Information Theory with Applications, Math6397
Lecture Notes from November 04, 2014

taken by Ilknur Telkes

Last Time

- Shannon-McMillan for continuous sources
- Relative entropy and mutual information for continuous sources
- Lossy compression for continuous sources
- Gaussian channel as worst case

5.6.24 Definition. A continuous memoryless channel (CMC) is a random map $\gamma : \mathbb{R} \times \Omega \to \mathbb{R}$ which is characterized by its transition kernel $\xi\mathbb{W}(\cdot\,|x)_{x\in\mathbb{R}}$ i.e. a family of probability densities on $\mathbb{R}$ for each $x \in \mathbb{R}$.

Standard strategy of a communication system:

$$\underbrace{X \in \mathbb{R}}_{\substack{\text{X:sent signal coming}\\ \text{from analog device}}} \longrightarrow \underbrace{\boxed{\mathrm{Q}}}_{\text{quantizer}} \longrightarrow \underbrace{\hat{X} \in \hat{\mathbb{A}}}_{\substack{\text{X : bits and bytes}\\ \text{A : reproduction alphabet}}} \longrightarrow \underbrace{\boxed{\phi_n}}_{\substack{\text{preprocessing before}\\ \text{transmitting}\\ \text{through antenna}}} \longrightarrow \underbrace{C_{\in\mathbb{R}^n}}_{\substack{\text{signal:broadcast out}\\ \text{through antenna}}} \longrightarrow \underbrace{\boxed{\gamma}}_{\text{cont. channel}}$$

$$\text{cont. channel} \downarrow Y \in \mathbb{R}^n \downarrow \boxed{\Psi_n} \downarrow Z \in \mathbb{R}$$

Important points about the system:

- We are interested in the size of alphabet for the input signal X to $\hat{X}$.
- At preprocessing step, sequence is important.
- S is the average power constraint.
- Usual capacity defined as max. mutual info. business includes codewords to be transmitted.

We recall the definition of capacity for γ,

$$C(S) = \max_{\mathbb{P}_x, \mathbb{E}[x^2] \le S} I(X;Y)$$

with $Y = \gamma(X)$

5.6.25 Remark. By definition, $C(S)$ is increasing using convexity argument. Similar as before it can be shown to be strictly increase.

5.6.26 Theorem. *For any $\varepsilon \in$ (0,1), there is τ, $0 < \tau < 2\varepsilon$ and a code sequence $\{ \mathscr{C}_n \}$ of sizes $|\mathscr{C}_n| = m_n$ such for all sufficiently large n,*

$$\frac{1}{n}\ln m_n > C(S) - \tau$$

and for each $c \in \mathscr{C}_n$,

$$\frac{1}{n}\sum_{j=1}^{n} c_j^2 \leq S \text{ and } P_e < \varepsilon$$

Proof. If C(S) = 0, choose m_n=1, so P_e=0. So, assume C(S) > 0

Step-1: Choose $0 < \tau < min\{2\varepsilon, C(S)\}$. Pick $\xi > 0$, $\xi <$ S such that
2(C(S)- C(S-ξ)) $< \tau$ (This exits because C is strictly increasing). Thus,

$$2C(S-\xi) + \frac{\tau}{2} > 2C(S)$$
$$C(S-\xi) - \frac{\tau}{2} > C(S) - \tau > 0$$

Pick m_n for sufficiently large n such that

$$C(S-\xi) - \frac{\tau}{2} > \frac{1}{n}\ln m_n > C(S) - \tau$$

Let $\delta = \frac{\tau}{8}$ and $\mathbb{P}_x$ the measure for the continuous source that achieves $C(S-\xi)$, i.e. $\mathbb{E}[X^2] \leq S-\xi$ and $I(X;Y) = C(S-\xi)$.

Step-2: Randomly draw m_n codewords according to $\mathbb{P}_x^{\otimes n}$. By the strong law of large numbers, a sequence of chosen codewords satisfies

$$\frac{1}{n}\sum_{j=1}^{n} c_j^2 \xrightarrow{n\to\infty} \mathbb{E}[X^2] \leq \text{S} - \xi \text{ a.s.}$$

