EECS 769 Homework 3

Elise McEllhiney

Oct. 4, 2016

0.1 Data processing

Let $X_1 \to X_2 \to X_3 ... \to X_n$ form a Markov chain in this order; i.e., let

$$p(x_1, x_2, ..., x_n) = p(x_1)p(x_2|x_1)...p(x_n|x_{n-1})$$

Reduce $I(X_1; X_2, ..., X_n)$ to it's simplest form

We can use the chain rule

$$I(X_1; X_2, ..., X_n) = I(X_1; X_2) + I(X_1; X_3|X_2) + ... + I(X_1; X_n|X_2, ..., X_{n-2})$$

By the we know that, for a Markov chain, X_i and X_{i+2} are conditionally independent given X_{i+1} we know that everything in the above equation goes to zero except $I(X_1, X_2)$. This can also be said as $I(X_1; X_i) \subseteq I(X_1; X_2)$ for $i \geq 2$ Hence,

$$I(X_1; X_2, ..., X_n) = I(X_1, X_2)$$

0.2 Markov's inequality for probabilities

Let $p(x)$ be a probability mass function. Prove, for all $d \geq 0$.

$$Pr\{p(X) \geq d\} log(\frac{1}{d}) \geq H(X)$$

$$\begin{aligned} P(p(X)) < d) log(\frac{1}{d}) &= \sum_x p(x) log(\frac{1}{d}) \\ &\text{where } p(x) < d \\ &\leq \sum_x p(x) log(\frac{1}{p(x)}) \\ &\leq \sum_x p(x) log(\frac{1}{p(x)}) \\ &= H(X) \end{aligned} \tag{1}$$

0.3 Fano

We are given the following joint distribution on (X, Y)

X \ Y	a	b	c
1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

Let $\hat{X}(Y)$ be an estimator for X (based on Y) and let $P_e = Pr\{\hat{X}(Y) \neq X\}$

0.3.1 Find the minimum probability of error estimator $\hat{X}(Y)$ and the associated P_e

$$\hat{X}(a) = 1$$

$$\hat{X}(b) = 2$$

$$\hat{X}(c) = 3$$

$$\begin{aligned} \therefore P_e &= Pr(1,b) + Pr(1,c) + Pr(2,a) + Pr(2,c) + Pr(3,a) + Pr(3,b) \\ &= \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} + \frac{1}{12} \\ &= \frac{1}{2} \end{aligned} \tag{2}$$

0.3.2 Evaluate Fano's inequality for this problem and compare

$$H(P_e) + P_e log(|\mathcal{X}| - 1) \geq H(X|Y)$$

$$H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) = -\frac{1}{2}\log\frac{1}{2} - \frac{2}{4}\log\frac{1}{4} = \frac{1}{2} + 1 = 1.5bits$$

$$H(P_e) = -\frac{6}{12}\log\frac{1}{12} = 1.79bits$$

$$\begin{aligned} H(X|Y) &= H(X|Y=a)Pr(y=a) + H(X|Y=b)Pr(y=b) + H(X|Y=c)Pr(y=c) \\ &= H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})Pr(y=a) + H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})Pr(y=b) + H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})Pr(y=c) \\ &= H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4})(Pr(y=a) + Pr(y=b) + Pr(y=c)) \\ &= H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) \\ &= 1.5bits \end{aligned} \tag{3}$$

$$H(P_e) + P_e log(|\mathcal{X}| - 1) \geq H(X|Y)$$

$$1.79bits + \frac{1}{2}log(3-1) \geq 1.5$$

$$1.79bits + \frac{1}{2}log(3-1) \geq 1.5$$

$$1.79 + 0.5 \geq 1.5$$

$$2.29 \geq 1.5$$

We can see that Fano's inequality holds true

0.4 Fano's inequality

Let $Pr(X = i) = p_i, i = 1, 2, ..., m$ and let $p_i \geq p_2 \geq p_3 \geq ... \geq p_m$. The minimal probability of error predictor of X is $\hat{X} = 1$, with resulting probability or error $P_e = 1 - p_1$. Maximize $H(\mathbf{p})$ subject to the constraint $1 - p_1 = P_e$ to find a bound on P_e in terms of H. This is Fano's inequality in the absence of conditioning

$$\begin{aligned} H(\mathbf{p}) &= -p_1 \log p_1 - \sum_{i=2}^{m} p_i \log p_i \\ &= -p_1 \log p_1 - \sum_{i=2}^{m} P_e \frac{p_i}{P_e} \log \frac{p_i}{P_e} - P_e \log P_e \\ &= H(P_e) + P_e H(\frac{p_2}{P_e}, \frac{p_3}{P_e}, ..., \frac{p_m}{P_e}) \\ &\leq H(P_e) + P_e \log(m-1) \end{aligned} \tag{4}$$

