SEQUENTIAL DECISION PROBLEMS FOR PROCESSES WITH CONTINUOUS TIME PARAMETER. TESTING HYPOTHESES[1]

A. DVORETZKY, J. KIEFER AND J. WOLFOWITZ

Hebrew University, Jerusalem, Cornell University and University of California at Los Angeles

Summary. The purpose of the present paper is to contribute to the sequential theory of testing hypotheses about stochastic processes with a continuous parameter (say, t which one may think of as time). Sequential decision problems about such processes seem not to have been treated before. Subsequently we shall treat problems of point and interval estimation and general sequential decision problems for such processes. The results, in addition to their interest per se and their practical importance, also shed light on the corresponding results for discrete stochastic processes. The subjects of sequential analysis and the theory of decision functions were founded by Wald, and we treat our present subjects in the spirit of his approach. The general results of decision theory, such as the complete class theorem, carry over to sequential problems about stochastic processes with continuous time parameter. As specific examples we treat the Wiener and Poisson processes and obtain, for example, the exact power function. (For discrete processes the corresponding known results, due to Wald, are approximations).

1. Introduction. Let $\{x_1(t), t \geqq 0\}$ and $\{x_2(t), t \geqq 0\}$ be two different stochastic processes. The statistician observes continuously, beginning at $t = 0$, a process $\{x(t), t \geqq 0\}$ which is either $\{x_1(t)\}$ or $\{x_2(t)\}$, and wishes to decide, as soon as possible, whether $\{x(t)\}$ is $\{x_1(t)\}$ or $\{x_2(t)\}$. "As soon as possible" means the following here. Let T be the time when he reaches a decision (in general this may be a chance variable and need not be a constant). Let E_iT denote the expected value of T when $\{x(t)\} = \{x_i(t)\}$, $i = 1, 2$. Let α_1, α_2 be two positive constants, $\alpha_1 + \alpha_2 < 1$. Subject to the requirement that the probability of an incorrect decision when $\{x(t)\} = \{x_i(t)\}$ be at most α_i, the problem is to give a procedure for deciding between $\{x_1(t)\}$ and $\{x_2(t)\}$ such that $E_i(T)$ is a minimum for $i = 1, 2$. This is simply the same formulation for stochastic processes with continuous parameter as was originally given by Wald ([3], [4]) for stochastic processes with a discrete parameter.

In this paper we shall limit ourselves to stochastic processes which fulfill the following conditions. For every $t \geqq 0$, $x(t)$ is a sufficient statistic for the process, that is, the conditional distribution of the chance function $x(\tau), 0 \leqq \tau \leqq t$, given $x(t)$, is, with probability one for every t, the same for the processes $\{x_1(t)\}$

Received 11/4/52.

[1] This work was sponsored by the Office of Naval Research under a contract with Columbia University and under a National Bureau of Standards contract with the University of California at Los Angeles.

and $\{x_2(t)\}$. For every t and x, both $x_1(t)$ and $x_2(t)$ have frequency functions, say $f_1(x, t)$ and $f_2(x, t)$, respectively. Let

$$\dotplus(t) = \log \frac{f_2(x(t), t)}{f_1(x(t), t)} \qquad (\dotplus(0) = 0). \tag{1.1}$$

Finally we postulate that the $\dotplus(t)$ process is one of stationary independent increments, that is, a) for every positive integral k, every $h > 0$, and every sequence $t_1 < t_2 < \cdots < t_k \leqq t$, $\dotplus(t + h) - \dotplus(t)$ is distributed independently of $\dotplus(t_1), \cdots, \dotplus(t_k)$; b) the distribution of $\dotplus(t + h) - \dotplus(t)$ depends only upon h and not upon t.

