Information Theory with Applications, Math6397
Lecture Notes from October 23, 2014

taken by Tianxiao Jiang

Last time

If $\frac{1}{n}\ln m_n \leq \rho(D)+\tau < \rho(D)+4\epsilon$, then we could show the existence of a code such that

$$\mathbb{E}[\frac{1}{n}d_n(X;\phi_n(x))] \leq D+\epsilon$$

So

$$R(D+\epsilon) \leq \rho(D)+4\epsilon$$

For the lower bound, we had derived

$$\rho(\frac{1}{n}\mathbb{E}[d_n(X;\phi_n(X))]) \leq \frac{1}{n}\ln m_n \tag{1}$$

By the assumption $\lim_{n\to\infty}\sup\frac{1}{n}\ln m_n$, for all sufficient large n, there exists $\tau > 0$ such that

$$\frac{1}{n}\ln m_n < \rho(D)-\tau \tag{2}$$

Combined with (1), thus

$$\rho(\underbrace{\frac{1}{n}\mathbb{E}[d_n(X;\phi_n(X))]}_{D'}) < \rho(D)-\tau$$

Recall properties of ρ

- $\rho(D)=0$ for all $D \geq D_0$
- ρ is convex
- ρ is decreasing by definition

This implies that ρ is continuous and strictly decreasing. Because of this, $D' > D$, so there is $\epsilon > 0$ such that for all sufficiently large n

$$\frac{1}{n}\mathbb{E}[d_n(X;\phi_n(X))] > D+\epsilon$$

We have thus shown that achieving the rate $\rho(D)$ requires an expected distortion of at least $D+\epsilon$, meaning $\rho(D) \leq R(D+\epsilon)$.

5 Source and channels with continuous alphabets

5.1 Differential entropy

$Recall$: For source $X : \Omega \to \mathbb{A}$, with discrete alphabet, the entropy

$$H(X) = -\sum_{a \in \mathbb{A}} \mathbb{P}(x = a) \ln \mathbb{P}(x = a)$$

is the minimum average code length for lossless source coding

5.1.1 Question. What about the entropy of sources with continuous alphabet $X : \Omega \to \mathbb{R}$?

5.1.2 Example. $X : \Omega \to [0, 1)$, which induces uniform probability measure on Borel sets, characterized by $\mathbb{P}(a \leq X \leq b) = b - a$,for all $0 \leq a < b \leq 1$.
Approximate X by $Y_m = \frac{j}{m}$ if $\frac{j-1}{m} \leq X < \frac{j}{m}$, with $1 \leq j \leq m$
So $Y_m = f_m(X)$ where f_m is a step function approximation of the identity X.
$Sketch$:

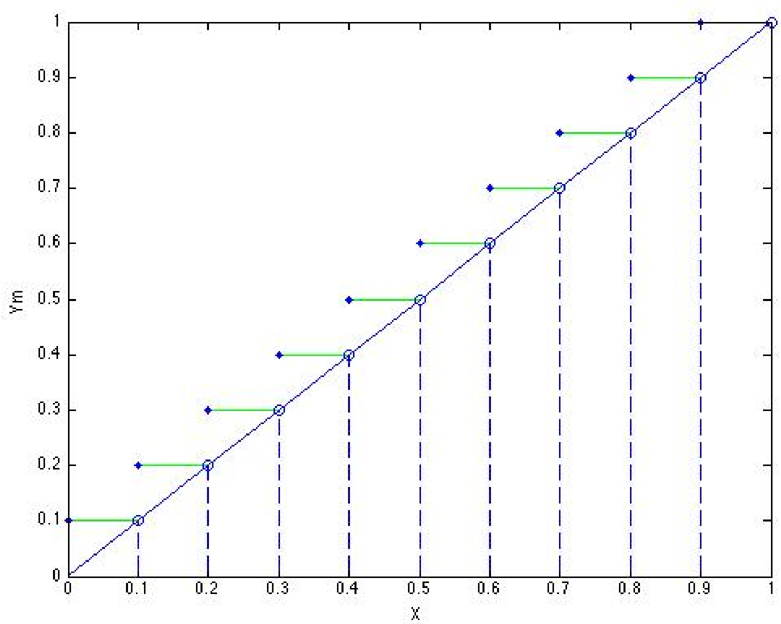

Figure 1: Maximum quantization error: $Err_{max} = \frac{1}{m}$

We have that the measure induced by Y_m is uniform on $\{\frac{1}{m}, \frac{2}{m}, ..., 1\}$, so

