A STUDY ON THE APPLICATION OF STATISTICS AND INFORMATION THEORY TO PROBLEMS OF COMMUNICATION WITH SPECIAL REFERENCE TO RADAR

B. S. RAGHURAM

Scientific Evaluation Group, Research & Development Organisation, New Delhi

ABSTRACT

Information theory, which originated in Tele-communication studies, is a branch of Mathematical statistics with many applications of statistical inference. The three fundamental problems are: (*i*) Development of statistical measures of information capacity in a Communication system, (*ii*) the transmission problem of information in a system, and (*iii*) analytical study of reception from a statistical decision point of view. This paper is an attempt to present a comprehensive study of all three aspects. In addition, application of sequential analysis, specially with reference to radar signal detection and range estimation has been briefly discussed. Finally, from the point of view of signal reception in the case of a radar, the problem has been considered as a statistical decision study. In conclusion, the computational problem as well as certain comparative studies have been briefly touched upon. Illustrative examples are given and graphs are shown wherever necessary.

INTRODUCTION

The concept of "information" was introduced by Fisher in Statistical theory. While developing the theory of estimation by maximum likelihood method, he defined the word "information" on probability grounds. We shall give his approach, in brief, as follows:

Consider a probability function $f(x/\theta)$, where θ is an estimable parameter, defined in a non-degenerate interval I. Let $t = t(x_1\ x_{21} \ldots x_n)$ be a sample estimate of θ based on a sample of size "n" and let $E(t) = \theta + b(\theta)$ where $b(\theta)$ is the bias in the estimate t. Under certain regularity conditions it could be shown that

$$E(t-\theta)^2 \geqslant \frac{\left[1+\left(\frac{db}{d\theta}\right)\right]^2}{n\int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}\log f\right)^2 f\,dx} \tag{1}$$

where f denotes $f(x/\theta)$ — the frequency or the probability function. Setting $b(\theta) = 0$ we get

$$E(t-\theta)^2 = \text{Var}(t) \tag{2}$$

and so we have

$$\text{Var}(t) \geqslant \frac{1}{n\int_{-\infty}^{\infty}\left(\frac{\partial}{\partial\theta}\ l\ g\ f\right)^2 f\,dx} \tag{3}$$

Fisher called this limiting value as the "information limit" to the variance and denoted the same by

$$\frac{1}{I} \tag{4}$$

Thus,

$$I = n \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \theta} \log f \right)^2 f \, dx \tag{5}$$

is the amount of information available with regard to the parameter "θ", based on a sample of size "n".

Based on probability considerations and on similar procedure as discussed above, Hartley, Shannon and others developed quantitative measures of information capacity in a communication system. In matters of derivation, these measures do not differ much from Fisher's measure. The only additional point is that Fisher's measure is just a statistical quantity, while the others have physical meaning behind them; as for instance, the measure due to Shannon being called as "Entropy" which is measurable as a physical quantity in bits per second and also in terms of thermodynamic units (degrees Kelvin).

THE GENERAL PROBLEM

The advent of Information Theory and its applications in the field of communication have made it imperative that the electronic engineer should have at least a basic understanding of "Engineering statistics". The general class of problems in communication work, having applications of Information Theory, may be put as follows :

(*i*) The first considers the fundamentals of communication—information theory.

(*ii*) The second considers a description of the statistics (measures) which are most applicable in communication problems.

(*iii*) The third one considers more specifically how mathematical methods of Statistics can be utilized in the development of formulæ for use in communication problems.

Communication theory centres round the problem that the events going on in Nature and Technological systems can be represented as time dependent stochastic processes. The concept of information refers to the inner meaning of these events forming the process. Probability plays an important role both in the generation and in the transmission of a message. Thus, statistically, information could be defined as the "Freedom of Choice" or the degree of uncertainty associated with the selection and transmission of a message waveform.

