An introduction to Information Theory

Adrish Banerjee

Department of Electrical Engineering
Indian Institute of Technology Kanpur
Kanpur, Uttar Pradesh
India

Sept. 6, 2016

Lecture #14B: Problem solving session-IV

Measure of Information

- **Problem # 1:** Prove that entropy is the only function that satisfies the four conditions for information measure.

Measure of Information

- **Problem # 1:** Prove that entropy is the only function that satisfies the four conditions for information measure.
- Assume a random source X comprising of M elements with probabilities $p_i (i = 1, \cdots, M)$, namely

Measure of Information

- **Problem # 1:** Prove that entropy is the only function that satisfies the four conditions for information measure.
- Assume a random source X comprising of M elements with probabilities $p_i (i = 1, \cdots, M)$, namely
  - (Axiom 1:) If all the probabilities were equal, the function $H(1/M, 1/M, \cdots, 1/M) = f(M)$ is a monotonically increasing function of M $(M = 1, 2, \cdots)$

Measure of Information

- **Problem # 1:** Prove that entropy is the only function that satisfies the four conditions for information measure.
- Assume a random source X comprising of M elements with probabilities $p_i (i = 1, \cdots, M)$, namely
  - (Axiom 1:) If all the probabilities were equal, the function $H(1/M, 1/M, \cdots, 1/M) = f(M)$ is a monotonically increasing function of M $(M = 1, 2, \cdots)$
  - (Axiom 2:) For independent sources X and Y with M and L elements respectively, $f(ML) = f(M) + f(L)$ $(M, L = 1, 2, \cdots)$

Measure of Information

- **Problem # 1:** Prove that entropy is the only function that satisfies the four conditions for information measure.
- Assume a random source X comprising of M elements with probabilities $p_i (i = 1, \cdots, M)$, namely
  - (Axiom 1:) If all the probabilities were equal, the function $H(1/M, 1/M, \cdots, 1/M) = f(M)$ is a monotonically increasing function of M $(M = 1, 2, \cdots)$
  - (Axiom 2:) For independent sources X and Y with M and L elements respectively, $f(ML) = f(M) + f(L)$ $(M, L = 1, 2, \cdots)$
  - (Axiom 3:) $H(p_1, p_2, \cdots, p_M) = H(p_1 + \cdots + p_r, p_{r+1} + \cdots + p_M)$
    $+(p_1 + \cdots + p_r) H\left(\frac{p_1}{\sum_{i=1}^{r} p_i}, \cdots, \frac{p_r}{\sum_{i=1}^{r} p_i}\right)$
    $+(p_{r+1} + \cdots + p_M) H\left(\frac{p_{r+1}}{\sum_{i=r+1}^{M} p_i}, \cdots, \frac{p_M}{\sum_{l=r+1}^{M} p_i}\right)$

Measure of Information

- **Problem # 1:** Prove that entropy is the only function that satisfies the four conditions for information measure.
- Assume a random source X comprising of M elements with probabilities $p_i (i = 1, \cdots, M)$, namely
  - (Axiom 1:) If all the probabilities were equal, the function $H(1/M, 1/M, \cdots, 1/M) = f(M)$ is a monotonically increasing function of M $(M = 1, 2, \cdots)$
  - (Axiom 2:) For independent sources X and Y with M and L elements respectively, $f(ML) = f(M) + f(L)$ $(M, L = 1, 2, \cdots)$
  - (Axiom 3:) $H(p_1, p_2, \cdots, p_M) = H(p_1 + \cdots + p_r, p_{r+1} + \cdots + p_M)$
    $+(p_1 + \cdots + p_r) H\left(\frac{p_1}{\sum_{i=1}^{r} p_i}, \cdots, \frac{p_r}{\sum_{i=1}^{r} p_i}\right)$
    $+(p_{r+1} + \cdots + p_M) H\left(\frac{p_{r+1}}{\sum_{i=r+1}^{M} p_i}, \cdots, \frac{p_M}{\sum_{l=r+1}^{M} p_i}\right)$
  - (Axiom 4:) $H = H(p_1, p_2, \cdots, p_N)$ is a continuous function of the probability set p_i.

Problem # 1 (contd.)

- **Solutions:** We will prove that only function satisfying these four axioms is of the form

$$H(p_1, \cdots, p_M) = -C \sum_{i=1}^{M} p_i \log p_i$$

Problem # 1 (contd.)

- **Solutions:** We will prove that only function satisfying these four axioms is of the form

$$H(p_1, \cdots, p_M) = -C \sum_{i=1}^{M} p_i \log p_i$$

a) $f(M^k) = kf(M)$ for all positive integers M and k.

Problem # 1 (contd.)

