# Information Theory with Applications, Math6397
# Lecture Notes from October 2, 2014

taken by Nathaniel Hammen

Last Time:

- Huffman's Code, also for k-ary trees
- Discrete Memoryless channels
- Channel coding with fixed length transition codes
- Jointly typical sequences

# 3 Coding for discrete channels (continued)

## 3.1 The discrete memoryless channel (continued)

**3.1.9 Theorem.** *Given a discrete memoryless source* $\{X_j\}_{j=1}^{\infty}$ *and a discrete memoryless channel* $\gamma : \mathbb{A} \times \Omega \to \mathbb{B}$*, denoting* $Y_j = \gamma(X_j)$ *for all* $j \in \mathbb{N}$*, then if* $\ln(\mathbb{P}_{X_1}(X_1))$*,* $\ln(\mathbb{P}_{Y_1}(Y_1))$ *and* $\ln(\mathbb{P}_{X_1,Y_1}(X_1, Y_1))$ *are integrable, we have*

$$-\frac{1}{n}\ln(\mathbb{P}_{X_1,\dots,X_n}(X_1,\dots,X_n)) \to H(X_1)$$

$$-\frac{1}{n}\ln(\mathbb{P}_{Y_1,\dots,Y_n}(Y_1,\dots,Y_n)) \to H(Y_1)$$

*and*

$$-\frac{1}{n}\ln(\mathbb{P}_{X_1,\dots,X_n,Y_1,\dots,Y_n}(X_1,\dots,X_n,Y_1,\dots,Y_n)) \to H(X_1,Y_1)$$

*Proof.* By independence,

$$\ln(\mathbb{P}_{X_1,\dots,X_n}(X_1,\dots,X_n)) = \ln\left(\prod_{j=1}^{n}\mathbb{P}_{X_j}(X_j)\right) = \sum_{j=1}^{n}\ln\left(\mathbb{P}_{X_j}(X_j)\right)$$

$$\ln(\mathbb{P}_{Y_1,\dots,Y_n}(Y_1,\dots,Y_n)) = \ln\left(\prod_{j=1}^{n}\mathbb{P}_{Y_j}(Y_j)\right) = \sum_{j=1}^{n}\ln\left(\mathbb{P}_{Y_j}(Y_j)\right)$$

$$\ln(\mathbb{P}_{X_1,\dots,X_n,Y_1,\dots,Y_n}(X_1,\dots,X_n,Y_1,\dots,Y_n)) = \ln\left(\prod_{j=1}^{n}\mathbb{P}_{X_j,Y_j}(X_j,Y_j)\right)$$
$$= \sum_{j=1}^{n}\ln\left(\mathbb{P}_{X_j,Y_j}(X_j,Y_j)\right)$$

We then get convergence by the strong law of large numbers:

$$-\frac{1}{n}\sum_{j=1}^{n}\ln\left(\mathbb{P}_{X_j}(X_j)\right) \to -\mathbb{E}\left[\ln\left(\mathbb{P}_{X_1}(X_1)\right)\right] = H(X_1)$$

$$-\frac{1}{n}\sum_{j=1}^{n}\ln\left(\mathbb{P}_{Y_j}(Y_j)\right) \to -\mathbb{E}\left[\ln\left(\mathbb{P}_{Y_1}(Y_1)\right)\right] = H(Y_1)$$

$$-\frac{1}{n}\sum_{j=1}^{n}\ln\left(\mathbb{P}_{X_j,Y_j}(X_j,Y_j)\right) \to -\mathbb{E}\left[\ln\left(\mathbb{P}_{X_1,Y_1}(X_1,Y_1)\right)\right] = H(X_1,Y_1)$$

□

Since the strong implies the weak law of large numbers, we have an immediate consequence for the probability of $F_\delta^n$.

