A SOURCE CODING THEOREM FOR DOUBLE DISTORTION MEASURE

by Y. D. MATHUR, *Department of Mathematics, University of Delhi, Delhi 110007*

(Communicated by F. C. Auluck, F.N.A.)

(*Received 25 September 1974*)

In most of communication models, the coding is done on the basis of an input distribution, but in practice, the input distribution is usually different from the ideal one. From this point of view Sharma and Mathur (*in press*) have earlier defined the rate distortion function under true and inaccurate input distribution and have studied communication under true and inaccurate input probability schemes. In this paper a source coding theorem is proved that induces the rate distortion function defined on the basis of the two probability schemes, with its operational significance.

1. INTRODUCTION

In most of communication models, the coding is done on the basis of an input distribution, but in practice the input distribution is mostly different from the ideal one. From this point of view Sharma and Mathur (*in press*) have earlier studied communication under true and inaccurate input probability distributions, which we shall distinguish from the ideal one by adding the word "practical".

Given a source output $X = \{X_t\}$ where X_t ; $t = 0, \pm 1, \pm 2, \ldots,$ are a family of discrete random variables assuming values only in a finite alphabet A_M, we associate with it a true input distribution $\{\theta\}$ and an inaccurate distribution $\{\alpha\}$. For other notations, we follow Berger (1971). Thus let $\{X_t, \theta, \alpha\}$ be a practical discrete memoryless source. The output is encoded into a code $B = \{\underset{\sim}{y}_1\, \underset{\sim}{y}_2 \ldots . \underset{\sim}{y}_k\}$ of size K and block length n, using a distortion measure $\rho_n(\underset{\sim}{x}\,;\underset{\sim}{y})$ in such a way that source word $\underset{\sim}{x}$ is encoded in that code word $\underset{\sim}{y} \in B$ which minimizes $\rho_n(\underset{\sim}{x}\,;\underset{\sim}{y})$. As in Berger (1971)

$$\rho_n(\underset{\sim}{x}/B) = \min_{\underset{\sim}{y} \in B} \rho_n(\underset{\sim}{x}\,;\underset{\sim}{y}) \qquad \ldots(1.1)$$

is the resulting minimum of $\underset{\sim}{x}$ for the code B. However, there would be two average distortions of B given by

$$\rho_\theta(B) = \sum_{\underset{\sim}{x}} \theta(\underset{\sim}{x})\, \rho_n(\underset{\sim}{x}/B) \qquad \ldots(1.2)$$

and

$$\rho_\alpha(B) = \sum_{\underset{\sim}{x}} \alpha(x)\, \rho_n(\underset{\sim}{x}/B). \qquad \ldots(1.3)$$

Our aim is to prove a source coding theorem for communication under true and inaccurate input probability schemes.

The generalized rate of information is given by Sharma and Mathur (*in press*) and Mathur (1974)

$$R = I(X\,;\,Y:\theta/\alpha) = \sum_{i,j} p_i \,.\, q_{j|i} \,.\, \log \frac{q_{j|i}}{\sum_k \alpha_k\, q_{j|k}}\,. \qquad ...(1.4)$$

The source coding theorem would be proved by random coding argument and the source codes would be chosen according to some joint probability distribution.

$$Q(B) = Q(y_1, y_2, y_3, \ldots y_k) \qquad ...(1.5)$$

The overall distortions associated with the code B are given by :

$$\bar{\rho}_\theta = \sum_{a\,\|\,\underline{x}} \theta(\underline{x}) \sum_{a\,\|\,B} Q(B)\,\rho_n(\underline{x}/B) \qquad ...(1.6)$$

$$= \sum_{a\,\|\,\underline{x}} \alpha(\underline{x}) \sum_{a\,\|\,B} Q(B)\,.\,\rho_n(\underline{x}/B). \qquad ...(1.7)$$

The role of Shannon's entropy is played in this practical situation by Kerridge's (1961) inaccuracy which, with θ, α scheme, may be written as

$$H(X\,;\,\theta/\alpha) = -\sum_{\underline{x}} \theta(\underline{x}) \log \alpha(\underline{x}). \qquad ...(1.8)$$