- **Solutions:** We will prove that only function satisfying these four axioms is of the form

$$H(p_1, \cdots, p_M) = -C \sum_{i=1}^{M} p_i \log p_i$$

a) $f(M^k) = kf(M)$ for all positive integers M and k.
We will prove this using mathematical induction.

Problem # 1 (contd.)

- **Solutions:** We will prove that only function satisfying these four axioms is of the form

$$H(p_1, \cdots, p_M) = -C \sum_{i=1}^{M} p_i \log p_i$$

a) $f(M^k) = kf(M)$ for all positive integers M and k.
We will prove this using mathematical induction.

- If M is fixed integer, then it is true for $k = 1$.

Problem # 1 (contd.)

- **Solutions:** We will prove that only function satisfying these four axioms is of the form

$$H(p_1, \cdots, p_M) = -C \sum_{i=1}^{M} p_i \log p_i$$

a) $f(M^k) = kf(M)$ for all positive integers M and k.
We will prove this using mathematical induction.
- If M is fixed integer, then it is true for $k = 1$.
- Since $f(M^k) = f(M \cdot M^{k-1}) = f(M) + f(M^{k-1})$. Assume it is true for $k-1$, then $f(M^k) = f(M) + (k-1)f(M) = kf(M)$.

Problem # 1 (contd.)

b)
- $f(M) = C \log M$ $(M = 1, 2, \cdots)$, where C is a positive number.

Problem # 1 (contd.)

b)

- $f(M) = C \log M$ $(M = 1, 2, \cdots)$, where C is a positive number.
- Let M=1, then $f(1) = f(1 \cdot 1) = f(1) + f(1)$, and hence $f(1) = 0$.

Problem # 1 (contd.)

b)

- $f(M) = C \log M$ $(M = 1, 2, \cdots)$, where C is a positive number.
- Let M=1, then $f(1) = f(1 \cdot 1) = f(1) + f(1)$, and hence $f(1) = 0$.
- Let M be a positive integer greater than 1. Let r be an arbitrary positive integer, such that 2^r lies between two powers of M, i.e. $M^k \leq 2^r < M^{k+1}$. We have from Axiom 1, $f(M^k) \leq f(2^r) < f(M^{k+1})$.

Problem # 1 (contd.)

b)

- $f(M) = C \log M$ $(M = 1, 2, \cdots)$, where C is a positive number.
- Let M=1, then $f(1) = f(1 \cdot 1) = f(1) + f(1)$, and hence $f(1) = 0$.
- Let M be a positive integer greater than 1. Let r be an arbitrary positive integer, such that 2^r lies between two powers of M, i.e. $M^k \leq 2^r < M^{k+1}$. We have from Axiom 1, $f(M^k) \leq f(2^r) < f(M^{k+1})$.
- Thus we have $kf(M) \leq rf(2) < (k+1)f(M)$ or $k/r \leq f(2)/f(M) < (k+1)/r$.

Problem # 1 (contd.)

b)

- $f(M) = C \log M$ $(M = 1, 2, \cdots)$, where C is a positive number.
- Let M=1, then $f(1) = f(1 \cdot 1) = f(1) + f(1)$, and hence $f(1) = 0$.
- Let M be a positive integer greater than 1. Let r be an arbitrary positive integer, such that 2^r lies between two powers of M, i.e. $M^k \leq 2^r < M^{k+1}$. We have from Axiom 1, $f(M^k) \leq f(2^r) < f(M^{k+1})$.
- Thus we have $kf(M) \leq rf(2) < (k+1)f(M)$ or $k/r \leq f(2)/f(M) < (k+1)/r$.
- Logarithm is a monotone increasing function, hence $\log M^k \leq \log 2^r < \log M^{k+1}$ or we have $k \log M \leq r \log 2 < (k+1) \log M$, or $k/r \leq (\log 2)/(logM) < (k+1)/r$.

Problem # 1 (contd.)

b)

- Now we have

$$\left|\frac{\log 2}{\log M} - \frac{f(2)}{f(M)}\right| < \frac{1}{r}$$

Problem # 1 (contd.)

b)

- Now we have

$$\left|\frac{\log 2}{\log M} - \frac{f(2)}{f(M)}\right| < \frac{1}{r}$$

- Since M is fixed and r is arbitrary, we may let $r \to \infty$ and we get

$$(\log 2)/(\log M) = f(2)/f(M)$$

or $f(M) = c \log M$, where $c = f(2)/\log 2$

Problem # 1 (contd.)

c) $H(p, 1-p) = -C[p \log p + (1-p)\log(1-p)]$ if p is a rational number.

Problem # 1 (contd.)

c) $H(p, 1-p) = -C[p \log p + (1-p)\log(1-p)]$ if p is a rational number.

