Information Theory with Applications, Math6397
Lecture Notes from October 30, 2014

taken by John Haas

Last Time

- Relationship between differential entropy and entropy of quantized random variables
- Other properties of differential entropy

5 Sources and Channels With Continuous Alphabets, continued

We begin with basic source coding results for continuous random variables. Because they are analogous to the those of the discrete case, we present them in an abridged format.

5.3.15 Theorem. *Let* $\{X_j\}_{j=1}^{\infty}$ *be a CMS with density* $p_X := p_{X_1}$ *on* $\mathbb{R}$ *such that* $p_X \ln p_X \in L^1(\mathbb{R})$*, then*

$$-\frac{1}{n} p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) \to h(X)$$

as $n \to \infty$ *almost surely.*

Proof. This follows by the Strong Law of Large Numbers. □

What follows is the generalization of typical sets to the continuous case. Because these sets are infinite, we replace the notion of size by the measure of the set, their volume (i.e. the Lebesgue measure).

5.3.16 Definition. For $\delta > 0$ and $n \in \mathbb{N}$, define

$$A_\delta^n := \{x \in \mathbb{R}^n : |\frac{1}{n} p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) + h(X_1)| < \delta\}.$$

Furthermore, define its volume as

$$vol(A_\delta^n) = \int_{\mathbb{R}^n} 1_{A_\delta^n}(x_1, \ldots, x_n) d_{x_1} \ldots d_{x_n}.$$

Next we state the continuous version of the A.E.P.

5.3.17 Theorem. *Let $\{X_j\}_{j=1}^{\infty}$ be a CMS with density $p_X := p_{X_1}$ such that $p_X \ln p_X \in L^1(\mathbb{R})$ and let $\delta > 0$, then for sufficiently large n, one has*

1. $\mathbb{P}((A_\delta^n)^c) < \delta$
2. $vol(A_\delta^n) \leq \exp(n(h(X) + \delta))$
3. $vol(A_\delta^n) \geq (1 - \delta) \exp(n(h(X) - \delta))$

Proof. The proof is analogous to that of the discrete version. For the volume bounds, the Hölder inequality for sums is simply replaced by the one for integrals. □

5.4 Relative (differential) entropy and mutual information

As with discrete case, the idea of mutual information is essential to channel coding and rate distortion theory.

5.4.18 Definition. The relative entropy between random variables X and Y with densities p_X and p_Y on $\mathbb{R}$ is defined as

$$D(X||Y) = \int_{\mathbb{R}} p_X(x) \ln \frac{p_X(x)}{p_Y(y)} dx.$$

If X and Y have joint density $p_{X,Y}$, then their mutual information is

$$I(X;Y) = \int_{\mathbb{R}} p_{X,Y}(x,y) \ln \frac{p_{X,Y}(x,y)}{p_X(x) p_Y(y)} dx dy.$$

5.4.19 Remark. In the definition above, if the densities p_X, p_Y, and the function $p_X \ln \frac{p_X}{p_Y}$ are all Riemann integrable (and so WLOG continuous), then

$$\begin{aligned} D(X^\Delta||Y^\Delta) &= -\sum_j p_j \ln \frac{p_j}{q_j} \\ &= -\sum_j (p_X(x_j)\Delta) \ln \frac{p_X(x_j)\Delta}{p_Y(x_j)\Delta} \\ &\longrightarrow -\int p_X(x) \ln \frac{p_X(x)}{p_Y(y)} dx = D(X||Y) \end{aligned}$$

as $\Delta \to 0$. Similarly, if $p_{X,Y}$ and $p_{X,Y} \ln \frac{p_{X,Y}}{p_X p_Y}$ are Riemann integrable, then

$$I(X^\Delta; Y^\Delta) \to I(X;Y)$$

as $\Delta \to 0$.

These remarks allow us to lift the following corollary by a limiting procedure from the discrete setting to the continuous one.

5.4.20 Corollary. *Given random variables X and Y with joint density $p_{X,Y}$ and marginal densities p_X and p_Y, then*

1. $D(X||Y) \geq 0$ *and equality holds if and only if* $p_X = p_Y$ *almost everywhere.*
2. $I(X;Y) \geq 0$ *and equality holds if and only if* X *and* Y *are independent.*

5.5 Lossy Compression For Continuous Sources

Instead of following the same order as we did with discrete random variables, we now cover distortion theory instead of channel coding. After all, rate distortion is a form of source coding, i.e. lossy compression.

