EE/Ma 126b Information Theory - Homework Set #4

Ling Li, ling@cs.caltech.edu

January 31, 2001

4.1 *One bit quantization of a single Gaussian random variable.*[*] Let the boundary be x. We use one symbol to represent all $t \le x$ and another symbol for $t > x$. Since squared error measure is used, the two conditional expectations should be the reproduction points. That is, $x_0 = \int_{-\infty}^{x} \frac{f(t)}{1/2+A(x)} t dt$ for $t \le x$ and $x_1 = \int_{x}^{\infty} \frac{f(t)}{1/2-A(x)} t dt$ for $t > x$, where $f(t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{t^2/2\sigma^2}$ is the probability density function and $A(x) = \int_0^x f(t)dt$. Thus the distortion D is the weighted sum of two conditional variances:

$$
\begin{aligned}
D(x) &= (1/2 + A(x)) \left[\int_{-\infty}^{x} \frac{f(t)}{1/2+A(x)} t^2 dt - \left(\int_{-\infty}^{x} \frac{f(t)}{1/2+A(x)} t dt \right)^2 \right] \\
&\quad + (1/2 - A(x)) \left[\int_{x}^{\infty} \frac{f(t)}{1/2-A(x)} t^2 dt - \left(\int_{x}^{\infty} \frac{f(t)}{1/2-A(x)} t dt \right)^2 \right] \\
&= \int_{-\infty}^{\infty} f(x) t^2 dt - \frac{\sigma^4 f^2(x)}{1/2+A(x)} - \frac{\sigma^4 f^2(x)}{1/2-A(x)} \\
&= \sigma^2 - \frac{4\sigma^4 f^2(x)}{1-4A^2(x)} = \sigma^2 - \frac{2\sigma^2 e^{-x^2/\sigma^2}}{\pi(1-4A^2(x))}. \qquad (1)
\end{aligned}
$$

Now we want to prove $e^{-x^2/\sigma^2} \le 1 - 4A^2(x)$. Without loss of generality, assume $x \ge 0$.

$$
\begin{aligned}
4A^2(x) &= \left(\int_{-x}^{x} f(t)dt \right)^2 = \int_{-x}^{x} \int_{-x}^{x} f(u) f(v) du dv \\
&\le \frac{1}{2\pi\sigma^2} \int_{-\sqrt{2}x}^{\sqrt{2}x} \left(\int_{-\sqrt{2x^2-u^2}}^{\sqrt{2x^2-u^2}} e^{-\frac{u^2+v^2}{2\sigma^2}} dv \right) du \qquad (2) \\
&= \frac{1}{2\pi\sigma^2} \int_0^{2\pi} d\theta \int_0^{\sqrt{2}x} e^{-\frac{r^2}{2\sigma^2}} r dr = 1 - e^{-x^2/\sigma^2}.
\end{aligned}
$$

The reason for (2) is that $f > 0$ and the square of $[-x, x] \times [-x, x]$ is inside the circle with center $(0, 0)$ and radius $\sqrt{2}x$. So $e^{-x^2/\sigma^2} \le 1 - 4A^2(x)$ with equality iff $x = 0$. Then from (1), $D(x) \ge \frac{\pi-2}{\pi}\sigma^2$ with equality iff $x = 0$. So, the minimum distortion is $\boxed{\frac{\pi-2}{\pi}\sigma^2}$, with $x = 0$ and $x_{0,1} = \pm 2 \int_{-\infty}^{0} f(t) t dt = \boxed{\pm\sqrt{\frac{2}{\pi}}\sigma}$.

[*]Prof. Malik Magdon-Ismail gave me the idea of first setting the boundary, instead of setting the two reproduction points. The last way wasted me more than 12 hours.

4.2 *Rate distortion for uniform source with hamming distortion.* The distortion with the Hamming measure is $\bar{D} = Ed(X, \hat{X}) = \Pr\{d(X, \hat{X}) = 1\}$. We have

$$\begin{aligned}
I(X;\hat{X}) &= H(X) - H(X|\hat{X}) && \text{(definition of } I(X;\hat{X})\text{)} \\
&= H(X) - H(X - \hat{X}|\hat{X}) && \text{(previous homework)} \\
&= H(X) - H(X - \hat{X}, d(X,\hat{X})|\hat{X}) && (d(X,\hat{X}) \text{ is a function of } X - \hat{X}) \\
&= H(X) - H(d(X,\hat{X})|\hat{X}) - H(X - \hat{X}|d(X,\hat{X}), \hat{X}) && \text{(chain rule of entropy)} \\
&\geq H(X) - H(d(X,\hat{X})) - H(X - \hat{X}|d(X,\hat{X}), \hat{X}) && \text{(conditioning reduces entropy)} \\
&= H(X) - H(\bar{D}) - H(X - \hat{X}|\hat{X}, d(X,\hat{X}) = 1)\bar{D} && (\bar{D} = \Pr\{d(X,\hat{X}) = 1\} \text{ and} \\
& && \quad d(X,\hat{X}) = 0 \Rightarrow X - \hat{X} = 0) \\
&\geq \log m - H(\bar{D}) - \bar{D}\log(m-1). && (p(X) \text{ is uniform and for given } \hat{X} \\
& && \quad \left|\{X - \hat{X} : X \neq \hat{X}\}\right| = m - 1)
\end{aligned}$$