- Let $p = r/s$ where r and s are positive integers. We consider

$$\begin{aligned} f(s) &= H\left(\frac{1}{s}, \cdots, \frac{1}{s}\right) \\ &= H\left(\frac{r}{s}, \frac{s-r}{s}\right) + \frac{r}{s} f(r) + \frac{(s-r)}{s} f(s-r) \end{aligned}$$

Problem # 1 (contd.)

c) $H(p, 1-p) = -C[p \log p + (1-p)\log(1-p)]$ if p is a rational number.

- Let $p = r/s$ where r and s are positive integers. We consider

$$\begin{aligned} f(s) &= H\left(\frac{1}{s}, \cdots, \frac{1}{s}\right) \\ &= H\left(\frac{r}{s}, \frac{s-r}{s}\right) + \frac{r}{s} f(r) + \frac{(s-r)}{s} f(s-r) \end{aligned}$$

- We have

$$C \log s = H(p, 1-p) + Cp \log r + C(1-p)\log(s-r)$$

Problem # 1 (contd.)

c) $H(p, 1-p) = -C[p \log p + (1-p)\log(1-p)]$ if p is a rational number.

- Let $p = r/s$ where r and s are positive integers. We consider

$$\begin{aligned} f(s) &= H\left(\frac{1}{s}, \cdots, \frac{1}{s}\right) \\ &= H\left(\frac{r}{s}, \frac{s-r}{s}\right) + \frac{r}{s} f(r) + \frac{(s-r)}{s} f(s-r) \end{aligned}$$

- We have

$$C \log s = H(p, 1-p) + Cp \log r + C(1-p)\log(s-r)$$

- Thus we have

$$\begin{aligned} H(p, 1-p) &= -C\left[p \log r - \log s + (1-p)\log(s-r)\right] \\ &= -C\left[p \log \frac{r}{s} + (1-p)\log \frac{s-r}{s}\right] \\ &= -C\left[p \log p + (1-p)\log(1-p)\right] \end{aligned}$$

Problem # 1 (contd.)

d) $H(p, 1-p) = -C[p \log p + (1-p)\log(1-p)]$ for all p. This follows from continuity axiom.

e) $H(p_1, \cdots, p_M) = -C \sum_{i=1}^{M} p_i \log p_i$ $(M = 1, 2, \cdots)$.

- Using mathematical induction, we know the above equation is true for $M = 1, 2$.
- If $M > 2$, by Axiom 3, we get

$$H(p_1, \cdots, p_M) = H(p_1 + \cdots + p_{M-1}, p_M)$$
$$+(p_1 + \cdots + p_{M-1}) H\left(\frac{p_1}{\sum_{i=1}^{M} p_i}, \cdots, \frac{p_{M-1}}{\sum_{i=1}^{M} p_i}\right) + p_M H(1)$$

Problem # 1 (contd.)

- Assuming the formula is valid for positive integers upto $M-1$, we obtain

$$\begin{aligned} H(p_1, \cdots, p_M) &= -C\left[(p_1 + \cdots + p_{M-1}) \log(p_1 + \cdots + p_{M-1}) + p_M \log p_M\right] \\ &-C(p_1 + \cdots + p_{M-1}) \left[\frac{p_1}{\sum_{i=1}^{M} p_i} \log \frac{p_1}{\sum_{i=1}^{M} p_i} + \cdots + \frac{p_{M-1}}{\sum_{i=1}^{M} p_i} \log \frac{p_{M-1}}{\sum_{i=1}^{M} p_i}\right] \\ &+p_M(0) \\ = \quad & -C\left[\left(\sum_{i=1}^{M-1} p_i\right) \log\left(\sum_{i=1}^{M-1} p_i\right) + p_M \log p_M\right] \\ & -C\left[\sum_{i=1}^{M} p_i \log p_i - \left(\sum_{i=1}^{M-1} p_i \log \sum_{i=1}^{M-1} p_i\right)\right] \\ = \quad & -C \sum_{i=1}^{M} p_i \log p_i \end{aligned}$$

Measure of Information

- **Problem # 2:** Suppose one has 12 coins, among which there may or may not be one counterfeit coin. If there is a counterfeit coin, it may be either heavier or lighter than the other coins. The coins are to be weighed by a balance. In three weightings, you will have to find the counterfeit coin (if any) and correctly declare it to be heavier or lighter.

Measure of Information

- **Problem # 2:** Suppose one has 12 coins, among which there may or may not be one counterfeit coin. If there is a counterfeit coin, it may be either heavier or lighter than the other coins. The coins are to be weighed by a balance. In three weightings, you will have to find the counterfeit coin (if any) and correctly declare it to be heavier or lighter.
- **Solutions:** For 12 coins we have 23 possible outcomes, corresponding to the case when one of the 12 coins is heavier or one of the 12 coins is lighter or all coins are of same weight. We denotes numbers $\{-12, -11, \cdots, -1, 0, 1, \cdots, 11, 12\}$ in ternary number system with alphabets $\{-1, 0, 1\}$.