5.5.21 Theorem. *Let* $\{X_j\}_{j=1}^{\infty}$ *be a CMS with mean zero, variance* σ^2*, and* $d(a, \hat{a}) = (a - \hat{a})^2$*, then*

$$R(D) = \begin{cases} \frac{1}{2} \ln \frac{\sigma^2}{D}, & 0 \leq D \leq \sigma^2 \\ 0, & D > \sigma^2 \end{cases}$$

and equality holds if and only if $X := X_1$ *is Gaussian.*

Proof. By the same proof as before (see 10-16-14), generalized to d being unbounded,

$$R(D) = \min_{\{\hat{X}: \mathbb{E}[d(X,\hat{X})] \leq D\}} I(X; \hat{X}),$$

so for any $\hat{X}$, we have $R(D) \leq I(X; \hat{X})$. Let $0 \leq D \leq \sigma^2$, choose a Gaussian random variable Y with mean zero and variance $\tilde{\sigma}^2 = D - \frac{D^2}{\sigma^2}$ which is independent of X, and set $\hat{X} = (1 - \frac{D}{\sigma^2})X + Y$. Now we verify that $\hat{X}$ is admissible whenever $D \leq \sigma^2$.

$$\begin{aligned} \mathbb{E}[(X - \hat{X})^2] &= \mathbb{E}[(\frac{D}{\sigma^2} X + Y)^2] \\ &= \frac{D^2}{\sigma^4} \mathbb{E}[X^2] + \mathbb{E}[Y^2] \\ &= \frac{D^2}{\sigma^2} + D - \frac{D^2}{\sigma^2} \\ &= D \end{aligned}$$

Thus, $\hat{X}$ is included in the minimization resulting in $R(D)$. Next,

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

Computing the variance of $\hat{X}$ gives

$$\begin{aligned}
\mathbb{E}[X^2] &= \mathbb{E}[(1-\frac{D}{\sigma^2})X+Y)^2] \\
&= \mathbb{E}[(1-\frac{D}{\sigma^2})^2X^2]+\mathbb{E}[Y^2] \\
&= \sigma^2(1-2\frac{D}{\sigma^2}+\frac{D^2}{\sigma^4})+D-\frac{D^2}{\sigma^2} \\
&= \sigma^2-D
\end{aligned}$$

Therefore, $h(\hat{X}) = \frac{!}{2}\ln(2\pi e(\sigma^2-D))$ (see example in 10-22-14 notes), so substituting this into the previous inequality gives

$$R(D) \le \frac{1}{2}\ln(2\pi e(\sigma^2-D)) - \frac{1}{2}\ln(2\pi e(D-\frac{D^2}{\sigma^2})) = \frac{1}{2}\ln\frac{\sigma^2}{D}.$$

On the other hand, if $D > \sigma^2$, then taking $\hat{X} = 0$ gives

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

and

$$I(X;\hat{X}) = 0$$

by independence. Hence, $R(D) = 0$. To see that equality holds if and only if X is Gaussian, note that

$$h(\hat{X}) \le \frac{1}{2}\ln(2\pi e(\sigma^2-D))$$

is saturated if and only if $\hat{X}$ is Gaussian. Rearranging the defining equation of $\hat{X}$ yields

$$X = \frac{1}{1-\frac{D}{\sigma^2}}(\hat{X}-Y),$$

which is Gaussian as a linear combination of Gaussians. Thus, if $\hat{X}$ is a Gaussian which maximizes the differential entropy, then X is Gaussian by the independence of X and Y.

□

5.6 Channel Coding For Continuous Channels and Discrete Alphabets

Usually, a channel input is subject to some constraint.

5.6.22 Definition. If $\mathbb{E}[X^2] \le s$ for some s, then we say that it has an average power bounded by s.

5.6.23 Definition. The capacity of a continuous channel is given by

$$C(s) = \max_{\{\mathbb{P}_X : \mathbb{E}[X^2] \le s\}} I(X;Y).$$