The Annals of Mathematical Statistics
1968, Vol. 39, No. 6, 1953–1977

SEQUENTIAL SELECTION OF EXPERIMENTS[1]

By K. B. Gray, Jr.[2]

Bank Administration Institute

0. Summary. The problem of sequential selection of experiments, with fixed and optional stopping, is considered. Conditions are given which allow selection, stopping and terminal action rules to be based on a sequence $\{T_j\}$ of statistics, where T_j is a function of past observations $\mathbf{X}^j = (X_1, \cdots, X_j)$ and experiment selections $\mathbf{E}^j = (E_1, \cdots, E_j)$. Randomized stopping, selection, and terminal action rules are allowed, and all probability distributions are defined by densities relative to σ-finite measures over Euclidean spaces.

Here we give a heuristic description of the principal results for the case of optional stopping. At each time j the random variable X_j is observed and a decision is made to stop or continue. If the procedure is stopped, a terminal action A is taken. If it is continued, an experiment E_{j+1}, to be performed at time $j + 1$, is chosen. At time j, all decisions are based on $\mathbf{X}^j$, $\mathbf{E}^j$, the past observations and experiment selections. Upon stopping, and taking action A, a loss $L(\theta, A)$, where θ is the unknown state of nature, is incurred. The sampling cost of stopping at j is $C_j(\theta, \mathbf{X}^j, \mathbf{E}^j)$. Let the random variable N denote the random stopping time. A *selection rule* $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \cdots)$ is defined by the sequence of conditional densities $\gamma_j(e_{j+1} \mid \mathbf{x}^j, \mathbf{e}^j)$, a *stopping rule* $\boldsymbol{\phi} = (\phi_0, \phi_1, \cdots)$ by the probabilities $\phi_j(\mathbf{x}^j, \mathbf{e}^j) = P\{N = j \mid N \geqq j, \mathbf{x}^j, \mathbf{e}^j\}$, and a *terminal action rule* $\boldsymbol{\delta} = (\delta_0, \delta_1, \cdots)$ by the conditional densities $\delta_j(a \mid \mathbf{x}^j, \mathbf{e}^j)$. Definition of the *population densities* $f_\theta(x_{j+1} \mid \mathbf{x}^j, \mathbf{e}^{j+1})$ for $j = 0, 1, 2, \cdots$ completely fixes the probability structure.

Define $\{T_j\}$ to be *parameter sufficient* (PARS) if, for $j = 0, 1, 2, \cdots$, $\mathrm{Dist}_{\theta,\boldsymbol{\gamma}}\,(\mathbf{X}^j, \mathbf{E}^j \mid T_j)$ is independent of θ for all $\boldsymbol{\gamma}$ and *policy sufficient* (POLS) if, for $j = 0, 1, 2, \cdots$, $\mathrm{Dist}_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}\,(T_{j+1} \mid T_j, E_{j+1}, N \geqq j + 1)$ is independent of $\boldsymbol{\phi}$, $\boldsymbol{\gamma}$ for all θ.

Theorem. *If* $\{T_j\}$ *is* PARS; *then the class of policies* $\{\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\delta}^0\}$, *where* $\boldsymbol{\delta}^0$ *is based on* $\{T_j\}$, *is essentially complete.*

Theorem. *If* $\{T_j\}$ *is* PARS *and* POLS, *and the sampling cost is of the form* $C_j(\theta, T_j)$, *then the class of policies* $\{\boldsymbol{\phi}^0, \boldsymbol{\gamma}^0, \boldsymbol{\delta}^0\}$, *where* $\boldsymbol{\phi}^0, \boldsymbol{\gamma}^0, \boldsymbol{\delta}^0$ *are based on* $\{T_j\}$, *is essentially complete.*

Conditions are given to aid in the verification of PARS and POLS. The theorems are applied to examples, including versions of the two armed bandit problem.

1. Introduction. The concept of statistical sufficiency is of recognired value in simplifying the search for optimal decision procedures in problems where the ex-

---

Received 13 March 1967; revised 5 June 1968.

[1] Based on doctoral dissertation at UCLA.

[2] Affiliated with Hughes Research Laboratories during preparation of the original draft.

periments to be performed have been fixed in advance. In the sequential selection (some authors use the word "design") of experiments one must remember not only past observations, but past experiment selections. It is natural to consider extension of the concept of sufficiency to the sequential selection of experiments, where the statistics are functions of both the past observations and experiments.

The theory of the sequential selection of experiments is appropriate to a broad range of practical problems. The list of applications, other than sequential design in the narrow sense, includes inventory management, adaptive process control, learning theory, and the design of teaching machines.

Our principal conclusions are conditions under which stopping, experiment selection, and terminal decision rules can be based on sufficient statistics of the past observations and experiments.

The basic reference for the general theory of sufficiency for sequential and nonsequential decision problems, without experiment selection, is Bahadur [2]. His paper contains an exposition of the theory of sufficiency, for the nonsequential case, as well as a generalization to the sequential case. Our results in the optional stopping situation are extensions of those of Bahadur to the case of experiment selection. The technique employed, throughout the present paper, for constructing rules based on sufficient statistics, is an extension of the Blackwell-Rao theorem [15]. Shiryaev [12] has developed a Bayesian theory of sufficiency for the sequential selection problem. He shows that there is a Bayes procedure which depends on past observations and experiment selections only through a posterior probability distribution.

There is a substantial body of work on the sequential design of experiments that relates to the present study. Robbins [11] defines the problem of sequential design of experiments, with optional stopping, and illustrates the general situation with a version of the two armed bandit problem. Blackwell [4] considers the special case where one of two possible experiments is strongly preferable in the sense that the class of terminal decision rules, based on only one of the experiments, is complete. In the more general situation, one experiment may be preferable for certain states of nature, and another preferable for other states. Thus, accumulated partial knowledge about the state of nature influences selection of experiments. Chernoff's work [7] on the sequential design of experiments, when many experiments are available, is concerned with the determination of stopping, selection, and terminal decision procedures that are asymptotically Bayes in the sense that the difference between their risk and the Bayes risk approaches zero as the sampling cost approaches zero.

Several papers [3], [5], [6], [14], deal with versions of the celebrated two armed bandit problem.

In the engineering literature, Aoki [1] discusses a Markovian control problem. He considers the determination of control policies that are optimal in the Bayes sense. It is shown that such Bayes policies can be based on sufficient statistics, rather than a complete past history of process and control variables, if a certain condition is satisfied. This condition is a specialization of the concept of policy sufficiency which is developed here. In another paper, Striebel [13] investigates the adequacy of ordinary sufficiency for stochastic dynamic programming problems in which there are no unknown parameters.

2. Selection with fixed stopping.

2.1. *Problem.* We consider a process in which experimentation is conducted at times $j = 1, 2, \cdots, J$. The sequence is initiated at time $j = 0$, when an experiment E_1, to be performed at $j = 1$, is selected. Then at $j = 1$, the experiment E_2, is selected. Finally, at time $J - 1$, the experiment E_J is chosen. The selection of each experiment is determined by a randomized selection rule. At each time j, the performance of the experiment E_j allows observation of the random variable X_j, the distribution of which depends on the past observations $X_1, \cdots, X_{j-1}$, past experiment choices $E_1, \cdots, E_j$, and an unknown state of nature θ.

After the procedure is terminated at time J, the statistician takes a terminal action A, chosen according to a randomized terminal action rule. The terminal loss of taking action $A = a$, when θ is the state of nature, is $L(\theta, a)$. In addition, there is a sampling cost $C(\theta, \mathbf{x}^J, \mathbf{e}^J)$ that depends on θ, the past observation values $\mathbf{x}^J = (x_1, \cdots, x_J)$, and past experiment selections $\mathbf{e}^J = (e_1, \cdots, e_J)$. The total loss accrued is

$$(1) \qquad L(\theta, a) + C(\theta, \mathbf{x}^J, \mathbf{e}^J).$$

Note that throughout we employ the convention of using lower case letters to denote the values realized by the corresponding upper case random variables.

We assume that $\theta \varepsilon \Theta$, X_j takes values $x_j \varepsilon \mathfrak{X}$, E_j takes values $e_j \varepsilon \mathcal{E}$, and the terminal action random variable A takes values $a \varepsilon \mathfrak{A}$, where the sets Θ, $\mathfrak{X}$, $\mathcal{E}$, and $\mathfrak{A}$ are Euclidean, i.e., Borel sets in Euclidean spaces. It is convenient to denote the absence of any observations of X, E by $\mathbf{X}^0$, $\mathbf{E}^0$.

We begin by describing the probability structure of the problem heuristically. The sequence of population distributions $\text{Dist}_\theta\,(X_{j+1} \mid \mathbf{X}^j, \mathbf{E}^{j+1})$, for $j = 0, 1, 2, \cdots, J - 1$, is implied by the nature of the problem or process. Each $\text{Dist}_\theta\,(X_{j+1} \mid \mathbf{X}^j, \mathbf{E}^{j+1})$ describes the probability distribution of the observation X_{j+1} given the past observations $\mathbf{x}^j$ and that the experiments $\mathbf{e}^{j+1}$ have been selected. Both selection and terminal action rules are chosen by the statistician. His choice of these rules will, of course, be affected by the particular population distributions assumed. The selection rule $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \cdots, \gamma_{J-1})$ corresponds to the sequence of conditional distributions $\text{Dist}\,(E_{j+1} \mid \mathbf{X}^j, \mathbf{E}^j)$, each of which is determined by the conditional density γ_j and assumed to be independent of θ and the terminal action rule. The conditional distribution $\text{Dist}\,(A \mid \mathbf{X}^J, \mathbf{E}^J)$, which is also independent of θ, is the terminal action rule. It is determined by the conditional density $\delta(a \mid \mathbf{x}^J, \mathbf{e}^J)$.

Next we formalize the probability structure suggested above for the case that the random variables have both Euclidean domains and ranges. This assumption simplifies the discussion, and, as indicated in Lehmann [10], reduces each random variable to a "carrier of its distribution." The book by Lehmann [10] is used as a reference for fundamental measure theoretic and probabilistic facts.

We begin by defining the σ-finite measure spaces

$$(\mathfrak{X}, \mathcal{B}_X, \mu_X), \qquad (\mathcal{E}, \mathcal{B}_E, \mu_E), \tag{2}$$

which correspond to the observation and selection variables, respectively. In addition, there is a σ-finite terminal action space

$$(\mathcal{A}, \mathcal{B}_A, \mu_A). \tag{3}$$

In terms of the spaces (2) and (3), the basic sample space

$$\Omega = \mathfrak{X}^J \times \mathcal{E}^J \times \mathcal{A} \tag{4}$$

can be defined. It is convenient to introduce the sub σ-fields (of sets in Ω)

$$\mathfrak{F}_n = \mathcal{B}(\mathbf{X}^n, \mathbf{E}^n), \qquad \mathcal{G}_{n-1} = \mathcal{B}(\mathbf{X}^{n-1}, \mathbf{E}^n), \tag{5}$$

for $n = 1, 2, \cdots, J$, of the product σ-field

$$\mathfrak{J} = \mathcal{B}(\mathbf{X}^J, \mathbf{E}^J, A). \tag{6}$$

These σ-fields satisfy the nesting relationship

$$\mathcal{G}_0 \subset \mathfrak{F}_1 \subset \cdots \subset \mathcal{G}_{n-1} \subset \mathfrak{F}_n \subset \cdots \subset \mathfrak{F}_J \subset \mathfrak{J}. \tag{7}$$

We next define in a formal way the component densities that yield the joint probability measure $P_{\theta,\gamma,\delta}$ on $\mathfrak{J}$. The *selection rule* $\gamma = (\gamma_0, \gamma_1, \cdots, \gamma_{J-1})$, which the statistician is free to choose, is composed of non-negative $\mathcal{G}_j$-measurable components $\gamma_j(e_{j+1} \mid \mathbf{x}^j, \mathbf{e}^j)$, each of which satisfies

$$\int \gamma_j(e_{j+1} \mid \mathbf{x}^j, \mathbf{e}^j)\, d\mu_E(e_{j+1}) = 1 \quad (\text{a.s. } \mathfrak{F}_j). \tag{8}$$

The *terminal action rule*, also specified by the statistician, is a non-negative $\mathfrak{J}$-measurable function $\delta(a \mid \mathbf{x}^J, \mathbf{e}^J)$ which satisfies

$$\int \delta(a \mid \mathbf{x}^J, \mathbf{e}^J)\, d\mu_A(a) = 1 \quad (\text{a.s. } \mathfrak{F}_J). \tag{9}$$

The definition of the probability measure on $\mathfrak{J}$ is completed by specifying a sequence of *population densities* $\{f_\theta(x_{j+1} \mid \mathbf{x}^j, \mathbf{e}^{j+1})\}$, composed of non-negative $\mathfrak{F}_{j+1}$-measurable functions $f_\theta(x_{j+1} \mid \mathbf{x}^j, \mathbf{e}^{j+1})$ for which

$$\int f_\theta(x_{j+1} \mid \mathbf{x}^j, \mathbf{e}^{j+1})\, d\mu_X(x_{j+1}) = 1 \quad (\text{a.s. } \mathcal{G}_j). \tag{10}$$

Notice that this definition depends on θ but is independent of γ and δ. The requirement that the population densities be independent of γ and δ is appropriate in applications to sequential design of experiments, control theory, and the theory of inventory. In many design of experiments problems, the densities of the population distributions take the form $f_\theta(x_j \mid e_j)$. This is a type of independence. A control process with a density of the form $f_\theta(x_j \mid x_{j-1}, e_j)$ can be called Markovian. In the terminology of control theory, the density $f_\theta(x_j \mid \mathbf{x}^{j-1}, \mathbf{e}^j)$ defines the dynamics of the process.