MEASURES OF INFORMATION

A communication network is a collection of points between which information is transmitted. The observations made in the form of a stochastic process constitute information. Now, what kinds of problems can there be in a sequence of outcome data ? and secondly, how do we use information theory to detect them ?. The first problem is one of "Frequency Analysis". Here we study the entropy content of information which describes the change of available energy in a communication system.

The problem of defining a quantitative measure of information capacity, whereby two different systems transmitting information could be compared, was firstly considered by Hartley[1] and he gave the following law :

$$C = log_e \; n \tag{6}$$

as a measure of information capacity in a system where n is the number of choices in the selection of a message. Computation problem is very much simplified by consideration of logarithmic measure. C, in the above law, gives information in natural units, while the same quantity with reference to base 2 in the logarithm gives information in bits (binary digit units), which is the most common and often adopted unit as far as machine studies are concerned. From the point of view of relationship between these units we have,

1 Natural unit $=$ 1·443 bits

or 1 bit $=$ 0·69 natural unit.

Consider a case of 'm' alternatives that are all equally likely. Then H, the amount of uncertainty associated with any event, is given by

$$H = log_2 \; m \tag{7}$$

$$= - \; log_2 \; p \tag{8}$$

where 'p' is the probability of any one alternative being selected. Generalizing, we have,

$$H = - \sum_{i=1}^{m} p_i \; log_2 \; p_i \tag{9}$$

as "Shannon-Wiener measure" of information. This gives the minimum number of bits into which an event may be coded. H, in the above equation, is also called as the "Entropy" and is a word borrowed from "Statistical Mechanics". Entropy is having physical meaning in that it describes the randomness and its pattern in a physical system. The quantity

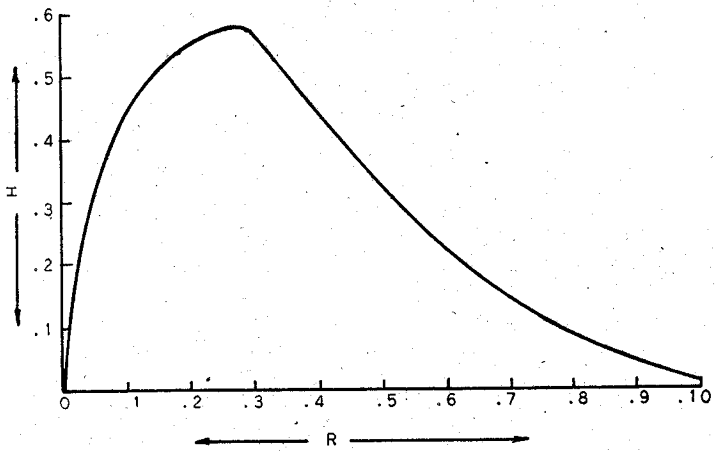
$$H' = \frac{H}{H \; (MAX)} \tag{10}$$

is called as "Relative Entropy" in a set of events and its complementary function

$$R = 1 - H' \tag{11}$$

is termed as "Redundancy" in a set of events. This is a very useful measure describing statistically the auto correlations in a set of events forming a stochastic process.

Correlation is present in all languages and in all systems of coding. If a certain letter of an alphabet occurs the probability that certain others will follow the letter is not the same as the prior probabilities that these letters would have occurred. Thus, in a word commencing with "q" the probability of 'u' following the letter is probably the highest. Such correlations as these discussed now may be called as "inter-symbol constraints" and these lead to redundancy. This inter-symbol constraint which is present in the message waveform makes the stochastic process a Markov processs. Best possible waveforms should have very little redundancy and the decision in matters of coding to remove the redundancy, or to minimize the same, is quite a complicated one depending chiefly upon experience and past history than on mathematical considerations. The following diagram Fig. 1 gives the relationship between H and R. The language and code



RELATIONSHIP BETWEEN H & R

FIG. 1

redundancy in any meassage waveform could be computed if the data were to be mady available as regards the letter frequencies and code repetitions. For example, the entrope or the information content of the English language shall have its maximum value as

$$log_2 \cdot 26 = 4 \cdot 7 \text{ bits}$$

per letter, assuming that all the letters from A to Z would have the same letter frequency of occurrence in the language. In the most general case, it is not so and the value of the entropy content shall be much below this figure.

