PROCEEDINGS *of the* FIFTH BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY

Held at the Statistical Laboratory
University of California
June 21–July 18, 1965
and
December 27, 1965–January 7, 1966

with the support of
University of California
National Science Foundation
National Institutes of Health
Air Force Office of Scientific Research
Army Research Office
Office of Naval Research

VOLUME IV

BIOLOGY AND PROBLEMS OF HEALTH

EDITED BY LUCIEN M. LE CAM
AND JERZY NEYMAN

UNIVERSITY OF CALIFORNIA PRESS
BERKELEY AND LOS ANGELES
1967

SEQUENTIAL MODELS FOR CLINICAL TRIALS

HERMAN CHERNOFF
STANFORD UNIVERSITY

1. Introduction

This paper presents a model, variations of which have been considered by Anscombe [1] and Colton [4] and others, which is relevant to the problem of sequential testing in clinical trials. This model is the same as one discussed by Chernoff and Ray [3] and by Wurtele [7] in a sampling inspection problem and is naturally related to a *one armed bandit* problem. The object of this paper is to demonstrate that techniques exist for dealing with some of the technical problems raised by these and similar models. A few of the insights derived from the results on the one armed bandit problem will be described in terms of nominal significance levels corresponding to the rejection of a new drug.

The model is oversimplified for many practical applications. Alternative models, including a two armed bandit problem are described. An important element in most of these models is the *horizon* consisting of the total number of anticipated patients to be treated.

2. The model

Suppose that a new drug is produced to treat an illness for which the treatment in the past has been a standard drug with known properties. We shall assume here that the result of the use of the drug can be classified simply as a success or failure in the treatment, and once one drug is applied, treatment cannot shift to the other. Then the known drug is characterized by a known probability p_0 of success while the new drug has unknown probability p of success. If it is anticipated that a horizon of N patients will have to be treated by one drug or another, the expected number of successes given that the new drug is used n times, is $np + (N - n)p_0 = Np_0 + n(p - p_0)$.

Clearly, the expected number of successes attains a maximum of Np_0 if $p < p_0$ (with $n = 0$) and Np if $p > p_0$ (with $n = N$). In view of the ignorance of p, it is desired to select a sequential procedure to maximize the expected number of successes which is equal to

$$Np_0 + E[n(p - p_0)], \tag{2.1}$$

where n is possibly a random quantity determined by the procedure. Since p_0 is known, it is apparent that a reasonable procedure ought to consist of sampling

the new drug until the data indicates it to be inferior, at which time one reverts to the old drug. If the data indicates the new drug to be superior one never reverts to the old one. The reader should bear in mind that the word reasonable is applied in the context of the model which incidentally neglects the cost of administering the new drug and the possible need for control groups. The problem that remains is to describe what constitutes sufficient indication to stop using the new drug. In this situation a variety of factors suggests that it is appropriate to study the problem from a Bayesian point of view especially when N is large. Then the problem of finding a sequential procedure to maximize

$$E\{n(p - p_0)\}, \qquad n \leqq N, \tag{2.2}$$

where p has a given prior distribution, is a well determined optimization problem.

Exactly this problem arises also in the following rectifying sampling inspection problem. A lot of N items has been produced by a process such that distinct items are independently defective with unknown probability p. The cost of inspecting an item is c. If a defective item is found, it is replaced from a pile of good items. The cost of sending a customer a defective item is k times more than the cost of replacing it. If $p < c/k$, it would be preferable to send out the lot without inspection, while if $p > c/k$, 100 per cent inspection is desirable. In fact a sampling plan where n items are inspected leads to an expected cost of

$$nc + (N - n)pk = k\left[Np + n\left(\frac{c}{p} - p\right)\right]. \tag{2.3}$$

It is desired to maximize $E\{n(p - p_0)\}$ where $p_0 = c/k$. It is in this context that the problem was first discussed by Wurtele [7], and subsequently by Chernoff and Ray [3].