Probability measures on the σ-fields

$$\mathcal{G}_0, \mathfrak{F}_1, \cdots, \mathfrak{F}_n, \mathcal{G}_n, \cdots, \mathfrak{F}_J, \mathfrak{J} \tag{11}$$

are defined recursively, using a method described by Doob [8] (Supplement, Example 2.5). The procedure is begun by letting

$$dP_{\gamma_0}(e_1) = \gamma_0(e_1)\, d\mu_E(e_1) \tag{12}$$

on $\mathcal{G}_0$, and

$$dP_{\theta,\gamma_0}(x_1, e_1) = f_\theta(x_1 \mid e_1)\, dP_{\gamma_0}(e_1)\, d\mu_X(x_1) \tag{13}$$

on $\mathfrak{F}_1$. Given the probability measure $P_{\theta,\gamma^{n-1}}$ on $\mathfrak{F}_n$, we define the measure

$$dP_{\theta,\gamma^n}(\mathbf{x}^n, \mathbf{e}^{n+1}) = \gamma_n(e_{n+1} \mid \mathbf{x}^n, \mathbf{e}^n)\, dP_{\theta,\gamma^{n-1}}(\mathbf{x}^n, \mathbf{e}^n)\, d\mu_E(e_{n+1}) \tag{14}$$

on $\mathcal{G}_n$, and then the measure

$$dP_{\theta,\gamma^n}(\mathbf{x}^{n+1}, \mathbf{e}^{n+1}) = f_\theta(x_{n+1} \mid \mathbf{x}^n, \mathbf{e}^{n+1})\, dP_{\theta,\gamma^n}(\mathbf{x}^n, \mathbf{e}^{n+1})\, d\mu_X(x_{n+1}) \tag{15}$$

on $\mathfrak{F}_{n+1}$. The definition procedure is continued for $n = 1, 2, \cdots, J-1$. Finally, the probability measure $P_{\theta,\gamma,\delta}$ on $\mathfrak{Z}$ is defined by

$$dP_{\theta,\gamma,\delta}(a, \mathbf{x}^J, \mathbf{e}^J) = \delta(a \mid \mathbf{x}^J, \mathbf{e}^J)\, dP_{\theta,\gamma}(\mathbf{x}^J, \mathbf{e}^J)\, d\mu_A(a). \tag{16}$$

The latter probability can be written in the closed form

$$dP_{\theta,\gamma,\delta}(a, \mathbf{x}^J, \mathbf{e}^J) = \left[\prod_{j=0}^{J-1} f_\theta(x_{j+1} \mid \mathbf{x}^j, \mathbf{e}^{j+1})\gamma_j(e_{j+1} \mid \mathbf{x}^j, \mathbf{e}^j)\right]\delta(a \mid \mathbf{x}^J, \mathbf{e}^J)\, d\tau, \tag{17}$$

where $\tau = \mu_A \times \mu_X^J \times \mu_E^J$.

It is assumed that the terminal loss $L(\theta, a)$ is $\mathcal{B}(A)$-measurable for all θ and that the sampling cost $C(\theta, \mathbf{x}^J, \mathbf{e}^J)$ is $\mathfrak{F}_J$-measurable for all θ. Furthermore, in order to insure the existence of the risk

$$R(\theta, \gamma, \delta) = E_{\theta,\gamma,\delta}[L(\theta, A) + C(\theta, \mathbf{X}^J, \mathbf{E}^J)], \tag{18}$$

it is assumed that both L and C are simultaneously bounded for all points in their domains. Using the probability measures defined above, the risk takes the form

$$R(\theta, \gamma, \delta) = \int [L(\theta, a) + C(\theta, \mathbf{x}^J, \mathbf{e}^J)]\, dP_{\theta,\gamma,\delta}(a, \mathbf{x}^J, \mathbf{e}^J). \tag{19}$$

For a given sequence of population distributions it is the statistician's goal to choose selection and terminal action rules that are in some sense optimal.

Example. *Two armed bandit.* At each stage j, of a J-stage game, the player plays one of two slot machines M_0 and M_1. Let the machine he selects at j be denoted by M_{e_j}, $e_j \in \{0, 1\}$. The selection is made on the basis of past experience in playing the machines, so the selection at time j is a random variable E_j, taking values e_j. After playing the machine E_j, the player observes the random variable X_j, which takes values $x_j \in \{0, 1\}$. We say that $X_j = 1$ if the machine pays off and $X_j = 0$ if it does not. Suppose the machine M_i has probability θ_i of paying off. The unknown state of nature $\theta = (\theta_0, \theta_1)$ is a point in $\Theta = [0, 1]^2$.

We assume that the machines are constructed so the X_j is independent of $X_1, \cdots, X_{j-1}$ and $E_1, \cdots, E_{j-1}$, or in other words, the population distributions have the form

(20) $$f_\theta(x_j \mid \mathbf{x}^{j-1}, \mathbf{e}^j) = f_\theta(x_j \mid e_j).$$

For our example,

(21) $$f_\theta(x_j \mid e_j) = \theta_{e_j}^{x_j}(1 - \theta_{e_j})^{1-x_j},$$

relative to counting measure on $\{0, 1\}$. The product of population distributions, that appears in (17) has the form

(22) $$\prod_{j=1}^n f_\theta(x_j \mid e_j) = \theta_0^{\Sigma_1^n x_j(1-e_j)}(1 - \theta_0)^{\Sigma_1^n (1-x_j)(1-e_j)}\theta_1^{\Sigma_1^n x_j e_j}(1 - \theta_1)^{\Sigma_1^n (1-x_j)e_j}.$$

We consider two possible loss situations. In the first it costs the player \$1 each time he plays. If the machine pays off, or $X = 1$, he wins \$2. If $x = 0$, there is no payoff. There is no terminal loss, and hence no terminal action is necessary. The sampling cost is the negative net winnings. It is given by

(23) $$C(\theta, \mathbf{x}^J, \mathbf{e}^J) = J - 2\sum_1^J x_j.$$

The first term is the cost of playing J times and the second term is the winnings. The player desires a selection rule γ that will, in some sense, maximize his net earnings.

In another conceivable loss situation, the player's goal is to estimate the value of $\theta \,\varepsilon\, \Theta$. It costs him $C_i > 0$ each time he plays machine M_i. An appropriate terminal loss for such a problem is

(24) $$L(\theta, a) = (\theta_0 - a_0)^2 + (\theta_1 - a_1)^2,$$

where $a = (a_0, a_1) \,\varepsilon\, \mathcal{A}$, a subset of 2-space. The sampling cost is

(25) $$C(\theta, \mathbf{x}^J, \mathbf{e}^J) = C_1\sum_1^J e_j + C_0\sum_1^J (1 - e_j).$$

2.2. *Concepts of sufficiency.* A *statistic* T_j *based on* $\mathbf{X}^j$, $\mathbf{E}^j$ is defined to be an $\mathfrak{F}_j$-measurable function taking $\mathfrak{X}^j \times \mathcal{E}^j$ into a Euclidean space. A sequence $\{T_j\}$, of statistics with each T_j based on $\mathbf{X}^j$, $\mathbf{E}^j$, is said to be *parameter sufficient* (PARS) if for each j, all γ^{j-1}, and all $F_j \,\varepsilon\, \mathfrak{F}_j$, there exists a version of $P_{\theta,\gamma^{j-1}}\{F_j \mid t_j\}$ that is independent of θ. Parameter sufficiency can be characterized according to the factorizability of the joint density $f_{\theta,\gamma}(\mathbf{x}^n, \mathbf{e}^n)$.

THEOREM 2.1. *The sequence* $\{T_j\}$ *is* PARS *if and only if, for all n, there exists a non-negative,* $\mathcal{B}(T_n)$*-measurable function* $g_{\theta,\gamma}(t_n)$, *and a non-negative* $\mathfrak{F}_n$*-measurable function* $h_\gamma(\mathbf{x}^n, \mathbf{e}^n)$ *such that*

(26) $$f_{\theta,\gamma}(\mathbf{x}^n, \mathbf{e}^n) = g_{\theta,\gamma}[T_n(\mathbf{x}^n, \mathbf{e}^n)]h_\gamma(\mathbf{x}^n, \mathbf{e}^n) \quad (\text{a.s. } \mathfrak{F}_n).$$

PROOF. The theorem follows directly from Corollary 1, p. 49 of Lehmann [10]. The preceding condition for PARS can be specialized for use in the sequential selection problem by the use of (17).

COROLLARY. *The sequence* $\{T_j\}$ *is* PARS *if, and only if, for each n, there is a non-negative* $\mathcal{B}(T_n)$*-measurable function* $g_\theta(t_n)$ *and a non-negative,* $\mathfrak{F}_n$*-measurable function* $h(\mathbf{x}^n, \mathbf{e}^n)$ *such that*

(27) $$\prod_{j=0}^{n-1} f_\theta(x_{j+1} \mid \mathbf{x}^j, \mathbf{e}^{j+1}) = g_\theta(t_n)h(\mathbf{x}^n, \mathbf{e}^n) \quad (\text{a.s. } \mathfrak{F}_n).$$

EXAMPLE. *Two armed bandit.* Let the sequence $\{T_j\}$ be defined by

$$T_j = (\textstyle\sum_1^j X_i,\ \sum_1^j E_i,\ \sum_1^j X_iE_i) = (T_{1,j},\ T_{2,j},\ T_{3,j}). \tag{28}$$

Using this definition of $\{T_j\}$ to substitute in (22), we see that

$$\textstyle\prod_{j=1}^n f_\theta(x_j \mid e_j) = \theta_0^{t_{1,n}-t_{3,n}}\cdot(1-\theta_0)^{n-t_{1,n}-t_{2,n}+t_{3,n}}\cdot\theta_1^{t_{3,n}}\cdot(1-\theta_1)^{t_{2,n}-t_{3,n}}, \tag{29}$$

which is of the form $g_\theta(t_n)$. It follows from the Corollary of Theorem 2.1 that the sequence $\{T_j\}$ is PARS.

An additional concept of sufficiency is required. The sequence $\{T_j\}$ is said to be *policy sufficient* (POLS) if, for each n, for all $\theta \varepsilon \Theta$, and for all $B_{n+1} \varepsilon \mathcal{B}(T_{n+1})$, there is a version of $P_{\theta,\gamma^n}\{B_{n+1} \mid t_n,\ e_{n+1}\}$ that is independent of γ^n. The next theorem provides sufficient conditions for POLS.