Thus our theory will include the following problems: 1) testing hypotheses about the parameter of a continuous Poisson process with stationary independent increments (to be discussed in detail below in Section 3); 2) testing hypotheses about the mean of a Wiener process (to be discussed in detail below in Section 4); 3) testing hypotheses about the value of $p(0 < p < 1)$ in the following process with stationary independent increments (called the negative binomial): the probability that $x(t) = k$ for every nonnegative integer k is

$$\Gamma(t + k)p^t(1 - p)^k/\Gamma(k + 1)\Gamma(t);$$

4) testing hypotheses about the value of $\theta(\theta > 0)$ in the following process with stationary independent increments (called the Gamma process): the probability density of $x(t)$ at $x(x \geqq 0)$ is given by $x^{t-1}e^{-x/\theta}/\Gamma(t)\theta^t$.

In practice it is, of course, impossible to observe without error a sample function of a continuous process such as the Poisson process or the Wiener process. Yet in many cases these processes do constitute an excellent approximation to physical reality. For example, the incidence of mesons on a Geiger counter is generally assumed to follow a Poisson process. If the recording lag and the dead time of the Geiger counter are very small, a physicist could use the present theory to decide between two possible values of meson density. In this case continuous observation means simply exact registration of incidence times. As another example, our method, or a modification of it, may be applied to problems of life testing.

Moreover, there are several distinct advantages of the continuous parameter procedure over the discrete one. These are as follows.

The expected duration of observing the process before reaching a decision about which hypothesis to adopt can obviously only be shortened by allowing continuous observation.

Moreover, there are many cases, notably the Poisson and Wiener processes, in which an exact determination of the optimal procedure is possible in the continuous case, while in the discrete case so far only approximations have been derived. Thus, even when treating the discrete case, the continuous case, which is easier to treat, may be used to derive approximations when the unit of time is small.

There may also be other advantages in special problems. Thus it is seen in

Section 3 that in the continuous Poisson process the solution does not depend, as in the discrete case, on the values of the two parameters λ_1 and λ_2, but only on their ratio λ_2/λ_1.

2. Application of the Wald sequential procedure. Optimum character of the test. A careful examination of the results of [5] and [6] shows that their conclusions in no way require that the processes be discrete in time, and under the assumptions about the processes made in the preceding section the following results hold.

i) Let a and b, $b < 0 < a$, be given numbers, and let us employ the Wald sequential probability ratio test as follows. As long as $\ell(t)$ lies between b and a, continue observing $\{x(t)\}$. As soon as $\ell(t) \leqq b$, stop observing $\{x(t)\}$ and decide $\{x(t)\} = \{x_1(t)\}$. As soon as $\ell(t) \geqq a$, stop observing $\{x(t)\}$ and decide $\{x(t)\} = \{x_2(t)\}$. Let $\alpha_i(a, b)$ be the probability of error and $E_i(T \mid a, b)$ be the expected value of T when $\{x(t)\} = \{x_i(t)\}$, $i = 1, 2$. For any other procedure with respective probabilities of error α_1^* and α_2^* and respective expected values E_1^*T an E_2^*T, we have that $\alpha_i^* \leqq \alpha_i$, $i = 1, 2$ implies $E_i^*T \geqq E_i(T \mid a, b)$, that is, the optimum character of the Wald sequential probability ratio test (with respect to all randomized as well as nonrandomized procedures).

ii) Let c, W_1 and W_2 be positive numbers, and let g_i be the a priori probability that $\{x(t)\} = \{x_i(t)\}$, $i = 1, 2$ (cf. remarks about a priori probability distributions in [5] and [6]). There exist two numbers $a(c, W_1, W_2, g_1, g_2)$ and $b(c, W_1, W_2, g_1, g_2)$ such that, if the statistician continues to observe $\{x(t)\}$ until either $\ell(t) \leqq b$ or $\ell(t) \geqq a$, and then decides respectively that $\{x(t)\} = \{x_1(t)\}$ or $\{x(t)\} = \{x_2(t)\}$, he will minimize $g_1(\alpha_1 W_1 + cE_1T) + g_2(\alpha_2 W_2 + cE_2T)$ with respect to all possible procedures for deciding between $\{x_1(t)\}$ and $\{x_2(t)\}$, where E_iT is the expected value of T when $\{x(t)\} = \{x_i(t)\}$, $i = 1, 2$. (It is of course assumed that $a \geqq b$, with the equality sign not excluded. Also $\ell(0) = 0$. Thus if $a = b$, or $a \leqq 0$ or $b \geqq 0$, the decision will always be made at time $t = 0$.)

