An introduction to Information Theory

Adrish Banerjee

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

July 25, 2016

Lecture #3B: Block to variable length coding-II: Bounds on optimal codelength

Outline of the lecture

- Coding a single random variable

Coding a single random variable Rooted tree with probabilities Shannon-Fano coding

Outline of the lecture

- Coding a single random variable
- Rooted tree with probabilities

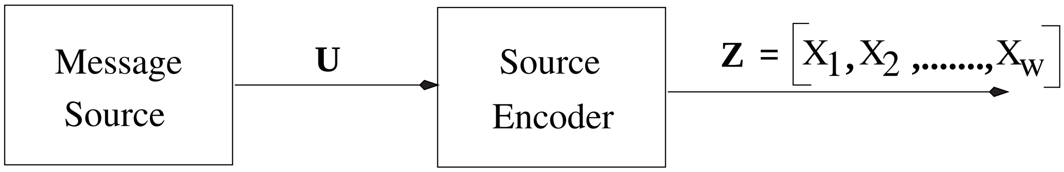
Adrish Banerjee Department of Electrical Engineering Indian Institute of Technology Kanpur Kanpur, Uttar Pradesh India
An introduction to Information Theory

Outline of the lecture

- Coding a single random variable
- Rooted tree with probabilities
- Shannon-Fano coding

Outline

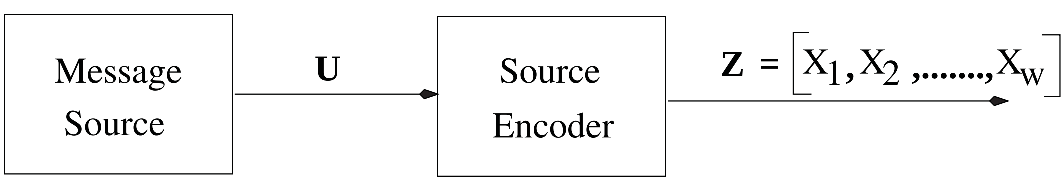
Coding a single random variable



Variable length coding scheme

- **U** is a $K-$ary random variable.

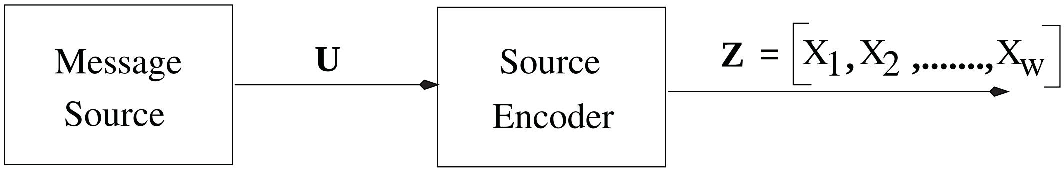
Coding a single random variable



Variable length coding scheme

- **U** is a $K-$ary random variable.
- X_i takes on values in the $D-$ary alphabet.

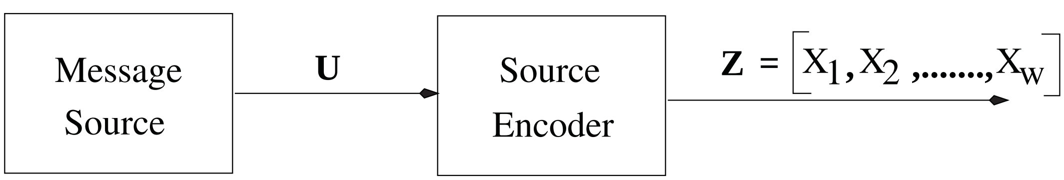
Coding a single random variable



Variable length coding scheme

- **U** is a $K-$ary random variable.
- X_i takes on values in the $D-$ary alphabet.
- **W** is a random variable, i.e. **Z** is variable length.

Coding a single random variable



Variable length coding scheme

- **U** is a $K-$ary random variable.
- X_i takes on values in the $D-$ary alphabet.
- **W** is a random variable, i.e. **Z** is variable length.
- A list $(z_1, z_2, \cdots, z_K)$ of D-ary sequences is a codeword of $U = [u_1, u_2, \cdots, u_K]$.

