[2] ARYEH DVORETZKY, "On stochastic approximation," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 39-56, University of California Press, 1956.
[3] J. KIEFER and J. WOLFOWITZ, "Stochastic estimation of the maximum of a regression function," *Ann. Math. Stat.*, Vol. 23 (1952), pp. 462–466.
[4] MICHEL LOÈVE, *Probability Theory*, D. Van Nostrand Company, New York, 1955.
[5] HERBERT ROBBINS and SUTTON MONRO, "A stochastic approximation method," *Ann. Math. Stat.*, Vol. 22 (1951), pp. 400–407.
[6] JEROME SACKS, "Asymptotic Distribution of Stochastic Approximation Procedures," *Ann. Math. Stat.*, Vol. 29 (1958), pp. 373–405.
[7] J. WOLFOWITZ, "On Stochastic Approximation Methods," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 1151–1155.

ON A PROBLEM OF ROBBINS

BY J. R. ISBELL

University of Washington

1. Introduction. This note concerns a sequential decision problem raised by Herbert Robbins [2]. The problem is not solved; in fact, it is not known if there is a uniformly best procedure. A procedure is given here which is uniformly better than the one proposed in [2] and is best at least in a special case.

The nature of the problem is this: given two coins with unknown probabilities p_1, p_2, of coming up heads, to prescribe a rule for making an infinite sequence of tosses, choosing the coin for the nth toss as a function of the history of the sequence since the $(n - r)$-th toss (inclusive). The memory length r is fixed. The aim is to maximize the frequency of heads.

The rule proposed here is best in case p_1 or p_2 is 0. We cannot say *the* best, since many rules have the same effects in this case. The rule may be briefly stated: "*Change coins when one coin shows tails r successive times, or when $r - 1$ tails with one coin are followed by a single toss with the other coin, which is tails*". Robbins' rule [2] calls for changing in these cases and further whenever the first toss with a new coin is tails. For $r \leqq 2$, the rules coincide. Otherwise the present rule is better except in two trivial cases, $p_1 = p_2$ and $\max (p_1, p_2) = 1$.

2. Formulation. The description of the memory requires some amplification for the case $n < r$. (None is given in [2].) Here we shall regard the sequence of tosses as a Markov process with 4^r states, namely the states of the memory. We consider that the process may begin in any state, and we propose to evaluate any procedure according to the results it yields starting from the worst possible state.

This is an artificial description which one might prefer to avoid. On the other hand, any decision procedure which might be optimal according to some other version of the problem but disqualified by our artificial start could be described

Received July 8, 1958; revised January 5, 1959.

as unstable; an experimenter using it could not afford to make finitely many errors in recording his results. This matter will be illustrated with an example at the end of the paper.

Given a rule R, an initial state i, and probabilities p_1 and p_2 of obtaining heads with each coin, one has a definite stationary Markov process. Then for each of the 4^r states s, the frequency of state s converges with probability 1 to a definite limit. The probability of obtaining heads on a toss in state s is either p_1 or p_2 (depending on s and R). The tosses are independent events, and therefore the frequency of heads converges with probability 1 to a definite limit $f(R, i, p_1, p_2)$.

We propose to measure the worth of a rule R by the function $w(R, p_1, p_2)$ defined as the smaller of $\min_i f(R, i, p_1, p_2)$ and $\min_i f(R, i, p_2, p_1)$. Presumably the permutation of p_1 and p_2 here has the same effect as Robbins' requirement [2] that f be symmetric in p_1 and p_2. (This is not an exact statement of Robbins' requirement, and cannot be exact, because of the vagueness concerning the start which we have already noted.) This still does not give us a numerical value for each R, nor does a numerical value seem appropriate. One may hope that for each memory length r there is a rule with maximum worth w for all p_1, p_2.

The rule R_r^* which is the subject of this note prescribes tossing the same coin as on the toss immediately preceding, in all but 4 of the 4^r states; one changes coins when the memory records r tails with the same coin, or $r - 1$ tails with one coin followed by a tail with the other coin.

The worth $w(R_r^*, p_1, p_2)$ is readily computed. For $r = 1$ and $r = 2$ the rule is identical with Robbins' R_r, and the worth is the value given in [2]:

$$\frac{p_1 q_2^r + p_2 q_1^r}{q_1^r + q_2^r} \qquad (r = 1, 2)$$

where $q_j = 1 - p_j$ for $j = 1, 2$. Here Robbins can use an argument on blocks of consecutive tosses with the same coin (untroubled by the vagueness we have noted), because these blocks are independent events. For $r > 2$, using R_r^*, those blocks of tosses are not independent. We have established the existence of $f(R_r^*, i, p_1, p_2)$ by considering 4^r states of the process. For computing its value it is convenient to consider four special states S_i, L_i $(i = 1, 2)$ called *marked* states. S_i is the state in which R_r^* prescribes that we terminate a short sequence (one toss long) of tosses of coin i, L_i the corresponding long sequence-ending state. Except in trivial cases, with probability 1, each marked state is succeeded by 0 or more unmarked states and then a next marked state. The subsequence consisting of the marked states is again a stationary Markov process. Let σ_i, λ_i, be the respective relative frequencies of the marked states. We have $\sigma_1 + \lambda_1 = \frac{1}{2} = \sigma_2 + \lambda_2$, $\sigma_1 = q_1 \lambda_2$, and $\sigma_2 = q_2 \lambda_1$. The proportion $\sigma_1 : \lambda_1 : \sigma_2 : \lambda_2$ is then $q_1 p_2 : p_1 : q_2 p_1 : p_2$. Now the occurrences of S_1 are *followed* by blocks of tosses with coin 2 which terminate precisely when tails first comes up r times; their expected length [2] is $(1 - q_2^r) / p_2 q_2^r$. The blocks following L_1 are governed exactly by Robbins' R_r, and have expected length $1/q_2^r$ [2]. We find w to be

$$\frac{p_1 q_2^r(1 - q_1^r q_2) + p_2 q_1^r(1 - q_1 q_2^r)}{q_2^r(1 - q_1^r q_2) + q_1^r(1 - q_1 q_2^r)}$$

for $r > 2$.

3. The results. The rule R_r^* may, for anything I know to the contrary, be uniformly best for each r. What will be proved here is this:

The worth $w(R_r^*, p_1, p_2)$ is greater than or equal to $w(R_r, p_1, p_2)$ for all values of r, p_1, and p_2, with equality in the three cases $r \leqq 2$, $p_1 = p_2$, $\max(p_1, p_2) = 1$. It is greater than or equal to $w(R, p_1, p_2)$ for any rule R using the same memory length r, in at least two cases: the case $r = 1$, and the case $\min(p_1, p_2) = 0$.

The assertions comparing R_r^* with R_r follow from the previous computation. The assertion concerning $r = 1$ follows from simple computations which the reader may supply. (There are 4 states and 16 rules in this case.)

Consider next the case that $p_1 = 0$ and $r \geqq 3$. We may assume p_2 is not 0 (all rules would have worth 0) nor 1 (R_r^* would have worth 1). Let R be any rule using memory length r. We shall consider subcases, assuming first (a) when the memory records r successive tails with coin 1, R prescribes the use of coin 1 again. But if this is the initial state, then with probability 1 the process consists entirely of tosses of coin 1 and entirely of tails. Thus $w(R, 0, p_2) = 0 < w(R_r^*, 0, p_2)$. Now suppose (b) when the memory records r successive tails with coin 2, R prescribes the use of coin 2 again. The argument under (a) shows that $w(R, p_2, 0) = 0$, and this is the same as $w(R, 0, p_2)$ from the definition of w. There remains subcase (c): neither (a) nor (b). Then with probability 1 each coin is used infinitely often. As in [2], define x_k for $k = 1, 2, \cdots$, as the length of the kth block of consecutive tosses of coin 1, and y_k as the length of the kth block of consecutive tosses of coin 2. Every x_k is at least 1. The expected value of y_k given $x_1, \cdots, x_k, y_1, \cdots, y_{k-1}$, is at most $\lambda = (1 - q_2^r) / p_2 q_2^r$; for this is the expected length of a sequence of tosses terminated precisely at the first run of r consecutive tails, and the present sequence must terminate at or before that run. Then the expected frequency of heads in the first $2k$ blocks is at most $\lambda p_2 / (1 + \lambda)$. We already know that the frequency of heads in the first n tosses converges with probability 1 to $f(R, i, 0, p_2)$, and therefore $f(R, i, 0, p_2) \leqq \lambda p_2 / (1 + \lambda) = w(R_r^*, 0, p_2)$. With a glance at the definition of w, the proof for this case is complete.

By the symmetry of w, the case $p_2 = 0$, $r \geqq 3$, is also accounted for. We need only complete the argument for the case $p_1 = 0$ and $r = 2$. As above, we may assume p_2 is not 0 nor 1; moreover, the subcases (a) and (b) are settled by the same reasoning as above.

We are now considering a rule R using memory length 2 which prescribes changing coins whenever the memory shows two tails with the same coin. Let us subdivide this case according to what R prescribes in case the memory records tails with coin 1 followed by tails with coin 2. (i) Suppose R prescribes changing back to coin 1 in this state. Observe that with probability 1, coin 1 comes up tails every time it is used. Then the argument under subcase (c) of the previous case ($r \geqq 3$, $p_1 = 0$) can be modified; every block of tosses with coin 1 is at least one toss long, and the expected length of any block of tosses with coin 2 is at most $1/q_2^r$. This leads to the conclusion $w(R, 0, p_2) \leqq w(R_2^*, 0, p_2)$ for subcase (i).