$$H(Y_m) = -\sum_{j=1}^{m} \frac{1}{m} \ln \frac{1}{m} = \ln m \xrightarrow{m \to \infty} \infty$$

So this means as $Y_m \xrightarrow{m \to \infty} X, H(Y_m) \to \infty$

5.1.3 Definition. The **differential entropy** of a random variable with values in $\mathbb{R}$, inducing a measure p which is absolute continuous with respect to the $Lebesgue$ measure, so $d(p(x)) = p(x)dx$ is:

$$h(X) = \int_{\mathbb{R}} p(x) \ln p(x) dx, \textit{ if } p(x) \ln p(x) \textit{ is integrable}$$

5.1.4 Example.

- $p(X) = 1, 0 \leq X \leq 1$, then

$$h(X) = 0$$

- $p(X) = 2X$, $0 \leq X \leq 1$ then

$$h(X) = -\int_{x=0}^{1} 2x \ln 2x dx = -\frac{1}{2} \ln 2 < 0$$

5.1.5 Question. Is there a relation between $h(X)$ and entropies of approximation $r.v's$ (Y_m) ?

In the second of our previous example: $Y_m = \frac{j}{m}$,if $\frac{j-1}{m} \leq X < \frac{j}{m}$

then $\mathbb{P}(Y_m = \frac{j}{m}) = \int_{X=\frac{j-1}{m}}^{\frac{j}{m}} 2xdx = \frac{2j-1}{m^2}$, for $1 \leq j \leq m//$ and the entropy of Y_m is

$$\begin{aligned}
H(Y_m) &= -\sum_{j=1}^{m} \frac{2j-1}{m^2} \ln \frac{2j-1}{m^2} \\
&= -\sum_{j=1}^{m} \frac{1}{m} (\frac{2j-1}{m} (\ln \frac{2j-1}{m} + \ln \frac{1}{m})) \\
&= \underbrace{-\sum_{j=1}^{m} \frac{1}{m} (\frac{2j-1}{m} \ln \frac{2j-1}{m})}_{\xrightarrow{m\to\infty} -\int_{x=0}^{1} 2x \ln 2x dx} + \ln m
\end{aligned}$$

Message: Apart from trivial divergence ($\ln m$), $H(Y_m)$ contains a part that converges to $h(X)$. This is meaningful when comparing entropies for Y_m, Y'_m belonging to two random variables X, X' with fixed m because the $\ln m$ term is then the same for both.

5.1.6 Example. Differential entropy of a normal (gaussian) *r.v.* X with

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

with expected value μ and variance σ^2,

$$\begin{aligned}
h(X) &= \int_{\mathbb{R}} p(x)(\frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}(x-\mu)^2)dx \\
&= \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}\underbrace{\mathbb{E}[(x-\mu)^2]}_{\sigma^2} \\
&= \frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2} \\
&= \frac{1}{2}\ln(2\pi e\sigma^2)
\end{aligned}$$

In fact, if X has an expected value and a finite second moment and variance is fixed at $\mathbb{E}[(X-\mu)^2] = \sigma^2$, then this is the largest differential entropy possible.