The quantity "self-information" of any event in a message sequence may be measured by the logarithm of the reciprocal of its probability of occurrence. In this meaning self-information of a message sequence can be regarded as representing the number of bits required for complete identification of the waveform, whenever we wish to minimise the expected value of the number of bits. This result provides the operational significance for the measure of self-information of any event in a message sequence.

Consider the Normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \quad \exp \ - \ \frac{x^2}{2\sigma^2} dx \quad - \infty \leqslant x \leqslant \infty \tag{12}$$

with the mean located at the origin (0,0). The entropy associated with the parameter σ is given by

$$\text{H} = log_e \sqrt{2\pi e}\, \sigma \tag{13}$$

$$= [\, 2 \cdot 04 + log_e \ \sigma \,] \, log_2{}^e \text{ bits} \tag{14}$$

$$= 2 \cdot 94 + 1 \cdot 443 \ log_2 \sigma \text{ bits} \tag{15}$$

Since the information or the entropy content of a message is the minimum capacity required for storage, the best possible code must depend on prior probabilities of the occurrence of the message. Prior probabilities of occurrence plays quite an important role in information storage as well as in the statistical analysis of the properties of waveform such as decision as to what particular message had been transmitted and so on.

A message sequence, since it is characterised by the prior probabilities of occurrence, can very well be studied as a stochastic process. There are many time dependent stochastic processes; as for example, telephone calls, radio active disintegratism, pulses reflected and received in a radar system and so on. The concept of ergodicity is associated with such processes where a change can occur at any time. Ergodicity is usually a necessary condition for the application of information measures to a sequence study. Under certain circumstances a stochastic process converges to a limiting form which is independent of the initial position from which it started. This property is called as ergodicity and any stochastic process with such a convergence is called an ergodic process.

INFORMATION TRANSFER

Information transfer is a function of one's degree of choice in communicating a message. Here not only transmission but more than that, the dependability one can have on transmitted signals is important. Thus, maximum dependability on transmission is the same as transmission with minimum probability of error. The three problems involved in information transfer are :—

(*i*) How can one measure the rate of information production and transmission ?

(*ii*) How much information can be sent through a channel ? This calls for a study of channel capacity with a given signal power and a noise power.

(*iii*) What is the part played by 'Noise' on 'Reception' ? This calls for a detailed study of statistical decision methods as also time series analysis involving the spectral and Fourier analysis of the waveforms.

Two of the important factors involved in the analysis of any communication network are : (*i*) the signal-to-noise ratio and (*ii*) the bandwidth. Bandwidth is a measure of the rate at which frequency changes can occur. Wider the bandwidth better shall be the amount of information that could be transmitted. This principle is not possibly to be generalized; for in the case of the study of radar signals, wider bandwidth is not the only criterion but the more important criterion is the maximization of the signal-to-noise ratio.

Consider a case of message transmission with error. Let there be 'n' possible messages, say 1,2,3,..........n. Obviously these could be described as values of a variate X, the message sent. Let their *a priori* probabilities of being sent be p_1, p_2..........p_n where p_i could be put as $p\,(X = i)$. When the message 'i' is actually being sent, due to transmission errors there is a chance of a different message 'j' being received. Let q_{ij} be such a conditional probability. These conditional probabilities which occur in information transfer problems are termed as "transitional probabilities" and in the case of an errorless transmission $q_{ij} = q_{ii} = 1$.

Consider the set of signals Si, $i = 1, 2,$ .... with time durations t_j, $j = 1, 2,$ .... at the receiving end of a communication channel. The channel is defined by the signals S_i and their time durations t_j. Let all the signals be independent and T be the total duration of all signals combined together. If $N(T)$ gives the number of distinct messages which will differ in the order of the signals, then 'C', the capacity of the communicating channel, could be defined as,

$$C = k \lim_{T \to \infty} \log_2 \frac{N(T)}{T} \tag{16}$$

in bits per second, where k is Boltzmann's constant.

