X^n, \hat{X}_2^n) \leq D_2$, as shown here:

$$X^n \longrightarrow i(X^n) \longrightarrow (\hat{X}_1^n(i), \hat{X}_2^n(i))$$

$$D_1 = Ed_1(X_1^n, \hat{X}_1^n)$$

$$D_2 = Ed_2(X_1^n, \hat{X}_2^n).$$

Here $i(\cdot)$ takes on 2^{nR} values. What is the rate distortion function $R(D_1, D_2)$?

Solution: Rate distortion with two constraints

$$R(D_1, D_2) = \min_{\substack{p(\hat{x}_1, \hat{x}_2|x) : \\ Ed_1(\hat{X}_1, X) \leq D_1 \\ Ed_2(\hat{X}_2, X) \leq D_2}} I(X; \hat{X}_1, \hat{X}_2) \tag{10}$$

This can be easily extended from the single distortion problem: Define a vector distortion function as

$$\mathbf{d}(x, \hat{\mathbf{x}}) = (d_1(x, \hat{x}_1), d_2(x, \hat{x}_2))$$

with the vector reconstruction symbol $\hat{\mathbf{x}} = (\hat{x}_1, \hat{x}_2)$. Since $E\mathbf{d}(X, \hat{\mathbf{X}})$ is a linear function of $p(\hat{\mathbf{x}}) = p(x_1, x_2)$, from rate distortion theory for a single distortion,

$$R(D_1, D_2) = R(\mathbf{D})$$

$$= \min_{p(\hat{\mathbf{x}}|x): E\mathbf{d}(X, \hat{\mathbf{X}}) \leq \mathbf{D}} I(X; \hat{\mathbf{X}})$$

$$= \min_{\substack{p(\hat{x}_1, \hat{x}_2|x) : \\ Ed_1(\hat{X}_1, X) \leq D_1 \\ Ed_2(\hat{X}_2, X) \leq D_2}} I(X; \hat{X}_1, \hat{X}_2).$$

For the purist, the achievability can be proved from the usual arguments of average over random codebooks and joint typicality encoding. As for the converse, first note that $R(D_1, D_2)$ defined in (10) is convex and elementwise non-increasing in (D_1, D_2) . Now the usual converse for the single distortion case can be easily applied line by line.

Remark: Note that $\max(R_1(D_1), R_2(D_2)) \leq R(D_1, D_2) \leq R_1(D_1) + R_2(D_2)$, which can be shown either algebraically from (10) or information theoretically as follows: We have the lower bound since the rate $R(D_1, D_2)$ should be large enough to describe X (optimally) under either distortion constraint. On the other hand, having both (optimal) descriptions for each distortion constraint suffices to satisfy both distortion constraints at the same time, hence the upper bound. Note that both bounds can be

made tight. Consider $d_1 = d_2$ for the lower bound. For the upper bound, consider a vector source $\mathbf{X} = (X_1, X_2)$ with X_1, X_2 independent and distortions $d_1(\mathbf{x}; \hat{x}_1) = d_1(x_1, \hat{x}_1)$ and $d_2(\mathbf{x}; \hat{x}_2) = d_2(x_2, \hat{x}_2)$.

Finally note that the optimization over the set of distributions of the form $p(\hat{x}_1, \hat{x}_2|x)$ is, in general, required, which is larger than the set of conditionally independent distributions $p(\hat{x}_1|x)p(\hat{x}_2|x)$. As a simple example, consider a Gaussian source with the same squared error distortion for both d_1 and d_2 and the same distortion bounds, i.e., $D_1 = D_2$. In this case the minimum rate is achieved when $\hat{X}_1 = \hat{X}_2$, which immediately necessitates the conditional dependence.

6. Adding a column to the distortion matrix.

Let $R(D)$ be the rate distortion function for an i.i.d. process with probability mass function $p(x)$ and distortion function $d(x, \hat{x}), x \in \mathcal{X}, \hat{x} \in \hat{\mathcal{X}}$. Now suppose that we add a new reproduction symbol \hat{x}_0 to $\hat{\mathcal{X}}$ with associated distortion $d(x, \hat{x}_0), x \in \mathcal{X}$. Does this increase or decrease $R(D)$ and why?

Solution: Adding a column to the distortion matrix

Let the new rate distortion function be denoted as $\tilde{R}(D)$, and note that we can still achieve $R(D)$ by restricting the support of $p(x, \hat{x})$, i.e., by simply ignoring the new symbol. Thus, $\tilde{R}(D) \leq R(D)$.

Finally note the duality to the problem in which we added a row to the channel transition matrix to have no smaller capacity (EE376A).

7. Bounds on the rate distortion function for squared error distortion.

For the case of a continuous random variable X with mean zero and variance σ^2 and squared error distortion, show that

$$h(X) - \frac{1}{2} \log(2\pi e)D \leq R(D) \leq \frac{1}{2} \log \frac{\sigma^2}{D}.$$

What does this imply about how easy or hard Gaussian random variables are to describe? Comment on *both* inequalities.

Solution:

Bounds on the rate distortion function for squared error distortion.

We assume that X has zero mean and variance σ^2 . To prove the lower bound, we use the same techniques as used for the Gaussian rate distortion function. Let (X, \hat{X}) be

random variables such that $E(X - \hat{X})^2 \leq D$. Then

$$I(X; \hat{X}) = h(X) - h(X|\hat{X}) \quad (11)$$

$$= h(X) - h(X - \hat{X}|\hat{X}) \quad (12)$$

$$\geq h(X) - h(X - \hat{X}) \quad (13)$$

$$\geq h(X) - h(\mathcal{N}(0, E(X - \hat{X})^2)) \quad (14)$$

$$= h(X) - \frac{1}{2} \log(2\pi e) E(X - \hat{X})^2 \quad (15)$$

$$\geq h(X) - \frac{1}{2} \log(2\pi e) D. \quad (16)$$

To prove the upper bound, we construct a variable \hat{X} that achieves this bound by adding Gaussian noise to X and scaling appropriately to form the best linear estimate of X given the sum. Let Z be $N(0, N)$, independent of X . Define

$$\hat{X} = \frac{\sigma^2}{\sigma^2 + N}(X + Z).$$

This produces a distortion $D = \frac{\sigma^2 N}{\sigma^2 + N}$. Thus, for this distortion we have,

$$\begin{aligned} R(D) &= I(X; \hat{X}) \\ &= h(\hat{X}) - h(\hat{X}|X) \\ &= h(\hat{X}) - \frac{1}{2} \log \left(2\pi e \left(\frac{\sigma^2}{\sigma^2 + N} \right)^2 N \right) \\ &\leq \frac{1}{2} \log \left(2\pi e \left(\frac{\sigma^2}{\sigma^2 + N} \right)^2 (\sigma^2 + N) \right) - \frac{1}{2} \log \left(2\pi e \left(\frac{\sigma^2}{\sigma^2 + N} \right)^2 N \right) \\ &= \frac{1}{2} \log \left(\frac{\sigma^2 + N}{N} \right) \\ &= \frac{1}{2} \log \left(\frac{\sigma^2}{D} \right) \end{aligned}$$

For a Gaussian source X , both bounds are tight. So, in comparison with other sources of the same variance, Gaussian sources are the hardest to describe (with respect to squared-error distortion). However, in comparison with other sources of the same differential entropy, Gaussian sources are the easiest to describe.