605099

ON THE ROLE OF DYNAMIC PROGRAMMING IN STATISTICAL COMMUNICATION THEORY

By

Richard Bellman and Robert Kalaba

P-949

Revised

December 19, 1956

H.C. $1.00

M.F. $0.50

The RAND Corporation

1700 MAIN ST. • SANTA MONICA • CALIFORNIA

SUMMARY

In this paper we wish to show that the fundamental problem of determining the utility of a communication channel in conveying information can be interpreted as a problem within the framework of multi-stage decision processes of stochastic type, and as such may be treated by means of the theory of dynamic programming.

We shall begin by formulating some aspects of the general problem in terms of multi-stage decision processes, with brief descriptions of stochastic allocation processes and learning processes. Following this, as a simple example of the applicability of the techniques of dynamic programming, we shall discuss in detail a problem posed recently by Kelly. In this paper, it is shown by Kelly that under certain conditions, the rate of transmission, as defined by Shannon can be obtained from a certain multi-stage decision process with an economic criterion. Here we shall complete Kelly's analysis in some essential points, using functional equation techniques and considerably extend his results.

ON THE ROLE OF DYNAMIC PROGRAMMING IN STATISTICAL COMMUNICATION THEORY

1. Introduction.

In this paper we wish to show that the fundamental problem of determining the utility of a communication channel in conveying information can be interpreted as a problem within the framework of multi-stage decision processes of stochastic type, and as such may be treated by means of the theory of dynamic programming, [1], [2] , [3] .

This paper is to be envisaged as a step in the direction of a broad theory of communication, as contemplated by N. Wiener in his recent article, [12] , and following the pioneering efforts of Rice, [8] , and Shannon, [10] . Among other steps along this path, we would like to cite the recent articles of Bussgang and Middleton, [5] , and Middleton and Van Meter, [7] , which employ the modern theory of statistical decision functions and sequential analysis, due to Wald, [11].

We shall begin by formulating some aspects of the general problem in terms of multi-stage decision processes, with brief descriptions of stochastic allocation processes and learning processes. Following this, as a simple example of the applicability of the techniques of dynamic programming, we shall discuss in detail a problem posed recently by Kelly, [6] . In this paper, it is shown by Kelly that under certain conditions, the rate of transmission, as defined by Shannon, [10] , can be obtained from a certain multi-stage decision process with an economic criterion. Here we shall complete Kelly's analysis in some essential points, using functional equation techniques and considerably extend his results.

We shall consider, in addition to the original problem of Kelly, a time-dependent case, a process involving correlated signals, and a multi-signal case, in both discrete and continuous versions. It will be seen that the logarithmic criterion function plays an extremely important role, since its special functional properties permit us to obtain explicit representations for both maximum return and optimal policy.

Finally we discuss briefly a functional equation arising from the general question of defining the "value" of a communication channel in a fashion which is independent of its use.

2. The Underlying Model.

Let us begin by constructing a simple model of one aspect of the general communication problem. It will be reasonably clear from what follows how more intricate models may be constructed to take account of more complicated systems.

Consider a source S which produces at discrete times* a sequence of pure signals, together with noise, which may be of either stochastic or deterministic type, depending upon our further assumptions concerning the structure of the system. The combined signal is fed into a black box which we call a "communication channel", which, in turn, emits a signal. This output signal is observed.

On the basis of the observation of the output signal, it is desired to make various deductions concerning the properties of

*The case of continous signal emission can also be treated by the methods outlined below, at the expense of the introduction of more sophisticated concepts. We prefer to keep the mathematical level moderate in this first discussion; however, see § 13.

the original pure signal.

Schematically,

Source → | Communication Channel | → Observer

In mathematical terms, let

(1) x = the pure signal emanating from S

r = the noise associated with the signal

$x' = F(x,r)$, the input to the communication system

y = the signal transmitted to the observer by the communication channel.

Let us further write

(2) $$y = T(x') = T(F(x,r)),$$

where T represents the transformation of the input signal x' due to the communication channel.

Consider the set of all communication systems, or, equivalently, the space of all associated transformations, T. We wish to introduce an ordering, or, what is much more satisfactory when possible, a metric which will enable us to compare two communication systems, and to evaluate their performance, see §12.