Consider a set of signals $S_1 \ldots S_2 \ldots\ldots\ldots S_n$ on a channel with their respective durations $t_1, t_2 \ldots t_n$ seconds on the channel. The average information content per signal will be maximum when $p_j = e^{-\beta t_j}$ (17)

where p_j is the prior probability of the occurrence of S_j, β is a constant determined by the equation

$$\sum_j e^{-\beta t_j} = 1 \tag{18}$$

From equation (17) it would be clear that most probable signals should have a relatively short duration on the channel, a condition which is true even from a commonsense point of view.

Statistically, communication problem could be treated as a parameter estimation problem. A parameter θ defined over a multidimensional domain could be visualized as the true transmitted message. A variable 'Y' represented as a random variable could be taken as a probability function depending on the parameter θ. The simplest statistical problem is to estimate 'θ' when the sample 'Y' is given. If 'X' and 'Y' represent the transmitted and the received messages, then, a function $f(x,y)$ could be constructed to give the description of the situation of the transmitted message. Thus, given the 'Y' value of a joint sample (x_i, y_i), we should estimate the 'X' value. Hence, alternatively the problem may be thought of as one of testing of hypothesis, as to which 'X' has been obtained at the time of observation. Following these arguments three questions of statistical interest arise in communication problems and they are :—

(*i*) By what criterion shall the various estimates of 'X' be comparable ?

(*ii*) Given the criterion, what is the best estimate of 'X' which could be formed and how good it shall be ?

(*iii*) How do competing methods of estimating 'X' compare with the best ?

Let us consider a communication system in which 'B' bits of information are stored in the encoder. Disregarding any delay in the channel, let us assume that T seconds elapse during the time a digit (code) is reproduced by the decoder. This is the time corresponding to the transmission of 'B' bits of information. The probability of error per digit (code) in transmission depends on 'B' as also on the rate of transmission 'R', and the channel capacity 'C'. For a channel with finite memory the probability of error has the form

$$p(e) = K \cdot 2^{-BM/R} \tag{19}$$

where K is a slowly varying function of B and R. M is a coefficient independent of B, but a function of R as well as the characteristics of the channel.

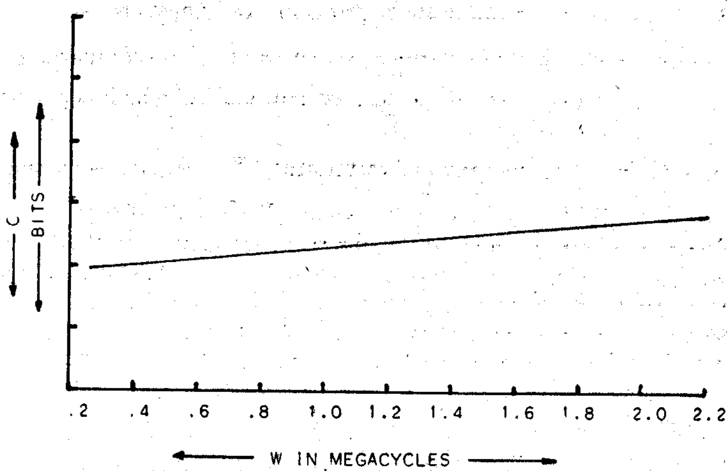
From this we have :

$$\lim_{B \to \infty} \log_2 \frac{p(e)}{B} = -\frac{M}{R} \qquad (20)$$

Thus, the probability of error can be reduced by increasing either B or M/R. The value of M/R increases with increasing channel capacity for a fixed transmission rate 'R'. The complexity of encoding and decoding devices increases with B. Thus, either a larger channel capacity or greater equipment complexity can bring down the probability of error in transmission.

Wolfowitz[2] has investigated the problem of statistical inference on channel statistics in the case of fixed channels and constant rate of transmission. He has shown that inference is possible at the receiving end of the channel only from what is received and has reasoned out further that this additional uncertainty does not affect the channel capacity.

