PLAY-THE-WINNER SAMPLING FOR SELECTING THE BETTER OF TWO BINOMIAL POPULATIONS

by

Milton Sobel* and George H. Weiss+

University of Minnesota Technical Report 123

Imperial College, London

July, 1969

* This author was supported in part by a National Institutes of Health Special Fellowship and in part by National Science Foundation Grant GP-9018 while on leave from University of Minnesota, Minneapolis, Minnesota.

+ On leave from National Institutes of Health, Bethesday, Maryland.

ABSTRACT

The sequential allocation of treatments by an experimenter is considered for determining which of two binomial populations has a larger probability of success. Of particular interest in this study is a "Play-the-Winner" (PW) sampling rule which prescribes that one continues with the same population after each success and one switches to the opposite population after each failure. The performance of the PW rule is examined for the selection problem, i.e., for selecting the better population with probability P^* when the two single-trial probabilities, p and p', differ by at least Δ^*, where P^* and Δ^* are prescribed. A comparison is made between PW-sampling and Vector-at-a-time (VT) sampling. In comparing results a criterion used is the expected number of failures that could have been avoided by using the better population throughout. It is shown for a particular common termination rule that with Δ^* close to zero the PW sampling is superior to VT sampling if and only if $\frac{1}{2}(p + p') > 3/4$. Other comparisons are also discussed.

1. Introduction. On the problem of selecting the better of two treatments for the same disease, the past work appears to fall into two different approaches, which we might call the allocation problem and the limiting behavior of the 2-arm bandit problem. After a brief critical discussion of these two, we discuss a third approach which assumes that (i) we wish to have a probability of at least P^* of selecting the better treatment when it is sufficiently better and (ii) that we are interested in minimizing the number of people put on the poorer treatment, or equivalently minimizing the expected number of failures that could have been avoided by using the better treatment throughout.

In the allocation problem there is a known total number N of patients (called the horizon), each patient to be treated with one or the other (but not both) treatments. The first n of the patients are used for selecting the better drug and the remaining $N - n$ patients are treated with the drug selected as better. Aspects of this formulation have been studied by Armitage [1], Anscombe [2], Colton [3] and Zelen [4]. It is assumed in all these studies that the results of each treatment are dichotomous (success or failure) and can be observed without delay, although Zelen [4] does discuss procedures that can be used when the observations may be delayed.

In the 2-arm bandit problem the goal in the literature has been to maximize the limiting output of the process, where the output is defined to be 1 for a success and zero for a failure. This problem has been treated by Robbins [5], Isbell [6], Bradt, Johnson and Karlin [7], Feldman [8], Smith and Pyke [9] and Samuels [10]. In this problem testing goes on forever and a useful procedure is one for which the probability of observing the better population after m trials tends to 1 as $m \to \infty$.

While these two formulations described above lead to interesting theoretical problems, it can be argued in the first case that any forecast of the patient horizon N is very hypothetical, even when the prediction is phased in terms of an a priori distribution. In the second case it is unlikely that any experiment would be allowed to continue indefinitely without deciding between the two treatments. If these objections are valid, then it follows that the problem of real interest involves a finite termination of the trials, together with an infinite (or at least an unknown) patient horizon. Phrased in another way, the problem is that of choosing the better of two binomial populations. This formulation has been considered by several authors (see e.g. Bechhofer, Kiefer and Sobel [11]) and a good set of references can be found at the end of [11].

Much of the previous work on the selection problem involves "Vector-at-a-time" (VT) sampling, which implies an equal number of observations on each treatment. In this paper we wish to compare this sampling rule with the "play-the-winner" (PW) sampling rule first suggested by Robbins [5] in which the treatment used for any trial depends only on the previous trial and its result; a success generates another trial on the same treatment and a failure generates a switch to the other treatment. In particular it is assumed that the results of each treatment are observed without delay.