It must be stressed at the very outset of an investigation of this type that it should be possible to accomplish this aim in an unlimited number of ways, dependent upon the source, channel, nature of the observer, and the personal philosophies involved, that is to say, upon the utility scales employed.

In the following sections, we shall present two alternate methods for evaluating a communication system. Although each is a particular case of a more general scheme, which we shall discuss subsequently, it is worthwhile to present them separately first, as they occur in important applications. In this way, we hope to avoid the usual risk of obscuring the issue by extreme generality.

3. A Stochastic Allocation Process.

Let us assume that the observer has a sum of money, or resources of other types, which we denote by the vector x, called the state vector. Upon receiving a y-signal, the observer is required to make an allocation of resources to various activities. The effect of this allocation is to change x into R(x,y), a stochastic vector whose distribution we shall assume here to be known. The case in which the distribution is not known is closely allied with the second model we shall discuss.

The process is now repeated N times, where N may be fixed, the simplest case, or the number of stages may depend upon the process itself as a consequence of a preassigned stop rule. Let us again consider only the simplest case, that of fixed N.

Further, let us suppose that the purpose of the observer in carrying out this process is to maximize the expected value of some function of his final state vector, the state attained after N stages of the process.

Let f_T denote this maximum expected value, and f_I the maximum expected value when the transformation T is the identity transformation, the case in which we have a distortionless communication channel.

Let us then agree to measure the worth of the original communication system by means of a preassigned function of f_T and f_I. In this fashion we introduce a metric into the space of transformations T, and thus into the set of communication channels. The simplest cases are those where we use a function of $f_I - f_T$, or a function of f_T/f_I.

We shall discuss a simple case of a process of the above type in later sections. For the formulation and mathematical discussion of some particular processes of this general type we refer to [9] and [4].

4. A Stochastic Learning Process.

Let us now consider a different type of stochastic process. The observer is required to make a decision concerning the nature of the pure signal emitted by S. He can observe as many samples of the signal emitted by the communication system as he wishes, subject to constraints imposed by the costs of observation, and by limitations of time.

As a result of these decisions, he makes an estimate concerning properties of the pure signal, and thereby incurs a cost dependent upon the deviation of this estimate from the actual situation.

The problem is to carry out the process of first observation and then estimation so as to minimize the expected total cost, where the total cost is a given function of the costs of observation and the cost of deviation.

The theory of sequential analysis is devoted to one aspect of this general problem. Other aspects arise in the theory of learning processes, cf. [6] and [4].

It is clear that we can define the worth of the communication channel in a manner completely analogous to the procedure discussed above.

5. A More General Process.

It is clear that both processes are particular cases of a more general process where neither the structure of, nor the transformation due to, the communication channel is completely known. Each stage of the process yields a certain return, which may be negative, in resources, and yields additional information, which may also be negative, concerning the intrinsic structure of the combined system.

The problem is to carry out the sequence of decisions so as to maximize some function, which may not necessarily be completely known, of the total returns and the information pattern.

It is interesting to observe that posed in this way, we encounter one of the basic problems of experimental research.

6. Discussion.

For the above approach to be fruitful, and to represent more technology than tautology, one must possess mathematical techniques capable of formulating in precise terms, and treating, processes of the kind described above.

The theory of sequential analysis developed by Wald, Wolfowitz, Blackwell and Girshick provides an approach to one class of problems of this type, while a general approach to these multi-stage decision processes is provided by the theory of dynamic programming of [1], [2], [3].

As an application of these general methods, we shall consider a simple interesting model proposed recently by Kelly, [6], and some generalizations.

7. The Model of Kelly.

Let us begin by treating the first problem posed by Kelly.

A gambler receives advance information concerning the outcomes of a sequence of independent sporting events over a noisy communication channel. We assume that the outcome of each event is the result of play between two evenly matched teams, and that p is the probability of a correct transmission, and $q = (1-p)$ the probability of an incorrect transmission.

Assuming that the gambler starts with an initial amount x and bets on the outcome of each event so as to maximize his expected capital at the end of N stages of play, it is clear that he wagers his entire fortune on each play if $p > 1/2$, and nothing if $p < 1/2$. If $p = 1/2$, it makes no difference what policy he employs. (We are supposing that the gambler must bet on the received signal, if at all. It is easy to see that if we allow him complete freedom in placing bets, then, in the case where $p < 1/2$, his bet will always be contrary to the information he receives.)