It is to be understood that any procedure which the statistician will employ should be such that the quantities α_1, α_2, E_1T, and E_2T will be well defined. The consideration of questions of measurability is a little more involved for our problem than it is in [5] and [6], but because of the assumptions on the processes made in the preceding section it can be carried out without difficulty. We shall therefore omit consideration of such questions.

From the remarks at the end of Section 1 and well known results of sequential analysis (see Stein [2]), it follows that $E_iT^k < \infty$ for any sequential probability ratio test and any positive k.

Other important results of sequential analysis established for discrete processes apply also to the continuous parameter case. For example, let $\{z(t), t \geqq 0\}$ $(z(0) = 0)$, be a process with stationary independent increments. Assume that $Ez(1)$ exists and denote it by h. Suppose that one has any stopping rule, that is, there is defined a positive chance variable T such that the set of chance functions for which $T = t$ is defined only by conditions on $z(\tau)$, $0 \leqq \tau \leqq t$. Then Wald's equation ([3], [7])

$$Ez(T+) = hE(T) \tag{2.1}$$

holds. Suppose also that $Ee^{uz(1)}$ exists for all real u, and denote it by $\phi(u)$. Then Wald's fundamental identity ([4], p. 159)

$$Ee^{uz(T+)}(\phi(u))^{-T} = 1 \tag{2.2}$$

holds for many stopping rules, including in particular the rule where $T = t$ if $z(t) \geqq a$ or $z(t) \leqq b$, while $b < z(\tau) < a$ for $\tau < t$. Here a and b are constants, $a > 0, b < 0$. The simplest way of proving these results is to derive them as immediate consequences of a theorem of J. L. Doob on martingales with a continuous parameter ([1], Chap. VII, Theorem 11.8). For (2.1) the martingale process is $\{z(t) - ht\}$, and for (2.2) the martingale process is $\{e^{uz(t)}(\phi(u))^{-t}\}$. Another, more laborious way, of proving these results is to consider the process $\{z(t)\}$ only at time intervals which are integral multiples of Δ, proceed as in [4] or [7], and then let Δ approach zero. This is, however, a laborious way of proving a special case of the martingale theorem.

3. The Wiener process. Let $\{x_1(t)\}$ and $\{x_2(t)\}$ be Wiener processes ($t \geqq 0$, $x_1(0) = x_2(0) = 0$) each with a variance which without loss of generality we take to be one per unit of time. Let m_1 and m_2 ($m_1 \neq m_2$) be the mean values per unit time of $\{x_1(t)\}$ and $\{x_2(t)\}$, respectively. Thus we have the following situation: $t \geqq 0$ is a continuous (time) parameter. For any a_1, $a_2(0 < a_1 < a_2)$, $x_i(a_2) - x_i(a_1)$ is normally distributed with mean $m_i(a_2 - a_1)(i = 1, 2)$ and variance $(a_2 - a_1)$. For any integral k and sequence $a_1^1 < a_2^1 \leqq a_1^2 < a_2^2 \leqq \cdots \leqq a_1^k < a_2^k$, the k chance variables $x_i(a_2^j) - x_i(a_1^j), j = 1, \cdots, k, i = 1, 2$, are independently distributed. The statistician observes continuously, beginning at $t = 0$, a process $\{x(t)\}$ which is either $\{x_1(t)\}$ or $\{x_2(t)\}$, and wishes to decide whether $\{x(t)\} = \{x_1(t)\}$ or $\{x(t)\} = \{x_2(t)\}$.