Transmission and receipt of information in the case of a radar system may be visualized as a 'game situation' where one player selects a suitable receiving filter and the other a function describing noise power. The pay-off function shall be taken as the signal-to-noise ratio. In detecting a known signal under conditions of Gaussian noise during a finite and fixed time interval, one is lead to consider for the receiver a wider network followed by a circuit, sampling the output and deciding if the signal is present or not. The error probability, computed at a fixed false-alarm rate, depends on a single entity that incorporates the relevant properties of signal and noise. If the signal has an unknown phase the radar receiver should be modified such that the input signal is effectively the complete envelope of the received signal. The signal-to-noise ratio determines the detection performance. Typically, one of the two possible signals is transmitted within a time interval and the receiver should decide as to which is most likely to be present. When both the signals are completely known, it is possible to use one filter and derive an expression for the error probability that is the same for both the signals and depends on a signal-to-noise ratio. The difference between the given pulses corresponds to the radar pulse.



DEPENDENCE OF C ON W

FIG 2

Consider Shannon's formula for the channel capacity :

$$C = W \log_2 \frac{P + N}{N} \qquad (21)$$

where C, in bits per second, is the upper limit which cannot be exceeded in transmission : W, P, and N are the bandwidth (in Megacycles), the signal power used on an average in transmission (in watts) and the average noise power for which the channel is subject to (in watts), respectively. We can write the equation also as

$$C = W \log_2 (1 + P/N) \qquad (22)$$

We see in the above equation C to depend on W and also on P/N. Keeping one of these two a constant we can show graphically the variation of C with regard to the other variable. The figures 2 & 3 show the dependence of C on W and P/N.

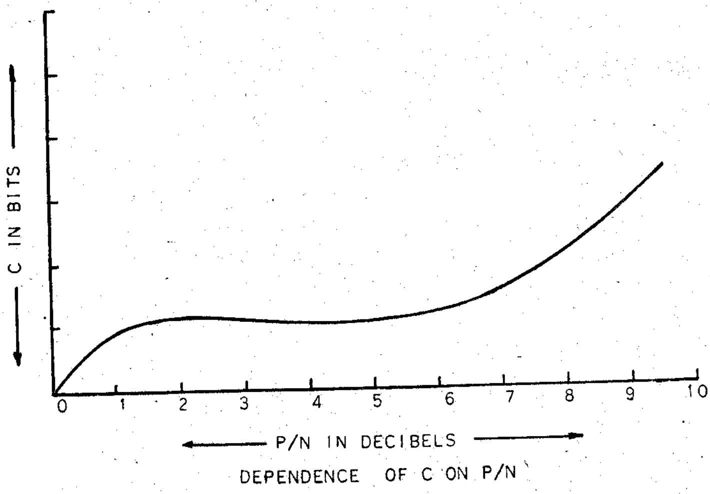


FIG 3

Thus, it could be seen that the channel capacity increases with increasing signal-to-noise ratio and also with increasing bandwidth. The most important interpretation of Shannon's result is that it is possible to exchange bandwidth for signal-to-noise ratio in a communication channel. If bandwidth is too small, by increasing the signal power sufficiently we may maintain the same rate of transmission in the channel. The principle underlying here is of great importance in the radar, where, performance efficiency is dependent both on P/N and W, but more on P/N than on W.

THE STATISTICAL PROBLEM OF RECEPTION

We shall consider here the influence of noise on the reception of signals; i.e., the problem of noise control and statistical decision in communications. Statistical problems of reception chiefly deal with the decision problem in signal detection and selection or extraction of the correct signal.

Noise, in information theory, refers to random disturbances that occur in communication work. Examples of such studies are found in the case of radar signal reception, telephony, and so on. Any random sequence of impulses with a finite average sequence of occurrence and a finite strength is defined as "white noise" and is a stochastic process.