It might appear that the PW-rule offers the possibility of reaching a decision (with probability of a correct selection $\geq P^*$) with fewer trials on the poorer treatment and indeed this is the motivation for studying this type of procedure in the drug-testing context. We analyze this question for the particular procedures R_S and R_S' both of which terminate when the absolute difference in the number of successes for

the two treatments first reaches a predetermined integer; R_S uses PW sampling and R_S' uses VT sampling. In this case it is not true uniformly in the parameter space that the PW-rule leads to a smaller number of trials on the poorer treatment. In fact for Δ^* close to zero the PW rule is superior to the VT-rule if and only if $\frac{1}{2}(p + p') > 3/4$, where p and p' are the single-trial probabilities of A and B, respectively. The question of whether PW-sampling is uniformly better than VT-sampling for other termination rules remains to be investigated.

2. Calculation of the OC and ASN for Procedures R_S and R_S'.

Let S_A (resp., S_B) denote the current number of successes with treatment A (resp., B). Let P^* with $\frac{1}{2} < P^* < 1$ and Δ^* with $0 < \Delta^* < 1$ denote specified constants. Without loss of generality we regard A as the better player and use $\Delta > 0$ to denote the true difference $p - p'$. We wish to determine to the smallest integer r such that

$$P\{CS\} \geq P^* \quad \text{whenever} \quad \Delta \geq \Delta^*; \tag{2.1}$$

here CS denotes a correct selection which is clearly the selection of A when $\Delta > 0$.

Let NT = A denote that the next treatment is A. Let

$$\begin{aligned} P_n &= P\{A \text{ is selected as better} \mid S_A - S_B = n, \ NT = A\}, \\ Q_n &= P\{A \text{ is selected as better} \mid S_A - S_B = n, \ NT = B\}. \end{aligned} \tag{2.2}$$

The first treatment at the outset is chosen at random and hence the $P\{CS\}$ at termination is given by

$$P\{CS\} = \tfrac{1}{2}(P_0 + Q_0). \tag{2.3}$$

Using the PW-sampling rule and (2.2) and letting $q = 1 - p$, $q' = 1 - p'$, we obtain

$$\begin{aligned} P_n &= pP_{n+1} + qQ_n \\ Q_n &= p'Q_{n-1} + q'P_n \end{aligned} \tag{2.4}$$

and the boundary conditions are

$$P_r = 1,\ Q_{-r} = 0. \tag{2.5}$$

A single equation for Q_n is easily obtained by solving for P_n in the first equation of (2.4) and substituting it into the second equation of (2.4); this gives

$$pQ_{n+1} - (p + p')\,Q_n + p'Q_{n-1} = 0 \tag{2.6}$$

with boundary conditions replaced by

$$Q_{-r} = 0,\ Q_r - p'Q_{r-1} = q'. \tag{2.7}$$

The solution to (2.6), and hence to (2.4), which satisfies these boundary conditions is easily shown to be

$$Q_n = \frac{q'(1 - \lambda^{r+n})}{q' - q\lambda^{2r}}\ ,\quad P_n = \frac{q' - q\lambda^{r+n}}{q' - q\lambda^{2r}}\ , \tag{2.8}$$

where $\lambda = p'/p < 1$. Using these results to obtain the $P\{CS\}$ in (2.3), we then set the result equal to P^* to obtain

$$q' - \tfrac{1}{2}(q + q')\lambda^r = P^*(q' - q\lambda^{2r}) \tag{2.9}$$

as the equation determining r. Solving this quadratic in λ^r gives

$$\lambda^r = (2qP^*)^{-1}\{\frac{q + q'}{2} - \sqrt{(\frac{q + q'}{2})^2 - 4\,qq'P^*(1 - P^*)}\ \}. \tag{2.10}$$

Equations (2.9) and (2.10) do not give a numerical solution if p and p' are unknown. In this case we have to set the minimum of the $P\{CS\}$ for $\Delta \geq \Delta^*$ equal to P^*; this is carried out in section 3 below.