At time t_0 the quantity $x(t_0)$ is sufficient for deciding between $\{x_1(t)\}$ and $\{x_2(t)\}$, that is, it is unnecessary to know the previous history of the process. The likelihood ratio $L(x(t), t))$ at time t is given by

$$L(x(t), t) = \frac{\dfrac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}((x(t)-m_2 t)^2/t)}}{\dfrac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}((x(t)-m_1 t)^2/t)}}.$$

Hence

$$\ell(t) = x(t)\,(m_2 - m_1) - \frac{t}{2}(m_2^2 - m_1^2).$$

The sample functions of the $\{x(t)\}$ process are continuous with probability one. We choose a and b, $b < 0 < a$, such that the statistician will continue to observe $\{x(t)\}$ only until $\ell(t) = a$ or $\ell(t) = b$. In the first case he will decide that $\{x(t)\} =$

$\{x_2(t)\}$, in the second case that $\{x(t)\} = \{x_1(t)\}$. We shall now find $\alpha_i(a, b)$, $E_i(T \mid a, b)$, and the distribution function of T. The same problem for the discrete stochastic process when one observes $\{x(t)\}$ only at $t = 1, 2, \cdots$ has been studied by Wald ([3], [4]) who gave, inter alia, approximations for these quantities. An examination of his argument shows that, in his problem, his results are approximate only because he neglects the excess of ł(T) over a or b. In our problem this excess is zero with probability one, and Wald's formulae cease to be mere approximations and become exact. Thus we have, for example, ([4], p. 50, equation (3.42))

$$\alpha_1(a, b) = \frac{1 - e^b}{e^a - e^b} \tag{3.1}$$

$$\alpha_2(a, b) = \frac{e^b(e^a - 1)}{e^a - e^b}. \tag{3.2}$$

For any Wiener process with variance one per unit of time, not necessarily either $\{x_1(t)\}$ or $\{x_2(t)\}$, the probability that ł(t) will reach b before reaching a is given *exactly* by [4], page 50, equation (3.43). Call this probability H. Then, for any Wiener process with variance one per unit of time, not necessarily $\{x_1(t)\}$ or $\{x_2(t)\}$, $ET = (Hb + (1 - H)a)/h$ ([4], page 53, equation (3.57)). These results can be derived from (2.1) and (2.2) by Wald's methods. Also the density function of T is given exactly by formula (A:194) on page 195 of [4].

In practice one has to find a and b to correspond to given values α_1 and α_2. Solving (3.1) and (3.2) we obtain

$$a = \log \frac{1 - \alpha_2}{\alpha_1}, \tag{3.3}$$

$$b = \log \frac{\alpha_2}{1 - \alpha_1}. \tag{3.4}$$

All of the above results are exact because the excess of ł(T) over the boundaries a and b is zero with probability one. For the same reason one may already infer the optimal character of the Wald sequential probability ratio test for testing hypotheses about the mean of a Wiener process from the approximations and heuristic arguments given by Wald on pages 196–199 of [4].

One may raise the question how to test hypotheses about the variance of a Wiener process. However, a scrutiny of the problem shows that from a knowledge of a sample function in any interval, no matter how small, one can, with probability one, determine the variance to any arbitrary accuracy, so that the problem is trivial. For suppose $\{x(t)\}$ is a Wiener process with mean value m and variance v, both per unit of time. Suppose the process has been observed from $t = 0$ to $t = \mathrm{H}_0$, where H_0 is any positive number. Let N be any integer which will later approach infinity, and write $t_i = iH_0/N$, $i = 0, 1, \cdots, N$. For any i from 1 to N we have

$$E(x(t_i) - x(t_{i-1}))^2 = v\frac{H_0}{N} + m^2\frac{H_0^2}{N^2}.$$

Now, for $i = 1, \cdots, N$, the chance variables

$$\left\{(x(t_i) - x(t_{i-1}))^2 - v\frac{H_0}{N} - m^2\frac{H_0^2}{N^2}\right\}$$

are identically and independently distributed, with variance of order N^{-2} and fourth moment of order N^{-4}. Hence the fourth moment of

$$\frac{\sum_{i=1}^{N}(x(t_i) - x(t_{i-1}))^2}{H_0} - v - \frac{m^2H_0}{N}$$

is of order N^{-2}. Consequently, for any $\epsilon > 0$ we have that

$$P\left\{\left|\frac{\sum_{i=1}^{N}(x(t_i) - x(t_{i-1}))^2}{H_0} - v - \frac{m^2H_0}{N}\right| > \epsilon\right\} < \frac{C}{\epsilon^4N^2}$$

where C is a suitable constant. Since the series $\sum N^{-2}$ converges it follows immediately from the Borel-Cantelli lemma that $(\sum_{i=1}^{N}(x(t_i) - x(t_{i-1}))^2)/H_0$ converges to v with probability one as $N \to \infty$.

