OPTIMAL STOPPING AND APPLICATIONS

Chapter 1. STOPPING RULE PROBLEMS

The theory of optimal stopping is concerned with the problem of choosing a time to take a given action based on sequentially observed random variables in order to maximize an expected payoff or to minimize an expected cost. Problems of this type are found in the area of statistics, where the action taken may be to test an hypothesis or to estimate a parameter, and in the area of operations research, where the action may be to replace a machine, hire a secretary, or reorder stock, etc. In this chapter, we introduce the problem mathematically and give a number of examples of applications.

Historically, the problem arose in the sequential analysis of statistical observations with Wald's theory of the sequential probability ratio test in Wald (1945) and the subsequent books, *Sequential Analysis* (1947) and *Statistical Decision Functions* (1950). The Bayesian perspective on these problems was treated in the basic paper of Arrow, Blackwell and Girshick (1948). The generalization of sequential analysis to problems of pure stopping without statistical structure was made by Snell (1952). In the 1960's, papers of Chow and Robbins (1961) and (1963) gave impetus to a new interest and rapid growth of the subject. The book, *Great Expectations: The Theory of Optimal Stopping* by Chow, Robbins and Siegmund (1971), summarizes this development.

§**1.1 The Definition of the Problem.** Stopping rule problems are defined by two objects,

(i) a sequence of random variables, $X_1, X_2, \ldots,$ whose joint distribution is assumed known, and

(ii) a sequence of real-valued reward functions,

$$y_0, y_1(x_1), y_2(x_1, x_2), \ldots, y_\infty(x_1, x_2, \ldots).$$

Given these two objects, the associated stopping rule problem may be described as follows. You may observe the sequence $X_1, X_2, \ldots$ for as long as you wish. For each $n = 1, 2, \ldots,$ after observing $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n,$ you may stop and receive the known reward $y_n(x_1,\ldots,x_n)$ (possibly negative), or you may continue and observe X_{n+1}. If you choose not to take any observations, you receive the constant amount, y_0. If you never stop, you receive $y_\infty(x_1,x_2,\ldots)$. (We shall allow the rewards to take the value $-\infty$; but we shall assume the rewards are uniformly bounded above by a random variable with finite expectation so that all the expectations below make sense.)

Your problem is to choose a time to stop to maximize the expected reward. You are allowed to use randomized decisions. That is, given that you reach stage n having observed $X_1 = x_1,\ldots,X_n = x_n$, you are to choose a probability of stopping that may depend on these observations. We denote this probability by $\phi_n(x_1,\ldots,x_n)$. A (randomized) **stopping rule** consists of the sequence of these functions,

$$\boldsymbol{\phi} = (\phi_0, \phi_1(x_1), \phi_2(x_1,x_2),\ldots), \tag{1}$$

where for all n and $x_1,\ldots,x_n$, $0 \le \phi_n(x_1,\ldots,x_n) \le 1$. The stopping rule is said to be non-randomized if each $\phi_n(x_1,\ldots,x_n)$ is either 0 or 1.

Thus, ϕ_0 represents the probability that you take no observations at all. Given that you take the first observation and given that you observe $X_1 = x_1$, $\phi_1(x_1)$ represents the probability you stop after the first observation, and so on. The stopping rule, $\boldsymbol{\phi}$, and the sequence of observations, $\boldsymbol{X} = (X_1, X_2,\ldots)$, determines the random time N at which stopping occurs, $0 \le N \le \infty$, where $N = \infty$ if stopping never occurs. The probability mass function of N given $\boldsymbol{X} = \boldsymbol{x} = (x_1,x_2,\ldots)$ is denoted by $\boldsymbol{\psi} = (\psi_0,\psi_1,\psi_2,\ldots,\psi_\infty)$, where

$$\begin{aligned} \psi_n(x_1,\ldots,x_n) &= \mathrm{P}(N = n|\boldsymbol{X} = \boldsymbol{x}) \quad \text{for} \quad n = 0,1,2,\ldots, \\ \psi_\infty(x_1,x_2,\ldots) &= \mathrm{P}(N = \infty|\boldsymbol{X} = \boldsymbol{x}). \end{aligned} \tag{2}$$

This may be related to the stopping rule $\boldsymbol{\phi}$ as follows:

$$\begin{aligned} \psi_0 &= \phi_0 \\ \psi_1(x_1) &= (1-\phi_0)\phi_1(x_1) \\ &\ \vdots \\ \psi_n(x_1,\ldots,x_n) &= [\prod_1^{n-1}(1-\phi_j(x_1,\ldots,x_j))]\phi_n(x_1,\ldots,x_n) \\ &\ \vdots \\ \psi_\infty(x_1,x_2,\ldots) &= 1 - \sum_0^\infty \psi_j(x_1,\ldots,x_j). \end{aligned} \tag{3}$$

$\psi_\infty(x_1,x_2,\ldots)$ represents the probability of never stopping given all the observations.