A related problem is the following. Let $X_1, X_2, \cdots, X_n$ be independent identically distributed random variables with unknown mean μ (and otherwise known distribution). Given a prior distribution for μ select $n \leqq N$ sequentially, so as to maximize

$$E(X_1 + X_2 + \cdots + X_n) = E\{n\mu\}. \tag{2.4}$$

Viewing $X_1, X_2, \cdots$ as the winnings from a "one armed bandit" by a player who can stay at most long enough to play N games, we may call this problem a *one armed bandit* problem. The two previous problems correspond to the special case where $X_i = 1 - p_0$ or $-p_0$ with probabilities p and $1 - p$, respectively.

The normal continuous time version of this problem is particularly interesting. Let $X(t)$, representing the gambler's gain at time t, be a Wiener process with unknown drift μ and known variance σ^2 per unit time. That is to say, for $t_1 < t_2$, $X(t_2) - X(t_1)$ is normally distributed with mean $\mu(t_2 - t_1)$ and variance $\sigma^2(t_2 - t_1)$ and is independent of the path $X(t)$ for $0 \leqq t \leqq t_1$. Then the continuous time one armed bandit problem consists of finding a sequential procedure for selecting a stopping time $T \leqq N$, so as to maximize $E\{X(T)\} = E\{T\mu\}$ where the

unknown drift is assumed to have a normal prior distribution of mean μ_0 and variance σ_0^2.

To relate the continuous one armed bandit problem with the discrete one, observe that for integer values of t_1 and t_2, $X(t_2) - X(t_1)$ corresponds to the sum of the observations from $t_1 + 1$ to t_2. In our model for clinical trials, $p - p_0$ corresponds to μ while σ^2 could be thought of being approximately $p_0(1 - p_0)$. One anticipates that the solution for the continuous time one armed bandit problem would serve as a reasonable approximation to the discrete versions, especially if N is large.

Chernoff and Ray [3] characterized certain asymptotic properties of the solution of the continuous problem and indicated a rough approximation to the optimal procedure. More refined approximations can be carried out by the use of a numerical computation involving backward induction.

3. The solution

To describe the solution of the continuous time version of the one armed bandit problem, consider first the limiting case where the normal prior distribution has variance $\sigma_0^2 = \infty$ corresponding to what has been termed vague prior knowledge. The asymptotic results of [3] combined with some freehand interpolation and a backward induction suggest the approximation of the solution presented in table I. Here $\tilde{x}(t)$ represents the boundary of the optimal stopping

TABLE I

APPROXIMATION SOLUTION OF THE CONTINUOUS TIME ONE ARMED BANDIT PROBLEM

β = nominal significance level = $\int_{-\infty}^{\tilde{\alpha}} (2\pi)^{-1/2} \exp(-u^2/2)\, du$.

t/N	$\tilde{\alpha} = \tilde{x}/\sigma t^{1/2}$	β
1.00	−0.0	0.50
0.90	−0.20	0.42
0.75	−0.36	0.36
0.50	−0.56	0.29
0.25	−0.78	0.22
0.10	−1.08	0.14
0.01	−2.05	0.02
10^{-4}	−3.55	2.10^{-4}
10^{-6}	−4.61	2.10^{-6}

region. That is, the optimal procedure calls for stopping if $X(t) \leqq \tilde{x}(t)$. The quantities $\tilde{\alpha}$ and β correspond to a nominal significance level as follows. Suppose that at time t the player stopped and decided to test (one tail) the hypothesis $\mu = 0$. The observation $X(t)$ would correspond to $\alpha = X(t)/\sigma t^{1/2}$ standard deviations from the mean 0. Thus, $\tilde{\alpha}$ is the number of standard deviations required for the game to be stopped at time t and β is the corresponding *nominal signif-*

icance level. From a classical point of view one can regard the player as continuously testing the hypothesis $\mu = 0$. As time t varies from 0 to N, the nominal significance level becomes less stringent increasing from 0 to 1/2. If and when he rejects $\mu = 0$ (in favor of $\mu < 0$), he stops playing. Although this nominal significance level serves as a convenient description of the procedure, its use should not be confused with that of the standard significance level which is not applicable here.

