J. Kelly, 1956.

N tampered coin flips : H – # of "heads"; T – # of "tails"
$H + T = N$, for $N \to +\infty$

$\frac{H}{N} \to p$, $\frac{T}{N} \to q = 1 - p$ ← no "standing on the side or edge";

$p = \frac{1}{2} + \varepsilon$, $\varepsilon > 0$ – heads is slightly more probable

1) Betting it all V_0 – initial amount of money

1.	$2 \cdot V_0$	win ;
2.	$2^2 \cdot V_0$	win ;
3.	$2^3 \cdot V_0$	win ;
4.	ϕ	loss.

exponential growth followed by ruin ("gamblers ruin").

2) Fractional Betting

Bet $l \cdot V_0$ only

cash balance: win $-l \cdot V_0 + V_0 + 2 \cdot l \cdot V_0 = V_0 \cdot (1 + l)$
loss $-l \cdot V_0 + V_0 \qquad = V_0 \cdot (1 - l)$

After N tosses:

$$\begin{cases} V_N = V_0 \cdot (1+l)^W \cdot (1-l)^L \\ W + L = N \end{cases}$$

make the process by finite as $N \to \infty$

$$\begin{cases} \dfrac{V_N}{V_0} = (1+\ell)^{p\cdot N} \cdot (1-\ell)^{q\cdot N} \\ p+q=1 \end{cases}$$

only $G \equiv \frac{1}{N} \log \frac{V_N}{V_0}$ is N-independent:

$$G = p\cdot \log(1+\ell) + q\cdot \log(1-\ell) \to \max$$
$$p+q=1$$

$$\frac{p}{1+\ell} = \frac{q}{1-\ell} \Rightarrow \ell^* = \frac{p-q}{p+q} = 2p-1 = 2\cdot\varepsilon$$

$$\boxed{\ell^* = 2\cdot\varepsilon}$$

For 2% - edge coin bet 4% of your equity.

A New Interpretation of Information Rate

reproduced with permission of AT&T

By J. L. KELLY, JR.

(Manuscript received March 21, 1956)

If the input symbols to a communication channel represent the outcomes of a chance event on which bets are available at odds consistent with their probabilities (i.e., "fair" odds), a gambler can use the knowledge given him by the received symbols to cause his money to grow exponentially. The maximum exponential rate of growth of the gambler's capital is equal to the rate of transmission of information over the channel. This result is generalized to include the case of arbitrary odds.

Thus we find a situation in which the transmission rate is significant even though no coding is contemplated. Previously this quantity was given significance only by a theorem of Shannon's which asserted that, with suitable encoding, binary digits could be transmitted over the channel at this rate with an arbitrarily small probability of error.

INTRODUCTION

Shannon defines the rate of transmission over a noisy communication channel in terms of various probabilities.[1] This definition is given significance by a theorem which asserts that binary digits may be encoded and transmitted over the channel at this rate with arbitrarily small probability of error. Many workers in the field of communication theory have felt a desire to attach significance to the rate of transmission in cases where no coding was contemplated. Some have even proceeded on the assumption that such a significance did, in fact, exist. For example, in systems where no coding was desirable or even possible (such as radar), detectors have been designed by the criterion of maximum transmission rate or, what is the same thing, minimum equivocation. Without further analysis such a procedure is unjustified.

The problem then remains of attaching a value measure to a communication

[1] C.E. Shannon, A Mathematical Theory of Communication, B.S.T.J., **27**, pp. 379-423, 623-656, Oct., 1948.

system in which errors are being made at a non-negligible rate, i.e., where optimum coding is not being used. In its most general formulation this problem seems to have but one solution. A cost function must be defined on pairs of symbols which tells how bad it is to receive a certain symbol when a specified signal is transmitted. Furthermore, this cost function must be such that its expected value has significance, i.e., a system must be preferable to another if its average cost is less. The utility theory of Von Neumann[2] shows us one way to obtain such a cost function. Generally this cost function would depend on things external to the system and not on the probabilities which describe the system, so that its average value could not be identified with the rate as defined by Shannon.