Coding a single random variable

- If $\mathbf{z}_i = [x_{i1}, x_{i2}, \cdots, x_{iw_i}]$ is the codeword for u_i, and w_i is the length of this codeword, then average codeword length is defined as

$$E[W] = \sum_{i=1}^{K} w_i P_U(u_i)$$

Coding a single random variable Rooted tree with probabilities Shannon-Fano coding

Coding a single random variable

- If $\mathbf{z}_i = [x_{i1}, x_{i2}, \cdots, x_{iw_i}]$ is the codeword for u_i, and w_i is the length of this codeword, then average codeword length is defined as

$$E[W] = \sum_{i=1}^{K} w_i P_U(u_i)$$

- Smallness of average codeword length is a measure of goodness of the code.

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

Outline

Rooted tree with probabilities

- By rooted tree with probabilities, we mean a finite rooted tree with probabilities assigned to each vertex such that

Rooted tree with probabilities

- By rooted tree with probabilities, we mean a finite rooted tree with probabilities assigned to each vertex such that
  (1) the root is assigned probability 1, and

Rooted tree with probabilities

- By rooted tree with probabilities, we mean a finite rooted tree with probabilities assigned to each vertex such that
  (1) the root is assigned probability 1, and
  (2) the probability of every node is the sum of the probabilities of the nodes and leaves at depth 1 in the subtree stemming from this intermediate node.

Rooted tree with probabilities

- By rooted tree with probabilities, we mean a finite rooted tree with probabilities assigned to each vertex such that
  (1) the root is assigned probability 1, and
  (2) the probability of every node is the sum of the probabilities of the nodes and leaves at depth 1 in the subtree stemming from this intermediate node.
- Path length lemma: In a rooted tree with probabilities, the average depth of the leaves is equal to the sum of the probabilities of the nodes (including the root).

Rooted tree with probabilities

Sketch of proof:

- The probability of each node is the sum of the probabilities of the leaves in the subtree stemming from that node.

Rooted tree with probabilities

Sketch of proof:

- The probability of each node is the sum of the probabilities of the leaves in the subtree stemming from that node.
- A leaf at depth d appears in d nodes on the path from the root to the leaf.

Coding a single random variable Rooted tree with probabilities Shannon-Fano coding

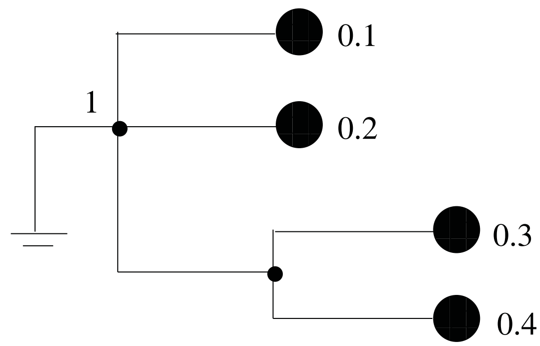
Rooted tree with probabilities

Sketch of proof:

- The probability of each node is the sum of the probabilities of the leaves in the subtree stemming from that node.
- A leaf at depth d appears in d nodes on the path from the root to the leaf.
- The sum of probabilities of the nodes equals the sum of the products of each leaf's probability and its depth, which is the average depth of the leaves.

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

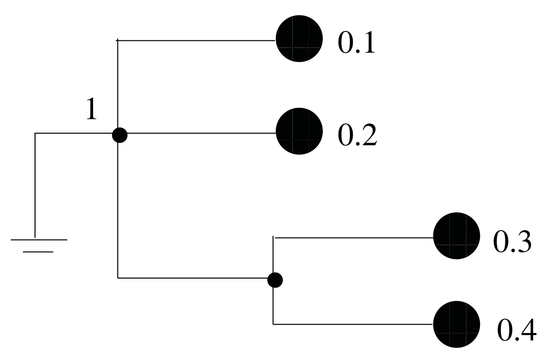
Rooted tree with probabilities



- By Path Length Lemma, average depth of the leaves is $1 + 0.7 = 1.7$.