Your problem, then, is to choose a stopping rule $\boldsymbol{\phi}$ to maximize the expected return, $V(\boldsymbol{\phi})$, defined as

$$\begin{aligned} V(\boldsymbol{\phi}) &= \mathrm{E}\, y_N(X_1,\ldots,X_N) \\ &= \mathrm{E} \sum_{j=0}^{=\infty} \psi_j(X_1,\ldots,X_j) y_j(X_1,\ldots,X_j) \end{aligned} \tag{4}$$

where the "$=\infty$" above the summation sign indicates that the summation is over values of j from 0 to ∞, including ∞. In terms of the random stopping time N, the stopping rule $\boldsymbol{\phi}$ may be expressed as

$$\phi_n(X_1,\ldots,X_n) = \mathrm{P}(N = n | N \geq n, \boldsymbol{X} = \boldsymbol{x}) \quad \text{for} \quad n = 0, 1, \ldots. \tag{5}$$

The notation used is that of Section 7.1 of Ferguson (1967).

Remarks

1. LOSS VS. REWARD. Often, the structure of the problem makes it more convenient to consider a loss or a cost rather than a reward. Although one may use the above structure by letting y_n denote the negative of the loss, clarity is gained in such cases by letting y_n denote the loss incurred by stopping at n, and considering the problem to be one of choosing a stopping rule to *minimize* $V(\boldsymbol{\phi})$.

2. RANDOM REWARD SEQUENCES. For some applications, the reward sequence is more realistically described as a sequence of random variables $Y_0, Y_1, \ldots, Y_\infty$ whose joint distribution with the observations $X_1, X_2, \ldots$ is known. The actual value of Y_n may not be known precisely at time n when the decision to stop or continue must be made. However, allowing returns to be random does not represent a gain in generality because, since the decision to stop at time n may depend on $X_1, \ldots, X_n$, we may replace the sequence of random rewards Y_n by the sequence of reward functions $y_n(x_1, \ldots, x_n)$ for $n = 0, 1, \ldots, \infty$, where

$$y_n(x_1,\ldots,x_n) = \mathrm{E}\{Y_n | X_1 = x_1, \ldots, X_n = x_n\}. \tag{6}$$

Any stopping rule $\boldsymbol{\phi}$ for the payoff sequence $Y_0, Y_1, \ldots, Y_\infty$ would give the same expected return for the sequence $y_0, y_1, \ldots, y_\infty$.

3. THE INCREASING SEQUENCE OF SIGMA-FIELDS APPROACH. There is a simpler, more widely used, notation to model stopping rule problems that we describe here. Let $(\Omega, \mathcal{B}, \mathrm{P})$ denote the probability space on which all our random variables are defined, and let $\mathcal{F}_n$ denote the sub-σ-field of $\mathcal{B}$ generated by $X_1, \ldots, X_n$ (the smallest σ-field containing the sets $\{X_1 \leq x_1, \ldots, X_n \leq x_n\}$ for all $x_1, \ldots, x_n$). With $\mathcal{F}_0 = \{\Omega, \emptyset\}$ and $\mathcal{F}_\infty =$ the σ-field generated by $\cup \mathcal{F}_n$,

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \ldots \subset \mathcal{F}_n \subset \ldots \subset \mathcal{F}_\infty \subset \mathcal{B} \tag{7}$$

represents an increasing sequence of σ-fields. For an arbitrary random variable Z, the conditional expectation of Z given $X_1, \ldots, X_n$ may be denoted by

$$\mathrm{E}(Z|\mathcal{F}_n) = \mathrm{E}(Z|X_1,\ldots,X_n). \tag{8}$$

The stopping rule problem may be stated in terms of the sequence (7), without mention of the random variables $X_1, X_2, \ldots$, as being defined by the two objects,

(i$'$) the increasing sequence of σ-fields (7) and, (ii′) a sequence of reward random variables, $Y_0, Y_1, \ldots, Y_n, \ldots, Y_\infty$.

Remark 2 above implies that we may assume without loss of generality that Y_n is $\mathcal{F}_n$-measurable. (Being a function of $X_1, \ldots, X_n$ is essentially equivalent to being $\mathcal{F}_n$-measurable.) In particular, we may assume that Y_∞ is $\mathcal{F}_\infty$-measurable. A stopping rule is defined to be a random variable N taking values in $\{0, 1, \ldots, \infty\}$, such that the event $\{N = n\}$ is in $\mathcal{F}_n$. (This is equivalent to saying that the decision to stop at time n can depend only on $X_1, \ldots, X_n$ and not otherwise on future observations, $X_{n+1}, \ldots$.) The problem is to choose a stopping rule N to maximize the expected return, $\mathrm{E}(Y_N)$.

This approach is somewhat more general than the approach using (i) and (ii) because there exist σ-fields that are not generated by any sequence of random variables. It may appear that some generality has been lost by this approach because the stopping rules defined by this method are non-randomized. However, we may restrict attention to non-randomized stopping rules without loss of generality. This may be seen by attaching to each X_j an independent uniform (0,1) random variable, U_j. For a given stopping rule $\boldsymbol{\phi}$ we could form an equivalent non-randomized stopping rule by stopping at j when we reach it if $U_j < \phi_j(X_1, \ldots, X_j)$.