A much more difficult process arises if we take p to be a fixed, but unknown quantity which must be determined on the basis of the observed results of betting. This leads to a "learning process." An expository treatment containing a number of additional references may be found in Robbins, [9], while a treatment by dynamic programming of a similar problem may be found in [4]. If $1/2 < p < 1$, with probability one the gambler will go broke following such a policy.

Let us now assume that the above mode of play appears too hazardous to the gambler, and that he wishes to pursue a more conservative policy, one that will prevent him from ever being wiped out. He may then proceed to maximize the expected value of the logarithm of his capital at the end of N stages of play, see § 14.

For the one-stage process, he is faced with the problem of maximizing

(1) $E_1(y) = p \log(x+y) + q \log(x-y)$

over all y in $[o,x]$. Here y is the amount wagered, fair odds being assumed. It is easy to see that, if $p > q$, we have

(2) $y = (p-q)x$,

and for that value of y

(3) $E_1 = \log x + \log 2 + p \log p + q \log q$.

If $p \leq q$, the maximum is at $y = o$.

It is not difficult to show that if we consider N-stage processes, where we restrict ourselves to wagering policies which require the wagered amount to be a <u>fixed</u> proportion of the total capital at each stage, then the policy described above is optimal. This result was established by Kelly [6], in a very ingenious fashion.

Let us now demonstrate that this policy is optimal within the class of all wagering policies.

8. Economic Forecasting.

It is clear that the above mathematical model is abstractly identical with problems that arise in connection with economic forecasting, in particular, and with forecasting, in general, as for example weather prediction.

In these cases, the physical world is the source and the scientific corps, both experimental and theoretical, the communication channel. Sometimes, the theorist or experimenter is also the observer; at other times, it is the business man or politician who must decide to what extent he trusts his communication channel.

9. Dynamic Programming Approach.*

Let us begin by formulating the problem in dynamic programming terms. Define the following sequence of functions,

(1) $f_N(x)$ = expected value of the logarithm of the final capital obtained from an N-stage process starting with an initial capital x and using an optimal policy.

An optimal policy is here defined as one which maximizes the expected value of the logarithm of the capital at the end of N stages. Using the principle of optimality, [2], we obtain the recurrence relations

*The results contained in this section answer the fundamental question posed by Kelly on p. 926 of [6].

(2) $f_1(x) = \log x + K,$

$$f_N(x) = \underset{0 \le y \le x}{\text{Max}} \left[pf_{N-1}(x+y) + (1-p)f_{N-1}(x-y) \right], \ N \geq 2,$$

where

(3) $$K = \begin{cases} \log 2 + p \log p + q \log q, & p > q, \\ 0, & p \leq q, \end{cases}$$

Let us now demonstrate the

Theorem: For $N \geq 1$, we have

(4) $f_N(x) = \log x + NK,$

where K is defined as above. The optimal policy is unique, and independent of N. It consists of choosing

(5) (a) $y = (p-q)x, \ p > q,$

(b) $y = 0, \qquad p \leq q.$

Proof. Let us proceed inductively, beginning with the known result for $N = 1$. Assuming that the result holds for N, we have,

(6) $$f_{N+1}(x) = \underset{0 \le y \le x}{\text{Max}} \left[p\left[\log(x+y) + NK\right] + (1-p) \left[\log(x-y) + NK\right] \right]$$

$$= \underset{0 \le y \le x}{\text{Max}} \left[p \log(x+y) + (1-p)\log(x-y) \right] + NK.$$

(7) $f_{N+1}(x) = (\log x + K) + NK = \log x + (N+1)K.$

The statement concerning the form of the optimal policy follows from the analytic form of $f_N(x)$.

Now that the "best" performance of the noisy channel has been determined, it may be compared in various possible ways with the performance of a perfect channel.

10. Generalizations I. Time Dependent Case.

Before proceeding to more general cases, let us consider a simple extension of the above model.

To begin with, let us suppose that at the k^{th} stage, the probability of correct transmission is p_k, and of incorrect transmission $q_k = 1 - p_k$. For fixed N, define the sequence of functions

(1) $f_k(x)$ = expected value of the logarithm of the final capital obtained from the remaining k stages of the original N-stage process, starting with an initial capital x, and using an optimal policy.