We next turn to calculations relevant to a specific (expected) loss function L related to the number of failures on the poorer treatment. Since there would have been failures even if the better treatment were used exclusively, L is chosen to be the difference in the expected number of successes (before reaching a decision) between a conceptual set of trials in which the better treatment is always used and the actual set of trials. Thus if N_B is the number of trials in which treatment B is used, then L is given by

$$L = (p - p')\, E\{N_B\}. \tag{2.11}$$

We calculate $E\{N_B\}$ by finding recurrence relations between two sets of variables R_n and S_n defined by

$$\begin{aligned} R_n &= E\{N_B \mid S_A - S_B = n,\ NT = A\}, \\ S_n &= E\{N_B \mid S_A - S_B = n,\ NT = B\}. \end{aligned} \tag{2.12}$$

The desired value of L is given by

$$L = \tfrac{1}{2}(p - p')(R_0 + S_0). \tag{2.13}$$

As in (2.4) and (2.5), the PW sampling rule gives

$$\begin{aligned} R_n &= p\, R_{n+1} + q\, S_n, \\ S_n &= p' S_{n-1} + q' R_n + 1, \end{aligned} \tag{2.14}$$

with the boundary conditions

$$R_r = S_{-r} = 0. \tag{2.15}$$

The solution to (2.14) that satisfies (2.15) is

$$R_n = \frac{q(r-n)}{p-p'} - \frac{q(p+2qr)\lambda^r(\lambda^r - \lambda^n)}{(p-p')(q\lambda^{2r} + p\lambda - 1)},$$

(2.16)

$$S_n = \frac{p+q(r-n)}{p-p'} - \frac{(p+2qr)\lambda^r[q\lambda^r - (1-p\lambda)\lambda^n}{(p-p')(q\lambda^{2r} + p\lambda - 1)}.$$

Hence by (2.13) the (expected) loss L is given by

$$L = \frac{(p+2qr)(1-\lambda^r)(1-p\lambda-q\lambda^r)}{2(1-p\lambda-q\lambda^{2r})}. \tag{2.17}$$

An expression can also be found for the expected total number of trials needed to reach a decision; for this we also add one to the first equation of (2.14), just as was done in the second equation. The result denoted by $E\{N_{PW}\}$ is given by

$$E\{N_{PW}\} = \frac{(1-\lambda^r)(1-p\lambda-q\lambda^r)}{(1-\lambda)(1-p\lambda-q\lambda^{2r})}\left(1 + \frac{r}{p}[2 - p(1+\lambda)]\right). \tag{2.18}$$

For the rest of this section we consider the procedure R'_S which uses the VT sampling rule. We now regard the situation after each vector of observations as a stage and our analysis takes us from one stage to another. If P_n is the $P\{CS|S_A - S_B = n\}$ and we stop when $|S_A - S_B| = s$, then P_n satisfies the equation

$$P_n = pq'P_{n+1} + qp'P_{n-1} + (pp' + qq')P_n, \tag{2.19}$$

with boundary conditions

$$P_s = 1, \quad P_{-s} = 0. \tag{2.20}$$

The solution to (2.19) that satisfies (2.20) is

$$(2.21) \qquad P_n = \frac{1 - \delta^{s+n}}{1 - \delta^{2s}} ,$$

where

$$(2.22) \qquad \delta = \frac{p'q}{pq'} < 1.$$

In the next section we show that Min P_0 for $\Delta \geq \Delta^*$ is attained by setting $\Delta = \Delta^*$ and $p = \frac{1}{2}(1 + \Delta^*)$, so that the required s is the solution of

$$(2.23) \qquad \left(\frac{1 - \Delta^*}{1 + \Delta^*}\right)^{2s} = \frac{1 - P^*}{P^*} .$$

For procedure R'_S the expected number of trials on the poorer treatment $E\{N_B\}$ is just the total expected number of vectors until a decision is reached. Let U_n denote $E\{N_B | S_A - S_B = n\}$; then

$$(2.24) \qquad U_n = pq'U_{n+1} + qp'U_{n-1} + (pp' + qq')\, P_n + 1,$$

and the boundary conditions are

$$(2.25) \qquad U_s = U_{-s} = 0.$$

The solution to (2.24) that satisfies (2.25) is

$$(2.26) \qquad U_n = \frac{s(1 + \delta^{2s} - 2\delta^{s+n})}{(1 - \delta^{2s})\Delta} - \frac{n}{\Delta} .$$