4. The Poisson process. In this section we treat the problem of deciding which of two values given in advance represents the correct mean occurrence time of a Poisson process with stationary independent increments.

The probability that a Poisson process with mean occurrence time λ will result in exactly k occurrences between times $t = 0$ and $t = T$ is

$$\frac{(\lambda T)^k}{k!}e^{-\lambda T} \qquad (k = 0, 1, 2, \cdots). \tag{4.1}$$

Let $H_i (i = 1, 2)$ denote the hypothesis that $\lambda = \lambda_i$, where λ_1 and λ_2 are any two different positive numbers. It is clear that the two corresponding processes satisfy the conditions imposed in the introduction. Hence, given two positive numbers α_1, α_2, $(\alpha_1 + \alpha_2 < 1)$, the optimal test procedure for deciding between H_1 and H_2 which satisfies the condition that the probability of a wrong decision when H_i is true does not exceed $\alpha_i (i = 1, 2)$ is given by a Wald sequential probability ratio test.

More specifically, in view of (4.1) we have

$$\psi(t) = x(t)\log\frac{\lambda_2}{\lambda_1} + (\lambda_2 - \lambda_1)t. \tag{4.2}$$

Thus, assuming $\lambda_2 > \lambda_1$, the best decision rule is specified by two numbers $a, b (b < 0 < a)$ in the manner described in the introduction.

Suppose now that α_1 and α_2 are the actual probabilities of error. According to Wald ([4], p. 196) we have

$$\frac{1 - \alpha_2}{\alpha_1} = \frac{P_2(H_2)}{P_1(H_2)}, \qquad \frac{\alpha_2}{1 - \alpha_1} = \frac{P_2(H_1)}{P_1(H_1)}, \tag{4.3}$$

where $P_i(H_j)$ is the probability that hypothesis H_j is accepted when hypothesis H_i is true. By the argument used by Wald we have

$$e^{\inf_i I(T)} = \inf_i e^{I(T)} \leqq \frac{P_2(H_i)}{P_1(H_i)} \leqq \sup_i e^{I(T)} = e^{\sup_i I(T)} \tag{4.4}$$

the $\sup_i$ and $\inf_i$ being taken over all values of $\dotplus(T)$ where the observation is stopped at time T with the decision to adopt H_i. In our case we know that if the decision to accept H_2 is adopted at time T we must have $\dotplus(T) \geqq a$, while $\dotplus(t) < a$ for $t < T$. Since (see (4.2)) $\dotplus(t + 0) - \dotplus(t) \leqq \log \lambda_2/\lambda_1$ with probability 1 we have from (4.3) and (4.4)

$$e^a \leqq \frac{1 - \alpha_2}{\alpha_1} \leqq \frac{\lambda_2}{\lambda_1} e^a. \tag{4.5}$$

Similarly if at time T we decide to terminate observation and adopt H_1 we must have $\dotplus(T) \leqq b$ and $\dotplus(t) > b$ for $t < T$. Since with probability 1 we have $\dotplus(t) \geqq \dotplus(t - 0)$ we find that $\dotplus(T) = b$ with probability 1. Therefore

$$\frac{\alpha_2}{1 - \alpha_1} = e^b. \tag{4.6}$$

We see here one of the advantages of continuous observation over observation at discrete times only. If we were treating the problem in the conventional manner we would have (4.6) replaced by an inequality, while only the first of the inequalities (4.5) could be derived in the above manner.

Thus we have

$$b = \log \frac{\alpha_2}{1 - \alpha_1} \tag{4.7}$$

and

$$\log \frac{\lambda_1}{\lambda_2} + \log \frac{1 - \alpha_2}{\alpha_1} \leqq a \leqq \log \frac{1 - \alpha_2}{\alpha_1}. \tag{4.8}$$

We now proceed to give a method for the exact computation of a. Without additional effort we shall also find the power function of the test.