We know that maximum entropy is achieved through a uniform distribution. Hence,

$$H(X) \leq H(P_e) + P_e \log(m-1)$$

Since $H(P_e)$ is ≥ 1

$$H(X) \leq 1 + P_e \log(m-1)$$

$$H(X) - 1 \leq P_e \log(m-1)$$

$$\frac{H(X) - 1}{\log(m-1)} \leq P_e$$

0.5 Maximum entropy

Find the probability mass function $p(x)$ that maximizes the entropy $H(X)$ of a non-negative inter-valued random variable X subject to the constraint

$$EX = \sum_{n=0}^{\infty} np(n) = A$$

for a fixed value $A > 0$. Evaluate this maximum $H(X)$.

If we let $q_i = \alpha(\beta)^i$

$$\begin{aligned} -\sum_{n=0}^{\infty} p(n) \log p(n) &\leq -\sum_{n=0}^{\infty} p(n) \log q(n) \\ &= -\log(\alpha) \sum_{n=0}^{\infty} p(n) - \log(\beta) \sum_{n=0}^{\infty} np(n) \\ &= -\log(\alpha) - A \log(\beta) \end{aligned} \tag{5}$$

This is true $\forall$ α and β such that

$$1 = \sum_{i=0}^{\infty} \alpha\beta^i = \alpha(\frac{1}{1-\beta})$$

Also, from the question, we know that:

$$A = \sum_{i=0}^{\infty} i\alpha\beta^i = \alpha\frac{\beta}{(1-\beta)^2}$$

$$\begin{aligned} A &= \alpha\frac{\beta}{(1-\beta)^2} \\ &= (\frac{\alpha}{1-\beta})(\frac{\beta}{1-\beta}) \\ &= (1)(\frac{\beta}{1-\beta}) \\ &= (\frac{\beta}{1-\beta}) \end{aligned} \tag{6}$$

$$\begin{aligned} \Rightarrow A &= (\frac{\beta}{1-\beta}) \\ A(1-\beta) &= \beta \\ A - A\beta &= \beta \\ A &= A\beta + \beta \\ A &= \beta(A+1) \\ \beta &= \frac{A}{A+1} \end{aligned}$$

$$\begin{aligned} 1 &= \alpha(\frac{1}{1-\beta}) = \alpha(\frac{1}{1-\frac{A}{A+1}}) \\ \Rightarrow \alpha &= 1 - \frac{A}{A+1} = \frac{A+1}{A+1} - \frac{A}{A+1} \\ \alpha &= \frac{1}{A+1} \end{aligned}$$

$$H(X) = -\log\alpha - A\log\beta = (A+1)\log(A+1) - A\log A$$

0.6 Uniquely decodable and instantaneous codes

Let $L = \sum_{i=1}^{m} p_i l_i^{70}$ be the expected value of the 70th power of the word lengths associated with an encoding of the random variable X. Let $L_1 = minL$ over all instantaneous codes; and let $L_2 = minL$, over all uniquely decodable codes. What inequality relationship exists between L_1 and L_2?

Since all instantaneous code is by definition also uniquely decodable, we know:

$$L_1 \geq L_2$$

By the Kraft-Mcmillan Inequality we know that any uniquely decodable code must satisfy the Kraft Inequality.

$$\sum_i 2^{-l_i} \leq 1$$

We also know that given a set of codeword lengths that satisfy the Kraft inequality, it is possible to construct an instantaneous code with these codeword lengths. Hence:

$$L_1 = L_2$$

0.7 Axiomatic definition of entropy

If we assume certain axioms for our measure of information, then we will be forced to use a logarithmic measure like entropy. Shannon used this to justify his initial definition of entropy.
If a sequence of symmetric functions $H_m(p_1, p_2, ..., p_m)$ satisfies the following properties,

- **Normalization:** $H_2(\frac{1}{2}, \frac{1}{2}) = 1$**,**
- **Continuity:** $H_2(p, 1-p)$ **is a continuous function of** p**,**
- **Grouping:** $H_m(p_1, p_2, ..., p_m) = H_{m-1}(p_1 + p_2, p_3, ..., p_m) + (p_1 + p_2) H_2(\frac{p_1}{p_1+p_2}, \frac{p_2}{p_1+p_2})$**,**

prove that H_m **must be of the form**

$$H_m(p_1, p_2, ..., p_m) = -\sum_{i=1}^{m} p_i \log p_i, m = 2, 3, ...$$

There are various other axiomatic formulation which also result in the same definition of entropy. See, for example, the book by Csiszar and Korner.