The cost function approach is, of course, not limited to studies of communication systems, but can actually be used to analyze nearly any branch of human endeavor. The author believes that it is too general to shed any light on the specific problems of communication theory. The distinguishing feature of a communication system is that the ultimate receiver (thought of here as a person) is in a position to profit from any knowledge of the input symbols or even from a better estimate of their probabilities. A cost function, if it is supposed to apply to a communication system, must somehow reflect this feature. The point here is that an arbitrary combination of a statistical transducer (i.e., a channel) and a cost function does not necessarily constitute a communication system. In fact (not knowing the exact definition of a communication system on which the above statements are tacitly based) the author would not know how to test such an arbitrary combination to see if it were a communication system.

What can be done, however, is to take some real-life situation which seems to possess the essential features of a communication problem, and to analyze it without the introduction of an arbitrary cost function. The situation which will be chosen here is one in which a gambler uses knowledge of the received symbols of a communication channel in order to make profitable bets on the transmitted symbols.

THE GAMBLER WITH A PRIVATE WIRE

Let us consider a communication channel which is used to transmit the results of a chance situation before those results become common knowledge, so that a gambler may still place bets at the original odds. Consider first the case of a noiseless binary channel, which might be

[2] Von Neumann and Morgenstein, Theory of Games and Economic Behavior, Princeton Univ. Press, 2nd Edition, 1947.

used, for example, to transmit the results of a series of baseball games between two equally matched teams. The gambler could obtain even money bets even though he already knew the result of each game. The amount of money he could make would depend only on how much he chose to bet. How much would he bet? Probably all he had since he would win with certainty. In this case his capital would grow exponentially and after N bets he would have 2^N times his original bankroll. This exponential growth of capital is not uncommon in economics. In fact, if the binary digits in the above channel were arriving at the rate of one per week, the sequence of bets would have the value of an investment paying 100 per cent interest per week compounded weekly. We will make use of a quantity G called the exponential rate of growth of the gambler's capital, where

$$G = \lim_{N \to \infty} \frac{1}{N} \log \frac{V_N}{V_0}$$

where V_N is the gambler's capital after N bets, V_0 is his starting capital, and the logarithm is to the base two. In the above example $G = 1$.

Consider the case now of a noisy binary channel, where each transmitted symbol has probability, p, of error and q of correct transmission. Now the gambler could still bet his entire capital each time, and, in fact, this would maximize the expected value of his capital, $\langle V_N \rangle$, which in this case would be given by

$$\langle V_N \rangle = (2q)^N V_0$$

This would be little comfort, however, since when N was large he would probably be broke and, in fact, would be broke with probability one if he continued indefinitely. Let us, instead, assume that he bets a fraction, ℓ, of his capital each time. Then

$$V_N = (1 + \ell)^W (1 - \ell)^L V_0$$

where W and L are the number of wins and losses in the N bets. Then

$$\begin{aligned} G &= \lim_{N \to \infty} \left[\frac{W}{N} \log(1 + \ell) + \frac{L}{N} \log(1 - \ell) \right] \\ &= q \log(1 + \ell) + p \log(1 - \ell) \text{ with probability one} \end{aligned}$$

Let us maximize G with respect to ℓ. The maximum value with respect to the Y_i of a quantity of the form $Z = \sum X_i \log Y_i$, subject to the constraint $\sum Y_i = Y$, is obtained by putting

$$Y_i = \frac{Y}{X} X_i,$$

where $X = \sum X_i$. This may be shown directly from the convexity of the logarithm.

Thus we put

$$(1 + \ell) = 2q$$
$$(1 - \ell) = 2p$$

and

$$G_{max} = 1 + p \log p + q \log q$$
$$= R$$

which is the rate of transmission as defined by Shannon.