The ultimate goal in communication problems, though a hypothetical one, is to produce a capacity of receiving and understanding a message which has been transmitted. If the pattern of noise interference could be fully determined for a future time, it would be possible to build a device to cancel out its effects. But, as noise interference is of a random nature, it cannot be predicted with certainty. However, if the interference is to be given as a "time series" it could be analysed by means ofstructural equations involving auto-correlation functions of the type

$$\phi_r\ (t) = \frac{1}{T}\int_{0}^{T} f(t)\,.\,f(t+r)dt \tag{23}$$

where $f(t)$ is the waveform based on the time parameter. T is the total time duration and r is the time lag. Alternatively spectral analysis also helps in the study of these interferences being represented as time series.

Periodic signals are useful for power transmission. Consider the Fourier transform of an input function $x(t)$ as

$$X(t) = \int_{-\infty}^{\infty} x(t) \exp - 2\pi i\, f(t)\; dt \tag{24}$$

From this, the inverse Fourier transform $x(t)$ could be written as,

$$x(t) = \int_{-\infty}^{\infty} X(t)\, e^{2\pi i\, f(t)}\; dt \tag{25}$$

If $X(f)$ and $Y(f)$ are the Fourier transforms of the input and output signals; i.e. the transmitted and received signals and, if $G(f)$ is the complex transfer function given by

$$Y(f) = G(f)\; X(f) \tag{26}$$

and also if $\beta(f)$ is the phase shift in the received or output signal, then $y(t)$ the output signal waveform shall be given by the Fourier integral

$$y(t) = \int_{-\infty}^{\infty} G(f)\; X(f) e^{2\pi i\, f\,(t)}\; df \tag{27}$$

The outstanding obstacle in formulating a completely satisfactory theory of reception that could be applied to observation systems such as radar, is the question of prior probability distributions whether they are for the message state "$X\,(t)$" or not. The central problem is the determination of the state of the message, say $X(t_o)$ from a consideration of noise and signal data say $Y(t)$. For example, consider the prior probability of observing an aircraft on a radar set at a range of 25 kilometres at 09·00 Hrs. on a certain day. If the radar set is situated at an airfield where all the regular services operate, statistical analysis of the past performance of the set might provide as with the required probability, on the assumption that the organisation of air-traffic in the domain concerned follows the pattern of a stationary stochastic process.

The problems associated with radar signal reception are relatively simple where certain idealisations are made. Consider the problem of measurement of the range of the target by means of the timing of its echo signal. Here we can formulate the stochastic processes $W(t)$ and $Y(t)$ and define them as the waveform which is transmitted, echoed and picked up again after suffering a time-lag T. The pattern of the received signal shall be given by

$$Y(t) = W(t - T) + N \tag{28}$$

where N is the noise component. Analysis of this function $Y(t)$ is very important in the case of such reception problems.

In the case of pulsed radar one is not generally interested in receiving targets at all ranges. For example, a target at a distant range of 1000 kilometres is never likely to be compared with another target at a range of 50 kilometres. In all such studies regarding range estimation, the pulse pattern with suitable pulse recurrence frequency is, therefore a more acceptable waveform for range, than a single pulse would be. Search radars search the sky for possible appearance of a target. For such ones, signal detection lies in determining the presence or not of a pulse, while the pulse shape is of secondary importance. Tracking radars require information as to the precise time of reception and here, however, the shape of the pulse (pattern) is equally important. In the case of a radar, for acceptably small probabilities of decision error, weak radar signals require longer pulse-integration times while strong ones need much less. On the assumption of an additive nature of signal and noise and knowing that a deterministic or entirely random signal is actually present, the method of maximum likelihood can be used in the estimation of one or more of the signal parameters; or, in the case of random continuous signals, it may be used to find out the signal waveform.

SEQUENTIAL ANALYSIS APPLICATION

Sequential analysis has many applications in radar signal detection problems. To detect a target at a specified location, a periodic succession of pulses are transmitted in the proper direction. If a target is present, the echo signals will be received mixed with noise. The receiver must choose between the hypothesis H_o of signal absence and H_1 — of signal presence. The reliability of the decision is given by Q_o, the probability of a false alarm and Q_d the probability of detecting the signal. Then, we can also define a probability Q_d , as the probability of detecting a signal of some standard strength, usually preassigned.