**3.1.10 Corollary.** *Let $\epsilon > 0$. By choosing $n$ sufficiently large, we can achieve $\mathbb{P}(F_\delta^n) > 1-\epsilon$*

**3.1.11 Theorem** (Shannon-McMillan for channels)**.** *Given a discrete memoryless source $\{X_j\}_{j=1}^{\infty}$, and a discrete memoryless channel $\gamma : \mathbb{A}\times\Omega \to \mathbb{B}$, $\ln(\mathbb{P}_{X_1}(X_1))$, $\ln(\mathbb{P}_{Y_1}(Y_1))$ and $\ln(\mathbb{P}_{X_1,Y_1}(X_1,Y_1))$ integrable, and $\delta > 0$, then for all sufficiently large $n$, $F_\delta^n$ satisfies*

1. $\mathbb{P}_{X_1,\dots,X_n,Y_1,\dots,Y_n}((F_\delta^n)^c) < \delta$
2. $(1-\delta)exp(n(H(X,Y)-\delta)) < |F_\delta^n|$
3. $|F_\delta^n| < exp(n(H(X,Y)+\delta))$
4. *If $(x_1,\dots,x_n,y_1,\dots,y_n) \in F_\delta^n$, then*

$$exp(-n(H(X,Y)+\delta)) < \mathbb{P}(x_1,\dots,x_n,y_1,\dots,y_n) < exp(-n(H(X,Y)-\delta))$$

*Proof.* 1. By the preceding corollary.

2. To prove the second assertion, use the above to get

$$1-\delta < \mathbb{P}_{X_1,\dots,X_n,Y_1,\dots,Y_n}(F_\delta^n) = \sum_{(x_1,\dots,x_n,y_1,\dots,y_n)\in F_\delta^n}\mathbb{P}_{X_1,\dots,X_n,Y_1,\dots,Y_n}(x_1,\dots,x_n,y_1,\dots,y_n)$$

By the definition of a jointly typical sequence, if $(x_1,\dots,x_n,y_1,\dots,y_n) \in F_\delta^n$, then

$$-\frac{1}{n}\mathbb{P}_{X_1,\dots,X_n,Y_1,\dots,Y_n}(x_1,\dots,x_n,y_1,\dots,y_n) > H(X_1,Y_1)-\delta$$

so, by exponentiating,

$$\begin{aligned} 1-\delta <& \sum_{(x_1,\ldots,x_n,y_1,\ldots,y_n)\in F_\delta^n} \mathbb{P}_{X_1,\ldots,X_n,Y_1,\ldots,Y_n}(x_1,\ldots,x_n,y_1,\ldots,y_n) \\ <& \sum_{(x_1,\ldots,x_n,y_1,\ldots,y_n)\in F_\delta^n} e^{-n(H(X_1,Y_1)-\delta)} \\ =& |F_\delta^n| e^{-n(H(X_1,Y_1)-\delta)} \end{aligned}$$

and hence $(1-\delta)exp(n(H(X,Y)-\delta)) < |F_\delta^n|$.

3. By the definition of a jointly typical sequence, if $(x_1,\ldots,x_n,y_1,\ldots,y_n)\in F_\delta^n$, then

$$-\frac{1}{n}\mathbb{P}_{X_1,\ldots,X_n,Y_1,\ldots,Y_n}(x_1,\ldots,x_n,y_1,\ldots,y_n) < H(X_1,Y_1)+\delta$$

so, by exponentiating,

$$\begin{aligned} 1 \geq& \sum_{(x_1,\ldots,x_n,y_1,\ldots,y_n)\in F_\delta^n} \mathbb{P}_{X_1,\ldots,X_n,Y_1,\ldots,Y_n}(x_1,\ldots,x_n,y_1,\ldots,y_n) \\ >& \sum_{(x_1,\ldots,x_n,y_1,\ldots,y_n)\in F_\delta^n} e^{-n(H(X_1,Y_1)+\delta)} \\ =& |F_\delta^n| e^{-n(H(X_1,Y_1)+\delta)} \end{aligned}$$

and thus $|F_\delta^n| < exp(n(H(X,Y)+\delta))$.

4. The last assertion follows directly from the definition of a jointly typical sequence, and was used in proving both parts of the second assertion.

□

*3.1.12 Question.* What is the transmission rate of a channel?

**3.1.13 Definition** (capacity)**.** Given a discrete memoryless channel $\gamma$ with transition probabilities $\{\mathbb{W}(\bullet|a)\}_{a\in\mathbb{A}}$, then we define the capacity of $\gamma$ to be $C=\max_{\mathbb{P}_X} I(X;Y)$ where $Y=\gamma(X)$.