Coding a single random variable Rooted tree with probabilities Shannon-Fano coding

Rooted tree with probabilities



- By Path Length Lemma, average depth of the leaves is $1 + 0.7 = 1.7$.
- As a check, average depth of the leaves is $1(0.1) + 1(0.2) + 2(0.3) + 2(0.4) = 1.7$.

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

Coding a single random variable

- Leaf entropy: Rooted tree with T leaves whose probabilities are $p_1, p_2, \cdots, p_T$, then

$$H_{\text{leaf}} = -\sum_{i:p_i \neq 0} p_i \log p_i$$

Coding a single random variable | Rooted tree with probabilities | Shannon-Fano coding

Coding a single random variable

- Leaf entropy: Rooted tree with T leaves whose probabilities are $p_1, p_2, \cdots, p_T$, then

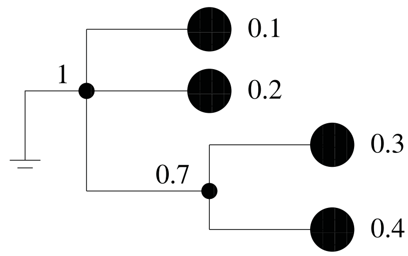
$$H_{\text{leaf}} = -\sum_{i:p_i \neq 0} p_i \log p_i$$

- Branching entropy: Suppose that $q_{i1}, q_{i2}, \cdots, q_{iL}$ are the probabilities of the nodes and leaves at the ends of the L_i branches stemming outward from the node whose probability is P_i. Then the branching entropy H_i at this node is given by

$$H_i = -\sum_{j:q_{ij} \neq 0} \frac{q_{ij}}{P_i} \log \frac{q_{ij}}{P_i}$$

where $\frac{q_{ij}}{P_i}$ is the conditional probability of choosing the j^{th} of these branches given we are at the given node.

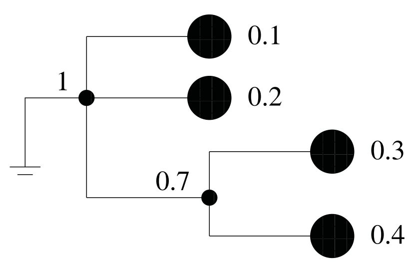
Coding a single random variable



- Leaf entropy,

$$H_{\text{leaf}} = -\sum_{i=1}^{4} p_i \log p_i = 1.846 \text{ bits}$$

Coding a single random variable



- Leaf entropy,

$$H_{\text{leaf}} = -\sum_{i=1}^{4} p_i \log p_i = 1.846 \text{ bits}$$

- Branching entropy

$$\begin{aligned} H_1 &= -.1 \log .1 - .2 \log .2 - .7 \log .7 = 1.157 \text{ bits} \\ H_2 &= -\frac{3}{7} \log \frac{3}{7} - \frac{4}{7} \log \frac{4}{7} = .985 \text{ bits} \end{aligned}$$

Coding a single random variable

- Leaf entropy theorem: The leaf entropy of a rooted tree with probabilities equals to the sum over all the nodes (including the root) of the branching entropy of that node weighted by the node probabilitity, i.e.

$$H_{\text{leaf}} = \sum_{i=1}^{N} P_i H_i$$

Sketch of Proof:

Coding a single random variable

- Leaf entropy theorem: The leaf entropy of a rooted tree with probabilities equals to the sum over all the nodes (including the root) of the branching entropy of that node weighted by the node probabilitity, i.e.

$$H_{\text{leaf}} = \sum_{i=1}^{N} P_i H_i$$

Sketch of Proof:

- By definition of rooted tree with probabilities

$$P_i = \sum_{j=1}^{L_i} q_{ij}$$

Coding a single random variable

- Leaf entropy theorem: The leaf entropy of a rooted tree with probabilities equals to the sum over all the nodes (including the root) of the branching entropy of that node weighted by the node probabilitity, i.e.