If any $c \in \mathscr{C}_n$ violates $\frac{1}{n}\sum_{j=1}^{n} c_j^2 \leq S$ then replace c by 0. Next, define the encoding map $\phi_n : \{1,2,...,k\} \to \mathscr{C}_n$. When receiving a sequence y, let the decoder Ψ_n be given by

$$\Psi_n(y) = \begin{cases} k, & (\phi_n(k), y) \in F_\delta^n \text{and there is no other} k' \text{with} (\phi_n(k'), y) \in F_\delta^n \\ 1, & \text{else} \end{cases}$$

where

$$
\begin{aligned}
F_\delta^n = \{(x,y)\in\mathbb{R}^n\times\mathbb{R}^n : \\
&|\frac{1}{n}p_{(X,Y)^{\otimes n}}(x,y)+h(X_1,Y_1)|<\delta,\\
&|\frac{1}{n}p_{X^{\otimes n}}(x)+h(X_1)|<\delta,\\
&|\frac{1}{n}p_{Y^{\otimes n}}(y)+h(Y_1)|<\delta\}
\end{aligned}
$$

Step-3: Let λ_K be the error probabilities for k-th transmitted codeword. Define,

$$G=\{x\in\mathbb{R}^n:\frac{1}{n}\sum_{j=1}^{n}x_j^2>S\}$$

Then as before, averaging w.r.t. choice of k-th codeword,

$$\mathbb{E}[\lambda_K]\le\mathbb{P}_{X^{\otimes n}}(G)+P_{(X,Y)^{\otimes n}}((F_\delta^n)^c)+\sum_{k'=1,k'\neq k}^{m_n}\int_{\mathbb{R}^n}\int_{\{y\in\mathbb{R}^n:(ck',y)\in F_\delta^n}P_{(X,Y)^{\otimes n}}(c,y)dydc$$

and,

$$
\begin{aligned}
\mathbb{E}[P_e]\overset{\text{Holder ineq.}}{\le}&\mathbb{P}_{X^{\otimes n}}(G)+P_{(X,Y)^{\otimes n}}((F_\delta^n)^c)\\
&+(m_n-1)\exp(-n(h(X)-\delta))\exp(-n(h(Y)-\delta))\exp(-n(h(X,Y)+\delta))\\
&\mathbb{P}_{X^{\otimes n}}(G)+P_{(X,Y)^{\otimes n}}((F_\delta^n)^c)+\exp(-n\delta),
\end{aligned}
$$

similar as in the case of channel coding proof.
By choosing n sufficiently large, we can bound the error probability smaller than,

$$\mathbb{E}[P_e]\le\delta+\delta+\delta=3\delta=3\frac{\tau}{8}<3\frac{\varepsilon}{4}<\varepsilon$$

Since the expected value for P_e is smaller than ε, there is at least one choice for $\mathscr{C}_n$ which has $P_e<\varepsilon$ a.s. (w.r.t. channel)

□

5.6.27 Example. Capacity for additive white Gaussian noise (AWGN), let $\{X_j\}_{j=1}^{\infty}$ be the channel input, then the memoryless additive channel produces outputs

$$Y_j=X_j+N_j\text{ with}\{X_j,N_j\}\text{ independent and}\{N_j\}_{j=1}^{\infty}\text{ i.i.d.}$$

We call a memoryless additive channel and AWGN channel if N_1 is normal (Gaussian).

5.6.28 Theorem. *Given an AWGN channel with mean-zero noise $\{N_j\}_{j=1}^{\infty}$ with variance $\sigma^2=\mathbb{E}[N_1^2]>0$, subject to the average power and $\mathbb{E}[X^2]\le S$, then*

$$C(S)=\frac{1}{2}\ln(1+\frac{S}{\sigma^2})$$

Proof. We compute

$$
\begin{aligned}
C(S) &= \max_{\mathbb{P}_x:\mathbb{E}[x^2]\leq S} I(X;Y) \\
&= \max(h(Y) - h(Y|X)) \\
&= \max(h(Y) - h(X+N|X)) \\
&= \max(h(Y) - h(N|X)) \\
&= \max(h(Y) - h(N)) \text{ (coming from indep.)} \\
&= \max_{\mathbb{P}_x:\mathbb{E}[x^2]\leq S} h(Y) - h(N) \text{ (Here, we are adding extra noise with } \sigma^2\text{)}
\end{aligned}
$$

So,

$$
C(S) = \frac{1}{2}\ln\left(2\pi e(S+\sigma^2)\right) - \frac{1}{2}\ln\left(2\pi e\sigma^2\right) = \frac{1}{2}\ln\left(1+\frac{S}{\sigma^2}\right)
$$

□

We note that the essential quantity is simply the ratio of the signal power to that of the noise.