We put

$$R(t) = \frac{\dotplus(t)}{\log \dfrac{\lambda_2}{\lambda_1}} = x(t) - ct \tag{4.9}$$

where $c = (\lambda_2 - \lambda_1)/\log (\lambda_2/\lambda_1)$. Together with the process $\{x(t)\}$ we have to consider also processes differing from it by a constant; that is, we consider processes with arbitrary $x(0)$.

For given a and b, let $V_\lambda(r)$ be the probability that the procedure described above will terminate with the adoption of H_2 when the Poisson parameter is really λ and $R(0) = r$. We then have

$$R(\Delta t) = \begin{cases} r - c\Delta t \\ r + 1 - c\Delta t \text{ with probability} \\ \text{any other value} \end{cases} \begin{cases} 1 - \lambda\Delta t + o(\Delta t) \\ \lambda\Delta t + o(\Delta t) \\ o(\Delta t) \end{cases}$$

where the $o(\Delta t)$ terms are all smaller than $\lambda^2\Delta t^2$ for $0 < \Delta t < 1/\lambda$.

Putting

$$K = \frac{b}{\log \frac{\lambda_2}{\lambda_1}}, \qquad J = \frac{a}{\log \frac{\lambda_2}{\lambda_1}} \tag{4.10}$$

we have

$$V_\lambda(r) = 0 \qquad \text{for } r \leqq K$$

$$V_\lambda(r) = 1 \qquad \text{for } r \geqq J$$

while for $K < r < J$ we have

$$V_\lambda(r) = (1 - \lambda\Delta t)V_\lambda(r - c\Delta t) + \lambda\Delta t V_\lambda(r + 1 - c\Delta t) + o(\Delta t) \tag{4.11}$$

with $|o(\Delta t)| < \lambda^2(\Delta t)^2$ for $0 < \Delta t < 1/\lambda$. It follows at once that $V_\lambda(r)$ is continuous for $K \leqq r < J$. (It will be discontinuous at $r = J$.) Rewriting (4.11) as

$$\frac{V_\lambda(r) - V_\lambda(r - c\Delta t)}{\Delta t} = -\lambda V_\lambda(r - c\Delta t) + \lambda V_\lambda(r + 1 - c\Delta t) + \frac{o(\Delta t)}{\Delta t} \tag{4.12}$$

and letting $\Delta t \to 0$ we see that $V_\lambda(r)$ is differentiable in the interval $K < r < J$ with the exception of the point $r = J - 1$ (in case $K < J - 1$). Thus we have the difference-differential equation

$$cV_\lambda'(r) + \lambda V_\lambda(r) = \lambda V_\lambda(r + 1) \tag{4.13}$$

for $K < r < J$ and $r \neq J - 1$. The unique solution in $K < r < J$ is determined by the conditions: (i) $V_\lambda(r)$ continuous for $\lambda < J$, (ii) $V_\lambda(K) = 0$, (iii) $V_\lambda(r) = 1$ for $r \geqq J$.

Let $n(r)$ be the integer such that

$$J - r - 1 \leqq n(r) < J - r. \tag{4.14}$$

It is easy to verify that, for $K \leqq r < J$,

$$V_\lambda(r) = 1 + Ce^{-(\lambda/c)r} \sum_{i=0}^{n(r)} \frac{(-1)^i}{i!} \left[(J - r - i)\frac{\lambda}{c} e^{-\lambda/c}\right]^i \tag{4.15}$$

satisfies (4.13) and (i) for every choice of the constant of integration C. To satisfy also (ii) one has merely to choose

$$C = -e^{(\lambda/c)K} \Big/ \sum_{i=1}^{n(K)} \frac{(-1)^i}{i!} \left[(J - K - i)\frac{\lambda}{c} e^{-\lambda/c}\right]^i. \tag{4.16}$$