THEOREM 2.2. *If, for each θ and n, T_{n+1} is $\mathcal{B}(T_n,\ X_{n+1},\ E_{n+1})$-measurable and, for all $C_{n+1} \varepsilon \mathcal{B}(X_{n+1})$, there is a version of $P_{\theta,\gamma^n}\{C_{n+1} \mid t_n,\ e_{n+1}\}$ that is independent* of γ^n *(up to $\mathcal{G}_n$ equivalence), then $\{T_j\}$ is* POLS.

PROOF. We must show that

$$E_{\theta,\gamma^n,t_n,e_{n+1}}g(T_{n+1}) \tag{30}$$

is independent of γ^n, for all $\mathcal{B}(T_{n+1})$-measurable g and all θ. However, since T_{n+1} is a measurable function of T_n, X_{n+1}, E_{n+1}, it is sufficient to show that, for all $\mathcal{B}(X_{n+1})$-measurable functions g',

$$E_{\theta,\gamma^n,t_n,e_{n+1}}\, g'(X_{n+1}) \tag{31}$$

is independent of γ^n. This follows from the assumption that there is a version of $P_{\theta,\gamma^n}\{C_{n+1} \mid t_n,\ e_{n+1}\}$ which is independent of γ^n. The above argument holds for all n.

COROLLARY. *If, for all n, T_{n+1} is a $\mathcal{B}(T_n, X_{n+1}, E_{n+1})$-measurable function and the conditional population density $f_\theta(x_{n+1} \mid \mathbf{x}^n,\ \mathbf{e}^{n+1})$ is of the form $f_\theta(x_{n+1} \mid t_n,\ e_{n+1})$, then $\{T_j\}$ is* POLS.

EXAMPLE. *Two armed bandit.* It has already been shown that the statistic

$$T_n = (\textstyle\sum_1^n X_i,\ \sum_1^n E_i,\ \sum_1^n X_iE_i)$$

is PARS. The Corollary of the last theorem can be utilized to prove that $\{T_n\}$ is POLS. First, it is clear that T_{j+1} is a function of T_j, X_{j+1}, E_{j+1}. Moreover, the population distributions are of the form $f_\theta(x_j|e_j)$. Hence $f_\theta(x_j|t_{j-1},\ e_j) = f_\theta(x_j|e_j)$ and the hypotheses of the Corollary are satisfied.

EXAMPLE. *Exponential family.* Suppose the population distributions $f_\theta(x_j\, |\mathbf{x}^{j-1},\ \mathbf{e}^j)$ are of the form

$$f_\theta(x_j \mid e_j) = C(\theta,\ e_j) \exp\ [\textstyle\sum_{k=1}^m Q_k(\theta,\ e_j)S_k(x_j,\ e_j)]h(x_j,\ e_j) \tag{32}$$

and that $\mathcal{E} = \{\epsilon_1,\ \epsilon_2,\ \cdots,\ \epsilon_l\}$. Then,

$$\textstyle\prod_{j=1}^n f_\theta(x_j \mid e_j) = [\prod_{j=1}^n C(\theta,\ e_j)] \cdot\exp\ [\sum_{j=1}^n \sum_{k=1}^m Q_k(\theta,\ e_j)S_k(x_j,\ e_j)] \prod_{j=1}^n h(x_j,\ e_j). \tag{33}$$

Letting $I_\epsilon(e)$ equal 1 if $\epsilon = e$ and 0 otherwise, (33) can be written

$$(34)\quad [\prod_{i=1}^{l} C(\theta, \epsilon_i)^{\sum_{j=1}^{n} I_{\epsilon_i}(e_j)}] \cdot \exp[\sum_{k=1}^{m} \sum_{i=1}^{l} Q_k(\theta, \epsilon_i)(\sum_{j=1}^{n} I_{\epsilon_i}(e_j) S_k(x_j, \epsilon_i))] \prod_{j=1}^{n} h(x_j, e_j).$$

Let

$$(35)\quad T_n(i) = \sum_{j=1}^{n} I_{\epsilon_i}(e_j),$$

and

$$(36)\quad T_n(i, k) = \sum_{j=1}^{n} I_{\epsilon_i}(e_j) S_k(x_j, \epsilon_i),$$

for $i = 1, \cdots, l$ and $k = 1, \cdots, m$; then it is clear, by the Corollary of Theorem 2.1, that the statistics $T_n = (T_n(1), \cdots, T_n(l), T_n(1, 1), \cdots, T_n(l, m))$ define a PARS sequence $\{T_n\}$. The statistic T_n has $l(m + 1)$ components. The Corollary of Theorem 2.2 implies that $\{T_n\}$ is POLS.

The above example can be specialized to the case of geometric population distributions

$$(37)\quad f_\theta(x_j \mid e_j) = (1 - p(\theta, e_j))p(\theta, e_j)^{x_j},$$

where $0 < p(\theta, e) < 1$, $\mathcal{E} = \{0, 1\}$, and $\mathfrak{X} = \{0, 1, 2, \cdots\}$. The sequence defined by $T_n = (\sum_1^n X_i, \sum_1^n E_i, \sum_1^n X_i E_i)$ is PARS and POLS.

The same sequence $\{T_n\}$ is also PARS and POLS if the population distributions are Poisson with parameter $\lambda(\theta, e_j) > 0$ and $\mathcal{E} = \{0, 1\}$, $\mathfrak{X} = \{0, 1, 2, \cdots\}$.

For the exponential family, there are cases in which a PARS sequence $\{T_n\}$ does not include the components:

$T_n(i) = \sum_{j=1}^{n} I_{\epsilon_i}(e_j) =$ number of times experiment ϵ_i performed, for $i = 1, \cdots, l$. That is, there is a PARS sequence $\{T_n\}$ such that $T_n(1), \cdots, T_n(l)$ are not elements of T_n. For example, let

$$(38)\quad f_\theta(x_j \mid 0) = (1 - \theta)\theta^{x_j} I_{\mathfrak{X}}(x_j) \qquad \text{(Geometric)}$$

and

$$(39)\quad f_\theta(x_j \mid 1) = e^{-\lambda(\theta)}[\lambda(\theta)^{x_j}/x_j!] I_{\mathfrak{X}}(x_j) \qquad \text{(Poisson)},$$

where $0 < \theta < 1$, $\lambda(\theta) = -\log(1 - \theta)$, and $\mathfrak{X} = \{0, 1, 2, \cdots\}$. By definition $\lambda(\theta) > 0$. Now, (33) can be written with $C(\theta, 1) = C(\theta, 0)$ for $0 < \theta < 1$, and hence the sequence $\{T_n\}$ defined by

$$(40)\quad T_n = (\sum_{j=1}^{n} E_j X_j, \sum_{j=1}^{n} (1 - E_j) X_j)$$

is PARS. However, the random variable $\sum_{j=1}^{n} E_j$ is not determined by T_n.

It is interesting to compare the concept of fixed stopping POLS with the transitivity condition of Bahadur [2]. The present discussion is conducted at a heuristic level, and the notation Dist $(\cdot)$ is used to denote the "distribution" of the indicated random variable. Bahadur defines a sequence of statistics $\{T_j\}$, in the fixed experiment problem, to be transitive if $\text{Dist}_\theta(T_{n+1} \mid \mathbf{X}^n) =$

$\text{Dist}_\theta\,(T_{n+1} \mid T_n)$. In order to compare this definition with fixed stopping POLS, we specialize the latter concept to the fixed experiment case. The specialization reduces the definition to: $\{T_j\}$ is fixed experiment POLS if $\text{Dist}_{\theta,\gamma}\,(T_{n+1} \mid T_n)$ is independent of γ, for all θ. Of course, this always holds; hence, in the fixed experiment case, all sequences $\{T_j\}$ are fixed stopping POLS.

An alternative method of comparison is to extend the definition of transitivity to the experiment selection case. A natural generalization is to say that $\{T_j\}$ is *experimentally transitive* if, for each sequence $e_1, e_2, \cdots$ of experiments chosen, $\{T_j\}$ is transitive in the fixed experiment sense; that is $\{T_j\}$ is experimentally transitive if, for all θ, γ, and n,

$$\text{Dist}_{\theta,\gamma}\,(T_{n+1} \mid \mathbf{X}^n, \mathbf{E}^{n+1}) = \text{Dist}_{\theta,\gamma}\,(T_{n+1} \mid T_n, E_{n+1}). \tag{41}$$

It is easy to verify that experimental transitivity implies POLS. Since the left hand side of (41) is clearly independent of γ, the right hand side must also be independent of γ. This constitutes the definition of POLS.

It is of interest to note that experimental transitivity is a stronger condition than POLS. It suffices to give an example in the fixed experiment case. Bahadur [2] describes a sequence which is not transitive. Since, as mentioned, all sequences $\{T_j\}$ are POLS in the fixed experiment case, the sequence of Bahadur must be POLS.

2.3. *Essential completeness of rules based on* $\{T_j\}$. In this section the main conclusions for the fixed stopping selection problem are established. First, it is shown that if $\{T_j\}$ is PARS, then for any rule there is an equivalent rule for which the terminal action rule δ is based only on T_J. More importantly, if the sampling cost has the form $C(\theta, \mathbf{x}^J, \mathbf{e}^J) = \sum_{j=1}^{J} C_j'(\theta, t_j)$, where C_j' is $\mathcal{B}(T_j)$-measurable, and $\{T_j\}$ is PARS and POLS, then both the selection and terminal action rules can be based on $\{T_j\}$.

Formally, a *selection rule* γ *is said to be based on* $\{T_j\}$ if, for each j,

$$\gamma_j(e_{j+1} \mid \mathbf{x}^j, \mathbf{e}^j) = \gamma_j(e_{j+1} \mid t_j) \tag{42}$$

is $\mathcal{B}\,(E_{j+1}, T_j)$ - measurable. A *terminal action rule* δ *is based on* $\{T_j\}$ if

$$\delta(a \mid \mathbf{x}^J, \mathbf{e}^J) = \delta(a \mid t_J) \tag{43}$$

is $\mathcal{B}\,(A, T_J)$ - measurable.

The next lemma provides an essential tool for defining rules based on sufficient statistics in terms of given rules based on complete information. This technique is used repeatedly in the remainder of the paper.

LEMMA 2.1. *Suppose there is a probability space* $(\mathfrak{X}, \mathcal{B}_X, P)$ *and a* σ*-finite measure space* $(\mathcal{Y}, \mathcal{B}_Y, \mu)$. *Relative to a* $\mathcal{B}_{X,Y}$*-measurable conditional density* $f(y \mid x)$, *the probability measure* P_f, *on the product* σ*-field* $\mathcal{B}_{X,Y}$, *is defined by*

$$P_f\{B\} = \int_B f(y \mid x)\, d\mu(y)\, dP(x), \tag{44}$$

where $B \in \mathcal{B}_{X,Y}$. *Then, there is a version of the conditional density*

$$f^0(y \mid t) = E_t f(y \mid X), \tag{45}$$

relative to the $\mathcal{B}_X$*-measurable statistic* $T(X)$, such that

$$\text{(i)}\quad f^0(y \mid t) \quad \textit{is} \quad \mathcal{B}_{T,Y}\textit{-measurable}, \tag{46}$$

$$\text{(ii)}\quad \int f^0(y \mid t)\, d\mu(y) = 1 \qquad (\text{a.s. } P),$$

and

$$\text{(iii)}\quad \textit{for all} \quad B \,\varepsilon\, \mathcal{B}_{T,Y}, \qquad P_f\{B \mid t\} = P_{f^0}\{B \mid t\}. \tag{47}$$

Proof. By the Radon-Nikodym theorem there is a unique (up to $(\mathcal{B}_{X,Y}, \mu \times P)$ equivalence) function $f^*(y, t)$ which satisfies

$$\int_B f^*(y, t)\, dP(x) = \int_B f(y \mid x)\, dP(x) \quad (\text{a.e. } \mu), \tag{48}$$

for all $B \,\varepsilon\, \mathcal{B}_T$. Hence, by definition, $f^*(y, t)$ and $f^0(y \mid t)$ are $\mathcal{B}_{X,Y}$-equivalent since (48) is exactly the defining relationship for $E_t f(y \mid X)$.