Then

$$(2) \quad f_1(x) = \underset{0 \le y \le x}{\text{Max}} \left\{ p_N \log(x+y) + q_N \log(x-y) \right\},$$

$$f_k(x) = \underset{0 \le y \le x}{\text{Max}} \left\{ p_{N-k+1} f_{k-1}(x+y) + q_{N-k+1} f_{k-1}(x-y) \right\}, \quad N \ge k \ge 2.$$

As before, it follows inductively that

$$(3) \quad f_k(x) = \log x + k \log 2 + \sum_{r=N-k+1}^{N} \left[p_r \log p_r + q_r \log_q q_r \right],$$

provided that $p_k > 1/2$ for $k = 1,2,\ldots,N$. Whenever this condition fails, the term $p_k \log p_k + q_k \log q_k$ must be replaced by $(-\log 2)$.

11. <u>Generalizations II.</u> <u>Correlation</u>.

Let us now consider the case where the signals are not independent. The simplest case, perhaps, is that where the probability of correct transmission p_k depends upon whether or not the preceding signal was transmitted correctly. Although a large variety of questions of this type may be formulated, we feel that the following discussion will illustrate the uniform method which may be employed to treat them.

Let

(1) p_k = probability of correct transmission of the k^{th} signal, if the (k-1)st signal was transmitted correctly.

r_k = probability of correct transmission of the k^{th} signal if the (k-1)st signal was transmitted incorrectly.

Define the sequence of functions,

(2) $f_k(x)$ = expected value of the logarithm of the final capital obtained from the remaining k stages of the original N-stage process, starting with an initial capital x, and the information that the (k-1)st signal was transmitted correctly and using an optimal policy.

$g_k(x)$ = the corresponding function in the case where the (k-1)st signal was transmitted incorrectly.

Then, as above,

(3) $$f_k(x) = \underset{0 \le y \le x}{\text{Max}} \left[p_{N-k+1} f_{k-1}(x+y) + (1-p_{N-k+1}) g_{k-1}(x-y) \right]$$

$$g_k(x) = \underset{0 \le y \le x}{\text{Max}} \left[r_{N-k+1} f_{k-1}(x+y) + (1-r_{N-k+1}) g_{k-1}(x-y) \right].$$

It follows inductively, as before, that

(4) $f_k(x) = \log x + a_k$,

$g_k(x) = \log x + b_k$,

where the recurrence relations for the a_k and b_k are readily established.

12. Generalizations III. M-Signal Channels.

Let us now consider the situation in which the channel is called upon to transmit any of M different symbols. Upon receiving a symbol the gambler must make bets on what he believes the transmitted signal to have been. Assume that the gambler possesses the following information:

p_{ij} = the conditional probability that the j-signal was sent if the i-signal is received.

q_i = the probability of receiving the i-signal.

r_j = the return from a unit winning bet on signal j.

Finally, let us assume that the gambler is free to bet the amount $z_i \geq 0$ on the i^{th} signal, subject to the restriction that

$\sum_{i=1}^{M} z_i \leq x$. As before, the gambler proceeds so

as to maximize the expected value of the logarithm of his capital after N stages.

Defining the sequence $\{f_N(x)\}$ as above we obtain the relations

$$f_N(x) = \sum_{i=1}^{M} q_i \max_{\substack{\Sigma z_i \leq x \\ z_i \geq 0}} \sum_{j=1}^{M} p_{ij} f_{N-1}(r_j z_j + x - \sum_{s=1}^{M} z_s), N \geq 2,$$

and

$$f_1(x) = \sum_{i=1}^{M} g_i \max_{\substack{\Sigma z_i \leq x \\ z_i \geq 0}} \sum_{j=1}^{M} p_{ij} \log [r_j z_j + x - \sum_{s=1}^{M} z_s).$$

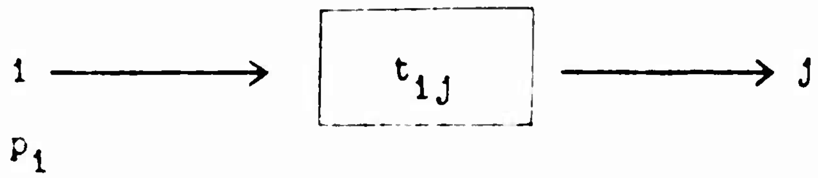
In this case we prove inductively that

$$f_N(x) = \log x + NK,$$

where

$$K = \sum_{i=1}^{M} q_i \max_{\substack{\Sigma z_i \leq 1 \\ z_i \geq 0}} \sum_{j=1}^{M} p_{ij} \log (r_j z_j + 1 - \sum_{x=1}^{M} z_s).$$