Problem # 2 (contd.)

- The representation of the positive numbers is shown below

	1	2	3	4	5	6	7	8	9	10	11	12	
3^0	1	-1	0	1	-1	0	1	-1	0	1	-1	0	$\sum_1 = 0$
3^1	0	1	1	1	-1	-1	-1	0	0	0	1	1	$\sum_2 = 2$
3^2	0	0	0	0	1	1	1	1	1	1	1	1	$\sum_3 = 8$

Problem # 2 (contd.)

- The representation of the positive numbers is shown below

	1	2	3	4	5	6	7	8	9	10	11	12	
3^0	1	-1	0	1	-1	0	1	-1	0	1	-1	0	$\sum_1 = 0$
3^1	0	1	1	1	-1	-1	-1	0	0	0	1	1	$\sum_2 = 2$
3^2	0	0	0	0	1	1	1	1	1	1	1	1	$\sum_3 = 8$

- Row sum can be made zero by negating columns 7, 9, 11 and 12. Thus we have

	1	2	3	4	5	6	7	8	9	10	11	12	
3^0	1	-1	0	1	-1	0	-1	-1	0	1	1	0	$\sum_1 = 0$
3^1	0	1	1	1	-1	-1	1	0	0	0	-1	-1	$\sum_2 = 0$
3^2	0	0	0	0	1	1	-1	1	-1	1	-1	-1	$\sum_3 = 0$

Problem # 2 (contd.)

- We will place the coin n on the left pan if $n_i = -1$, on the right pan if $n_i = 1$, and will keep it aside if $n_i = 0$.

Problem # 2 (contd.)

- We will place the coin n on the left pan if $n_i = -1$, on the right pan if $n_i = 1$, and will keep it aside if $n_i = 0$.
- The result of each weighting is -1 if the left pan is heavier, 1 if the right pan is heavier and 0 if both the pans are same.

Problem # 2 (contd.)

- We will place the coin n on the left pan if $n_i = -1$, on the right pan if $n_i = 1$, and will keep it aside if $n_i = 0$.
- The result of each weighting is -1 if the left pan is heavier, 1 if the right pan is heavier and 0 if both the pans are same.
- Three weighings of the coin can give us the ternary representation of the index of the odd coin.

Problem # 2 (contd.)

- We will place the coin n on the left pan if $n_i = -1$, on the right pan if $n_i = 1$, and will keep it aside if $n_i = 0$.
- The result of each weighting is -1 if the left pan is heavier, 1 if the right pan is heavier and 0 if both the pans are same.
- Three weighings of the coin can give us the ternary representation of the index of the odd coin.
- If the expansion is same as the column of the matrix, the coin is heavier, if it is of opposite sign, the coin is lighter.

Data Processing Lemma

- **Problem # 3:** Prove that $H(X_0|X_n)$ increases with n for any Markov Chain.

Data Processing Lemma

- **Problem # 3:** Prove that $H(X_0|X_n)$ increases with n for any Markov Chain.
- **Solutions:** For a Markov Source by data processing lemma, we have

$$I(X_0; X_n) \leq I(X_0; X_{n-1})$$

Hence we have

$$H(X_0) - H(X_0|X_n) \leq H(X_0) - H(X_0|X_{n-1})$$

or

$$H(X_0|X_n) \geq H(X_0|X_{n-1})$$

Data Processing Lemma

- **Problem # 4:** Under what conditions does $H(X|g(Y)) = H(X|Y)$? State the general condition, not just a special case.

Data Processing Lemma

- **Problem # 4:** Under what conditions does $H(X|g(Y)) = H(X|Y)$? State the general condition, not just a special case.
- **Solutions:** If $H(X|g(Y) = H(X|Y)$, then $H(X) - H(X|g(Y)) = H(X) - H(X|Y)$, i.e. $I(X; g(Y) = I(X; Y)$. This is the condition of equality in the data processing inequality.

Data Processing Lemma

- **Problem # 4:** Under what conditions does $H(X|g(Y)) = H(X|Y)$? State the general condition, not just a special case.
- **Solutions:** If $H(X|g(Y) = H(X|Y)$, then $H(X) - H(X|g(Y)) = H(X) - H(X|Y)$, i.e. $I(X; g(Y) = I(X; Y)$. This is the condition of equality in the data processing inequality.
- From the derivation of the inequality, we have equality iff $X \to g(Y) \to Y$ forms a Markov chain.

Data Processing Lemma

- **Problem # 4:** Under what conditions does $H(X|g(Y)) = H(X|Y)$? State the general condition, not just a special case.
- **Solutions:** If $H(X|g(Y) = H(X|Y)$, then $H(X) - H(X|g(Y)) = H(X) - H(X|Y)$, i.e. $I(X; g(Y) = I(X; Y)$. This is the condition of equality in the data processing inequality.
- From the derivation of the inequality, we have equality iff $X \to g(Y) \to Y$ forms a Markov chain.
- Hence $H(X|g(Y) = H(X|Y)$ iff $X \to g(Y) \to Y$.

Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Data Processing Lemma

- **Problem # 4:** Under what conditions does $H(X|g(Y)) = H(X|Y)$? State the general condition, not just a special case.
- **Solutions:** If $H(X|g(Y) = H(X|Y)$, then $H(X) - H(X|g(Y)) = H(X) - H(X|Y)$, i.e. $I(X; g(Y) = I(X; Y)$. This is the condition of equality in the data processing inequality.
- From the derivation of the inequality, we have equality iff $X \to g(Y) \to Y$ forms a Markov chain.
- Hence $H(X|g(Y) = H(X|Y)$ iff $X \to g(Y) \to Y$.
- This condition includes many special cases, such as g being one-to-one, and X and Y being independent. However, these two special cases do not exhaust all the possibilities.

Convex Function

- **Problem # 5:** Let $P_0(j/k)$ and $P_1(j/k)$, $0 \leq k \leq K-1$, $0 \leq j \leq J-1$, be two arbitrary sets of transition probabilities, and let $P(j/k) = \theta P_0(j/k) + (1-\theta)P_1(j/k)$. for an arbitrary θ, $0 \leq \theta < 1$.

Convex Function

- **Problem # 5:** Let $P_0(j/k)$ and $P_1(j/k)$, $0 \leq k \leq K-1$, $0 \leq j \leq J-1$, be two arbitrary sets of transition probabilities, and let $P(j/k) = \theta P_0(j/k) + (1-\theta)P_1(j/k)$. for an arbitrary θ, $0 \leq \theta < 1$.
- Let I_0, I_1, and I be the average mutual informations for these sets of transition probabilities, then to prove that the mutual information, $I(X;Y)$ is a convex function of $p(y/x)$ for fixed $p(x)$.

Convex Function

- **Problem # 5:** Let $P_0(j/k)$ and $P_1(j/k)$, $0 \leq k \leq K-1$, $0 \leq j \leq J-1$, be two arbitrary sets of transition probabilities, and let $P(j/k) = \theta P_0(j/k) + (1-\theta)P_1(j/k)$. for an arbitrary θ, $0 \leq \theta < 1$.
- Let I_0, I_1, and I be the average mutual informations for these sets of transition probabilities, then to prove that the mutual information, $I(X;Y)$ is a convex function of $p(y/x)$ for fixed $p(x)$.
- One has to prove that

$$\theta I_0 + (1-\theta)I_1 \geq I$$

Problem # 5 (contd.)

- Consider P_0 and P_1 as conditional on a binary variable Z

$$P_0(j/k) = P_{Y/XZ}(j|k,0) \quad P_1(j/k) = P_{Y/XZ}(j|k,1)$$

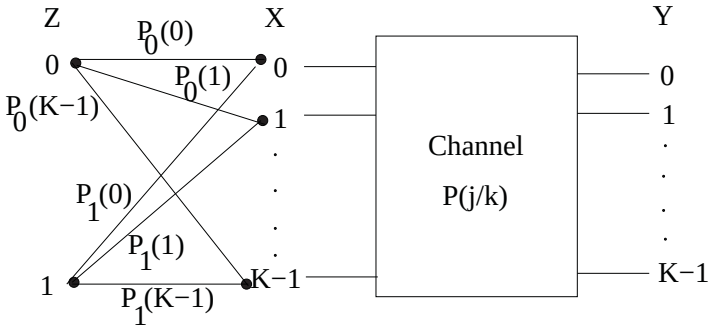


Figure: Figure for Problem 5

Problem # 5 (contd.)

- Consider P_0 and P_1 as conditional on a binary variable Z

$$P_0(j/k) = P_{Y/XZ}(j|k,0) \quad P_1(j/k) = P_{Y/XZ}(j|k,1)$$

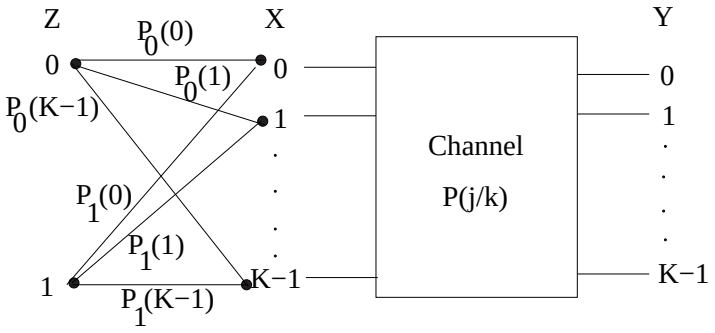


Figure: Figure for Problem 5

- Let $P_Z(0) = \theta$, $P_Z(1) = 1 - \theta$, and we define Z to be statistically independent of X.