The (expected) loss L' for procedure R'_S is therefore

$$(2.27) \qquad L' = \frac{s(1 - \delta^s)}{1 + \delta^s}$$

and the total number of trials N_{VT} needed to reach a decision has expectation

$$(2.28) \qquad E\{N_{VT}\} = \frac{2L'}{\Delta} = \frac{2s(1 - \delta^s)}{\Delta(1 + \delta^s)},$$

which is twice the expected number of vectors needed to reach a decision.

3. Comparison of Results.

Thus far we have presented exact results for procedures R_S and R_S', assuming that p and p' are both known. A proper comparison required us to put ourselves in the position of an experimenter who has no a priori knowledge of these parameters. The most conservative choice of p and p' that satisfies (2.1) will be called the least favorable (LF) configuration.

Consider first the procedure R_S which used the PW sampling rule and let $p_0 = \frac{1}{2}(p + p')$, so that $p = p_0 + \Delta/2$ and $p' = p_0 - \Delta/2$. It is easy to show for any fixed values of p_0, r and P^* that the right side of (2.10) is increasing in Δ and that $\lambda = p'/p$ on the left side of (2.10) is decreasing in Δ. Hence as a first step in obtaining the LF configuration we set $\Delta = \Delta^*$ and write

$$(3.1) \qquad p = p_0 + \Delta^*/2, \qquad p' = p_0 - \Delta^*/2,$$

where the range of p_0 is $(\Delta^*/2,\ 1 - \Delta^*/2)$. We concentrate our attention on the limit $P^* \to 1$ in which case (2.10) becomes

$$(3.2) \qquad \left(\frac{2p_0 - \Delta^*}{2p_0 + \Delta^*}\right)^r = \left(1 + \frac{\Delta^*}{2(1 - p_0)}\right)(1 - P^*)$$

with an error in λ^r that is $\mathcal{O}\{(1 - P^*)^2\}$. To find the 'worst' value of p_0 (for the second step) we maximize r with respect to p_0 in (3.2) and denote the resulting maximum by r_m. A straightforward differentiation yields a transcendental equation, but by taking logarithms in (3.2) we note that the term with $\ln(1 - P^*)$ dominates the right side of (3.2). It follows that the maximum occurs close to the maximum possible value of p_0, i.e., $p_0 = 1 - \Delta^*/2$. Hence, using both

factors on the right side of (3.2), we can approximate r_m by setting $p_0 = 1 - \Delta^*/2$, i.e.,

$$(3.3) \qquad r_m = \left\{\frac{\ln 2(1 - P^*)}{\ln(1 - \Delta^*)}\right\},$$

where $\{x\}$ denotes the smallest integer $\geq x$. To confirm this numerically, we calculated r_m from (3.3) for $\Delta = .1, .2$ and $.4$, $P^* = .99$ and $.999$ and found that the result agreed with the more accurate value $\{\max r(p_0)\}$, obtained by differentiation from (3.2).

Using (3.3) we can give a simple expression for the (expected) loss L in (2.17). If we neglect terms of order $O(1 - P^*)$ in (2.17) (i.e., the terms λ^r and λ^{2r}), then L reduces to

$$(3.4) \qquad L \sim \frac{p}{2} + qr_m \sim q\,\frac{\ln 2(1 - P^*)}{\ln(1 - \Delta^*)} \quad ,$$

where $\frac{p}{2}$ is also dropped since $r_m \to \infty$ as $P^* \to 1$.

