conditional and *marginal (individual)*, informations associated with the bivariate probability distribution (p_{ij}).

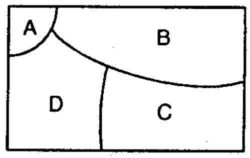

Fig. 26.6

We may define the marginal probability distributions of X and Y by

$$p_{i0} = \sum_{j=1}^{n} p_{ij} \text{ and } p_{0j} = \sum_{i=1}^{m} p_{ij} \text{ for all } i, j.$$

Then, obviously the *marginal entropies* of the two marginal distributions are given by

$$H(X) = -\sum_{i=1}^{m} p_{i0} \log p_{i0} \quad \text{and} \quad H(Y) = -\sum_{j=1}^{n} p_{0j} \log p_{0j}.$$

The entropy $H(X)$ measures the uncertainty of the message sent (irrespective of the message received) and $H(Y)$ performs the same role for the message received.

The *joint entropy* is the entropy of the joint distribution of the messages sent and received, and is therefore given by

$$H(X, Y) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log p_{ij}.$$

It may be observed that :

$$\text{Max } H(X, Y) = \log mn = \log m + \log n = \max H(X) + \max H(Y).$$

Theorem 26.2. *$H(X, Y) \le H(X) + H(Y)$ with equality, if and only if, X and Y are independent.*

Proof. We may write

$$H(X) + H(Y) = -\sum_{i=1}^{m} p_{i0} \log p_{i0} - \sum_{j=1}^{n} p_{0j} \log p_{0j}$$

$$= -\sum_{i=1}^{m}\left(\sum_{j=1}^{n} p_{ij}\right)\log p_{i0} - \sum_{j=1}^{n}\left(\sum_{i=1}^{m} p_{ij}\right)\log p_{0j}$$

$$= -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log (p_{i0}\, p_{0j}) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log q_{ij}, \text{ where } q_{ij} = p_{i0}\, p_{0j}. \quad \text{...(i)}$$

Also, by definition

$$H(X, Y) = -\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \log p_{ij}. \quad \text{...(ii)}$$

But, we observe that

$$\sum_{i=1}^{m}\sum_{j=1}^{n} q_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} p_{i0}\, p_{0j} = \left(\sum_{i=1}^{m} p_{i0}\right)\left(\sum_{j=1}^{n} p_{0j}\right) = 1 = \sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}$$

By virtue of *Theorem 26.1, (i)* and *(ii)*, it follows that

$$H(X, Y) \le H(X) + H(Y)$$

with equality, if and only if, $q_{ij} = p_{ij}$ for all i and j. The condition for equality reduces to $p_{i0}\, p_{0j} = p_{ij}$ meaning thereby X and Y are independent.

Q. 1. Show that $H(X)$ achieves its maximum if all the valuos of X are equi-probable. **[Delhi (OR) 92]**

2. Let X and Y be two discrete radom variables, each taking a finite number of values. Show that $H(Y \mid X) \le H(Y)$ with equality iff X and Y are independent

---

26.10–2. Conditional Entropies

Consider two finite discrete sample spaces S_1 and S_2 and their product space S. In S_1 and S_2, select complete sets of events in the sense of equations (26.17) and (26.18).

$$\{E\} = [E_1, E_2, \ldots, E_n] \quad \text{...(26.26)}$$

$$\{F\} = [F_1, F_2, \ldots, F_m] \quad \text{...(26.27)}$$

Thus, the complete set of events in the product space $S_1 \times S_2$ will be given by

$$\{EF\} = \begin{bmatrix} E_1F_1 & E_1F_2 & E_1F_3 & \dots & E_1F_m \\ E_2F_1 & E_2F_2 & E_2F_3 & \dots & E_2F_m \\ \vdots & \vdots & \vdots & \dots & \vdots \\ E_nF_1 & E_nF_2 & E_nF_3 & \dots & E_nF_m \end{bmatrix} \quad ...(26.28)$$

[By taking cartesian product of two sets $\{E\}$ and $\{F\}$]

For example, an event F_r may occur in conjunction with E_1, E_2, ..., or E_n.

$$F_r = \sum_{k=1}^{n} E_kE_r \quad ...(26.29)$$

$$P\{X = x_k \mid Y = y_r\} = \frac{P\{X = x_r \cap Y = y_r\}}{P\{Y = y_r\}} \quad ...(26.30)$$

or

$$p\{x_k \mid y_r\} = \frac{p\{k, r\}}{p\{y_r\}} \quad ...(26.31)$$

Now, consider the following probability scheme

$$\{E \mid F_r\} = [E_1 \mid E_r, E_2 \mid E_r, E_3 \mid E_r, \dots, E_n \mid E_r] \quad ...(26.32)$$

$$P\{E \mid F_r\} = \left[\frac{p\{1, r\}}{p\{y_r\}}, \frac{p\{2, r\}}{p\{y_r\}}, \frac{p\{3, r\}}{p\{y_r\}}, \dots, \frac{p\{n, r\}}{p\{y_r\}}\right] \quad ...(26.33)$$