This, unlike entropy may become infinite, however, for a source code of size K and length n ; the rate as usual would be taken as

$$R = \frac{1}{n} \log K. \qquad ...(1.9)$$

Source code B would be called to be $(D_1^*\,;\,D_2^*)$-admissible if $\rho_\theta(B) \leqslant D_1^*$ and $\rho_\alpha(B) \leqslant D_2^*$. We shall denote the smallest size of $(D_1^*\,;\,D''^*)$-admissible code by $K(n:\,D_1^*\,;D_2^*)$ and the generalised rate-distortion function by

$$R(D_1^*\,;D_2^*) = \text{Inf.}\;\; I(X\,;\;Y:\;\theta/\alpha) \qquad ...(1.10)$$

where infimum is taken over all admissible transition probabilities $\{q_{j|i}\}$.

2. Source Coding Theorem

Let the practical discrete memoryless source $\{X_t\,;\;\theta\,;\;\alpha\}$ and the single letter fidelity criterion $\rho(.,.)$ be given and let $R(.,.)$ denote the rate distortion function of $\{X_t\,;\,\theta\,;\;\alpha\}$. Then given any $\varepsilon > 0$ and any $D_1^* \geqslant 0$ and $D_2^* \geqslant 0$, an integer n can be found such that there exists a $(D_1^* + \varepsilon\,;\;D_2^* + \varepsilon)$-admissible code of block length n with rate $R < R\,(D_1^*\,;\;D_2^*) + \varepsilon$.

In other words, for sufficiently large n, we have

$$n^{-1} \log K(n\,;\;D_1^* + \varepsilon\,; D_2^* + \varepsilon) < R\,(D_1^*\,;\,D_2^*) + \varepsilon \qquad ...(2.1)$$

Proof : The proof is a modification of the one given by Berger (1971) for Shannon's rate and is based on a random coding argument. We construct the desired code ensemble by independent word selection, that is, code B is randomly selected with probability

$$Q(B) = \prod_{k=1}^{k} Q^{\alpha}(y_k) \qquad ...(2.2)$$

where $Q^{\alpha}(y_k)$ is the probability of y_k when input is governed by the distribution α. To each x, we assign a set of reproducing vectors

$$S(x) = \{y : \rho_n(x ; y) \leqslant D_1^* + \varepsilon$$

and

$$\rho_n(x ; y) \leqslant D_2^* + \delta\} \qquad ...(2.3)$$

where δ is a small positive number.

If $Q(S(x))$ is the probability that a code word chosen at random belongs to $S(x)$; then

$$Q(S(x)) = \sum_{y \in s(x)} Q(y) ; \qquad ...(2.4)$$

so that the probability that no words of a randomly chosen code of size K belongs to $S(x)$ is

$$Q^*(x) = [1 - Q(S(x))]^k. \qquad ...(2.5)$$

If

$$\rho_{\max} = \max_{j,k} \rho_{jk} \qquad ...(2.6)$$

then $\rho_{\max}$ denotes the largest possible value of $\rho_n(x ; y)$ of our code ensemble.

As is shown above, for fixed x the probability that

$$\rho_n(x/B) \leqslant D_1^* + \delta$$

and

$$\rho_n(x/B) \leqslant D_2^* + \delta$$

for a random choice of B is $1 - Q^*(x)$, while in the remaining cases $\rho_n(x/B)$ cannot exceed $\rho_{\max}$. This leads to upper bound on distortions $\bar{\rho}_{\theta}$ and $\bar{\rho}_{\alpha}$; which from (2.5) ; (1.6) and (1.7) are obtained as

$$\bar{\rho}_{\theta} \leqslant \sum_{\text{all } x} \theta(x) \, [(D_1^* + \delta)\,(1 - Q^*(x)) + \rho_{\max} Q^*(x)]^k$$

$$\leqslant D_1^* + \delta + \rho_{\max} \sum_{\text{all } x} \theta(x) \, [1 - Q(S(x))]^k \qquad ...(2.7)$$

and $$\bar{\rho}_{\alpha} \leqslant D_2^* + \delta + \rho_{\max} \cdot \sum_{\text{all } x} \alpha(x) \, [1 - Q(s(x))]^k. \qquad ...(2.8)$$

We now examine the situation in which the last terms on the right sides of (2.7) and (2.8) respectively vanish as $n \to \infty$.