From the expression for K it is clear that the optimal policy depends only on p_{ij} and r_j and not on the q_i, though the return itself does also depend on q_i. An interesting special case is that in which it is required that $\sum_{i=1}^{M} z_i = x$; that is, the gambler is required to wager all of his available funds. In this case the optimal policy depends only on (p_{ij}), i.e., on the communication channel, and not at all on q_i or on v_i. If we now introduce



p_i = probability of sending an i.

t_{ij} = the conditional probability that if an i is sent, then j is received,

we then have

$$p_i = \sum_{j=1}^{M} q_j \, p_{ji}$$

$$q_i = \sum_{j=1}^{M} p_j \, t_{ji} \, ,$$

and (t_{ij}) is the inverse of (p_{ij}).

Consequently, in this case the <u>optimal policy</u> is dependent only on (t_{ij}), which characterizes the communication channel and is independent of both the source characterized by p_i and the outside world, characterized by the odds, r_i. The <u>return</u>, however, does depend on all these quantities.

These considerations are significant, for they imply that the gambler's actions are controlled solely by the quality of the communication channel, though his ultimate return is determined by the situation in toto. This leads to the possibility of comparing two channels under the same conditions or of evaluating the performance of a given channel under various conditions.

Specializations to the unsymmetric binary channel are immediate.

13. Generalizations - IV. Continuum of Signals.

Consider now the case where there is a continuum of different signals. Let

(1) $dG(u,v)$ = the conditional probability that a signal with label between v and $v + dv$ is sent if the u - signal is received, $-\infty < u,\ v < \infty$,

and let

(2) $dH(u)$ = the probability that a signal with label between u and $u + du$ is received at any stage.

Then, considering the process corresponding to the special case discussed above, even bets being assumed for the sake of simplicity, we derive the recurrence relations

$$(3)\quad f_N(x) = \int_{-\infty}^{\infty} \left[\underset{z(v)}{\mathrm{Max}} \int_{-\infty}^{\infty} f_{N-1}(2z(v))dG(u,v) \right] dH(u),$$

$$f_1(x) = \int_{-\infty}^{\infty} \left[\underset{z(v)}{\mathrm{Max}} \int_{-\infty}^{\infty} \log(2z(v))dG(u,v) \right] dH(u).$$

In both cases, the maximization is over all functions $z(v)$ satisfying the conditions

(4) (a) $z(v) \geq 0$

(b) $\int_{-\infty}^{\infty} z(v)dv = x.$

As above, it is easily seen inductively that

(5) $f_N(x) = \log 2x + KN,$

where

(6) $$K = \int_{-\infty}^{\infty} \underset{z(v)}{\mathrm{Max}} \left[\int_{-\infty}^{\infty} \log z(v) dG(u,v) \right] dH(u),$$

and

(7) (a) $z(v) \geq 0,$

(b) $$\int_{-\infty}^{\infty} z(v)dv = 1.$$

14. Criterion Functions Yielding Invariant Policies.

We have seen above that the linear function yields an invariant policy at each stage, and likewise the logarithm. It is of interest to determine all criterion functions possessing this property.

The following version of the problem will be treated here. Let $\phi(x)$ be a monotone increasing concave function defined over $0 < x \leq 1$. Consider the one-stage process where we wish to maximize

(1) $E(y) = p\phi(x+y) + (1-p)\phi(x-y).$

The function $E(y)$ is concave as a function of y for $0 \leq y \leq x$, $0 < x \leq 1$, and thus has a unique maximum, unless $\phi(x)$ is linear and $p = 1/2$. Let us dismiss the case of linearity by requiring strict concavity, $\phi''(x) < 0$, and take $p > 1/2$.

Let us assume that, for all x in $0 < x \leq 1$, there is a solution of

(2) $$\frac{dE}{dy} = p\,\phi'(x+y) - (1-p)\,\phi'(x-y) = 0$$

having the form

(3) $y = r(p)x$,

where $r(p)$ is a nonnegative differentiable function of p for $1/2<p\leq 1$, possessing a continuous derivative.