It follows immediately from the definition of $f^*(y, t)$ that (i) and (iii) are true. Furthermore, because

$$\int_B \int_{\mathcal{Y}} f^*(y, t)\, d\mu(y)\, dP(x) = \int_B \int_{\mathcal{Y}} f(y \mid x)\, d\mu(y)\, dP(x), \tag{49}$$

for all $B \,\varepsilon\, \mathcal{B}_T$, and since the inner integral on the right hand side of (49) is equal to 1 (a.s. P) by the assumption that $f(y \mid x)$ is a conditional density, it follows that

$$\int f^*(y, t)\, d\mu(y) = 1 \quad (\text{a.s. } P). \tag{50}$$

This establishes the truth of (ii) and completes the proof of the lemma.

Theorem 2.3. *If* $\{T_j\}$ *is* PARS, *then the class of policies* $\{\gamma, \delta^0\}$, *where* δ^0 *is based on* T_J, *is essentially complete.*

Proof. Take any policy γ, δ. Let

$$\delta^0(a \mid t_J) = E_{\gamma,\delta,t_J}\delta(a \mid \mathbf{X}^J, \mathbf{E}^J). \tag{51}$$

Since $\{T_j]$ is PARS, the definition does not depend on θ.

By Lemma 2.1, δ^0 is a terminal action rule based on T_J and, for all $B \,\varepsilon\, \mathcal{B}(A)$,

$$P_{\gamma,\delta^0}\{B \mid t_J\} = P_{\gamma,\delta}\{B \mid t_J\} \quad (\text{a.s. } \mathcal{B}(T_J)). \tag{52}$$

Now, by (52),

$$E_{\theta,\gamma,\delta^0}L(\theta, A) = E_{\theta,\gamma}E_{\gamma,\delta,t_J}L(\theta, A). \tag{53}$$

It follows that

$$R(\theta, \gamma, \delta^0) = E_{\theta,\gamma,\delta}L(\theta, A) + E_{\theta,\gamma}C(\theta, \mathbf{X}^J, \mathbf{E}^J) \tag{54}$$

$$= R(\theta, \gamma, \delta).$$

and hence the policy (γ, δ^0) is equivalent to (γ, δ).

Example. *Two armed bandit.* Consider the problem of estimating θ. Recall that

$$L(\theta, a) = (\theta_0 - a_0)^2 + (\theta_1 - a_1)^2 \tag{55}$$

and

$$C(\theta, \mathbf{x}^J, \mathbf{e}^J) = 0.$$

We have shown that $\{T_j\}$ is PARS in the earlier discussion of the example. Therefore, Theorem 2.3 applies, and we know that δ can be based on $T_J = (\sum_1^J X_i, \sum_1^J E_i, \sum_1^J X_i E_i)$ without impairing the risk.

We next derive two lemmas that will aid in the proof that both δ and γ can be based on $\{T_j\}$.

Lemma 2.2. *If the sequence $\{T_j\}$ is* PARS *and* POLS *and if, for the given selection rule γ, a new rule γ^0 based on $\{T_j\}$ is defined by*

$$\gamma_j^0(e_{j+1} \mid t_j) = E_{\gamma, t_j} \gamma_j(e_{j+1} \mid \mathbf{X}^j, \mathbf{E}^j), \tag{56}$$

for $j = 0, 1, 2, \cdots, J - 1$, then, for all $B_n \,\varepsilon\, \mathcal{B}(T_n)$,

$$P_{\theta,\gamma^0}\{B_n\} = P_{\theta,\gamma}\{B_n\}, \tag{57}$$

for $n = 1, 2, \cdots, J$.

Proof. By Lemma 2.1, γ^0 is a selection rule based on $\{T_j\}$. We employ induction on n to verify (57). Since $\gamma_0^0 = \gamma_0$, the statement holds for $n = 1$. Suppose it is true for n, and consider its validity for $n + 1$.

Using $I_B(t)$ to denote the indicator function, which is 1 if $t \,\varepsilon\, B$ and 0 otherwise, we have that

$$P_{\theta,\gamma}\{B_{n+1}\} = E_{\theta,\gamma^n} E_{\theta,\gamma^n,t_n} E_{\theta,t_n,e_{n+1}} I_{B_{n+1}}(T_{n+1}), \tag{58}$$

since $\{T_j\}$ is POLS.

By Lemma 2.1 and the definition of γ_n^0, this is equal to

$$E_{\theta,\gamma^n} E_{\theta,\gamma_n 0,t_n} I_{B_{n+1}}(T_{n+1}), \tag{59}$$

which, by the induction hypothesis, can be written.

$$E_{\theta,\gamma^{n-1,0}} E_{\theta,\gamma_n{}^0,t_n} I_{B_{n+1}}(T_{n+1}) = P_{\theta,\gamma^0}\{B_{n+1}\}. \tag{60}$$

This completes the proof.

Lemma 2.3. *If $\{T_j\}$ is* PARS *and* POLS, *then, for the selection rule γ^0 and terminal action rule δ^0, defined by (56) and (51) respectively,*

$$P_{\theta,\gamma^0,\delta^0}\{B\} = P_{\theta,\gamma,\delta}\{B\}, \tag{61}$$

for all $B \,\varepsilon\, \mathcal{B}(A)$.

Proof. The probability (61) can be written

$$\begin{aligned} E_{\theta,\gamma^0,\delta^0} I_B(A) &= E_{\theta,\gamma^0} E_{\delta^0,t_J} I_B(A) && (\text{PARS and } POLS) \\ &= E_{\theta,\gamma^0} E_{\delta,t_J} I_B(A) && (\text{Lemma 2.1}) \\ &= E_{\theta,\gamma} E_{\delta,t_J} I_B(A) && (\text{Lemma 2.2}) \\ &= P_{\theta,\gamma,} \{B\}. \end{aligned} \tag{62}$$

Note that, in the second line of (62), Lemma 2.1 is employed as in Theorem 2.3.

We are now prepared to state the principal theorem for the fixed stopping selection problem.

THEOREM 2.4. *If* $\{T_j\}$ *is* PARS *and* POLS *and* $C(\theta, \mathbf{x}^J, \mathbf{e}^J)$ *is of the form* $\sum_{j=1}^J C_j'(\theta, t_j)$, *where for all* θ, *each* C_j' *is* $\mathcal{B}(T_j)$*-measurable, then the class of policies* $\{\gamma^0, \delta^0\}$, *where both* γ^0 *and* δ^0 *are based on* $\{T_j\}$, *is essentially complete.*

PROOF. For any policy γ, δ, define new rules γ^0, δ^0 as in (51) and (56). We know, by Theorem 2.3, that

$$R(\theta, \gamma, \delta) = R(\theta, \gamma, \delta^0) = E_{\theta,\gamma,\delta^0}L(\theta, A) + \textstyle\sum_{j=1}^J E_{\theta,\gamma}C_j'(\theta, T_j). \tag{63}$$

By Lemmas 2.3 and 2.2, this equals

$$E_{\theta,\gamma^0,\delta^0}L(\theta, A) + \textstyle\sum_{j=1}^J E_{\theta,\gamma^0}C_j'(\theta, T_j) = R(\theta, \gamma^0, \delta^0). \tag{64}$$

Thus, the policy γ^0, δ^0 is at least as good as γ, δ and the class of policies based on $\{T_j\}$ is essentially complete.

The preceding theorem is untrue if the assumption of policy sufficiency is removed. The following example demonstrates this. The concept of the example is similar to that of a counterexample given in Bahadur [2].

EXAMPLE. Let $\mathfrak{X} = \mathcal{E} = \{0, 1\}$, $\Theta = [0, 1]$, and $J = 2$. We define the population distributions

$$f_\theta(x_1 \mid e_1) = \tfrac{1}{2}I_{\mathfrak{X}}(x_1) \tag{65}$$

and

$$\begin{aligned} P_\theta\{X_2 = 1 \mid x_1, e_1, e_2\} &= \theta \quad \text{if} \quad x_1 = 0, \quad e_2 = 0, \\ &= \tfrac{1}{2} \quad \text{if} \quad x_1 = 0, \quad e_2 = 1, \\ &= \tfrac{1}{2} \quad \text{if} \quad x_1 = 1, \quad e_2 = 0, \\ &= \theta/2 \quad \text{if} \quad x_1 = 1, \quad e_2 = 1. \end{aligned} \tag{66}$$

Note that $f_\theta(x_2 \mid x_1, e_1, e_2)$ is independent of e_1. There is no terminal action space and the total loss is given by the sampling cost

$$C(\theta, \mathbf{x}^2, \mathbf{e}^2) = x_2. \tag{67}$$

Hence, the risk of employing a selection rule $\gamma = (\gamma_0, \gamma_1)$ is

$$R(\theta, \gamma) = P_{\theta,\gamma}[X_2 = 1]. \tag{68}$$

If we let

$$g(x_1) = P\{E_2 = 1 \mid X_1 = x_1\}, \tag{69}$$

for $x_1 = 0, 1$, then (68) reduces to

$$\tfrac{1}{2}[\theta(1 - g(0)) + \tfrac{1}{2}g(0) + \tfrac{1}{2}(1 - g(1)) + \tfrac{1}{2}\theta g(1)]. \tag{70}$$

The risk can then be written

$$R(\theta, \gamma) = \tfrac{1}{4}[2g(0)(\tfrac{1}{2} - \theta) + 2g(1)(\theta/2 - \tfrac{1}{2}) + 2\theta + 1]. \tag{71}$$

Clearly, $g(0)$ and $g(1)$ do not uniquely define a selection rule γ. However, the risk depends on γ only through $0 \leqq g(0) \leqq 1, 0 \leqq g(1) \leqq 1$.

We first show that there is a uniformly best rule based on a particular sufficient statistic. Then we exhibit a rule, based on complete information, which is superior for some values of θ.

Consider the sequence of statistics T_1 = constant, $T_2 = (\mathbf{X}^2, \mathbf{E}^2)$. Clearly, the sequence is parameter sufficient. Selection rules γ_1 based on T_1 must be constant in x_1, e_1, since T_1 is constant. Thus, for any rule γ_1 based on T_1, we must have that $g(0) = g(1) = b$, say, where $0 \leqq b \leqq 1$. The risk of such a policy is

$$R(\theta, \gamma) = \tfrac{1}{4}(2 - b)\theta + \tfrac{1}{4}. \tag{72}$$

Therefore, any policy $\gamma_1^0(e_2 \mid x_1, e_1)$, corresponding to $b = 1$, with

$$\textstyle\sum_{e_1} \gamma_1^0(1 \mid x_1, e_1)\gamma_0(e_1) = 1; \tag{73}$$

for $x_1 = 0, 1$, is uniformly best among the policies based on T_1. The risk for any such uniformly best policy is

$$R(\theta, \gamma^0) = \tfrac{1}{4}\theta + \tfrac{1}{4}. \tag{74}$$

Next we exhibit a rule γ_1', based on x_1, e_1, which is better than γ_1^0, for some θ. Define γ_1' by letting $g'(0) = 0$ and $g'(1) = 1$. The risk of this policy is

$$R(\theta, \gamma') = \tfrac{3}{4}\theta. \tag{75}$$

For $0 \leqq \theta < \frac{1}{2}$, $R(\theta, \gamma') < R(\theta, \gamma^0)$ and hence the selection rules based on T_1 do not form an essentially complete class. The relationship between the risks of γ' and γ^0 is shown in Figure 1.

EXAMPLE. *Two armed bandit.* For the estimation problem, where

$$L(\theta, a) = (\theta_0 - a_0)^2 + (\theta_1 - a_1)^2 \tag{76}$$

and the sampling cost is of the form (25), we show that the preceding theorem applies. Clearly the sampling cost is of the required form. In Section 2.2 it was shown that $\{T_j\}$, where $T_j = (\sum_1^j X_i, \sum_1^j E_i, \sum_1^j X_iE_i)$, is both PARS and POLS. Hence, by Theorem 2.4, both γ and δ can be based on $\{T_j\}$.