One might still argue that the gambler should bet all his money (make $\ell = 1$) in order to maximize his expected win after N times. It is surely true that if the game were to be stopped after N bets the answer to this question would depend on the relative values (to the gambler) of being broke or possessing a fortune. If we compare the fates of two gamblers, however, playing a nonterminating game, the one which uses the value ℓ found above will, with probability one, eventually get ahead and stay ahead of one using any other ℓ. At any rate, we will assume that the gambler will always bet so as to maximize G.

THE GENERAL CASE

Let us now consider the case in which the channel has several input symbols, not necessarily equally likely, which represent the outcome of chance events. We will use the following notation:

$p(s)$ the probability that the transmitted symbol is the s'th one.

$p(r/s)$ the conditional probability that the received symbol is the r'th on the hypothesis that the transmitted symbol is the s'th one.

$p(s, r)$ the joint probability of the s'th transmitted and r'th received symbol.

$q(r)$ received symbol probability.

$q(s/r)$ conditional probability of transmitted symbol on hypothesis of received symbol.

α_s the odds paid on the occurrence of the s'th transmitted symbol, i.e., α_s is the number of dollars returned for a one-dollar bet (including that one dollar).

$a(s/r)$ the fraction of the gambler's capital that he decides to bet on the occurrence of the s'th transmitted symbol *after* observing the r'th *received* symbol.

Only the case of independent transmitted symbols and noise will be considered. We will consider first the case of "fair" odds, i.e.,

$$\alpha_s = \frac{1}{p(s)}$$

In any sort of parimutuel betting there is a tendency for the odds to be fair (ignoring the "track take"). To see this first note that if there is no "track take"

$$\sum \frac{1}{\alpha_s} = 1$$

since all the money collected is paid out to the winner. Next note that if

$$\alpha_s > \frac{1}{p(s)}$$

for some s a bettor could insure a profit by making repeated bets on the s^{th} outcome. The extra betting which would result would lower α_s. The same feedback mechanism probably takes place in more complicated betting situations, such as stock market speculation.

There is no loss in generality in assuming that

$$\sum_s a(s/r) = 1$$

i.e., the gambler bets his total capital regardless of the received symbol. Since

$$\sum \frac{1}{\alpha_s} = 1$$

he can effectively hold back money by placing canceling bets. Now

$$V_N = \prod_{r,s} [a(s/r)\alpha_s]^{W_{sr}} V_0$$

where W_{sr} is the number of times that the transmitted symbol is s and the received symbol is r.

$$\log \frac{V_n}{V_o} = \sum_{rs} W_{sr} \log \alpha_s a(s/r)$$

$$G = \lim_{N\to\infty} \frac{1}{N} \log \frac{V_N}{V_0} = \sum_{rs} p(s, r) \log \alpha_s a(s/r) \tag{1}$$

with probability one. Since

$$\alpha_s = \frac{1}{p(s)}$$

here

$$G = \sum_{rs} p(s,r) \log \frac{a(s/r)}{p(s)}$$
$$= \sum_{rs} p(s,r) \log a(s/r) + H(X)$$

where $H(X)$ is the source rate as defined by Shannon. The first term is maximized by putting

$$a(s/r) = \frac{p(s,r)}{\sum_k p(k,r)} = \frac{p(s,r)}{q(r)} = q(s/r)$$

Then $G_{max} = H(X) - H(X/Y)$, which is the rate of transmission defined by Shannon.

WHEN THE ODDS ARE NOT FAIR

Consider the case where there is no track take, i.e.,

$$\sum \frac{1}{\alpha_s} = 1$$

but where α_s is not necessarily

$$\frac{1}{p(s)}$$

It is still permissible to set $\sum_s a(s/r) = 1$ since the gambler can effectively hold back any amount of money by betting it in proportion to the $1/\alpha_s$. Equation (1) now can be written

$$G = \sum_{rs} p(s,r) \log a(s/r) + \sum_s p(s) \log \alpha_s .$$

G is still maximized by placing $a(s/r) = q(s/r)$ and

$$G_{max} = -H(X/Y) + \sum_s p(s) \log \alpha_s$$
$$= H(\alpha) - H(X/Y)$$

where

$$H(\alpha) = \sum_s p(s) \log \alpha_s$$

Several interesting facts emerge here

(a) In this case G is maximized as before by putting $a(s/r) = q(s/r)$. That is, *the gambler ignores the posted odds* in placing his bets!