We will show that the term in question will be eradicated if K grows at an exponential rate slightly greater than $R(D_1^* ;\ D_2^*)$. Let us take a conditional probabilities assignment $Q(\underset{\sim}{y}/\underset{\sim}{x})$ and let $T(\underset{\sim}{x})$ be a set associated with a source word $\underset{\sim}{x}$ such that

$$T(\underset{\sim}{x}) = \left\{ \underset{\sim}{y} :\ n^{-1} \log \frac{Q(\underset{\sim}{y}/\underset{\sim}{x})}{Q^{\alpha}(\underset{\sim}{y})} \leqslant R(D ;\ D) + \delta \right\} \qquad \text{...(2.9)}$$

where $Q^{\alpha}(\underset{\sim}{y})$ is taken for $Q(\underset{\sim}{y})$ and is determined as an output probability for the inaccurate probability scheme. Clearly

$$Q(S(\underset{\sim}{x})) \geqslant Q(S(\underset{\sim}{x}) \cap T(\underset{\sim}{x})). \qquad \text{...(2.10)}$$

From (2.4), (2.9) and (2.10) ; it follows that

$$Q(S(\underset{\sim}{x})) \geqslant \exp\{-n[R(D_1^* ;\ D_2^*) + \delta]\}. \sum_{\underset{\sim}{y} \in S(\underset{\sim}{x}) \cap T(\underset{\sim}{x})} Q(\underset{\sim}{y}/\underset{\sim}{x}). \qquad \text{...(2.11)}$$

We now use the inequality

$$(1 - \alpha\beta)^k \leqslant 1 - \beta + e^{-k\alpha} ;\ (0 \leqslant \alpha ;\ \beta \leqslant 1) \qquad \text{...(2.12)}$$

which gives

$$[1 - Q(S(\underset{\sim}{x}))]^k \leqslant 1 - \sum_{\underset{\sim}{y} \in S(\underset{\sim}{x}) \cap T(\underset{\sim}{x})} Q(\underset{\sim}{y}/\underset{\sim}{x}) + e^{-K\exp\{-n[R(D_1^* ;\ D_2^*)+\delta]\}}. \qquad \text{...(2.13)}$$

Next, let us choose K ; the largest integer in

$$\exp\{n/R(D_1^* ;\ D_2^*) + 2\delta]\}$$

so that upon choosing n large enough, the last term in (2.13) can be made less than δ/ρ_{max} and (2.7) and (2.8) may be put as

$$\bar{\rho}_{\theta} \leqslant D_1^* + 2\delta + \rho_{max}[1 - \sum_{\alpha \| \underset{\sim}{x}} \theta(\underset{\sim}{x}) \sum_{\underset{\sim}{y} \in S(\underset{\sim}{x}) \cap T(\underset{\sim}{x})} Q(\underset{\sim}{y}/\underset{\sim}{x})] \qquad \text{...(2.14)}$$

and

$$\bar{\rho}_{\alpha} \leqslant D_2^* + 2\delta + \rho_{max}[1 - \sum_{\alpha \| \underset{\sim}{x}} \alpha(\underset{\sim}{x}) \sum_{\underset{\sim}{y} \in S(\underset{\sim}{x}) \cap T(\underset{\sim}{x})} Q(\underset{\sim}{y}/\underset{\sim}{x})] \qquad \text{...(2.15)}$$

Now let

$$S = \{(\underset{\sim}{x} ;\ \underset{\sim}{y}) :\ \underset{\sim}{y} \in S(\underset{\sim}{x})\} \qquad \text{...(2.16)}$$

$$T = \{(\underset{\sim}{x} ;\ \underset{\sim}{y}) : \underset{\sim}{y} \in T(\underset{\sim}{x})\} \qquad \text{...(2.17)}$$

$$\theta(\underset{\sim}{x} ; \underset{\sim}{y}) = \theta(\underset{\sim}{x})\, Q(\underset{\sim}{y}/\underset{\sim}{x}) \qquad \text{...(2.18)}$$

$$\alpha(\underset{\sim}{x} ; \underset{\sim}{y}) = \alpha(\underset{\sim}{x})\, Q(\underset{\sim}{y}/\underset{\sim}{x}) \qquad \text{...(2.19)}$$