**3.1.14 Theorem** (Channel coding)**.** *Consider a discrete memoryless channel $\gamma:\mathbb{A}\times\Omega\to\mathbb{B}$, and $C=\max_{\mathbb{P}_X} I(X;Y)$ with $Y=\gamma(X)$. Let $\epsilon>0$, $4\epsilon>\tau>0$. Then there is a sequence of $(n,m_n)$ fixed-length transmission codes $(\mathcal{C}_n,\phi_n,\psi_n)$ such that for sufficiently large $n$, $\frac{1}{n}\ln m_n > C-\tau$ and the averaged error probability $P_e(\psi_n\circ\gamma\circ\phi_n\neq id)<\epsilon$.*

*Proof.* For this proof, we will abbreviate $(X_1,\ldots,X_n)$ as $X^{\otimes n}$, $(Y_1,\ldots,Y_n)$ as $Y^{\otimes n}$, and $(X_1,\ldots,X_n,Y_1,\ldots,Y_n)$ as $(X,Y)^{\otimes n}$. By choosing $N_0$ appropriately, we have $\{m_n\}$ integers with $C-\tau<\frac{1}{n}\ln m_n\leq C-\frac{\tau}{2}$. Let $\delta=\frac{\tau}{8}$. Choose $\mathbb{P}_X$ which achieves the capacity, let $\mathbb{Q}_Y$ be the corresponding output distribution, i.e. $\mathbb{Q}_Y(b)=\sum_{a\in\mathbb{A}}\mathbb{P}_X(a)\mathbb{W}(b|a)$. We demonstrate the existence of a code sequence in 3 steps.

Step 1: Select $m_n$ input sequences $\{c_j\}_{j=1}^{m_n}$ of length $n$ from $\mathbb{A}$ randomly, according to $\mathbb{P}_{X^{\otimes n}}$. Note that duplicates are possible. Identify $\mathcal{C}_n=\{1,2,\ldots,m_n\}$ with codes sequences $\phi_n(j)=c_j$ and choose

$$\psi_n(y)=\begin{cases} j & \text{if } \phi(j)=c_j, (c_j,y)\in F_\delta^n, \text{ and } \nexists (c',y)\in F_\delta^n \text{ with } c'\neq c_j \\ \text{arbitrary} & \text{otherwise} \end{cases}$$

Step 2: We estimate the averaged error probability for the randomly generated code. Let $\Lambda(c_m)$ be the probability of having a decoding error given fixed input sequence $c_m$. The error could come from $(c_m, y) \notin F_\delta^n$ (not being typical) or from multiple source symbols mapped to $y$ within $F_\delta^n$. We have

$$\Lambda(c_m) \leq \sum_{\substack{y \in \mathbb{B}^n \\ (c_m, y) \notin F_\delta^n}} \mathbb{W}(y|c_m) + \sum_{m' \neq m} \sum_{\substack{y \in \mathbb{B}^n \\ (c_{m'}, y) \in F_\delta^n}} \mathbb{W}(y|c_m)$$

with the notation $y = (y_1, \ldots, y_n)$ with $\mathbb{W}(y|c_m) = \prod_{k=1}^n \mathbb{W}(y_k|(c_m)_k)$.

If we average with respect to the choice of codewords, governed by $\mathbb{P}_{X^{\otimes n}}$, then we obtain the expected error probability.

$$\begin{aligned}
\mathbb{E}_{X^{\otimes n}}[\Lambda(c_m)] &= \sum_{a \in \mathbb{A}^n} \mathbb{P}_{X^{\otimes n}}(a)\Lambda(a) \\
&\leq \sum_{a \in \mathbb{A}^n} \left( \sum_{\substack{y \in \mathbb{B}^n \\ (a, y) \notin F_\delta^n}} \mathbb{P}_{X^{\otimes n}}(a)\mathbb{W}(y|a) + \sum_{m' \neq m} \sum_{\substack{y \in \mathbb{B}^n \\ (c_{m'}, y) \in F_\delta^n}} \mathbb{P}_{X^{\otimes n}}(a)\mathbb{W}(y|a) \right) \\
&= \mathbb{P}_{(X,Y)^{\otimes n}}\left((F_\delta^n)^c\right) + \sum_{a \in \mathbb{A}^n} \sum_{m' \neq m} \sum_{\substack{y \in \mathbb{B}^n \\ (c_{m'}, y) \in F_\delta^n}} \mathbb{P}_{X^{\otimes n}, Y^{\otimes n}}(a, y) \\
&= \mathbb{P}_{(X,Y)^{\otimes n}}\left((F_\delta^n)^c\right) + \sum_{m' \neq m} \sum_{\substack{y \in \mathbb{B}^n \\ (c_{m'}, y) \in F_\delta^n}} \mathbb{P}_{Y^{\otimes n}}(y)
\end{aligned}$$

To be continued...