Similar but easier calculations show for any fixed values of p_0, s and P^* that δ on the left side in (2.22) is decreasing in Δ and hence we again set $\Delta = \Delta^*$ as the first step in obtaining the LF configuration. From (2.23) and (3.1) we obtain

$$(3.5) \qquad \left[\left(\frac{2p_0 - \Delta^*}{2p_0 + \Delta^*}\right)\left(\frac{2(1 - p_0) - \Delta^*}{2(1 - p_0) + \Delta^*}\right)\right]^s = \frac{1 - P^*}{P^*}$$

and for the second step we wish to find the value of p_0 that maximizes the solution in s of (3.5). It is easily verified that for any fixed s the left side of (3.5) attains its maximum value when $p_0 = \frac{1}{2}$ (see also p.270 of [11]). Hence for $P^* \to 1$ we can write

$$(3.6) \qquad s_m \sim \left\{\frac{\ln(1 - P^*)}{2\ln\left(\frac{1 - \Delta^*}{1 + \Delta^*}\right)}\right\},$$

accurate to within terms that are $\mathcal{O}(1 - P^*)$. From (2.27) and (3.5) we find that under procedure R'_S the (expected) loss L' is given by

$$(3.7) \qquad L' \sim s_m \sim \frac{\ln(1 - P^*)}{2\ln\left(\frac{1 - \Delta^*}{1 + \Delta^*}\right)} .$$

We now investigate the conditions under which $L/L' < 1$, i.e., under which the PW sampling reaches a decision (with the same P^* and) with a smaller number of trials on the poorer treatment. If p and Δ^* are held fixed then the ratio approaches

$$(3.8) \qquad \frac{L}{L'} \sim \frac{qr_m}{s_m} \sim 2q \frac{\ln\left(\frac{1 - \Delta^*}{1 + \Delta^*}\right)}{\ln(1 - \Delta^*)}$$

For small Δ^* the right side of (3.8) is less than 1 when

$$(3.9) \qquad p > \frac{\ln \frac{1 - \Delta^*}{(1 + \Delta^*)^2}}{\ln\left(\frac{1 - \Delta^*}{1 + \Delta^*}\right)^2} \sim \frac{3}{4} - \frac{\Delta^*}{8} + \mathcal{O}(\Delta^{*3}),$$

and hence for P^* close to 1 and Δ^* small the PW sampling is preferred when $p > \frac{3}{4} - \frac{\Delta^*}{8}$ and the VT sampling is preferred when the reverse inequality holds.

Another comparison of interest is that represented by the ratio of the expected total number of treatments needed to reach a decision by the two procedures. For $P^* \to 1$ we use (2.18) to obtain

$$(3.10) \qquad E\{N_{PW}\} \sim \left(\frac{q + q'}{\Delta}\right) r_m \sim \left(\frac{q + q'}{\Delta}\right) \frac{\ln 2(1 - P^*)}{\ln(1 - \Delta^*)} .$$

Furthermore from (2.27) as $P^* \to 1$

$$(3.11) \qquad E\{N_{VT}\} \sim \frac{2s_m}{\Delta} \sim \frac{\ln(1 - P^*)}{\Delta \ln\left(\frac{1 - \Delta^*}{1 + \Delta^*}\right)} ,$$

so that the limiting ratio is

$$(3.12) \qquad \frac{E\{N_{PW}\}}{E\{N_{VT}\}} \sim (q + q') \frac{\ln\left(\frac{1 - \Delta^*}{1 + \Delta^*}\right)}{\ln(1 - \Delta^*)} .$$

For small Δ^* the right side of (3.12) is close to $2(q + q')$ and this is less than 1 when

$$(3.13) \qquad \frac{p + p'}{2} > \frac{3}{4} ,$$

i.e., roughly speaking, when both treatments are quite good. If (3.13) does not hold then from the point of view of speed in reaching a decision it is preferable to sample the treatments equally.