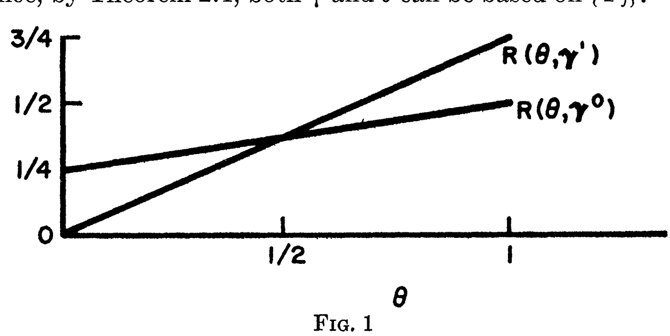

FIG. 1

For the other version of the problem, where $L(\theta, a) = 0$ and $C(\theta, \mathbf{x}^J, \mathbf{e}^J) = J - 2 \sum_1^j X_j$, the net gambling loss, it is desired to minimize the average net loss. The sampling cost is a function of $T_{1,J} = \sum_1^J X_i$, and, therefore, of the form of Theorem 2.4. It follows that γ and δ can be based on $\{T_j\}$.

Example. *Discrete control process.* To illustrate the application of the present theory to control problems, we consider a model that is a discrete approximation to many real world situations. Because of its suggestive value the terminology of control theory is used. However, the initial reference to a control theoretic term is followed, parenthetically, by its sequential selection equivalent.

The process starts at $j = 0$ with the process variable (observation) $X_0 = 0$. The process dynamics (population distributions) are defined by

$$X_j = X_{j-1} + E_j + Z_j, \tag{77}$$

for $j = 1, 2, \cdots, J$, where the "noise" process $Z_1, \cdots, Z_J$ is a sequence of independent, identically distributed, random variables with density

$$\begin{aligned} f_\theta(z) &= \theta^{(1+z)/2}(1-\theta)^{(1-z)/2} \quad \text{if} \quad z = \pm 1 \\ &= 0 \qquad\qquad\qquad\qquad\quad \text{otherwise,} \end{aligned} \tag{78}$$

determined by $\theta \, \varepsilon \, \Theta = [0, 1]$. The process variables X_j take values in $\mathfrak{X} = \{0, \pm 1, \pm 2, \cdots\}$ and the control variables (experiment selections) E_j take the values in $\mathcal{E} = \{-1, 0, +1\}$. There is no terminal loss, and the terminal action rule is not considered. We are interested in the determination of optimal control policies (selection rules) γ. The performance criterion (sampling cost) is defined to be

$$C(\theta, \mathbf{x}^J, \mathbf{e}^J) = \sum_1^J x_j^2 + C \sum_1^J e_j^2, \tag{79}$$

where $C > 0$ relates the cost e_j^2 of control corrections to the cost x_j^2 of the process variable deviating from its desired setting of zero.

The control model just described is an approximation to the performance of a heat exchanger with hot or cold "coolant." Temperatures are measured from a reference value so that the desired temperature is zero. The process variable X_j is the outgoing temperature of the fluid whose temperature is to be controlled. The control variable E_j corresponds to a valve that admits hot "coolant" if $E_j = +1$, is closed if $E_j = 0$, and admits cold "coolant" if $E_j = -1$. The system is diagrammed in Figure 2. The "noise" represents random variations in the incoming temperature of the fluid to be cooled.

The relationship (77), defining the process dynamics, can also be expressed by letting

$$\begin{aligned} &f_\theta(x_j \mid \mathbf{x}^{j-1}, \mathbf{e}^j) \\ &= f_\theta(x_j \mid x_{j-1}, e_j) \\ &= \theta^{(x_j - x_{j-1} - e_j + 1)/2}(1-\theta)^{(-x_j + x_{j-1} + e_j + 1)/2} I_{\{-1,+1\}}(x_j - x_{j-1} - e_j). \end{aligned} \tag{80}$$

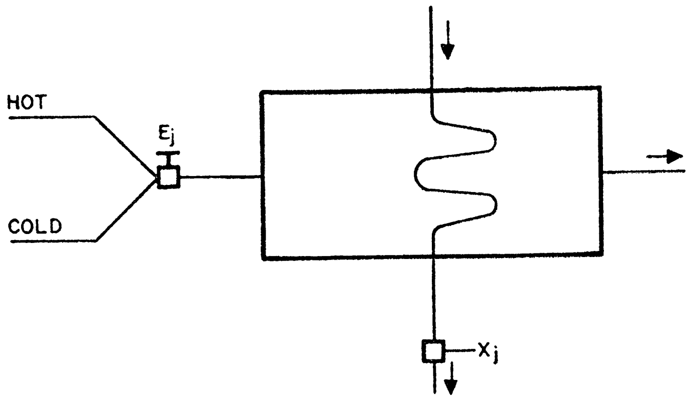

FIG. 2

It follows that

$$\begin{aligned} \textstyle\prod_{j=1}^{n} f_\theta(x_j \mid \mathbf{x}^{j-1}, \mathbf{e}^{j}) \\ (81) \qquad &= \textstyle\prod_{j=1}^{n} f_\theta(x_j \mid x_{j-1}, e_j) \\ &= \theta^{(x_n - \Sigma e_j + n)/2}(1-\theta)^{(-x_n + \Sigma e_j + n)/2} \textstyle\prod_{j=1}^{n} I_{\{-1,+1\}}(x_j - x_{j-1} - e_j), \end{aligned}$$

since $X_0 = 0$. Consider the statistic

$$T_n = (X_n, \textstyle\sum_1^n E_j). \tag{82}$$

From (81) it is clear that $\prod_{j=1}^{n} f_\theta(x_j \mid \mathbf{x}^{j-1}, \mathbf{e}^{j})$ is of the form $g_\theta(t_n)h(\mathbf{x}^n, \mathbf{e}^n)$. The Corollary of Theorem 2.1 implies, then, that $\{T_n\}$ is PARS. Also, T_{n+1} is a function of T_n, X_{n+1}, E_{n+1} and $f_\theta(x_j \mid x_{j-1}, e_j)$ is of the form $f_\theta(x_j \mid t_{j-1}, e_j)$. Hence, by the Corollary of Theorem 2.2, $\{T_n\}$ is POLS. In order to conclude, by Theorem 2.4, that the control policy γ can be based on $\{T_n\}$, we need to have that the performance criterion $C(\theta, \mathbf{x}^J, \mathbf{e}^J)$ is of the form $\sum_1^J C_j'(\theta, T_j)$. Now (79) can be written

$$\textstyle\sum_1^J (x_j^2 + Ce_j^2). \tag{83}$$

It is clear that while x_j^2 is a function of T_j, e_j^2 is not; moreover, $\sum_1^n e_j^2$ cannot be written in the form $\sum_{j=1}^n C_j(T_j)$. Therefore, Theorem 2.4 does not apply using $\{T_j\}$ and this performance criterion. However, if we consider the expanded statistic $T_n^* = (X_n, \sum_1^{n-1} E_j, E_n)$, then $\{T_n^*\}$ is still PARS and POLS. Moreover, $x_j^2 + Ce_j^2$ is a function of T_j^*; hence Theorem 2.4 implies that the control policy γ can be based on $\{T_n^*\}$.

EXAMPLE. *Continuous control process.* We provide an example of a control process in which the control and process variables are continuous. The model described here is proposed in [9]. Suppose the process is started by letting $X_0 = 0$. The process obeys the dynamical law

$$X_j = \theta_1 X_{j-1} + \theta_2 E_j + \theta_3 + Z_j, \tag{84}$$

for $j = 1, 2, \cdots, J$, where $\theta = (\theta_1, \theta_2, \theta_3)$ is a point in a parameter space Θ, and the random variables $Z_1, Z_2, \cdots, Z_J$ are independent and $\mathfrak{N}(0, 1)$. Suppose that the sampling cost is defined to be

$$C(\theta, \mathbf{x}^J, \mathbf{e}^J) = \sum_{j=1}^J (x_j - X^0)^2, \tag{85}$$

where X^0 is the desired setting, or set point, for the process. We do not consider the possibility of a terminal action. Thus, the risk of employing the control policy γ is

$$R(\theta, \gamma) = \sum_{j=1}^J E_{\theta,\gamma}(X_j - X^0)^2. \tag{86}$$

Consider the problem of finding a sequence of statistics $\{T_j\}$ which constitute a reduction of data and upon which the control policy γ can be based without adversely affecting the risk.

We employ Theorem 2.4. The process dynamics (84) can be stated in the form

$$\begin{aligned} f_\theta(x_j \mid \mathbf{x}^{j-1}, \mathbf{e}^j) &= f_\theta(x_j \mid x_{j-1}, e_j) \\ &= (2\pi)^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(x_j - \theta_1 x_{j-1} - \theta_2 e_j - \theta_3)^2\right]. \end{aligned} \tag{87}$$

Since

$$\prod_{j=1}^n f_\theta(x_j \mid x_{j-1}, e_j) = (2\pi)^{-n/2} \exp\left[-\tfrac{1}{2} \sum_{j=1}^n (x_j - \theta_1 x_{j-1} - \theta_2 e_j - \theta_3)^2\right] \tag{88}$$

and

$$\begin{aligned} &\sum_{j=1}^n (x_j - \theta_1 x_{j-1} - \theta_2 e_j - \theta_3)^2 \\ &\quad = \sum^n x_j^2 - 2\theta_1 \sum^n x_{j-1}x_j - 2\theta_2 \sum^n e_j x_j - 2\theta_3 \sum^n x_j \\ &\qquad + \theta_1^2 \sum^n x_{j-1}^2 + 2\theta_1\theta_2 \sum^n x_{j-1}e_j + 2\theta_1\theta_3 \sum^n x_{j-1} \\ &\qquad + \theta_2^2 \sum^n e_j^2 + 2\theta_2\theta_3 \sum^n e_j + n\theta_3^2, \end{aligned} \tag{89}$$

we have, by the Corollary of Theorem 2.1, that the sequence $\{T_j\}$, defined by

$$\begin{aligned} T_n = \Big(&\sum_{j=1}^n X_{j-1}, X_n, \sum_{j=1}^n E_j, \sum_{j=1}^n X_j X_{j-1}, \sum_{j=1}^n X_j E_j, \\ &\cdot \sum_{j=1}^n X_{j-1}E_j, \sum_{j=1}^n X_{j-1}^2, \sum_{j=1}^n E_j^2\Big) \end{aligned} \tag{90}$$

is PARS. Applying the Corollary of Theorem 2.2 we note that $\{T_j\}$ is also POLS. It happens that the sampling cost is of the required form since $(X_j - X^0)^2$ is a function of X_j, which is a component of T_j. Thus, the hypotheses of Theorem 2.4 are satisfied. Any search for optimal control policies, then, can be begun by restricting attention to policies based on $\{T_j\}$.

3. Selection with optional stopping.

3.1. *Problem.* If the structure of the preceding section is expanded to allow the experimenter to choose, at each instant, between stopping and continuing, then it is clear that such increased discretion will not enlarge the risk. That is, the best optional stopping rule is at least as good as any fixed sample rule. In this section the problem of sequential selection of experiments with optional stopping is considered as a generalization of the sequential selection problem. Frequently, instead of repeating a slightly modified version of a discussion given in the fixed stopping time case, we will simply indicate the changes that make the earlier comment valid for the optional stopping situation.

As before, it is assumed that Θ, $\mathcal{A}$, $\mathcal{X}$, and $\mathcal{E}$ are Euclidean. The notation $\mathbf{x}^j$, $\mathbf{e}^j$, etc. is retained. In addition infinite sequences of the form $\mathbf{x} = (x_1, x_2, \cdots)$ are considered. It is necessary to introduce a new random variable N, the stopping time, which takes values in $\{0, 1, 2, \cdots\}$. If N takes the value n, the procedure is terminated at time n and the experimenter incurs a total loss

$$L(\theta, a) + C_n(\theta, \mathbf{x}^n, \mathbf{e}^n), \tag{91}$$

where L is the terminal loss of taking action $a \in \mathcal{A}$ when $\theta \in \Theta$ is the true state of nature and $C_n(\theta, \mathbf{x}^n, \mathbf{e}^n)$ is the sampling cost of stopping at n, given θ, observations $\mathbf{X}^n = \mathbf{x}^n$, and selections $\mathbf{E}^n = \mathbf{e}^n$.