I found various papers on this proof online and was able to attempt to figure out some of the early steps that way, but I was unable to decipher the middle section, so there's a big gap where I don't understand how to put together the second half of the proof.

$$H_m(p_1, ...p_m) = H_{m-k}(p_1+p_2+...p_k, p_{k+1}, ..., p_m)+(p_1+p_2+...+p_k)H_k(\frac{p_1}{p_1+p_2+...+p_k}, ..., \frac{p_k}{p_1+p_2+...+p_k})$$

$$\begin{aligned}
H_m(p_1, ...p_m) &= H_{m-1}(\sum_{i=1}^{2} p_i, p_3, ..., p_m) + \sum_{i=1}^{2} p_i H(\frac{p_2}{\sum_{i=1}^{2} p_i}, 1 - \frac{p_2}{\sum_{i=1}^{2} p_i}) \\
&= H_{m-2}(\sum_{i=1}^{3} p_i, p_4, ..., p_m) + \sum_{i=1}^{3} p_i H(\frac{p_3}{\sum_{i=1}^{3} p_i}, 1 - \frac{p_3}{\sum_{i=1}^{3} p_i}) + \sum_{i=1}^{2} p_i H(\frac{p_2}{\sum_{i=1}^{2} p_i}, 1 - \frac{p_2}{\sum_{i=1}^{2} p_i}) \\
&... \\
&= H_{m-k-1}(\sum_{i=1}^{k} p_i, p_{k+1} ..., p_m) + \sum_{i=2}^{k} \sum_{j=1}^{i} p_j H(\frac{p_i}{\sum_{j=1}^{i} p_j})
\end{aligned} \tag{7}$$

$$H_k(\frac{p_1}{\sum_{i=1}^k p_i}, ..., \frac{p_m}{\sum_{i=1}^k p_i}) = H_2(\frac{\sum_{i=1}^{k-1} p_i}{\sum_{i=1}^k p_i}, \frac{p_k}{\sum_{i=1}^k p_i}) = \frac{1}{\sum_{i=1}^k p_i} \sum_{i=2}^{k} \sum_{j=1}^{i} p_j$$

$$H_m(p_1, ..., p_m) = H_{m-k}(\sum_{i=1}^{k} p_i, p_{k+1}, ..., p_m) + \sum_{i=1}^{k} p_i H_k(\frac{p_1}{\sum_{i=1}^k p_i}, ..., \frac{p_k}{\sum_{i=1}^k p_i})$$

Let $f(m) = H_m(\frac{1}{m}, \frac{1}{m}, ..., \frac{1}{m})$

$$\begin{aligned} f(mn) &= H_{mn}(\frac{1}{mn}, ..., \frac{1}{mn}) \\ &= H_{mn-n}(\frac{1}{m}, \frac{1}{mn}, ..., \frac{1}{mn}) + \frac{1}{m} H_n(\frac{1}{n}, .., \frac{1}{n}) \\ &... \\ &= H_m(\frac{1}{m}, ..., \frac{1}{m}) + H_n(\frac{1}{n}, ..., \frac{1}{n}) \\ &= f(m) + f(n) \\ &\Rightarrow f(m^k) = kf(m) \end{aligned} \tag{8}$$

$$\begin{aligned} f(m+1) &= H_{m+1}(\frac{1}{m+1}, ..., \frac{1}{m+1}) \\ &= H(\frac{1}{m+1}, 1 - \frac{1}{m+1}) + \frac{m}{m+1} H_m(\frac{1}{m}, ..., \frac{1}{m}) \\ &= H(\frac{1}{m+1}, 1 - \frac{1}{m+1}) + \frac{1}{m+1} f(m) \end{aligned} \tag{9}$$

$$H(\frac{1}{m+1}, 1 - \frac{1}{m+1}) = f(m+1) - \frac{m}{m+1} f(m)$$

...I can't figure out this section at all...

$$H_m(p_1, ..., p_m) = -\sum_{i=1}^{m} p_i \log p_i$$