Putting $r = 0$ to represent the start of the actual probability ratio test as used in applications, we have from (4.15) and (4.16) that the "OC" function corresponding to the given values of λ_1, λ_2, a and b is given by, say,

$$(4.17) \qquad g\left(\frac{\lambda}{c}\right) = 1 - e^{(\lambda/c)K} \frac{\displaystyle\sum_{i=0}^{n(0)} \frac{(-1)^i}{i!}\left[(J-i)\frac{\lambda}{c}e^{-\lambda/c}\right]^i}{\displaystyle\sum_{i=0}^{n(K)} \frac{(-1)^i}{i!}\left[(J-K-i)\frac{\lambda}{c}e^{-\lambda/c}\right]^i}.$$

(K is not displayed since it is given explicitly by (4.10).) Now J should be determined so that

$$(4.18) \qquad \begin{aligned} g\left(\frac{\lambda_1}{c}\right) &= g\left(\frac{\log \dfrac{\lambda_2}{\lambda_1}}{\dfrac{\lambda_2}{\lambda_1} - 1}\right) = \alpha_1; \\ g\left(\frac{\lambda_2}{c}\right) &= g\left(\frac{\log \dfrac{\lambda_2}{\lambda_1}}{1 - \dfrac{\lambda_1}{\lambda_2}}\right) = 1 - \alpha_2. \end{aligned}$$

Each of the equations (4.18) follows from the other and either one may be used to find J.

It should be noticed that the dependence of K and J on λ_1 and λ_2 is only through the ratio λ_2/λ_1. This follows from (4.10) and (4.17) and could also have been foreseen from the nature of the problem. This remark is useful in the numerical tabulation of the values of J and K or, equivalently, of a and b. (The fact that the λ_i are involved only through their ratio is due to the fact that they are not attached to a given time-unit. In the discrete parameter problem there is an absolute unit of time and hence the two λ_i enter as two parameters. The simplification mentioned above therefore does not occur.)

We now derive, in a manner similar to that used above, an expression for the moment generating function $M_\lambda(u; r) = Ee^{uT}$ of the observation time T necessary to reach a decision when $R(0) = r$ and the true Poisson parameter is λ. From a result of C. Stein [2] it follows that for given J, K and λ there is a positive number $u_0 = u_0(J, K, \lambda)$ such that $M_\lambda(u; r)$ is analytic and uniformly bounded in r for each complex u with real part smaller than u_0. By definition we have $M_\lambda(u; r) = 1$ for $r \leqq K$ or $r \geqq J$. In the same way as (4.11) was derived we obtain (for each u with real part smaller than u_0) for $K < r \leqq J - 1$

$$\begin{aligned} M_\lambda(u; r) &= (1 - \lambda\Delta t)E\{e^{ut} \mid R(0) = r, \quad R(\Delta t) = r - c\Delta t\} \\ &\qquad + \lambda\Delta t E\{e^{ut} \mid R(0) = r, \quad R(\Delta t) = r + 1 - c\Delta t\} + o(\Delta t) \\ &= (1 - \lambda\Delta t)e^{u\Delta t}E\{e^{u(t-\Delta t)} \mid R(\Delta t) = r - c\Delta t\} \\ &\qquad + \lambda\Delta t E\{e^{ut} \mid R(\Delta t) = r + 1 - c\Delta t\} + o(\Delta t), \end{aligned}$$

or

$$M_\lambda(u; r) = (1 - \lambda\Delta t)(1 + u\Delta t)M_\lambda(u; r - c\Delta t) + \lambda\Delta t M_\lambda(u; r + 1 - c\Delta t) + o(\Delta t). \tag{4.19}$$

This form is also valid for $J - 1 < r < J$ since $M_\lambda(u; r + 1 - c\Delta t) = 1 + o(1)$ for $r > J - 1$. Since the $o(\Delta t)$ term and $M_\lambda(u; r)$ are uniformly bounded in r we deduce, as in the case of $V_\lambda(r)$, that, considered as a function of r, $M_\lambda(u; r)$ is continuous for $r < J$, possesses a derivative for $K < r < J$ and $r \neq J - 1$ and satisfies in the last range the equation

$$c \frac{\partial}{\partial r} M_\lambda(u; r) + (\lambda - u)M_\lambda(u; r) = \lambda M_\lambda(u; r + 1). \tag{4.20}$$