5.1.7 Theorem. *Given a r.v. X with density p and mean μ. Let q be the density of a gaussian source Y with $\mu = \mathbb{E}[X] = \mathbb{E}[Y]$ and $\mathbb{E}[(X-\mu)^2] = \mathbb{E}[(Y-\mu)^2] = \sigma^2 < \infty$. Then $h(Y) \geq h(X)$.*

Proof. Note since:

$$\begin{aligned}
\int_{\mathbb{R}} p(x)\ln q(x)dx &= \int_{\mathbb{R}} p(x)(\frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}(x-\mu)^2)dx \\
&= \int_{\mathbb{R}} q(x)(\frac{1}{2}\ln(2\pi\sigma^2) + \frac{1}{2\sigma^2}(x-\mu)^2)dx \\
&= \int_{\mathbb{R}} q(x)\ln q(x)dx
\end{aligned}$$

Because $\int_{\mathbb{R}} q(x)dx = \int_{\mathbb{R}} p(x)dx = 1$ and $\mathbb{E}[(X-\mu)^2] = \mathbb{E}[(Y-\mu)^2] = \sigma^2$, changing from $p(x)$ to $q(x)$ does not affect the result of integral. Next,

$$\begin{aligned}
h(Y) - h(X) &= -\int_{\mathbb{R}} \underbrace{q(x)}_{p(x)}\ln q(x)dx + \int_{\mathbb{R}} p(x)\ln p(x)dx \\
&= -\int_{\mathbb{R}} p(x)(\ln q(x) - \ln p(x))dx
\end{aligned}$$

With $-\ln\frac{q(x)}{p(x)} \geq 1 - \frac{q(x)}{p(x)}$

$$h(Y) - h(X) \geq \int_{\mathbb{R}} p(x)(1 - \frac{q(x)}{p(x)})dx = 1 - 1 = 0$$

$\square$

5.2 A closer look at the meaning of differential entropy

5.2.1 Lemma. *Let X be a r.v. with density p on $\mathbb{R}$ and $p \ln p$ be Riemann Integrable which means*

$$\inf_{\substack{f \in C(\mathbb{R}) \\ f \geq p}} \int f dx = \sup_{\substack{f \in C(\mathbb{R}) \\ f \leq p}} \int f dx$$

Then, rounding X with stepsize $\Delta = 2^{-n}, n \in \mathbb{N}$, yields an entropy for $X^\Delta = \lceil \frac{X}{\Delta} \rceil \Delta$ which satisfies $H(X^\Delta) - \ln(2^n) = H(X^\Delta) - n \ln 2 \xrightarrow{n \to \infty} h(X)$

Proof. Without loss of generality, assume p is continuous, bounded and compact support. Let $t_j = j\Delta, \Delta = 2^{-n}, j \in \mathbb{Z}$. Choose $x_j = [t_{j-1}, t_j]$ by the mean value theorem for integration *s.th.*

$$\int_{t_{j-1}}^{t_j} p(x) dx = p(x_j)(t_j - t_{j-1}) = p(x_j)(t_j - t_{j-1}) = \Delta p(x_j) \, .$$

Now define the Riemann sum

$$h^\Delta(X) \equiv \sum_{j=-\infty}^{\infty} \underbrace{\Delta p(x_j)}_{prob with \Delta} \underbrace{\ln p(x_j)}_{\textit{one-point}}$$

then the Riemann integral is obtained as the limit

$$h^\Delta(X) \xrightarrow{\Delta \to 0} h(X)$$

□

This means for $\epsilon > 0$ there is N *s.th.* for all $n > N$,

$$|h(X) - h^\Delta(X)| < \epsilon$$

Compare this with

$$H(X^\Delta) = -\sum_{j=-\infty}^{\infty} \mathbb{P}_j \ln \mathbb{P}_j \qquad = -\sum_{j=-\infty}^{\infty} (p(x_j)\Delta) \ln(p(x_j)\Delta)$$

therefore

$$H(X^\Delta) - h^\Delta(X) = -\sum_{j=-\infty}^{\infty} p(x_j) \Delta \ln \underbrace{\Delta}_{\frac{1}{2^n}} = n \ln 2$$

By the convergence of h^Δ,

$$H(X^\Delta) - n \ln 2 = h^\Delta(X) \xrightarrow{\Delta \to 0} h(X)$$