Given $X(t) = x$, the posterior distribution of μ is normal with mean x/t and variance $= \sigma^2/t$. This fact permits us to reduce the solution of the problem, where the gambler with horizon N_0 has the normal prior distribution $\mathfrak{N}(\mu_0, \sigma_0^2)$, to the previous problem by simply initiating the Wiener process from the point

$$(x_0, t_0) = (\mu_0\sigma^2/\sigma_0^2, \sigma^2/\sigma_0^2) \tag{3.1}$$

and letting

$$N = N_0 + t_0 = N_0 + \sigma^2/\sigma_0^2. \tag{3.2}$$

The asymptotic results of [3] indicate that as $t/N \to 0$

$$\beta \approx 2t/N \tag{3.3}$$

and

$$\tilde{\alpha} = \frac{\tilde{x}}{\sigma t^{1/2}} \approx -\{2 \log (t/N) - \log [-16\pi \log (t/N)]\}^{1/2}. \tag{3.4}$$

As $t/N \to 1$,

$$\tilde{\alpha} \approx -0.639 \left(1 - \frac{t}{N}\right)^{1/2}. \tag{3.5}$$

When the horizon N_0 is large, the case $t/N \to 0$ is especially important. Here it has been shown that the expected gain is approximated by

$$N_0\sigma_0[\varphi(\alpha_0) + \alpha_0\Phi(\alpha_0)] - \frac{\sigma^2}{2\sigma_0}\left[\log\left(1 + \frac{\sigma_0^2 N_0}{2}\right)\right]^2 \varphi(\alpha_0), \tag{3.6}$$

where $\alpha_0 = \mu_0/\sigma_0$, while $\varphi(u) = (2\pi)^{-1/2}e^{u^2/2}$ and $\Phi(\alpha) = \int_{-\infty}^{\alpha} \varphi(u)\, du$. The first and larger term corresponds to the expected gain if μ were selected from the normal distribution with mean μ_0 and variance σ_0^2 and immediately *told* to the player who would proceed to play the entire allotted time N_0 if $\mu > 0$ and refuse to play otherwise. Thus, the second term represents the expected loss due to ignorance of μ and is of the order of magnitude of $(\log N_0)^2$, which increases slowly as N_0 becomes large.

Now let us relate the clinical trials problem posed in section 2 to the continuous time one armed bandit problem by assuming that the horizon N_0 is large. Assuming vague information about p, let

$$t = n/N_0 \tag{3.7}$$

and

$$\alpha = \frac{n^{1/2}(\hat{p} - p_0)}{[p_0(1 - p_0)]^{1/2}}, \tag{3.8}$$

where $\hat{p}$ is the usual estimate of p based on the first n trials. Then one may substitute directly in table I to determine the stopping time. When n is small compared to N_0, one has the asymptotic relation

$$\alpha = \frac{n^{1/2}(\hat{p} - p_0)}{[p_0(1 - p_0)]^{1/2}} \approx -\left\{2 \log \left(\frac{N_0}{n}\right) - \log \left[16\pi \log \left(\frac{N_0}{n}\right)\right]\right\}^{1/2} \tag{3.9}$$

at the boundary.

Consider the following problem. If each trial of the new drug leads to failure how many successive failures would be required before one should stop the use of the new drug? Clearly, the answer to this question should depend on N_0. Substituting in the above formula, we obtain

$$n_0 \approx 2 \frac{1 - p_0}{p_0} (\log N_0). \tag{3.10}$$

This result should not be taken too seriously since it is implicitly based on the normal approximation to the distribution of $\hat{p}$ which is not especially good for approximating the probability of n successive failures $(1 - p_0)^n$. There is reason to expect that the correct answer to the question posed would be between the answer suggested above and

$$n_0' = (\log N_0)/[-\log (1 - p_0)]. \tag{3.11}$$

In any case the order of magnitude $\log N_0$ is important as we shall see in our subsequent discussion.

If the investigator has some strong feelings about the unknown p which can be represented by a prior Beta distribution $B(a, b)$, for which the mean is $a/(a + b)$ and variance $ab(a + b + 1)/(a + b)$, then the above results are applicable after replacing N_0 by $N = N_0 + (a + b)$, and assuming that $n = a + b$ fictitious trials resulting in a successes had taken place.