As a point of secondary interest to this paper, we consider the corresponding identification problem in which the parameter values $p > p'$ are both known <u>ab initio</u> and the problem is to correctly identify which treatment is associated with p. We show in this case that the PW sampling is always better than VT sampling. In this situation we have as $P^* \to 1$

$$(3.14) \qquad \frac{E\{N_{PW}\}}{E\{N_{VT}\}} \sim q \frac{r'_m}{s'_m}$$

where r'_m and s'_m are analagous to r_m and s_m but not calculated from the LF configuration. From (2.10) and (2.23), respectively, we obtain for $P^* \to 1$

$$(3.15) \qquad r'_m \sim \frac{\ln(1 - P^*)}{\ln \lambda}, \quad s'_m \sim \frac{\ln(1 - P^*)}{\ln \delta}$$

and hence from (3.14)

$$(3.16) \qquad \frac{E\{N_{PW}\}}{E\{N_{VT}\}} \sim q \frac{\ln \delta}{\ln \lambda} = q \left[1 + \frac{\ln(q/q')}{\ln(p'/p)}\right] \leq 1.$$

To show the last inequality it has to be shown that

$$(3.17) \qquad p^p q^q \leq (q')^q (p')^p$$

Since the right side of (3.17) for fixed p has a minimum at $p' = p$, the inequality (3.17) holds; this proves the last inequality in (3.16).

In conclusion, we see that the PW sampling rule will be advantageous when a certain type of prior information is available and the procedure depends on the absolute difference $|S_A - S_B|$. Roughly speaking if the larger p is greater than $.75 - .125\Delta^*$, then the PW-rule is preferable. If nothing is known about p then our results suggest that a preliminary estimate of both p's should be made to estimate the larger p, but no procedure of this kind has been investigated. Extensions of the PW-rule to the Markovian generalizations suggested by Robbins [5], Isbell [6], Pyke and Smith [9] and Samuels [10] may offer improvements in peformance over the PW-rule, but this remains to be investigated. Further investigation of the PW-rule in connection with other sequential rules would seem to be of interest. This has been done for inverse sampling in [12] where PW sampling (procedure R_I) is shown to be preferable to VT sampling (procedure R'_I) for Δ^* close to zero. Although R_S and R'_S are preferable, i.e., have smaller $E\{N\}$ - and L - values than R_I and R'_I, respectively, the reverse is true for Δ close to zero, Δ^* fixed, and P^* sufficiently close to one. Hence there is no result based on $E\{N\}$ or L that is uniform in both Δ^* and P^*.

REFERENCES

[1] Armitage, P. (1960). Sequential Medical Trials. Springfield, Illinois, Thomas.

[2] Auscombe, F. J. (1963). "Sequential medical trials." Journal of the American Statistical Association Vol. 58, pp.365-383.

[3] Colton, T. (1963). "A model for selecting one of two medical treatments," Journal of the American Statistical Association, 58 pp. 388-400.

[4] Zelen, M. (1969). Play the winner rule and the controlled clinical trial. J. Amer. Statist. Assoc. 64 131-146.

[5] Robbins, H. (1956). "A sequential decision problem with a finite memory," Proceedings of the National Academy of Science, Vol. 42, pp. 920-923.

[6] Isbell, J. (1959). "One a problem of Robbins," Annals of Mathematical Statistics, Vol. 30, pp. 606-610.

[7] Bradt, R. N., Johnson, S. M., Karlin, S. (1956). "On sequential designs for maximizing the sum of n observations," Annals of Mathematical Statistics Vol. 27, pp.1060-1074.

[8] Feldman, D. (1962). "Contributions to the two-armed-bandit problem," Annals of Mathematical Statistics, Vol. 33, pp.847-856.

[9] Smith, C. V. Pyke, R. (1965). "The Robbins-Isbell two-armed-bandit problem with finite memory," Annals of Mathematical Statistics Vol.36, pp. 1375-1386.

[10] Samuels, S. M. (1968). "Randomized rules for the two-armed-bandit with finite memory," Annals of Mathematical Statistics Vol. 39, pp. 2103-2107.

[11] Bechhofer, R. E., Kiefer, J. Sobel, M., (1968). Sequential Identification and Ranking Problems, Chicago, Illinois, University of Chicago Press.

[12] Sobel, M., and Weiss, George H. (1969). "Play-the-winner rule and inverse sampling in selecting the better of two binomial populations," University of Minnesota, Department of Statistics, Technical Report No. 124.