$$H_{\text{leaf}} = \sum_{i=1}^{N} P_i H_i$$

Sketch of Proof:

- By definition of rooted tree with probabilities

$$P_i = \sum_{j=1}^{L_i} q_{ij}$$

- Using the result $\log(q_{ij}/P_i) = \log q_{ij} - \log P_i$ in the definition of branching entropy, H_i, we get

$$P_i H_i = - \sum_{j:q_{ij} \neq 0} q_{ij} \log q_{ij} + P_i \log P_i$$

Coding a single random variable

Sketch of Proof (contd.):

- A non-root k-th node will contribute $+P_k \log P_k$ to the sum (corresponding to $i = k$) and will contribute $-P_k \log P_k$ to the term for i such that node k is at the end of branch leaving node i. Hence net contribution to the sum is zero.

Coding a single random variable

Sketch of Proof (contd.):

- A non-root k-th node will contribute $+P_k \log P_k$ to the sum (corresponding to $i = k$) and will contribute $-P_k \log P_k$ to the term for i such that node k is at the end of branch leaving node i. Hence net contribution to the sum is zero.
- The root will contribute the term $P_1 \log P_1$ to the sum and that's zero as $P_1 = 1$.

Coding a single random variable Rooted tree with probabilities Shannon-Fano coding

Coding a single random variable

Sketch of Proof (contd.):

- A non-root k-th node will contribute $+P_k \log P_k$ to the sum (corresponding to $i = k$) and will contribute $-P_k \log P_k$ to the term for i such that node k is at the end of branch leaving node i. Hence net contribution to the sum is zero.
- The root will contribute the term $P_1 \log P_1$ to the sum and that's zero as $P_1 = 1$.
- The k-th leaf will contribute $-p_k \log p_k$ to the sum (corresponding to $i : q_{ij} = p_k$).

Coding a single random variable

Sketch of Proof (contd.):

- A non-root k-th node will contribute $+P_k \log P_k$ to the sum (corresponding to $i = k$) and will contribute $-P_k \log P_k$ to the term for i such that node k is at the end of branch leaving node i. Hence net contribution to the sum is zero.
- The root will contribute the term $P_1 \log P_1$ to the sum and that's zero as $P_1 = 1$.
- The k-th leaf will contribute $-p_k \log p_k$ to the sum (corresponding to $i : q_{ij} = p_k$).
- Hence,

$$\sum_{i=1}^{N} P_i H_i = -\sum_{k=1}^{T} p_k \log p_k$$

Coding a single random variable

Lower bound on E[W]

- Leaf entropy $H_{\text{leaf}} = H(U)$, and $H_i \leq \log D$.

Coding a single random variable

Lower bound on E[W]

- Leaf entropy $H_{\text{leaf}} = H(U)$, and $H_i \leq \log D$.
- Using leaf entropy theorem and above results, we get

$$H(U) \leq \log D \sum_{i=1}^{N} P_i$$

Coding a single random variable

Lower bound on E[W]

- Leaf entropy $H_{\text{leaf}} = H(U)$, and $H_i \leq \log D$.
- Using leaf entropy theorem and above results, we get

$$H(U) \leq \log D \sum_{i=1}^{N} P_i$$

- By path length lemma we know

$$E[W] = \sum_{i=1}^{N} P_i$$

Coding a single random variable

Lower bound on E[W]

- Leaf entropy $H_{\text{leaf}} = H(U)$, and $H_i \le \log D$.
- Using leaf entropy theorem and above results, we get

$$H(U) \le \log D \sum_{i=1}^{N} P_i$$

- By path length lemma we know

$$E[W] = \sum_{i=1}^{N} P_i$$

- Hence,

$$E[W] \ge \frac{H(U)}{\log D}$$

Outline

Shannon-Fano prefix-free code

- The length, w_i of the codeword u_i is chosen as

$$w_i = \lceil -\log_D P_U(u_i) \rceil$$

Shannon-Fano prefix-free code

- The length, w_i of the codeword u_i is chosen as

$$w_i = \lceil -\log_D P_U(u_i) \rceil$$

- Kraft's inequality is satisfied for this choice of w_i.