Let the input voltage function be given by $V(t)$. In a sequential detection system a likelihood function is formed after each pulse transmission on the basis of the information so far received. The k^{th} such likelihood ratio $z_k\,[V_k\,(t)]$ depends on the receiver input $V_k\,(t)$ during the first k observation periods. z_k is compared with two thresholds a and b, where $b<a$. If $z_k<b$ or $z_k>a$ the testing procedure stops; in the former case H_0 is chosen, and in the latter case H_1 is chosen. If $b<z_k<a$ another pulse is transmitted and the receiver input following it is incorporated with previous inputs into a new likelihood function x_{k+1} with which the same comparisons are repeated. The procedure continues until one hypothesis or the other is chosen. The thresholds a and b are given by

$$b = log_e \frac{(1 - Q^1_d)}{(1 - Q_o)} \tag{29}$$

and

$$a = log_e \left(\frac{Q'_d}{Q_o}\right) \tag{30}$$

following Wald's criteria of fundamental definitions in sequential analysis, provided that the expected number of tests is large.

In sequential detection system, the total number 'n' of tests is equal to the number of pulses transmitted, and this shall be a random variable.

Since electronically scanned antennae are being developed, use of sequential detection promises a significant saving in the time a radar system needs to search out the entire sky. As soon as the system has reliably decided whether a given sector of the sky contains any target or not, the antenna beam can be moved to a new position. Because in most sectors of the sky no targets are present, the average number of pulses transmitted into each sector will be smaller than the number required by an equally reliable radar whose antenna, scanning at a uniform rate, spends the same amount of time in each sector.

THE COMPUTATION PROBLEM

Here we deal with the problem of storage of information in a computer set up. The treatment of "memory channels" becomes important here. The digital computer is chiefly used for storage, for retrieving and analysing the information stored. Although in the ideal circumstances information may not be lost in a digital computer set up, yet, normally there will be some loss and thus the output of the machine will be slightly less than its input. Digital computers are best suited for operations when storage problems are on consideration. Ideal machines correspond to conservation of information; or in other words to the reversibility phenomenon.

The average information rate, in the flow of information in a system, must be the same in all parts of the system. In the case of multi-channel or link systems three situations might occur. After encoding, the signal corresponding to the message may have (*i*) the same amount of information as the message, (*ii*) less information and (*iii*) within a restricted time space more of information. If the encoded signal contains more information, the situation needs careful examination. Over a short time interval more information can come out of a machine than that goes in. However, the added information had been stored in the machine for later use, so that, over a long time interval no more information can leave a machine that enters in.

CONCLUSIONS

We have examined, in brief, the application of information theory methods in the problems of communication with special reference to radar. We have seen also how estimation in the theory of signal reception is exactly the counterpart of parameter estimation in statistical theory of decision and testing of hypothesis. Information theory and decision theory approaches, although slightly different, complement each other providing powerful tools for a comprehensive attack on the basic problems of system optimization, design and evaluation.

In a way, the terms, "Communication System" and "Industrial Organisation" appear analogous. All production activities are essentially communication channels that could be analysed by means of techniques of information theory. The analogue between a communication system and a manufacturing system reveals that we are interested in (*i*) measurement and testing, (*ii*) quality control. These can be subdivided further into measurement system and the filter or the decision system. Measurement and testing are in signal detection studies and the decision is in the extraction or selection of the signal. The information principle is useful in many engineering problems involving man-machine systems.

Acknowledgement—My sincere thanks are due to Major H.S. Subba Rao for his help and encouragement in the preparation of this paper.

REFERENCES

1. Hartley, R. V. L., *Bell System Tech. J.*, **7**, 535 (1928).
2. Wolfowitz, J., "Coding Theorems of Information Theory" (Prentice Hall), 1961.