We proceed to the formal probability structure for the optional stopping problem. As in the fixed stopping situation, there are σ-finite measure spaces

$$(\mathcal{X}, \mathcal{B}(X), \mu_X), \qquad (\mathcal{E}, \mathcal{B}(E), \mu_E) \tag{92}$$

and

$$(\mathcal{A}, \mathcal{B}(A), \mu_A), \tag{93}$$

corresponding to the indicated random variables. The observation and experiment spaces (92) are defined for times $j = 1, 2, 3, \cdots$. Moreover, a σ-finite measure space

$$(\mathfrak{N}, \mathcal{B}(N), \mu_N), \tag{94}$$

where μ_N is a counting measure on the non-negative integers $\mathfrak{N} = \{0, 1, 2, \cdots\}$, is assumed for the optional stopping variable N.

The basic sample space for the optional stopping situation, with a terminal decision rule, is

$$\Omega = \mathcal{X}^\infty \times \mathcal{E}^\infty \times \mathfrak{N} \times \mathcal{A}. \tag{95}$$

If the convenient notation

$$\mathcal{G}_{n-1} = \mathcal{B}(\mathbf{X}^{n-1}, \mathbf{E}^n), \qquad \mathcal{F}_n = \mathcal{B}(\mathbf{X}^n, \mathbf{E}^n), \tag{96}$$

for $n = 1, 2, 3, \cdots$, is employed, then the nesting relationship

$$\mathcal{G}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{G}_{n-1} \subset \mathcal{F}_n \subset \cdots \tag{97}$$

is satisfied. The notational conventions (96) are illustrated by observing that $\mathcal{F}_n = \mathcal{B}(\mathbf{X}^n, \mathbf{E}^n)$ is an abbreviation for the Borel field generated by sets of the form

$$B_1 \times C_1 \times \cdots \times B_n \times C_n(\textstyle\prod_{j=n+1}^\infty \mathcal{X})(\prod_{j=n+1}^\infty \mathcal{E}) \times \mathcal{A} \times \mathfrak{N},$$

where $B_j \in \mathcal{B}(X)$ and $C_j \in \mathcal{B}(E)$ for $j = 1, 2, \cdots, n$.

Define

$$\mathfrak{F} = \mathcal{B}(\mathbf{X}, \mathbf{E})$$

to be the minimal σ-field containing the sequence (97). Finally, we introduce the σ-fields

(98) $$\mathcal{S} = \mathcal{B}(\mathbf{X}, \mathbf{E}, N) \quad \text{and} \quad \mathfrak{I} = \mathcal{B}(\mathbf{X}, \mathbf{E}, N, A)$$

which, of course, satisfy

$$\mathfrak{F} \subset \mathcal{S} \subset \mathfrak{I}.$$

The various probability measures on the above σ-fields are given in terms of the conditional probability densities describing the probability structure and decision rules of the stopping version of the problem. These densities are given next.

As in the fixed stopping case, we assume that for each $\theta \varepsilon \Theta$, there is a sequence $\{f_\theta(x_{j+1} \mid \mathbf{x}^j, \mathbf{e}^{j+1})\}$ of *population densities*. Each conditional density $f_\theta(x_{j+1} \mid \mathbf{x}^j, \mathbf{e}^{j+1})$ is non-negative, $\mathfrak{F}_{j+1}$-measurable, and satisfies

$$\int f_\theta(x_{j+1} \mid \mathbf{x}^j, \mathbf{e}^{j+1})\, d\mu_X(x_{j+1}) = 1 \quad (\text{a.s. } \mathcal{G}_j)$$

for $j = 0, 1, 2, \cdots$. The class of population densities is determined in advance. It is the statistician's task to choose the selection, stopping, and terminal action rules according to a given criterion of optimality. The *selection rule* $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2, \cdots)$ is composed of $\mathcal{G}_j$-measurable, non-negative, conditional densities $\gamma_j(e_{j+1} \mid \mathbf{x}^j, \mathbf{e}^j)$, each of which satisfies

$$\int \gamma_j(e_{j+1} \mid \mathbf{x}^j, \mathbf{e}^j)\, d\mu_E(e_{j+1}) = 1 \quad (\text{a.s. } \mathfrak{F}_j)$$

for $j = 0, 1, 2, \cdots$. The *stopping rule* $\boldsymbol{\phi} = (\phi_0, \phi_1, \phi_2, \cdots)$ consists of $\mathfrak{F}_j$-measurable functions $\phi_j(\mathbf{x}^j, \mathbf{e}^j)$, which satisfy $0 \leqq \phi_j \leqq 1$. The non-negative, $\mathfrak{I}$-measurable densities $\delta_j(a \mid \mathbf{x}^j, \mathbf{e}^j)$ satisfying

$$\int \delta_j(a \mid \mathbf{x}^j, \mathbf{e}^j)\, d\mu_A(a) = 1 \quad (\text{a.s. } \mathfrak{F}_j)$$

for $j = 0, 1, 2, \cdots$, constitute the *terminal action rule* $\boldsymbol{\delta} = (\delta_0, \delta_1, \delta_2, \cdots)$.

For $n = 1, 2, \cdots$, the probability spaces

(99) $$(\Omega, \mathcal{G}_{n-1}, P_{\theta,\boldsymbol{\gamma}^{n-1}}), \qquad (\Omega, \mathfrak{F}_n, P_{\theta,\boldsymbol{\gamma}^{n-1}})$$

are defined analogously to the fixed stopping case (14), (15). By the Kolmogorov extension theorem (an application similar to the present one is discussed by Doob [8] in Example 2.6 of the Supplement) there is a unique probability measure space

(100) $$(\Omega, \mathfrak{F}, P_{\theta,\boldsymbol{\gamma}})$$

which is induced by the spaces (99).

By a similar application of the technique described in Doob, a probability measure $P_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}$ on $\mathcal{S}$ can be determined by the stopping rule $\boldsymbol{\phi}$ and probability $P_{\theta,\boldsymbol{\gamma}}$ on $\mathfrak{F}$. If the function

$$\psi_n(\mathbf{x}^n, \mathbf{e}^n) = \phi_n(\mathbf{x}^n, \mathbf{e}^n) \textstyle\prod_{j=0}^{n-1} (1 - \phi_j(\mathbf{x}^j, \mathbf{e}^j)) \tag{101}$$

is used to denote the conditional stoppng probability $P\{N = n \mid \mathbf{x}, \mathbf{e}, N \geqq m\}$, then the probability measure space

$$(\Omega, \mathcal{S}, P_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}) \tag{102}$$

is defined by

$$dP_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}(\mathbf{x}, \mathbf{e}, n) = \psi_n(\mathbf{x}^n, \mathbf{e}^n)\, dP_{\theta,\boldsymbol{\gamma}}(\mathbf{x}, \mathbf{e})\, d\mu_N(n),$$

for $n = 0, 1, 2, \cdots$. Utilizing this probability measure and the terminal action rule δ, the probability measure space

$$(\Omega, \mathfrak{I}, P_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma},\boldsymbol{\delta}}) \tag{103}$$

is defined by

$$dP_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma},\boldsymbol{\delta}}(\mathbf{x}, \mathbf{e}, n, a) = \delta_n(a \mid \mathbf{x}^n, \mathbf{e}^n)\, dP_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}(\mathbf{x}, \mathbf{e}, n)\, d\mu_A(a).$$

The definition of the last probability measure allows computation of the risk, given θ, of employing the policy $\boldsymbol{\phi}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$. We assume that the terminal loss $L(\theta, a)$ is $\mathcal{B}(A)$-measurable for all θ and that the sampling cost $C_n(\theta, \mathbf{x}^n, \mathbf{e}^n)$ is $\mathfrak{F}_n$-measurable for each n and all θ. It follows that the total loss

$$L(\theta, A) + C_N(\theta, \mathbf{X}^N, \mathbf{E}^N) \tag{104}$$

is $\mathfrak{I}$-measurable. If it is assumed that both $L(\theta, a)$ and $C_n(\theta, \mathbf{x}^n, \mathbf{e}^n)$ are simultaneously bounded, either above or below, for all n, then the risk

$$R(\theta, (\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\delta})) = E_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma},\boldsymbol{\delta}}L(\theta, A) + E_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}C_N(\theta, \mathbf{X}^N, \mathbf{E}^N) \tag{105}$$

is well defined for all θ and all policies $\boldsymbol{\phi}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$.

3.2. *Extension of sufficiency concepts.* As in the fixed stopping case, a statistic T_j is defined to be an $\mathfrak{F}_j$-measurable function taking $\mathfrak{X}^j \times \mathcal{E}^j$ into a Euclidean space. The fixed stopping definition of PARS is extended directly to the case of infinite sequences. A sequence $\{T_j\}$ is said to be PARS if, for $j = 0, 1, 2, \cdots$, all selection rules $\boldsymbol{\gamma}$ and all $F_j \in \mathfrak{F}_j$, there is a version of $P_{\theta,\boldsymbol{\gamma}}\{F_j \mid t_j\}$ that is independent of θ.

Since the present definition of PARS is identical, except for the range of n, with that given in Section 2, the Corollary of Theorem 2.1 still applies. It provides a sufficient condition for PARS. The condition is stated in terms of the conditional population densities.

In the present situation, we have defined the concept PARS solely in terms of the probability $P_{\theta,\boldsymbol{\gamma}}$ on $\mathfrak{F}$. In fact, PARS implies that $\{T_j\}$ is sequentially PARS, i.e., that $\text{Dist}_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}(\mathbf{X}^N, \mathbf{E}^N \mid T_N, N)$ is independent of θ for all $\boldsymbol{\phi}$, $\boldsymbol{\gamma}$. We state this fact formally.

THEOREM 3.1. *If* $\{T_j\}$ *is* PARS, *then for all* $\boldsymbol{\phi}$, $\boldsymbol{\gamma}$ *and all* $F_n \in \mathfrak{F}_n$, *there exist versions of* $P_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}\{F_n \mid t_n, N = n\}$ *and* $P_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma},\boldsymbol{\delta}}\{F_n \mid t_n, N \geqq n\}$ *that are independent of* θ.

PROOF. We consider only the first conclusion. The second conclusion is established by a similar argument. The probability

$$
\begin{aligned}
P_{\theta,\phi,\gamma}\{F_n \mid t_n, N = n\} & \\
= & P_{\theta,\phi,\gamma}\{F_n, N = n \mid t_n\}[P_{\theta,\phi,\gamma}\{N = n \mid t_n\}]^{-1} \\
= & E_{\theta,\phi,\gamma,t_n} I_{F_n}(\mathbf{X}^n, \mathbf{E}^n) I_n(N)[E_{\theta,\phi,\gamma,t_n} I_n(N)]^{-1} \\
= & E_{\theta,\gamma,t_n} I_{F_n}(\mathbf{X}^n, \mathbf{E}^n)\psi_n(\mathbf{X}^n, \mathbf{E}^n)[E_{\theta,\gamma,t_n}\psi_n(\mathbf{X}^n, \mathbf{E}^n)]^{-1}.
\end{aligned} \tag{106}
$$

The assumption that $\{T_j\}$ is PARS implies that both the numerator and denominator are independent of θ.

We introduce a concept of *policy sufficiency for the optional stopping case.* A sequence $\{T_j\}$ is said to be POLS if, for $n = 0, 1, 2, \cdots$, for all θ, and all $B_{n+1} \varepsilon \mathcal{B}(T_{n+1})$ there is a version of $P_{\theta,\phi,\gamma}\{B_{n+1} \mid t_n, e_{n+1}, N \geqq n + 1\}$ that is independent of ϕ, γ. Although the same abbreviation is used for fixed and optional stopping policy sufficiency, the intended meaning should be obvious from the context.