It is of some interest to tabulate the estimate $\tilde{p}$ of p for which the clinical investigator should stop the new drug. We have

$$\tilde{p} = p_0 + \tilde{\alpha}\left[\frac{p_0(1 - p_0)}{n}\right]^{1/2}, \tag{3.12}$$

and table II gives some insight when we consider that $2[p_0(1 - p_0)]^{1/2}$ is close to 1 for a broad range of p_0. The large entries corresponding to small n should

TABLE II

$(p_0 - \tilde{p})/2\,[p_0\,(1 - p_0)]^{1/2} = -\tilde{\alpha}/n^{1/2}$

$n\backslash N$	10^2	10^4	10^6
1	2.05	3.55	4.61
5	0.60	1.36	1.91
10	0.34	0.91	1.30
25	0.16	0.52	0.78
100	0	0.21	0.36
1000	—	0.03	0.09

not be taken too seriously since, the normal approximation to the binomial distribution, on which they are based, is not very good.

4. Alternative models

4.1. We shall discuss a variety of situations where the model presented in section 2 for clinical trials requires some modification. First let us consider the case where the two competing treatments being compared for a horizon of N patients are both of unknown efficacy. This corresponds to the two armed bandit problem where the player attempts to maximize

$$E\{X_1 + X_2 + \cdots + X_N\}, \tag{4.1}$$

where X_n represents the value of the outcome of the nth trial whose treatment is determined by past history. The continuous time normal version of this problem consists of maximizing

$$E\{X_1(T_1) + X_2(T_2)\} = E\{T_1\mu_1 + T_2\mu_2\}, \qquad T_1 + T_2 = N, \tag{4.2}$$

where X_1 and X_2 are Wiener processes with unknown drifts μ_1 and μ_2 per unit time and known variance σ^2 per unit time. Both μ_1 and μ_2 are assumed to have normal independent prior distributions. At any given moment one is entitled to observe either X_1 or X_2 depending on past history, T_i is the total time spent observing X_i, and $T_1 + T_2 = N$.

While a special version of the two armed bandit problem was solved by Feldman [5], little is known about the solution of this one. It is intuitively clear that the solution will call for using the arm (Wiener process) which has the higher estimated mean until a balance is struck between the difference in the estimated means and the amount of information accumulated on each arm and the remainder of the horizon. Thus, for a large horizon, the optimal procedure may call for the arm with the lower estimated mean drift if that estimate is based on a relatively small sample time.

4.2. Suppose two unknown treatments are being compared in a variety of locations, at each one of which the number of patients available is rather small. In that case control considerations suggest the model proposed by Anscombe [1] and Colton [4]. Here, patients are paired and one of each pair is randomly selected for one treatment while the other is given the second treatment. The number of pairs to be treated n is determined sequentially after which the remainder of the horizon $N - 2n$ are given the treatment estimated to be superior.

If μ is the mean of the difference X in the treatment effects, one seeks to maximize

$$E\{X_1 + X_2 + \cdots + X_n - (N - 2n)\mu\epsilon(\mu)\}, \tag{4.3}$$

where $\epsilon(\mu)$ is the probability of selecting the wrong treatment, given μ. Colton presents a detailed series of analyses of procedures which have certain optimality properties. One shortcoming is the restriction of his sequential procedures to

those of the Wald type which may lead to rather poor results when the horizon is very large. In particular if one treatment is substantially inferior to the other and N is very large, the Colton procedures call for sample sizes of order $N^{1/2}$ instead of the more reasonable log N.

Anscombe derives an "outer" bound on the optimal Bayes procedure. This bound corresponds to the use of a nominal significance level of n/N. This model is subject to the same technical approach as the one armed bandit problem and for n/N small, the nominal significance level should be approximately $2n/N$. In the sense that the difference in the α corresponding to these significance levels is small, the Anscombe result furnishes a good approximation to the optimal procedure.