We are left with subcase (ii): R prescribes using coin 2 when the memory records "coin 1, tails; coin 2, tails". I do not know an argument for this case which would avoid the computation of six estimates for $w(R, p_2, 0)$. The case hypothesis guarantees that blocks of tosses with coin 2 are at least two tosses long, unless the preceding block of tosses with coin 1 ended with heads. There are three memory states which may have non-zero frequency and which end "coin 1, heads" (only three since coin 2 here never shows heads). Then there are 2^3 possibilities as to what R prescribes in these states; but the eight reduce to six because some rules exclude some memory states.

Five of the six subcases seem absurd—stopping with heads. All I can say is that routine computations suffice to dispose of them. In the remaining case every block of tosses with coin 2 has length exactly 2; a block of tosses with coin 1 lasts at most until tails shows twice, the expected length is at most $(1 - q_2^2) / p_2 q_2^2$, and

$$w(R, p_2, 0) \leqq \frac{p_2(1 - q_2^2)}{1 - q_2^2(1 - 2p_2)} \leqq \frac{p_2(1 - q_2^2)}{1 - q_2^4} = w(R_2^*, p_2, 0).$$

4. Concluding remarks. In my review [1] of Robbins' paper [2], I stated that Robbins' rule "appears to be best for $r = 2$". This remark was based on computation of the effects of the rules which treat the coins symmetrically and never change coins when the last toss was heads; there are only eight of these. It would certainly be surprising if R_2 (which is R_2^*) were not uniformly best; but a proof is still lacking.

Finally, consider the following alternative formulation of the problem, which is consistent with the incomplete description in [2]. For the nth toss, $n \geqq r + 1$, the situation is as we have described it; but for the first r tosses the experimenter may use a special rule treating this part of the process as a collection of transient states. *With this formulation, for $r \geqq 4$, there is no uniformly best rule.* This may be established by checking that there is no single rule which both (a) does as well for $p_1 = 0$, $p_2 = \frac{3}{4}$, as R_r^*, and (b) does as well for $p_1 = 0$, $p_2 = \frac{1}{4}$, as a certain rule S_r described herewith. The effect of S_r will be (with probability 1) to set up an alternation of blocks of tosses with coin 1, one toss long, and blocks of tosses with coin 2 which end only when heads comes up r successive times. When p_2 is less than $\frac{1}{2}$, these are longer than the corresponding blocks which R_r^* gives and which end when tails comes up r successive times. An argument similar to those we have been using shows that this is the best one could possibly hope for. To see that it is possible, examine the rule S_4, which calls for using the coin used last in all but eight memory states: (1–2) four successive heads with either coin; (3–4) three heads with one coin, tails with the other; (5–6) tails twice with one coin, then twice with the other; (7–8) the transient states succeeding the first two tosses, in case the same coin was used and came up tails both times.

One can easily modify S_4 to obtain an S_r, as described above, for $r = 5, 6, \cdots$. A little study of S_r suggests objections to its use even if it is permitted; its worth is a discontinuous function of p_1 and p_2, less than the worth of R_r^* for almost all values. There is also the objection mentioned earlier; the limiting frequency of heads can be changed by finitely many errors.

REFERENCES

[1] J. ISBELL, Review of [2], *Math. Reviews* Vol. 18 (1957), p. 606.
[2] H. ROBBINS, "A sequential decision problem with a finite memory," *Proc. Nat. Acad. Sci.*, Vol. 42 (1956), pp. 920–923.

ACKNOWLEDGMENT OF PRIORITY

BY J. N. K. RAO

Iowa State College

I am grateful to I. M. Chakravarti of The Indian Statistical Institute, Calcutta, for kindly bringing to my attention that Theorem I in my note "A characterization of the normal distribution" (*Ann. Math. Stat.*, Vol. 29(1958), p. 914), had been derived under a less stringent condition by R. G. Laha in two notes, "On an extension of Geary's Theorem" (*Biometrika*, Vol. 40(1953), p. 228) and "On a characterization of the multi-variate normal distribution" (*Sankhya*, Vol. 14(1954), p. 367). I wish to acknowledge the priority of Dr. Laha's results, which were overlooked by me and Seymour Geisser (My note is a follow up of Geisser's "A note on the normal distribution" (*Ann. Math. Stat.*, Vol. 27 (1956), p. 858).

ADDENDUM

BY CHARLES E. CLARK AND G. TREVOR WILLIAMS

Booz, Allen and Hamilton and *The Johns Hopkins University*

The references listed in our "Distributions of the members of an ordered sample" (*Ann. Math. Stat.*, Vol. 29 (1958), pp. 862–870) should have included "Statistical treatment of censored data. I Fundamental formulae," by F. N. David and N. L. Johnson (*Biometrica*, Vol. 41 (1954), pp. 228–240). This earlier paper considers the basic problem of our paper, inter alia. Both papers use power series expansions of the inverse of the distribution function. Since the analysis of the earlier paper leads to expressions in powers of $(N + 2)^{-1}$ and our paper leads to reciprocals of factorials of $N + 2$, many results of the two papers are identical to terms of order N^{-1}; in other words both papers reproduce the classical approximations.