Then (2) is equivalent to the functional equation

(4) $p\phi'(x(1+r(p))) = (1-p)\phi'(x(1-r(p)))$,

for $0 < x \leq 1$, $1/2 < p \leq 1$.

Let $x(1+r(p)) = y$. Then (4) reduces to

(5) $$\frac{p}{1-p}\,\phi'(y) = \phi'\left(\frac{y(1-r(p))}{1+r(p))}\right)$$

We now differentiate first with respect to y, and then with respect to p, obtaining the two equations

(6) $$\frac{p}{1-p}\,\phi''(y) = \left(\frac{1-r(p)}{1+r(p)}\right)\,\phi''\left(\frac{y(1-r(p))}{(1+r(p))}\right),$$

$$\frac{1}{(1-p)^2}\,\phi'(y) = y\frac{d}{dp}\left(\frac{1-r(p)}{1+r(p)}\right)\phi''\left(\frac{y(1-r(p))}{1+r(p)}\right).$$

Dividing the two equations, we obtain

(7) $$\frac{y\phi''(y)}{\phi'(y)} = \frac{u(p)}{p(1-p)\ (du/dp)},$$

where $u(p) = (1-r(p))/(1+r(p))$.

Since the left side is a function of y and the right side a function of p, both sides must be constant. Setting

(8) $$\frac{y\phi''(y)}{\phi'(y)} = K, \quad K \leq 0,$$

we obtain

(9) $\log \phi'(y) = K \log y + c_1$.

Hence

(10) $\phi'(y) = c_2 y^K$.

Without loss of generality, let us normalize, so that $\phi'(1) = 1$. Then

(11) $\phi'(y) = y^K$.

If $K > -1$, we have

(12) $\phi(y) = \frac{y^{K+1}}{K+1} + c_1'$.

If $K = -1$, we have

(13) $\phi(y) = \log y + c'$.

It is clear that $K \geq -1$ is necessary for $\phi(y)$ to be non-negative for $y > 0$. Finally, without loss of generality, we can let $c' = c_1' = 0$.

15. Discussion.

In the foregoing pages, we have essayed to describe some applications of the concepts and techniques of the theory of dynamic programming t various aspects of communication theory. As simple illustrations we have considered a particular process discussed by J. Kelly and various generalizations. In subsequent papers, we propose to treat in greater detail some mathematical models of greater scope.

REFERENCES

1. R. Bellman, Dynamic Programming, Princeton University Press, to appear.

2. ________, The Theory of Dynamic Programming, Bull. Amer. Math. Soc. Vol. 60(1954), pp. 503-515.

3. ________, Some Functional Equations in the Theory of Dynamic Programming, I. Functions of Points and Point Transformations, Trans. Amer. Math. Soc., vol. 80(1955), pp. 51-71.

4. ________, A Problem in the Sequential Design of Experiments, Sankhya, vol. 16(1956), pp. 221-229.

5. J. Bussgang and D. Middleton, Optimum Sequential Detection of Signals in Noise, IRE Trans. on Information Theory, v. IT-1 (1955) pp. 5-18.

6. J. Kelly, A New Interpretation of Information Rate, Bell System Technical Journal, vol. 35(1956), pp. 917-926.

7. D. Middleton and D. Van Meter, Detection and Extraction of Signals in Noise from the Point of View of Statistical Decision Theory, J. Soc. Indust. App. Math., v. 3 (1955), pp. 142-253; and v. 4 (1956), pp. 86-119.

8. S. Rice, Mathematical Analysis of Random Noise, Bell Syst. Tech. Journal, v. 23 (1944), pp. 282-332; and v. 24(1945), pp. 46-156.

9. H. Robbins, Some Aspects of the Sequential Design of Experiments, Bull. Amer. Math. Soc., vol. 58 (1952), pp. 527-536.

10. C. Shannon, A Mathematical Theory of Communication, The Bell System Technical Journal, vol. 27 (1948), pp. 379-423 and pp. 623-656.

11. A. Wald, Statistical Decision Functions, Wiley, New York (1950).

12. N. Wiener, What is Information Theory? IRE Transactions on Information Theory, vol. IT-2 (1956), p. 48.