It can be verified that the solution of (4.20) satisfying the required boundary conditions is given for $K \leqq r < J$ by

$$\begin{aligned} M_\lambda(u; r) &= \left(\frac{\lambda}{\lambda - u}\right)^{n(r)+1} \\ &+ C(u)e^{-r(\lambda-u)/c} \sum_{i=0}^{n(r)} \frac{(-1)^i}{i!}\left[(J - r - i)\frac{\lambda}{c}e^{-(\lambda-u)/c}\right]^i - \frac{\lambda u}{(\lambda - u)^2} \\ &\cdot e^{(\lambda-u)(J-r-1)/c} \sum_{i=0}^{n(r)-1} \left(\frac{\lambda}{\lambda - u}e^{-(\lambda-u)/c}\right)^i \sum_{j=0}^{i} \frac{(-1)^j}{j!}\left[(J - r - i - 1)\frac{\lambda - u}{c}\right]^j \end{aligned} \tag{4.21}$$

with $C(u)$ determined so that $M_\lambda(u; K) \equiv 1$.

Let $Z_\lambda(r)$ be the expected length of time before a final decision is adopted. Then $Z_\lambda(r) = \partial/(\partial u)\, M_\lambda(u; r) \mid_{u=0}$. Since $C(0) = 0$ in (4.21) we obtain, on putting $C'(0) = C'$,

$$\begin{aligned} Z_\lambda(r) &= \frac{n(r) + 1}{\lambda} + C'e^{-(\lambda/c)r} \sum_{i=0}^{n(r)} \frac{(-1)^i}{i!}\left[(J - r - i)\frac{\lambda}{c}e^{-\lambda/c}\right]^i \\ &- \frac{1}{\lambda}e^{(\lambda/c)(J-r-1)} \sum_{i=0}^{n(r)-1} e^{-(\lambda/c)i} \sum_{j=0}^{i} \frac{(-1)^j}{j!}\left[(J - r - i - 1)\frac{\lambda}{c}\right]^j \end{aligned} \tag{4.22}$$

for $K \leqq r < J$ (of course $Z_\lambda(r) \equiv 0$ outside this range and C' is determined so that $Z_\lambda(K) = 0$).

(One could derive (4.22) without using the moment generating function by establishing the equation

$$cZ'_\lambda(r) + \lambda Z_\lambda(r) = 1 + \lambda Z_\lambda(r + 1)$$

for $K < r < J, r \neq J - 1$.)

If we write in a more explicit manner $Z_\lambda(r \mid \lambda_1, \lambda_2)$ for $Z_\lambda(r)$ with J and K determined as explained above, it is easily seen that

$$Z_{\alpha\lambda}(r \mid \alpha\lambda_1, \alpha\lambda_2) = \frac{1}{\alpha} Z_\lambda(r \mid \lambda_1, \lambda_2) \tag{4.23}$$

for every positive α.

It is possible to treat the negative binomial process in a manner essentially the same in which we have treated the Poisson process above. A complication is caused by the fact that the probability that the chance variable will exceed one in a small time interval is of the same order of magnitude as the probability that the chance variable will be one.

The authors are obliged to Professor J. L. Doob for several helpful remarks.

REFERENCES

[1] J. L. Doob, *Stochastic Processes*, John Wiley and Sons, 1953.
[2] C. Stein, "A note on cumulative sums," *Ann. Math. Stat.*, Vol. 17 (1946), pp. 498–499.
[3] A. Wald, "Sequential tests of statistical hypotheses, *Ann. Math. Stat.*, Vol. 16 (1945), pp. 117–186.
[4] A. Wald, *Sequential Analysis*, John Wiley and Sons, 1947.
[5] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, Vol. 19 (1948), pp. 326–329.
[6] A. Wald and J. Wolfowitz, "Bayes solutions of sequential decision problems," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 82–99.
[7] J. Wolfowitz, "The efficiency of sequential estimates and Wald's equation for sequential processes," *Ann. Math. Stat.*, Vol. 18 (1947), pp. 215–230.