The probability scheme is not only finite but also complete because the sum of elements of this matrix is unity. Therefore, an entropy will be given by

$$H\{X \mid y_r\} = -\sum_{k=1}^{n} \frac{p\{k, r\}}{p\{y_r\}} \log \frac{p\{k, r\}}{p\{y_r\}} \quad ...(26.34)$$

$$= -\sum_{k=1}^{n} p\{x_k \mid y_r\} \log p\{x_k \mid y_r\} \quad ...(26.35)$$

Now, take the average of this conditional entropy for all admissible values of y_r, so that a measure of average conditional entropy of the system can be obtained.

$$H\{X \mid Y\} = \overline{H\{X \mid y_r\}} = \sum_{r=1}^{m} p\{y_r\} \cdot H\{X \mid y_r\}$$

$$= -\sum_{r=1}^{m} p\{y_r\} \sum_{k=1}^{n} p\{x_k \mid y_r\} \log p\{x_k \mid y_r\} \quad ...(26.36)$$

$$H\{X \mid Y\} = -\sum_{r=1}^{m} \sum_{k=1}^{n} p\{y_r\} \cdot p\{x_k \mid y_r\} \log p\{x_k \mid y_r\} \quad ...(26.37)$$

Likewise, it is possible to otbain the expression for the average conditional entropy $H(Y \mid X)$, *i.e.*

$$H\{Y \mid X\} = -\sum_{k=1}^{n} \sum_{r=1}^{m} p\{x_k\} \cdot p\{y_r \mid x_k\} \log p\{y_r \mid x_k\} \quad ...(26.38)$$

Two conditional entropies may be expressed as

$$H\{X \mid Y\} = -\sum_{r=1}^{m} \sum_{k=1}^{n} p\{x_k, y_r\} \log p\{x_k \mid y_r\} \quad ...(26.39)$$

$$H\{Y \mid X\} = -\sum_{k=1}^{n} \sum_{r=1}^{m} p\{x_k, y_r\} \log p\{y_r \mid x_k\} \quad ...(26.40)$$

Note. It should be noted that all entropies are positive numbers as they are sum of positive numbers.

26.11. AN IMPORTANT THEOREM

Theorem 26.3. *Prove that*

$$H(X, Y) = H(X \mid Y) + H(Y) = H(Y \mid X) + H(X),$$

where $H(X) \geq H(X \mid Y)$. **[Delhi (O.R.) 92]**

Proof. By definition of entropy function H,

$$H(X) = -\sum_{k=1}^{n} p\{x_k\} \log p\{x_k\} \qquad ...(26.41)$$

$$H(Y) = -\sum_{r=1}^{m} p\{y_r\} \log p\{y_r\} \qquad ...(26.42)$$

The joint entropy function of X and Y will be given by

$$H\{X, Y\} = -\sum_{k=1}^{n} \sum_{r=1}^{m} p\{k, r\} \log p\{k, r\} \qquad ...(26.43)$$

where $p(k, r)$ is the joint probability for the occurrence of two events E_k and E_r simultaneously.

Rewriting expressions (26.39) and (26.40) for conditional probabilities, we have

$$H\{X \mid Y\} = -\sum_{r=1}^{m} \sum_{k=1}^{n} p\{x_k, y_r\} \log p\{x_k \mid y_r\} \qquad ...(26.39a)$$

$$H\{Y \mid X\} = -\sum_{k=1}^{n} \sum_{r=1}^{m} p\{x_k, y_r\} \log p\{y_r \mid x_k\} \qquad ...(26.40a)$$

Now, basic relationships among *marginal, joint* and *conditional* probabilities are :

$$p\{x_k, y_r\} = p\{x_k \mid y_r\}\, p\{y_r\} = p\{y_r \mid x_k\}\, p\{x_k\} \qquad ...(26.44)$$

$$\log p\{x_k, y_r\} = \log p\{x_k \mid y_r\} + \log p\{y_r\} = \log p\{y_r \mid x_k\} + \log p\{x_k\} \qquad ...(26.45)$$

(i) To prove $H(X, Y) = H(Y \mid X) + H(X)$ and $H(X, Y) = H(X \mid Y) + H(Y)$.

Directly substituting the relations (26.44) and (26.45) in defining equations (26.41), (26.42), (26.39a), (26.43) and (26.40a) of entropies, these two basic identities can be easily proved.

(ii) To prove $H(X) \geq H(X \mid Y)$.

For the proof of this inequality,

$$H(X \mid Y) - H(X) = \sum_{r=1}^{m} \sum_{k=1}^{n} p(x_k, y_r) \log \frac{p\{x_k\}}{p\{x_k \mid y_r\}}$$

$$\leq \sum_{r=1}^{m} \sum_{k=1}^{n} p\{x_k, y_r\} \left[\frac{p\{x_k\}}{p\{x_k \mid y_r\}} - 1 \right] \log e \qquad ...(26.46)$$

(since $\log x \leq (x-1) \log e$ by convexity of logarithmic function).