and

$$\theta(G) = \sum_{\underset{\sim}{x} ; \underset{\sim}{y} \in G} \theta(\underset{\sim}{x} ; \underset{\sim}{y}) \qquad \text{...(2.20)}$$

$$\alpha(G) = \sum_{\underset{\sim}{x} ; \underset{\sim}{y} \in G} \alpha(\underset{\sim}{x} ; \underset{\sim}{y}) \qquad \text{...(2.21)}$$

where G is an arbitrary subset of the n-fold product space. Then

$$1 - \sum_{\underset{\sim}{a} \| \underset{\sim}{e}} \theta(\underset{\sim}{x}) \sum_{\underset{\sim}{y} \in S(\underset{\sim}{e}) \cap T(\underset{\sim}{e})} Q(\underset{\sim}{y}/\underset{\sim}{x}) = 1 - \theta(S \cap T) \quad ...(2.22)$$

$$1 - \sum_{\underset{\sim}{a} \| \underset{\sim}{e}} \alpha(\underset{\sim}{x}) \sum_{\underset{\sim}{y} \in S(\underset{\sim}{e}) \cap T(\underset{\sim}{e})} Q(\underset{\sim}{y}/\underset{\sim}{x}) = 1 - \alpha(S \cap T). \quad ...(2.23)$$

If we denote the compliments of the sets S and T by $\bar{S}$ and $\bar{T}$ respectively, then

$$1 - \theta(S \cap T) = \theta(\bar{S} \cup \bar{T}) \leqslant o\, \theta(\bar{S}) + \theta(\bar{T}) \quad ...(2.24)$$

$$1 - \alpha(S \cap T) = \alpha(\bar{S} \cup \bar{T}) \leqslant \alpha(\bar{S}) + \alpha(\bar{T}). \quad ...(2.25)$$

Thus for n sufficiently large (2.14) and (2.15), in view of (2.22) and (2.23), gives

$$\bar{\rho}_\theta \leqslant D_1^* + 2\delta + \rho_{\max}[\theta(\bar{S}) + \theta(\bar{T})] \quad ...(2.26)$$

and

$$\bar{\rho}_\alpha \leqslant D_2^* + 2\delta + \rho_{\max}[\alpha(\bar{S}) + \alpha \bar{T})]. \quad ...(2.27)$$

We now make an appropriate selection of $Q(\underset{\sim}{y}/\underset{\sim}{x})$ as follows :

$$Q(\underset{\sim}{y}/\underset{\sim}{x}) = \prod_{t=1}^{n} Q^*(y_t/x_t) \quad ...(2.28)$$

where $Q^*(k/j)$ is that probability assignment which solves the variational problem defining $R(D_1^* ;\ D_2^*)$.

Also since $\{X_t;\ \theta\ ;\ \alpha\}$ is a practical discrete memoryless source, it follows that

$$\theta(\underset{\sim}{x}\ ;\ \underset{\sim}{y}) = \prod_{t=1}^{n} \theta\ .\ (x_t)\, Q(y_t/x_t) \quad ...(2.29)$$

$$\alpha(\underset{\sim}{x}\ ;\ \underset{\sim}{y}) = \prod_{t=1}^{n} \alpha\ .\ (x)\, Q(y_t/x_t). \quad ...(2.30)$$

By the law of large numbers the average of $\rho_n(x_t;\ y_t)$ with respect to the distributions $\theta(\underset{\sim}{x}\ ;\ \underset{\sim}{y})$ and $\alpha(\underset{\sim}{x}\ ;\ \underset{\sim}{y})$ respectively converge in probabilities to

$$E(\rho_\theta) = \sum_{j,k} \theta(j)\, Q(k/j)\ .\ \rho(j\ ;\ k) = D_1 = D_1^* \quad ...(2.31)$$

and

$$E(\rho_\alpha) = \sum_{j,k} \alpha(j)\, Q(k/j)\ .\ \rho(j\ ;k) = D_2 = D_2^* \quad ...(2.32)$$

and we, therefore have

$$\bar{S} = \{(\underset{\sim}{x}\ ;\underset{\sim}{y})\ ;\ \rho_n(\underset{\sim}{x}\ ;\underset{\sim}{y}) > D_1^* + \delta \text{ and } \rho_n(\underset{\sim}{x}\ ;\ \underset{\sim}{y}) > D_2^* + \delta\} \quad ...(2,33)$$

so that

$$\lim_{n\to\infty} \theta\,(\bar{S}) = 0 \qquad \text{...(2.34)}$$

and $\lim_{n\to\infty} \alpha\,(\bar{S}) = 0.$ ...(2.35)