The following sufficient condition for POLS is similar to that given in Theorem 2.2 for the fixed stopping case.

THEOREM 3.2. *If, for $n = 0, 1, 2, \cdots$, and for all $C_{n+1} \varepsilon \mathcal{B}(X_{n+1})$, there is a version of $P_{\theta,\phi,\gamma}\{C_{n+1} \mid t_n, e_{n+1}, N \geqq n + 1\}$ which is independent of ϕ, γ, for all θ, and T_{n+1} is a $\mathcal{B}(T_n, X_{n+1}, E_{n+1})$-measurable function, then $\{T_j\}$ is* POLS.

PROOF. Since T_{n+1} is $\mathcal{B}(T_n, X_{n+1}, E_{n+1})$-measurable, it suffices to show that for all $\mathcal{B}(X_{n+1})$-measurable functions g, there is a version of $E_{\theta,\phi,\gamma}\{g(X_{n+1}) \mid t_n, e_{n+1}, N \geqq n + 1\}$ which is independent of ϕ, γ. This follows immediately from the first hypothesis of the theorem.

COROLLARY. *If, for each n and all θ, T_{n+1} is $\mathcal{B}(T_n, X_{n+1}, E_{n+1})$-measurable and the conditional probability density $f_\theta(x_{n+1} \mid \mathbf{x}^n, \mathbf{e}^{n+1})$ is a $\mathcal{B}(T_n, X_{n+1}, E_{n+1})$-measurable function, say of the form $f_\theta(x_{n+1} \mid t_n, e_{n+1})$, then $\{T_j\}$ is* POLS.

PROOF. Using the last theorem, it is sufficient to show that, for each n and all $C_{n+1} \varepsilon \mathcal{B}(X_{n+1})$, there is a version of $P_{\theta,\phi,\gamma}\{C_{n+1} \mid t_n, e_{n+1}, N \geqq n + 1\}$ which is independent of ϕ, γ. Now,

$$
\begin{aligned}
P_{\theta,\phi,\gamma}\{C_{n+1} \mid t_n, e_{n+1}, N \geqq n + 1\} & \\
= & E_{t_n,e_{n+1}} \textstyle\prod_{j=0}^{n} (1 - \phi_j(\mathbf{X}^j, \mathbf{E}^j)) E_{\theta,\mathbf{x}^n,\mathbf{e}^n} I_{C_{n+1}}(X_{n+1}) \\
& \cdot [E_{t_n,e_{n+1}} \textstyle\prod_{j=0}^{n} (1 - \phi_j(\mathbf{X}^j, \mathbf{E}^j))]^{-1} \\
= & \int_{C_{n+1}} f_\theta(x_{n+1} \mid t_n, e_{n+1})\, d\mu_X(x_{n+1}),
\end{aligned} \tag{107}
$$

which clearly is independent of ϕ, γ.

We compare optional stopping POLS with Bahadur's [2] notion of transitivity. The latter condition is sufficient to allow stopping policies to be based on sufficient sequences of statistics in the fixed experiment, optional stopping, problem considered by Bahadur. In the present situation we generalize transitivity, as in the fixed stopping case, by defining a sequence $\{T_j\}$ to be *experimentally transitive* if, for all $B_{n+1} \varepsilon \mathcal{B}(T_{n+1})$, $P_{\theta,\gamma}\{T_{n+1} \varepsilon B_{n+1} \mid \mathbf{x}^n, \mathbf{e}^{n+1}\}$ is $\mathcal{B}(T_n, E_{n+1})$-measurable, for each n and all θ, γ. This statement implies that for every set $E_1, E_2, \cdots$ of experiment choices, the sequence $\{T_j\}$ is transitive in the Bahadur sense. The following theorem shows that, for the optional stopping problem, POLS and experimental transitivity are equivalent.

It can be concluded from this theorem that optional stopping POLS is a stronger condition than fixed stopping POLS. Recall that, in Section 2.2, an example was given showing that fixed stopping POLS does not imply experimental transitivity. Since optional stopping POLS is equivalent to experimental transitivity, it is clear that optional stopping POLS is not implied by fixed stopping POLS.

THEOREM 3.3. *For the optional stopping problem, a sequence of statistics* $\{T_j\}$ *is* POLS *if and only if it is experimentally transitive.*

PROOF. Suppose $\{T_j\}$ is experimentally transitive. Let $B_{n+1} \varepsilon \mathcal{B}(T_{n+1})$. Now,

$$\begin{aligned} P_{\theta,\boldsymbol{\phi},\gamma}\{T_{n+1} \varepsilon B_{n+1} \mid t_n, e_{n+1}, N \geqq n+1\} \\ = E_{\theta,\gamma,t_n,e_{n+1}} \textstyle\prod_{j=0}^{n} (1 - \phi_j(\mathbf{X}^j, \mathbf{E}^j)) I_{B_{n+1}}(T_{n+1}) \\ \cdot [E_{\theta,\gamma,t_n,e_{n+1}} \textstyle\prod_{j=0}^{n} (1 - \phi_j(\mathbf{X}^j, \mathbf{E}^j))]^{-1}. \end{aligned} \tag{108}$$

The numerator equals

$$E_{\theta,\gamma,t_n,e_{n+1}} \textstyle\prod_{j=0}^{n} (1 - \phi_j(\mathbf{X}^j, \mathbf{E}^j)) E_{\theta,t_n,e_{n+1}} I_{B_{n+1}}(T_{n+1}), \tag{109}$$

by the assumption of experimental transitivity. The expression can be written

$$[E_{\theta,t_n,e_{n+1}} I_{B_{n+1}}(T_{n+1})] E_{\theta,\gamma,t_n,e_{n+1}} \textstyle\prod_{j=0}^{n} (1 - \phi_j(\mathbf{X}^j, \mathbf{E}^j)). \tag{110}$$

It follows that (108) is

$$P_{\theta,\boldsymbol{\phi},\gamma}\{T_{n+1} \varepsilon B_{n+1} \mid t_n, e_{n+1}, N \geqq n+1\} = E_{\theta,t_n,e_{n+1}} I_{B_{n+1}}(T_{n+1}), \tag{111}$$

which is independent of both $\boldsymbol{\phi}$ and γ. Hence $\{T_j\}$ is POLS.

Now, consider the converse. Take any j. Let $\boldsymbol{\phi}^{(\infty)}$ be the stopping rule that never stops. For any $x_1, \cdots, x_j, e_1, \cdots, e_j$, let $\boldsymbol{\phi}^{(\mathbf{x}^j,\mathbf{e}^j)}$ be the stopping rule that stops at j if at least one of $X_1 \geqq x_1, \cdots, X_j \geqq x_j, E_1 \geqq e_1, \cdots, E_j \geqq e_j$ holds and otherwise never stops. If $B_{j+1} \varepsilon \mathcal{B}(T_{j+1})$, then since $\{T_j\}$ is POLS, we know that

$$\begin{aligned} P_{\theta,\boldsymbol{\phi}^{(\infty)},\gamma}\{T_{j+1} \varepsilon B_{j+1} \mid t_j, e_{j+1}, N \geqq j+1\} \\ = P_{\theta,\boldsymbol{\phi}^{(\mathbf{x}^j,\mathbf{e}^j)},\gamma}\{T_{j+1} \varepsilon B_{j+1} \mid t_j, e_{j+1}, N \geqq j+1\}, \end{aligned} \tag{112}$$

for all $\mathbf{x}^j$, $\mathbf{e}^j$. Utilizing the definitions of the policies $\boldsymbol{\phi}^{(\infty)}$ and $\boldsymbol{\phi}^{(\mathbf{x}^j,\mathbf{e}^j)}$, (112) is equivalent to

$$\begin{aligned} P_{\theta,\gamma}\{T_{j+1} \varepsilon B_{j+1} \mid t_j, e_{j+1}\} \\ = P_{\theta,\gamma}\{T_{j+1} \varepsilon B_{j+1} \mid t_j, e_{j+1}, \\ X_1 < x_1, \cdots, X_j < x_j, E_1 < e_1, \cdots, E_j < e_j\}, \end{aligned} \tag{113}$$

for all $\mathbf{x}^j$, $\mathbf{e}^j$. Since the sets of the form $\{X_1 < x_j, \cdots, X_j < x_j, E_1 < e_1, \cdots, E_j < e_j\}$ generate the Borel field $\mathcal{B}(\mathbf{X}^j, \mathbf{E}^j)$, we have that

$$P_{\theta,\boldsymbol{\gamma}}\{T_{j+1} \varepsilon B_{j+1} \mid \mathbf{x}^j, \mathbf{e}^{j+1}\} \tag{114}$$

is $\mathcal{B}(T_j, E_{j+1})$-measurable and, hence, that $\{T_j\}$ is experimentally transitive.

3.3. *Essential completeness of rules based on* $\{T_j\}$. In this section we exhibit conditions which are sufficient for basing stopping, selection, and terminal action rules on sequences of statistics $\{T_j\}$. The addition of optional stopping to the selection problem complicates the development. The main result (Theorem 3.5) is valid for a narrower class of sampling costs than the analogous result (Theorem 2.4) in the fixed stopping case.

A stopping rule $\boldsymbol{\phi}$ *is said to be based on* $\{T_j\}$ if, for each n,

$$\phi_n(\mathbf{x}^n, \mathbf{e}^n) = \phi_n(t_n) \tag{115}$$

is $\mathcal{B}(T_n)$-measurable, a *selection rule* $\boldsymbol{\gamma}$ *is said to be based on* $\{T_j\}$ if, for each n,

$$\gamma_n(e_{n+1} \mid \mathbf{x}^n, \mathbf{e}^n) = \gamma_n(e_{n+1} \mid t_n) \tag{116}$$

is $\mathcal{B}(T_n, E_{n+1})$-measurable, and a *terminal action rule* $\boldsymbol{\delta}$ *is said to be based on* $\{T_j\}$ if, for each n,

$$\delta_n(a \mid \mathbf{x}^n, \mathbf{e}^n) = \delta_n(a \mid t_n) \tag{117}$$

is $\mathcal{B}(A, T_n)$-measurable.

The hypotheses that allow the terminal action rule $\boldsymbol{\delta}$ to be based on a sufficient sequence are somewhat weaker than those required for basing the complete policy $\boldsymbol{\phi}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$ on $\{T_j\}$. For this reason we state separately the next result.

THEOREM 3.4. *If the sequence* $\{T_j\}$ *is* PARS, *then the class of policies* $\{\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\delta}^0\}$, *where* $\boldsymbol{\delta}^0$ *based on* $\{T_j\}$, *is essentially complete.*

PROOF. Given a policy $(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\delta})$, define a new terminal action rule by letting

$$\delta_n^0(a \mid t_n) = E_{\boldsymbol{\phi},\boldsymbol{\gamma},\boldsymbol{\delta}}\{\delta_n(a \mid \mathbf{X}^n, \mathbf{E}^n) \mid t_n, N = n\} \tag{118}$$

for $n = 0, 1, 2, \cdots$. Since $\{T_j\}$ is PARS, the definition is indeed independent of θ. We can apply Lemma 2.1 and conclude that $\boldsymbol{\delta}^0$ is a terminal action rule. Now,

$$E_{\theta,(\boldsymbol{\phi},\boldsymbol{\gamma},\boldsymbol{\delta})}L(\theta, A) = E_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}I_n(N)E_{\delta_n^0,t_n,N=n}L(\theta, A), \tag{119}$$

the last step holding true by the definition of $\boldsymbol{\delta}^0$. Hence, the risk

$$\begin{aligned} R(\theta, (\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\delta})) &= E_{\theta,(\boldsymbol{\phi},\boldsymbol{\gamma},\boldsymbol{\delta}^0)}L(\theta, A) + E_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}C_N(\theta, \mathbf{X}^N, \mathbf{E}^N), \\ &= R(\theta, (\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\delta}^0)), \end{aligned} \tag{120}$$

and the conclusion of the theorem is established.

We provide two lemmas which aid in the development of sufficient conditions for basing the entire policy on $\{T_j\}$.