Problem # 5 (contd.)

- Consider P_0 and P_1 as conditional on a binary variable Z

$$P_0(j/k) = P_{Y/XZ}(j|k,0) \quad P_1(j/k) = P_{Y/XZ}(j|k,1)$$

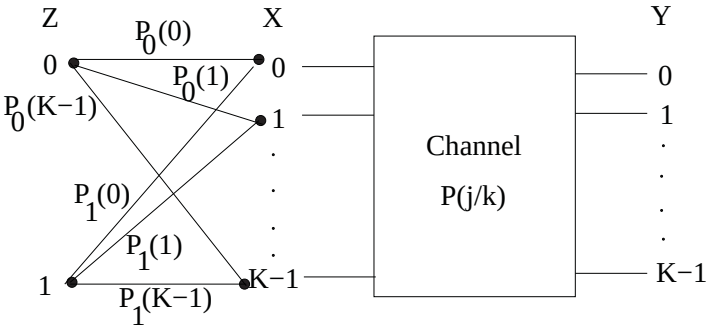


Figure: Figure for Problem 5

- Let $P_Z(0) = \theta$, $P_Z(1) = 1 - \theta$, and we define Z to be statistically independent of X.
- Use the above formulation to prove that the mutual information, $I(X;Y)$ is a convex function of $p(y/x)$ for fixed $p(x)$.

Problem # 5 (contd.)

- **Solutions:** Let $P_0(j/k)$ and $P_1(j/k)$, $0 \leq k \leq K-1$, $0 \leq j \leq J-1$, be two arbitrary sets of transition probabilities, and let $P(j/k) = \theta P_0(j/k) + (1-\theta)P_1(j/k)$. for an arbitrary θ, $0 \leq \theta < 1$.

Problem # 5 (contd.)

- **Solutions:** Let $P_0(j/k)$ and $P_1(j/k)$, $0 \leq k \leq K-1$, $0 \leq j \leq J-1$, be two arbitrary sets of transition probabilities, and let $P(j/k) = \theta P_0(j/k) + (1-\theta)P_1(j/k)$. for an arbitrary θ, $0 \leq \theta < 1$.
- Let I_0, I_1, and I be the average mutual informations for these sets of transition probabilities, then we need to prove that

$$\theta I_0 + (1-\theta)I_1 \geq I$$

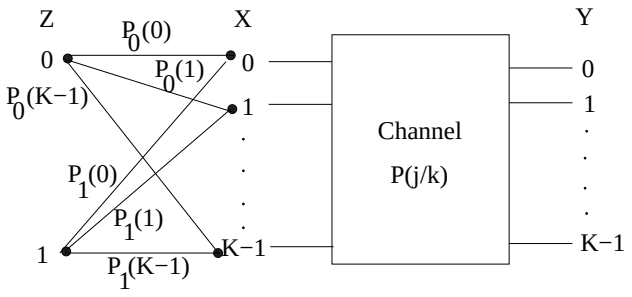
Problem # 5 (contd.)

- **Solutions:** Let $P_0(j/k)$ and $P_1(j/k)$, $0 \leq k \leq K-1$, $0 \leq j \leq J-1$, be two arbitrary sets of transition probabilities, and let $P(j/k) = \theta P_0(j/k) + (1-\theta)P_1(j/k)$. for an arbitrary θ, $0 \leq \theta < 1$.
- Let I_0, I_1, and I be the average mutual informations for these sets of transition probabilities, then we need to prove that

$$\theta I_0 + (1-\theta)I_1 \geq I$$

- We consider P_0 and P_1 as conditional on a binary variable Z

$$P_0(j/k) = P_{Y/XZ}(j|k,0) \quad P_1(j/k) = P_{Y/XZ}(j|k,1)$$



Problem # 5 (contd.)

- Let $P_Z(0) = \theta$, $P_Z(1) = 1 - \theta$, and defining Z to be statistically independent of X, then equation

$$\theta I_0 + (1 - \theta) I_1 \geq I$$

becomes

$$I(X; Y|Z) \geq I(X; Y)$$

Problem # 5 (contd.)

- Let $P_Z(0) = \theta$, $P_Z(1) = 1 - \theta$, and defining Z to be statistically independent of X, then equation

$$\theta I_0 + (1 - \theta) I_1 \geq I$$

becomes

$$I(X; Y|Z) \geq I(X; Y)$$

- By chain rule, we know that

$$\begin{aligned} I(X; YZ) &= I(X; Z) + I(X; Y|Z) \\ &= I(X; Y) + I(X; Z|Y) \end{aligned}$$

Problem # 5 (contd.)