4.3. If the experimental situation calls for controls but one treatment is well known, our problem may be regarded as maximizing

$$E\{X_1 + X_2 + \cdots + X_n\} = E\{n\mu\}, \qquad n \leqq N/2, \tag{4.4}$$

where n is the number of pairs treated before returning to the known treatment. Here μ is the mean of X, the difference in the effect between the new and old treatments. This problem is also a one armed bandit problem with the horizon N replaced by $N/2$, the available number of pairs.

5. Discussions and more models

The models discussed so far are oversimplified to say the least. In practical problems where specific shortcomings of the models are important, the models can be modified so as to be more meaningful and still capable of analysis.

One serious difficulty is the specification of the horizon N. Even if one can regard such a conception as meaningful because of anticipated changes in technology it is difficult to make the choice of a number to represent N. The fact that many of the important decisions and losses involve N only through log N can serve to give the experimenter the assurance that an incorrect specification of N will hardly affect the procedures. It is remarkable how little effect is due to changing N from 10^6 to 10^{10}.

Should one wish to conceive of N as infinite, it is possible to consider a model where the effects of future treatments are discounted. Thus, one may seek to maximize

$$E\{X_1 + \rho X_2 + \rho^2 X_3 + \cdots + \rho^{n-1} X_n + \cdots\}, \tag{5.1}$$

where X_i is the outcome of the ith treatment which can be one of the two alternatives and ρ is the discount factor between 0 and 1.

The problem of medical ethics seems to be unavoidable in experiments involving clinical treatments. It is difficult to imagine a reasonable experiment where one can be assured that subjects will never be given treatments estimated to be inferior. One should be prepared to try a new treatment again even if it fails on its first trial and the skimpy evidence is unfavorable to it. One approach to quantifying the cost of treating a patient with a drug currently estimated to be

inferior is to add a charge, real or imaginary, proportional to $p_0 - \hat{p}$ when a patient is given the treatment with estimated probability of success $\hat{p}$ if $\hat{p} < p_0$.

I am not well enough acquainted with the field to present a reasonable discussion of the ethical problem and shall not attempt to do so.

The Anscombe and Colton model represents a simplification of one presented previously by Maurice [6]. This more complicated model also had a cost for experimentation on the first n pairs of patients. The Anscombe and Colton model neglects the cost of experimentation. The opposite extreme version of the Maurice model would be to summarize the effect of giving the entire horizon the wrong treatment by a multiple of the mean difference. This extreme leads simply to the more standard problem of sequentially testing whether the unknown mean is positive or negative.

A variety of procedures and models are presented by Armitage [2]. While the procedures are not optimal Bayes procedures, the general impression one receives is that they are rather efficient for situations where sample sizes are expected to be moderate, but that there is substantial loss when very large samples are anticipated. Considerable work remains to be done in analyzing various models and comparing optimal with standard procedures.

As a final remark, I would like to add that one sometimes encounters a naive conception that situations where control is desired require pairing of treatments. In many instances control may be achieved when two treatments are given in ratios of 1 to 2, or 1 to 3, and so forth. Not only can such ratios give more efficient results but it is quite likely that ethical requirements would also point in this direction.

REFERENCES

[1] F. J. ANSCOMBE, "Sequential medical trials," *J. Amer. Statist. Assoc.*, Vol. 58 (1963), pp. 365–383.

[2] P. ARMITAGE, *Sequential Medical Trials*, Springfield, Thomas, 1960.

[3] H. CHERNOFF and S. N. RAY, "A Bayes sequential sampling inspection plan," *Ann. Math. Statist.*, Vol. 36 (1965), pp. 1387–1407.

[4] T. COLTON, "A model for selecting one of two medical treatments," *J. Amer. Statist. Assoc.*, Vol. 58 (1963), pp. 388–400.

[5] D. FELDMAN, "Contributions to the 'Two-armed Bandit' problem," *Ann. Math. Statist.*, Vol. 33 (1962), pp. 847–856.

[6] R. MAURICE, "A different loss function for the choice between two populations," *J. Roy. Statist. Soc. Ser. B*, Vol. 21 (1959), pp. 203–213.

[7] Z. WURTELE, "Continuous sampling plans from the risk point of view," unpublished Ph.D. thesis, Columbia University, 1951.