LEMMA 3.1. *If* $\{T_j\}$ *is* PARS *and* POLS *and if, for every policy* $\boldsymbol{\phi}$, $\boldsymbol{\gamma}$, *we define new policies* $\boldsymbol{\phi}^0$, $\boldsymbol{\gamma}^0$, *based on* $\{T_j\}$, *by letting*

$$\phi_0^{\,0} = \phi_0, \qquad \phi_n^{\,0}(t_n) = E_{\boldsymbol{\phi},\boldsymbol{\gamma},t_n,N\geqq n}\phi_n(\mathbf{X}^n, \mathbf{E}^n), \tag{121}$$

and

$$\gamma_0^{\,0} = \gamma_0, \qquad \gamma_n^{\,0}(e_{n+1} \mid t_n) = E_{\boldsymbol{\phi},\boldsymbol{\gamma},t_n,N\geqq n}\gamma_n(e_{n+1} \mid \mathbf{X}^n, \mathbf{E}^n), \tag{122}$$

for $n = 0, 1, 2, \cdots$, *then for all* $B_n \varepsilon \mathcal{B}(T_n)$,

$$P_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}\{B_n, N \geqq n\} = P_{\theta,\boldsymbol{\phi}^0,\boldsymbol{\gamma}^0}\{B_n, N \geqq n\}, \tag{123}$$

for $n = 0, 1, 2, \cdots$.

Proof. Since $\{T_j\}$ is PARS, we know by Theorem 3.1 that $\boldsymbol{\phi}^0$, $\boldsymbol{\gamma}^0$ are defined independently of θ, and, by Lemma 2.1, they are rules. Mathematical induction is employed to establish (123) for $n = 1, 2, 3, \cdots$. It is clear that (123) holds for $n = 1$.

Suppose now that (123) is true for $n - 1$. For $B_n \varepsilon \mathcal{B}(T_n)$, the probability

$$P_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}\{B_n, N \geqq n\} = E_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}I_{\geqq n}(N)E_{\theta,t_{n-1},e_n,N\geqq n}I_{B_n}(T_n), \tag{124}$$

where "$\geqq n$" denotes $\{N \geqq n\}$, since $\{T_j\}$ is POLS. Expanding (124) we have

$$E_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}I_{\geqq n}(N)E_{\gamma_{n-1}^0,t_n,N\geqq n}E_{\theta,t_{n-1},e_n,N\geqq n}I_{B_n}(T_n), \tag{125}$$

by the definition of γ_{n-1}^0 and Lemma 2.1. Changing the form again, the expression

$$\mathrm{E}_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}I_{\geqq n-1}(N)E_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma},t_{n-1},N\geqq n-1}I_{\geqq n}(N)E_{\theta,\gamma_{n-1}^0,t_{n-1},N\geqq n}I_{B_n}(T_n) \tag{126}$$

results. By the definition of ϕ_{n-1}^0 and Lemma 2.1, this reduces to

$$E_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}I_{\geqq n-1}(N)E_{\phi_{n-1}^0,t_{n-1},N\geqq n-1}I_{\geqq n}(N)E_{\theta,\gamma_{n-1}^0,t_{n-1},N\geqq n}I_{B_n}(T_n), \tag{127}$$

which, by the induction hypothesis, equals

$$E_{\theta,\boldsymbol{\phi}^0,\boldsymbol{\gamma}^0}I_{\geqq n-1}(N)E_{\phi_{n-1}^0,t_{n-1},N\geqq n-1}I_{\geqq n}(N)E_{\theta,\gamma_{n-1}^0,t_{n-1},N\geqq n}I_{B_n}(T_n), \tag{128}$$

and the lemma is established.

Lemma 3.2. *If the sequence* $\{T_j\}$ *is* PARS *and* POLS, *then for every policy* $(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\delta})$, *there is a policy* $(\boldsymbol{\phi}^0, \boldsymbol{\gamma}^0, \boldsymbol{\delta}^0)$, *based on* $\{T_j\}$, *such that for all* $C_n \varepsilon \mathcal{B}(T_n, A)$,

$$P_{\theta,(\boldsymbol{\phi},\boldsymbol{\gamma},\boldsymbol{\delta})}\{C_n, N = n\} = P_{\theta,(\boldsymbol{\phi}^0,\boldsymbol{\gamma}^0,\boldsymbol{\delta}^0)}\{C_n, N = n\}. \tag{129}$$

Proof. Let $\boldsymbol{\delta}^0$, $\boldsymbol{\phi}^0$, $\boldsymbol{\gamma}^0$ be defined by (118), (121) and (122) respectively. Then

$$P_{\theta,\boldsymbol{\phi},\boldsymbol{\delta}}\{C_n, N = n\} = E_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}I_n(N)E_{\boldsymbol{\phi},\boldsymbol{\gamma},\boldsymbol{\delta},t_n,N=n}I_{C_n}(t_n, A), \tag{130}$$

since $\{T_j\}$ is PARS. By the definition of $\delta_n^{\,0}$ and applying Lemma 2.1, we can write the expression in the form

$$E_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}I_n(N)E_{\delta_n^0,t_n,N=n}I_{C_n}(t_n, A). \tag{131}$$

This equals

$$E_{\theta,\boldsymbol{\phi},\boldsymbol{\gamma}}I_{\geqq n}(N)\phi_n^{\,0}(T_n)E_{\delta_n^0,t_n,N=n}I_{C_n}(t_n, A), \tag{132}$$

by the definition of ϕ_n^0. Finally, we employ Lemma 3.1 and get

$$E_{\theta,\phi^0,\gamma^0} I_{\geqq n}(N)\phi_n^0(T_n)E_{\delta_n^0,t_n,N=n} I_{C_n}(t_n, A_n). \tag{133}$$

The final theorem is an immediate result of this lemma.

THEOREM 3.5. *If* $\{T_j\}$ *is* PARS *and* POLS *and if, for each* n, $C_n(\theta, \mathbf{x}^n, \mathbf{e}^n) = C_n'(\theta, t_n)$ *is* $\mathcal{B}(T_n)$*-measurable for all* θ, *then the class of policies* $\{\boldsymbol{\phi}^0, \boldsymbol{\gamma}^0, \boldsymbol{\delta}^0\}$, *based on* $\{T_j\}$, *is essentially complete.*

PROOF. The risk of using the policy $(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\delta})$ is

$$R(\theta, (\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\delta})) = E_{\theta,(\boldsymbol{\phi},\boldsymbol{\gamma},\boldsymbol{\delta})}[L(\theta, A) + C_N'(\theta, T_N)]. \tag{134}$$

By Lemma 3.2, this risk equals

$$E_{\theta,(\boldsymbol{\phi}^0,\boldsymbol{\gamma}^0,\boldsymbol{\delta}^0)}[L(\theta, A) + C_N'(\theta, T_N)], \tag{135}$$

and, hence, the policy $(\boldsymbol{\phi}^0, \boldsymbol{\gamma}^0, \boldsymbol{\delta}^0)$ is at least as good as $(\boldsymbol{\phi}, \boldsymbol{\gamma}, \boldsymbol{\delta})$.

A further version of the two armed bandit problem is provided in order to illustrate the theorem.

EXAMPLE. *Two armed bandit.* Suppose the player is confronted by a pair M_0, M_1 of slot machines, with unknown payoff probabilities θ_0, θ_1. The space Θ of states of nature and the action space $\mathcal{A}$ are both assumed to be the unit square $[0, 1]^2$. As in the version of this problem discussed earlier, $\mathcal{X} = \mathcal{E} = \{0, 1\}$. The machines have positive payoffs P_0, P_1 and positive costs C_0, C_1 of playing.

It is assumed that the sampling cost, or net loss, is

$$\begin{aligned} C_j(\theta, \mathbf{x}^j, \mathbf{e}^j) = &-P_0 \textstyle\sum_{i=1}^j x_i(1 - e_i) - P_1 \sum_{i=1}^j x_i e_i \\ &+ C_0 \textstyle\sum_{i=1}^j (1 - e_i) + C_1 \sum_{i=1}^j e_i, \end{aligned} \tag{136}$$

if the player elects to stop at j. Now, the population distribution at j, if θ is the state of nature, is

$$f_\theta(x_j \mid e_j) = \theta_0^{x_j}(1 - \theta_0)^{1-x_j} I_0(e_j) + \theta_1^{x_j}(1 - \theta_1)^{1-x_j} I_1(e_j), \tag{137}$$

and

$$\textstyle\prod_{j=1}^n f_\theta(x_j \mid e_j) = \theta_0^{t_{1,n}-t_{2,n}}(1 - \theta_0)^{n-t_{1,n}-t_{2,n}+t_{3,n}}\theta_1^{t_{3,n}}(1 - \theta_1)^{t_{2,n}-t_{3,n}}, \tag{138}$$

where

$$T_n = (T_{1,n}, T_{2,n}, T_{3,n}) = (\textstyle\sum_1^n X_i, \sum_1^n E_i, \sum_1^n X_i E_i). \tag{139}$$

Hence $\{T_j\}$ is PARS by the Corollary of Theorem 2.1 and POLS by the Corollary of Theorem 3.2. Moreover, the sampling cost $C_j(\theta, \mathbf{x}^j, \mathbf{e}^j)$ can be written in the form

$$C_j'(\theta, t_j) = -P_0(t_{1,j} - t_{3,j}) - P_1 t_{3,j} + C_0(j - t_{2,j}) + C_1 t_{2,j}, \tag{140}$$

which is the form required by Theorem 3.5. Thus, Theorem 3.5 applies for this example and we have that the stopping, selection, and terminal action rules can be based on $\{T_j\}$.

Acknowledgment. The suggestions and contributions of my advisor, Professor Thomas S. Ferguson, are gratefully acknowledged. In addition to numerous helpful comments, he proposed the idea of the proof of Lemma 2.1 and the concept of experimental transitivity as a generalization of Bahadur's notion. The referee's remarks resulted in the improvement of the manuscript and the correction of several faulty arguments.

I also wish to thank Hughes Research Laboratories for its support while I performed this research. Revision of the original manuscript was supported by Booz, Allen Applied Research Inc.

REFERENCES

[1] AOKI, MASANAO (1967). *Optimization of Modern Control Systems*. Academic Press, New York.

[2] BAHADUR, R. R. (1954). Sufficiency and statistical decision functions. *Ann. Math. Statist.* **25** 423–462.

[3] BELLMAN, R. E. (1956). A problem in the sequential design of experiments. *Sankhyā* **16** 221–229.

[4] BLACKWELL, DAVID (1953). Equivalent comparisons of experiments. *Ann. Math. Statist.* **24** 265–272.

[5] BRADT, R. N., JOHNSON, S. M. and KARLIN, S. (1956). On sequential designs for maximizing the sum of n observations. *Ann. Math. Statist.* **27** 1060–1074.

[6] BRADT, R. N. and KARLIN, S. (1956). On the design and comparison of certain dichotomous experiments. *Ann. Math. Statist.* **27** 390–409.

[7] CHERNOFF, HERMAN (1959). Sequential design of experiments. *Ann. Math. Statist.* **30** 755–770.

[8] DOOB, J. L. (1953). *Stochastic Processes*. Wiley, New York.

[9] GRAY, K. B. (1966). Completely adaptive control. *Instrumentation in the Chemical and Petroleum Industries*, **2**. Plenum Press, New York.

[10] LEHMANN, E. L. (1959). *Testing Statistical Hypotheses*. Wiley, New York.

[11] ROBBINS, H. E. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 527–535.

[12] SHIRYAEV, A. N. (1964). On the theory of random functions and the control of processes with incomplete data. *Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*. Czechoslovak Academy of Sciences, Prague.

[13] STRIEBEL, C. (1956). Sufficient statistics in the optimum control of stochastic systems. *J. Math. Anal. Appl.* **12** 576–592.

[14] VOGEL, W. (1960). A sequential design for the two armed bandit. *Ann. Math. Statist.* **31** 430–443.

[15] WILKS, S. S. (1962). *Mathematical Statistics*. Wiley, New York.