Notice that

$$\frac{d[\log m - H(\bar{D}) - \bar{D}\log(m-1)]}{d\bar{D}} = \log\frac{\bar{D}}{1-\bar{D}} - \log(m-1),$$

which is less than 0 when $0 < \bar{D} < 1 - \frac{1}{m}$. Thus when $D \leq 1 - \frac{1}{m}$, we have

$$R(D) = \min_{\bar{D} \leq D} I(X;\hat{X}) \geq \log m - H(D) - D\log(m-1).$$

We can design distributions $p(X|\hat{X})$ and $p(\hat{X})$ to achieve the minimum $I(X;\hat{X})$. Let $p(\hat{X})$ be uniform distribution. For $0 \leq D \leq 1 - \frac{1}{m}$, set

$$p(X|\hat{X}) = \begin{cases} 1 - D, & X = \hat{X}; \\ D/(m-1), & X \neq \hat{X}. \end{cases}$$

Thus

$$\begin{aligned}
p(X = x) &= \sum_{\hat{x}} p(X = x|\hat{X} = \hat{x})p(\hat{X} = \hat{x}) \\
&= \frac{1}{m}\left(p(X = x|\hat{X} = x) + \sum_{\hat{x} \neq x} p(X = x|\hat{X} = \hat{x})\right) \\
&= \frac{1}{m}\left(1 - D + (m-1) \times \frac{D}{m-1}\right) = \frac{1}{m},
\end{aligned}$$

and the distortion is

$$\Pr\{X \neq \hat{X}\} = 1 - \Pr\{X = \hat{X}\} = 1 - \sum_{\hat{x}} p(X = \hat{x}|\hat{X} = \hat{x})p(\hat{X} = \hat{x}) = D.$$

So such distribution meets the requirements on the distribution of X and the distortion. And, now

$$\begin{aligned}
I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\
&= H(X) - H(1 - D, \frac{D}{m-1}, \ldots, \frac{D}{m-1}) \\
&= H(X) + (1 - D)\log(1 - D) + (m-1) \times \frac{D}{m-1}\log\frac{D}{m-1} \\
&= \log m - H(D) - D\log(m-1).
\end{aligned}$$

Thus we know $\boxed{R(D) = \log m - H(D) - D\log(m-1)}$ for $0 \le D \le 1 - \frac{1}{m}$. When $D > 1 - \frac{1}{m}$, we can send nothing and simply choose $\hat{X}$ at random. Thus the distortion is $\Pr\{X \ne \hat{X}\} = \frac{m-1}{m} < D$. So obviously $\boxed{R(D) = 0}$ when $D > 1 - \frac{1}{m}$.

4.3 *Erasure distortion.* Let $\{0, \mathcal{E}, 1\}$ denote the set $\hat{\mathcal{X}}$, where '$\mathcal{E}$' stands for erasure. Since $d(0,1) = d(1,0) = \infty$, we must have $p(0,1) = p(1,0) = 0$ for a finite distortion. Thus

$$D = p(0,\mathcal{E}) + p(1,\mathcal{E}) = p_{\hat{X}}(\mathcal{E}),$$

and

$$\begin{aligned} I(X;\hat{X}) &= H(X) - H(X|\hat{X}) \\ &= 1 - H(X|\hat{X} = \mathcal{E})p_{\hat{X}}(\mathcal{E}) \\ &\ge 1 - D. \end{aligned} \tag{3}$$

When $p(X|\hat{X} = \mathcal{E}) = \frac{1}{2}$ and $D \le 1$, we can set $p_{\hat{X}}(0) = p_{\hat{X}}(1) = \frac{1-D}{2}$ and $p_{\hat{X}}(\mathcal{E}) = D$. Then $p_X(x) = p_{\hat{X}}(x) + \frac{1}{2}p_{\hat{X}}(\mathcal{E}) = \frac{1}{2}$, meeting that $X \sim \text{Bernoulli}(\frac{1}{2})$. And now the equality of (3) holds. Thus $\boxed{R(D) = 1 - D}$ when $0 \le D \le 1$. When $D > 1$, obviously $\boxed{R(D) = 0}$.

A simple strategy to achieve $R(D)$ is to erase X at random with probability $1 - R(D)$. Since X is uniformly distributed, we have with such strategy, $p(0,\mathcal{E}) = p(1,\mathcal{E}) = \frac{1}{2}(1 - R(D))$ and $p(0,1) = p(1,0) = 0$. Thus from the above discussion, the rate is $1 - (1 - R(D)) = R(D)$.