But, $$\sum_{r=1}^{m} \sum_{k=1}^{n} [p\{x_k\} \cdot p\{y_r\} - p\{x_k, y_r\}] \log e = \sum_{r=1}^{m} [p\{y_r\} - p\{y_r\}] \log e = 0$$

Hence, $H(X \mid Y) - H(X) \leq 0$ or $H(X) \geq H(X \mid Y)$.

Similarly, it can be proved that $H(Y) \geq H(Y \mid X)$.

Q. 1. Define conditional entropy. In the usual notation show that $H(Y) + H(X) \geq H(X, Y)$.

2. Show that average amount of information increases if one of the events is partitioned. **[Delhi (OR) 92]**

3. Let X and Y be two discrete random variables, each taking finite number of values. Prove that
$H(X, Y) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y)$ **[Delhi (OR) 92]**

Example 10. *A transmitter has an alphabet consisting of five letters* $\{x_1, x_2, x_3, x_4, x_5\}$ *and the receiver has an alphabet consisting of four letters* $\{y_1, y_2, y_3, y_4\}$. *The joint probabilities for the communication are given below :*

$$\begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{array} \begin{array}{c} \begin{array}{cccc} y_1 & y_2 & y_3 & y_4 \end{array} \\ \begin{bmatrix} 0.25 & 0.00 & 0.00 & 0.00 \\ 0.10 & 0.30 & 0.00 & 0.00 \\ 0.00 & 0.05 & 0.10 & 0.00 \\ 0.00 & 0.00 & 0.05 & 0.10 \\ 0.00 & 0.00 & 0.05 & 0.00 \end{bmatrix} \end{array}$$

Determine the Marginal, Conditional and Joint entropies for this channel (Assume 0 log 0 ≡ 0).

Solution. The channel is described here by the joint probabilities p_{ij}, $i = 1, 2, \ldots, 5$ and $j = 1, 2, \ldots, 4$. The conditional and marginal probabilities are easily obtained from p_{ij}'s as given below :

$$p_{10} = 0.25 + 0.00 = 0.25,\ p_{20} = 0.10 + 0.30 = 0.40,\ p_{30} = 0.05 + 0.10 = 0.15,$$
$$p_{40} = 0.05 + 0.10 = 0.15,\ p_{50} = 0.05 + 0.00 = 0.05.$$

Similarly, $p_{01} = 0.35$, $p_{02} = 0.35$, $p_{03} = 0.20$, $p_{04} = 0.10$.

By using the result $(p_{j\mid i} = p_{ij}/p_{i0})$, the conditional probabilities are given in the following channel matrix :

Conditional Prob. Matrix $(p_{j\mid i})$

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 \\ 0 & 1/3 & 2/3 & 0 \\ 0 & 0 & 1/3 & 2/3 \\ 0 & 0 & 1 & 0 \end{bmatrix} \end{array}$$

Now various entropies associated with this channel are obtained (taking all logarithms to the base 2).

Marginal Entropies :

$$H(X) = -\sum_{i=1}^{5} p_{i0} \log p_{i0}$$
$$= -[(.25)\log(.25) + (.40)\log(.40) + 2(.15)\log(.15) + (.05)\log(.05)]$$
$$= \frac{1}{4}\log 4 + \frac{2}{5}\log\frac{5}{2} + \frac{3}{10}\log\frac{20}{3} + \frac{1}{20}\log 20 = 1.3260 \text{ bits.}$$

$$H(Y) = -\sum_{j=1}^{4} p_{0j}\log p_{0j} = -[2(.35)\log(.35) + (.20)\log(.20) + (.10)\log(.10)]$$
$$= \frac{7}{10}\log(2.857) + \frac{1}{5}\log 5 + \frac{1}{10}\log 10$$
$$= 1.8556 \text{ bits}$$

Conditional Entropies :

$$H(Y\mid X) = -\sum_{i=1}^{5}\sum_{j=1}^{4} p_{ij}\log p_{j\mid i}$$
$$= (.25)\log 1 + (.10)\log 4 + (.30)\log(4/3) + (.05)\log 3$$
$$+ (.10)\log(3/2) + (.05)\log 3 + (.10)\log(3/2) + (.05)\log 1$$
$$= \frac{1}{10}\log 4 + \frac{3}{10}\log 4 - \frac{3}{10}\log 3 + \frac{1}{10}\log 3 + \frac{1}{5}\log 3 - \frac{1}{5}\log 2$$
$$= 1/10\,[4\log 4 - 2\log 2] = 6/10 = 0.60 \text{ bits.}$$

By *Theorem* 26.3, we have

$$H(X\mid Y) = H(X) + H(Y\mid X) - H(Y) = 1.3260 + 0.6000 - 1.8556 = 0.0704 \text{ bits.}$$

Joint Entropies *:*

$$H(X, Y) = H(X) + H(Y\mid X) = 1.3260 + 0.6000 = 1.9260 \text{ bits.}$$

26.12. SET OF AXIOMS FOR AN ENTROPY FUNCTION

Assume the following four conditions as axioms :