- Let $P_Z(0) = \theta$, $P_Z(1) = 1 - \theta$, and defining Z to be statistically independent of X, then equation

$$\theta I_0 + (1 - \theta) I_1 \geq I$$

becomes

$$I(X; Y|Z) \geq I(X; Y)$$

- By chain rule, we know that

$$\begin{aligned} I(X; YZ) &= I(X; Z) + I(X; Y|Z) \\ &= I(X; Y) + I(X; Z|Y) \end{aligned}$$

- Since X and Z are statistically independent, we have $I((X; Z) = 0$ yielding

$$\begin{aligned} I(X; Y|Z) &= I(X; Y) + I(Z; X|Y) \\ I(X; Y|Z) &\geq I(X; Y) \end{aligned}$$

Uniquely Decodable Codes

- **Problem # 6:** Consider a source with source alphabet $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ in which the symbol probabilities are as follows:

$$p_1 = 0.27, p_2 = 0.09, p_3 = 0.23, p_4 = 0.11, p_5 = 0.15, p_6 = 0.15$$

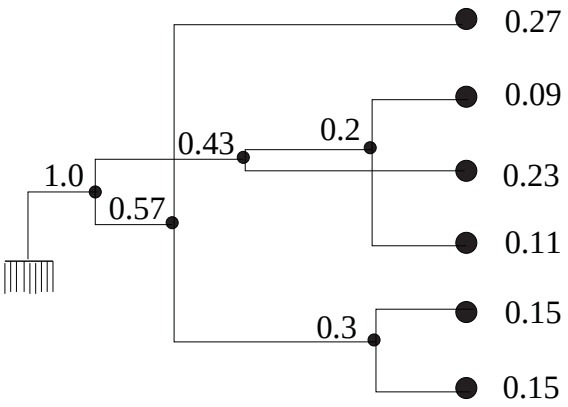
Find an optimal uniquely decodable code for this source that is not a Huffman code.

Uniquely Decodable Codes

- **Problem # 6:** Consider a source with source alphabet $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ in which the symbol probabilities are as follows:

  $$p_1 = 0.27, p_2 = 0.09, p_3 = 0.23, p_4 = 0.11, p_5 = 0.15, p_6 = 0.15$$

  Find an optimal uniquely decodable code for this source that is not a Huffman code.
- **Solutions:** We will first create a Huffman code for the source distribution.



Problem #6 (contd.)

- Huffman code can thus be written as

Source alphabet probability	Codeword
0.27	10
0.09	010
0.23	00
0.11	011
0.15	110
0.15	111

Problem #6 (contd.)

- Huffman code can thus be written as

Source alphabet probability	Codeword
0.27	10
0.09	010
0.23	00
0.11	011
0.15	110
0.15	111

- In Huffman code, two least likely codewords differ at the last bit. If we interchange the codeword for one of the source alphabet with probability 0.15 with the codeword for source alphabet with probability 0.11, we still get a uniquely decodable code, but it is not a Huffman code.

Problem #6 (contd.)

- Thus one example of uniquely decodable code that is not Huffman code for this source distribution is given by

Source alphabet probability	Codeword
0.27	10
0.09	010
0.23	00
0.11	110
0.15	011
0.15	111

Huffman Coding

- **Problem # 7:** Prove that for any binary Huffman code, if the most probable message symbol has probability $p_1 > 2/5$ and it is the only message symbol with this probability, then that symbol must be assigned a codeword of length 1.

Huffman Coding

- **Problem # 7:** Prove that for any binary Huffman code, if the most probable message symbol has probability $p_1 > 2/5$ and it is the only message symbol with this probability, then that symbol must be assigned a codeword of length 1.
- **Solutions:** Let's assume that the codeword length of the most probable message symbol is larger than one.

Huffman Coding

- **Problem # 7:** Prove that for any binary Huffman code, if the most probable message symbol has probability $p_1 > 2/5$ and it is the only message symbol with this probability, then that symbol must be assigned a codeword of length 1.
- **Solutions:** Let's assume that the codeword length of the most probable message symbol is larger than one.
- Then, since the most probable codeword must have the shortest codeword length, the codeword tree can be reduced to

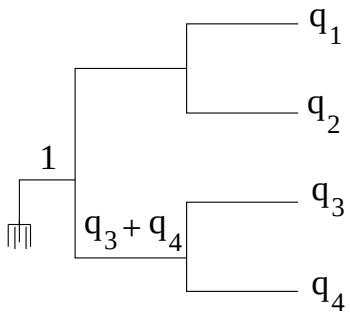


Figure: Huffman code tree

Problem # 7

- Furthermore, we know that $q_1 + q_2 + q_3 + q_4 = 1$. We consider two cases

Problem # 7

- Furthermore, we know that $q_1 + q_2 + q_3 + q_4 = 1$. We consider two cases

Case 1: $q_3, q_4 \geq q_1, q_2$. Since $q_1 > 2/5$, q_3, q_4 are also $> 2/5$, and $q_1 + q_2 + q_3 + q_4 > 6/5 > 1$, which is impossible.