$$\begin{aligned} \sum_{i=1}^{K} D^{-w_i} &\leq \sum_{i=1}^{K} D^{\log_D P_U(u_i)} \\ &= \sum_{i=1}^{K} P_U(u_i) = 1 \end{aligned}$$

Shannon-Fano prefix-free code

- The length, w_i of the codeword u_i is chosen as

$$w_i = \lceil -\log_D P_U(u_i) \rceil$$

- Kraft's inequality is satisfied for this choice of w_i.

$$\begin{aligned} \sum_{i=1}^{K} D^{-w_i} &\leq \sum_{i=1}^{K} D^{\log_D P_U(u_i)} \\ &= \sum_{i=1}^{K} P_U(u_i) = 1 \end{aligned}$$

- This ensures that a D-ary prefix-free code exists for this choice of w_i.

Shannon-Fano prefix-free code

- Using the relation

$$x \leq \lceil x \rceil < x + 1$$

we get

$$w_i < \frac{-\log P_U(u_i)}{\log D} + 1$$

Shannon-Fano prefix-free code

- Using the relation
$$x \leq \lceil x \rceil < x + 1$$
we get
$$w_i < \frac{-\log P_U(u_i)}{\log D} + 1$$
- Multiplying the above equation by $P_U(u_i)$ and summing over i, we get
$$E[W] < \frac{H(U)}{\log D} + 1$$

Coding theorem for a K-ary Random variable

- The average codeword length of an optimum D-ary prefix-free code for a K-ary random variable U satisfies
$$\frac{H(U)}{\log D} \leq E[W] < \frac{H(U)}{\log D} + 1$$
with equality on the left if and only if the probability of each value of U is some negative integer power of D.

Coding theorem for a K-ary Random variable

- The average codeword length of an optimum D-ary prefix-free code for a K-ary random variable U satisfies

$$\frac{H(U)}{\log D} \leq E[W] < \frac{H(U)}{\log D} + 1$$

  with equality on the left if and only if the probability of each value of U is some negative integer power of D.
- E[W] for Shannon-Fano coding also satisfies the above inequality.

Shannon-Fano prefix-free code

- Consider binary Shannon-Fano prefix-free coding for 4-ary random variable U for which $P_U(u_i)$ equals $0.4, 0.3, 0.2$ and 0.1 for i equal to $1, 2, 3, 4$ respectively.

Shannon-Fano prefix-free code

- Consider binary Shannon-Fano prefix-free coding for 4-ary random variable U for which $P_U(u_i)$ equals $0.4, 0.3, 0.2$ and 0.1 for i equal to $1, 2, 3, 4$ respectively.
- $w_1 = \lceil \log_2 \frac{1}{0.4} \rceil = 2, w_2 = \lceil \log_2 \frac{1}{0.3} \rceil = 2, w_3 = \lceil \log_2 \frac{1}{0.2} \rceil = 3, w_4 = \lceil \log_2 \frac{1}{0.1} \rceil = 4$

Shannon-Fano prefix-free code

- Consider binary Shannon-Fano prefix-free coding for 4-ary random variable U for which $P_U(u_i)$ equals $0.4, 0.3, 0.2$ and 0.1 for i equal to $1, 2, 3, 4$ respectively.
- $w_1 = \lceil \log_2 \frac{1}{0.4} \rceil = 2, w_2 = \lceil \log_2 \frac{1}{0.3} \rceil = 2, w_3 = \lceil \log_2 \frac{1}{0.2} \rceil = 3, w_4 = \lceil \log_2 \frac{1}{0.1} \rceil = 4$
- $H(U) = 1.846$ bits, and by path length lemma, $E[W] = 1 + 0.7 + 0.3 + 0.3 + 0.1 = 2.4$.