Further let $Q^{\alpha}(\underset{\sim}{y})$ be the marginal distribution induced by the joint ensemble $\alpha\,(\underset{\sim}{x}\,;\underset{\sim}{y})$ of (2.29) and (2.30), that is,

$$Q^{\alpha}(\underset{\sim}{y}) = \sum_{a\,\|\,\underset{\sim}{x}} \prod_{t=1}^{n} \alpha\,(x_t)\; Q\,(y_t/x_t) = \prod_{t=1}^{n} Q^{\alpha}(y_t) \qquad \text{...(2.36)}$$

where

$$Q^{\alpha}(\underset{\sim}{y}) = \sum_{x} \alpha\,(\underset{\sim}{x}),\; Q\,(\underset{\sim}{y}/\underset{\sim}{x}). \qquad \text{...(2.37)}$$

In view of the distributions in (2.28) and (2.36), we have ;

$$n^{-1} \log \left(\frac{Q\,(\underset{\sim}{y}/\underset{\sim}{x})}{Q^{\alpha}(\underset{\sim}{y})}\right) = n^{-1} \sum_{t=1}^{n} \log \frac{Q\,(y_t/x_t)}{Q^{\alpha}(y_t)}\,, \qquad \text{...(2.38)}$$

showing that expression on the left is the arithmetic average of the n independent identically distributed random variables on the right. Hence by the law of large numbers

$$n^{-1} \log \left(\frac{Q\,(\underset{\sim}{y}/\underset{\sim}{x})}{Q^{\alpha}(\underset{\sim}{y})}\right) \text{ converges in probability to}$$

$$\sum_{j,k} \theta\,(j)\, Q\,(k/j) \log \frac{Q\,(k/j)}{Q^{\alpha}(k)} = R\,(D_1^* \,;\, D_2^*).$$

Since

$$\bar{T} = \left\{(\underset{\sim}{x}\,;\underset{\sim}{y}) = n^{-1} \log \frac{Q\,(\underset{\sim}{y}/\underset{\sim}{x})}{Q^{\alpha}(\underset{\sim}{y})} > R\,(D_1^*\,;\, D_2^*) + \delta\right\}$$

it follows that

$$\lim_{n\to\infty} \theta\,(\bar{T}) = 0$$

and

$$\lim_{n\to\infty} \alpha\,(\bar{T}) = 0.$$

Thus it has been established that $\theta\,(\bar{S})$; $\theta\,(\bar{T})$; $\alpha\,(\bar{S})$ and $\alpha\,(\bar{T})$ are all less than δ/ρ_{max} where n is sufficiently large.

Now setting $\delta = \varepsilon/4$ in (2.26) and (2.27) ; we get

$$\bar{\rho}_\theta \leqslant D_1^* + \varepsilon$$

and

$$\bar{\rho}_\alpha \leqslant D_2^* + \varepsilon.$$

This completes the proof of the theorem.

Acknowledgement

The author is thankful to Dr. B. D. Sharma, Department of Mathematics, Delhi University, for the help and guidance in the preparation of this paper.

References

Berger, T. (1971). Rate distortion theory. Prentice Hall, N.J.

Kerridge, D. F. (1961). Inaccuracy and inference. *J. Roy. Soc.*, **23B**, 184–94.

Mathur, Y. D. (1974). Information improvement and computation of the rate distortion function under true and inaccurate input distributions. *Elektron. Inform. Kybernetik, West Germany.*

Shannon, C. E. (1960). Coding theorems for a discrete source with a fidelity criterion. *Information and Decision Processes*, ed. Machol., New York.

Sharma, B. D. and Mathur, Y. D. (*in press*), The rate distortion function under true and inaccurate input probability schemes. *Rev. Sci. Univ., Istanbul.*