Problem # 7

- Furthermore, we know that $q_1 + q_2 + q_3 + q_4 = 1$. We consider two cases

Case 1: $q_3, q_4 \geq q_1, q_2$. Since $q_1 > 2/5$, q_3, q_4 are also $> 2/5$, and $q_1 + q_2 + q_3 + q_4 > 6/5 > 1$, which is impossible.

Case 2: $q_1, q_2 \geq q_3, q_4$. We also have $q_3 + q_4 \geq q_1 > 2/5$. Since, $q_1 + q_2 + q_3 + q_4 = 1$ or $1 \geq 2q_1 + q_2 = 4/5 + q_2$. So, we get $q_2 < 1/5$. Since $q_1, q_2 \geq q_3, q_4$, this implies $q_3 + q_4 < 2q_2 < 2/5$. As $q_3 + q_4 \geq q_1$, this would imply $q_1 < 2/5$, which contradicts the condition given above ($q_1 > 2/5$). Hence this is also not possible.

Problem # 7

- Furthermore, we know that $q_1 + q_2 + q_3 + q_4 = 1$. We consider two cases

Case 1: $q_3, q_4 \geq q_1, q_2$. Since $q_1 > 2/5$, q_3, q_4 are also $> 2/5$, and $q_1 + q_2 + q_3 + q_4 > 6/5 > 1$, which is impossible.

Case 2: $q_1, q_2 \geq q_3, q_4$. We also have $q_3 + q_4 \geq q_1 > 2/5$. Since, $q_1 + q_2 + q_3 + q_4 = 1$ or $1 \geq 2q_1 + q_2 = 4/5 + q_2$. So, we get $q_2 < 1/5$. Since $q_1, q_2 \geq q_3, q_4$, this implies $q_3 + q_4 < 2q_2 < 2/5$. As $q_3 + q_4 \geq q_1$, this would imply $q_1 < 2/5$, which contradicts the condition given above ($q_1 > 2/5$). Hence this is also not possible.

- Thus the original assumption is false and the codeword length of the most probable message must be one.

Rate Distortion Theory

- **Problem # 8:** Let $X \sim N(0, \sigma^2)$ and let the distortion measure be squared error. Show that the optimum reproduction points for 1 bit quantization are $\pm\sqrt{\frac{2}{\pi}}\sigma$, and that the expected distortion for 1 bit quantization is $\frac{\pi-2}{\pi}\sigma^2$.

Rate Distortion Theory

- **Problem # 8:** Let $X \sim N(0, \sigma^2)$ and let the distortion measure be squared error. Show that the optimum reproduction points for 1 bit quantization are $\pm\sqrt{\frac{2}{\pi}}\sigma$, and that the expected distortion for 1 bit quantization is $\frac{\pi-2}{\pi}\sigma^2$.
- **Solutions:** Let X $\sim \mathcal{N}(0, \sigma^2)$ and let the distortion measure be squared error.

Rate Distortion Theory

- **Problem # 8:** Let $X \sim N(0, \sigma^2)$ and let the distortion measure be squared error. Show that the optimum reproduction points for 1 bit quantization are $\pm\sqrt{\frac{2}{\pi}}\sigma$, and that the expected distortion for 1 bit quantization is $\frac{\pi-2}{\pi}\sigma^2$.
- **Solutions:** Let X $\sim \mathcal{N}(0, \sigma^2)$ and let the distortion measure be squared error.
- With one bit quantization, the obvious reconstruction regions are positive and negative real axes.

Rate Distortion Theory

- The reconstruction point is the centroid of each region. for example, for the positive real line, the centroid a is

$$\begin{aligned} a &= \int_0^\infty x \frac{2}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \int_0^\infty \sigma\sqrt{\frac{2}{\pi}} e^{-y} dy = \sigma\sqrt{\frac{2}{\pi}}, \end{aligned}$$

using the substitution $y = x^2/2\sigma^2$.

Rate Distortion Theory

- The expected distortion for one bit quantization is

$$\begin{aligned} D &= \int_{-\infty}^0 \left(x + \sigma\sqrt{\frac{2}{\pi}}\right)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx + \int_0^\infty \left(x - \sigma\sqrt{\frac{2}{\pi}}\right)^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= 2\int_{-\infty}^\infty \left(x^2 + \sigma^2\frac{2}{\pi}\right) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx - 2\int_0^\infty \left(-2x\sigma\sqrt{\frac{2}{\pi}}\right) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \sigma^2 + \frac{2}{\pi}\sigma^2 - 4\frac{1}{\sqrt{2\pi}}\sigma^2\sqrt{\frac{2}{\pi}} \\ &= \sigma^2\frac{\pi - 2}{\pi}. \end{aligned}$$