(b) Since the minimum value of $H(\alpha)$ subject to

$$\sum_s \frac{1}{\alpha_s} = 1$$

obtains when

$$\alpha_s = \frac{1}{p(s)}$$

and $H(X) = H(\alpha)$, any deviation from fair odds helps the gambler.

(c) Since the gambler's exponential gain would be $H(\alpha) - H(X)$ if he had no inside information, we can interpret $R = H(X) - H(X/Y)$ as the increase of G_{max} due to the communication channel. When there is no channel, i.e., $H(X/Y) = H(X)$, G_{max} is *minimized* (at zero) by setting

$$\alpha_s = \frac{1}{p_s}$$

This gives further meaning to the concept "fair odds."

WHEN THERE IS A "TRACK TAKE"

In the case there is a "track take" the situation is more complicated. It can no longer be assumed that $\sum_s a(s/r) = 1$. The gambler cannot make canceling bets since he loses a percentage to the track. Let $b_r = 1 - \sum_s a(s/r)$, i.e., the fraction not bet when the received symbol is the r^{th} one. Then the quantity to be maximized is

$$G = \sum_{rs} p(s, r) \log[b_r + \alpha_s a(s/r)], \tag{2}$$

subject to the constraints

$$b_r + \sum_s a(s/r) = 1.$$

In maximizing (2) it is sufficient to maximize the terms involving a particular value of r and to do this separately for each value of r since both in (2) and in the associated constraints, terms involving different r's are independent. That is, we must maximize terms of the type

$$G_r = q(r) \sum_s q(s/r) \log[b_r + \alpha_s a(s/r)]$$

subject to the constraint

$$b_r + \sum_s a(s/r) = 1$$

Actually, each of these terms is the same form as that of the gambler's exponential gain where there is no channel

$$G = \sum_s p(s) \log[b + \alpha_s a(s)]. \tag{3}$$

We will maximize (3) and interpret the results either as a typical term in the general problem or as the total exponential gain in the case of no communication channel. Let us designate by λ the set of indices, s, for which $a(s) > 0$, and by λ' the set for which $a(s) = 0$. Now at the desired maximum

$$\begin{aligned} \frac{\partial G}{\partial a(s)} &= \frac{p(s)\alpha_s}{b + a(s)\alpha_s} \log e = k \qquad for\ s\epsilon\lambda \\ \frac{\partial G}{\partial b} &= \sum_s \frac{p(s)}{b + a(s)\alpha_s} \log e = k \\ \frac{\partial G}{\partial a(s)} &= \frac{p(s)\alpha_s}{b} \log e \leqq k \qquad for\ s\epsilon\lambda' \end{aligned}$$

where k is a constant. The equations yield

$$k = \log e, \quad b = \tfrac{1-p}{1-\sigma}$$

$$a(s) = p(s) - \frac{b}{\alpha_s} \qquad \text{for } s\epsilon\lambda$$

where $p = \sum_\lambda p(s)$, $\sigma = \sum_\lambda (1/\alpha_s)$, and the inequalities yield

$$p(s)\alpha_s \leqq b = \frac{1-p}{1-\sigma} \qquad \text{for } s\epsilon\lambda'$$

We will see that the conditions

$$\begin{aligned} &\sigma < 1 \\ p(s)\alpha_s &> \frac{1-p}{1-\sigma} \qquad \text{for } s\epsilon\lambda \\ p(s)\alpha_s &\leqq \frac{1-p}{1-\sigma} \qquad \text{for } s\epsilon\lambda' \end{aligned}$$

completely determine λ.

If we permute indices so that

$$p(s)\alpha_s \geqq p(s+1)\alpha_{s+1}$$