4.4 *Bounds on the rate distortion function for squared error distortion.* With D as the upper bound of the distortion, we have

$$\sigma^2(X - \hat{X}) = E(X - \hat{X})^2 - [E(X - \hat{X})]^2 = E(X - \hat{X})^2 \le D.$$

Thus $h(X - \hat{X}) \le \frac{1}{2}\log 2\pi e\sigma^2(X - \hat{X}) \le \frac{1}{2}\log(2\pi e D)$. So

$$\begin{aligned} I(X;\hat{X}) &= h(X) - h(X|\hat{X}) = h(X) - h(X - \hat{X}|\hat{X}) \\ &\ge h(X) - h(X - \hat{X}) \\ &\ge h(X) - \frac{1}{2}\log(2\pi e D), \end{aligned}$$

and

$$\boxed{R(D) = \min_{E(X-\hat{X})^2 \le D} I(X;\hat{X}) \ge h(X) - \frac{1}{2}\log(2\pi e D)}.$$

For $Z \sim \mathcal{N}(0, \frac{D\sigma^2}{\sigma^2 - D})$ independent with X and $\hat{X} = \frac{\sigma^2 - D}{\sigma^2}(X + Z)$, we have the distortion is

$$\begin{aligned} E(X - \hat{X})^2 &= E\left(\frac{D}{\sigma^2}X - \frac{\sigma^2 - D}{\sigma^2}Z\right)^2 \\ &= E\left(\frac{D}{\sigma^2}X\right)^2 + E\left(\frac{\sigma^2 - D}{\sigma^2}Z\right)^2 \\ &= \left(\frac{D}{\sigma^2}\right)^2\sigma^2 + \left(\frac{\sigma^2 - D}{\sigma^2}\right)^2\frac{D\sigma^2}{\sigma^2 - D} \\ &= D. \end{aligned}$$

And it is surprising to find out that for a constant a,[†]

$$h(X|aY) = h(X|Y), \quad I(X;aY) = I(X;Y).$$

You can find the proof in the footnote. Thus the mutual information is

$$\begin{aligned}
I(X;\hat{X}) &= I(X;X+Z) \\
&= h(X+Z) - h(X+Z|X) \\
&= h(X+Z) - h(Z) \\
&\leq \frac{1}{2}\log 2\pi e(\sigma^2 + \frac{D\sigma^2}{\sigma^2 - D}) - \frac{1}{2}\log 2\pi e\frac{D\sigma^2}{\sigma^2 - D} \qquad (4) \\
&= \frac{1}{2}\log\frac{\sigma^2}{D}.
\end{aligned}$$

The reason for (4) is that

$$\sigma^2(X+Z) = \sigma^2(X) + \sigma^2(Z) = \sigma^2 + \frac{D\sigma^2}{\sigma^2 - D}.$$

For such $\hat{X}$ and $I(X;\hat{X})$, the distortion is bounded by D. So we get

$$\boxed{R(D) \leq \frac{1}{2}\log\frac{\sigma^2}{D}}.$$

Since for Gaussian random variable with variance σ^2, $R(D) = \frac{1}{2}\log\frac{\sigma^2}{D}$ achieves the maximum of $R(D)$, it is harder to describe the Gaussian random variable than other random variables with the same variance.

4.5 *Properties of optimal rate distortion code.* The conditions of equalities are listed at the right

---

[†]Let $Z = aY$. Then $f_Z(z) = \frac{1}{|a|}f_Y(\frac{z}{a})$, and $f_{X,Z}(x,z) = \frac{1}{|a|}f_{X,Y}(x,\frac{z}{a})$. Thus with variable replacing,

$$\begin{aligned}
h(X|aY) &= -\int_S f_{X,Z}(x,z)\log\frac{f_{X,Z}(x,z)}{f_Z(z)}dxdz \\
&= -\int_S \frac{1}{|a|}f_{X,Y}(x,\frac{z}{a})\log\frac{f_{X,Y}(x,\frac{z}{a})}{f_Y(\frac{z}{a})}dxdz \\
&= -\int_S f_{X,Y}(x,y)\log\frac{f_{X,Y}(x,y)}{f_Y(y)}dxdy \\
&= h(X|Y).
\end{aligned}$$

This can also be proved by

$$I(X;aY) = h(aY) - h(aY|X) = h(Y) + \log|a| - h(Y|X) - \log|a| = I(X;Y).$$

side:

$$\begin{aligned}
nR &\geq H(\hat{X}^n) && (\hat{X}^n \text{ is uniformly distributed}) \\
&\geq H(\hat{X}^n) - H(\hat{X}^n|X^n) && (\hat{X}^n \text{ is a deterministic function of } X^n) \\
&= H(X^n) - H(X^n|\hat{X}^n) \\
&= \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i|\hat{X}^n, X_1^{i-1}) \\
&\geq \sum_{i=1}^{n} H(X_i) - \sum_{i=1}^{n} H(X_i|\hat{X}_i) && (\text{independent encoding among } X_i) \\
&= \sum_{i=1}^{n} I(X_i; \hat{X}_i) \\
&\geq \sum_{i=1}^{n} R(Ed(X_i, \hat{X}_i)) && (\text{optimal for each single } X_i) \\
&\geq nR\left(\frac{1}{n}\sum_{i=1}^{n} Ed(X_i, \hat{X}_i)\right) && (\text{same distortion on each } X_i) \\
&= nR(D)
\end{aligned}$$