§**1.2 Examples.** Here are a number of optimal stopping rule problems that have important applications. Since stopping rule problems are defined by the sequences (i) and (ii) of §1.1, we must specify in each case the observations, $X_1, X_2, \ldots$, their joint distributions, and the reward (or cost) function, $y_n(x_1, \ldots, x_n)$ for stopping at stage n. We often use Y_n to denote the random payoff for stopping at stage n,

$$Y_n = y_n(X_1, \ldots, X_n).$$

1. THE HOUSE-SELLING PROBLEM. Offers come in daily for an asset, such as a house, that you wish to sell. Let X_n denote the amount of the offer received on day n. You don't know the values of the offers before they come in but you feel you may assume that the offers are independent and all have the same distribution that you feel you know. Each offer costs an amount $c > 0$ to observe; one may think of c as a cost of living. When you receive an offer, X_n, you must decide whether accept it or to wait for a better offer. You know a better offer will eventually appear, but will the increased size of the offer compensate for the observational costs you will have to pay?

For (i) then, the observations are $X_1, X_2, \ldots$ assumed to be independent and identically distributed with known distribution. For (ii), we distinguish two problems with differing payoffs, depending on whether or not you are able to recall and accept a past offer after you have observed a subsequent one. If you may not recall past offers, then

$$\begin{aligned} y_0 &= 0 \\ y_n(x_1, \ldots, x_n) &= x_n - nc \quad \text{for} \quad n = 1, 2, \ldots \\ y_\infty(x_1, x_2, \ldots) &= -\infty. \end{aligned}$$

Thus, after paying to observe X_n, you may accept the offer and receive X_n or reject it and pay c to see the offer X_{n+1}. If you are allowed to recall past offers, then

$$\begin{aligned} y_0 &= 0 \\ y_n(x_1,\ldots,x_n) &= \max(x_1,\ldots,x_n) - nc \quad \text{for} \quad n = 1,2,\ldots \\ y_\infty(x_1,x_2,\ldots) &= -\infty. \end{aligned}$$

In this case if you decide to stop, you receive the largest outstanding offer. The problems with recall were introduced by MacQueen and Miller (1960), Derman and Sacks (1960) and Chow and Robbins (1961), and, with discount rather than cost, by Karlin (1962). The problems without recall were treated by Sakaguchi (1961) and Chow and Robbins (1961, 1963).

In the economics literature, this problem is called the job search problem, and is attributed to George Stigler, (1961, 1962). An unemployed worker is searching for a job. Each search costs a certain amount in time and lost wages. When an available job is found, conditions for employment, including salary, are announced. How many searches should the worker undertake before accepting the best offer so far found? For a review of this problem from this viewpoint, see Lippman and McCall (1976).

2. MAXIMIZING THE AVERAGE. You observe a fair coin being tossed repeatedly. You may stop observing at any time , and when you do you receive as a reward the average number of heads observed. Thus, if the first toss is heads, you should certainly stop since your payoff is one and you can never receive a higher payoff than that. On the other hand, the strong law of large numbers implies that the average number of heads converges almost surely to 1/2, so you would never stop at a time when the average number of heads is less than or equal to 1/2. What stopping rule should you employ to maximize your expected payoff? And how great an expected payoff can you obtain?

Problems of this sort were first studied by Y. S. Chow and H. Robbins (1965) who describe a stopping rule that achieves an expected payoff greater than .79 in the above problem. This problem was mentioned on page 314 of Ferguson (1967) as the problem of the experimenter who knows the probability of success is 1/2, but who is going to estimate the probability of success by the average number of successes and wants to bias his estimate as much as possible.

We may put the problem of maximizing the average in the form of a stopping rule problem as follows. Let $X_1, X_2, \ldots$ be independent identically distributed random variables with a known distribution having a finite mean μ, and let

$$\begin{aligned} y_0 &= \mu \\ y_n(x_1,\ldots,x_n) &= (x_1 + \ldots + x_n)/n \quad \text{for} \quad n = 1,2,\ldots \\ y_\infty(x_1,x_2,\ldots) &= \mu. \end{aligned}$$

This assumes then that if you don't take any observations, you receive μ. If you never stop, you receive $\lim_{n\to\infty} \overline{X}_n = \mu$ a.s.

3. BAYES SEQUENTIAL STATISTICAL DECISION PROBLEMS. Stopping rule problems originated in the theory of sequential statistical analysis as developed by Wald (1947). Bayes sequential decision problems provide examples of stopping rule problems with dependent $X_1, X_2, \ldots$.

In this problem, a parameter θ is chosen from a parameter space Θ according to some prior distribution τ. Eventually the statistician must choose an action a in a given action space $\mathcal{A}$ incurring a loss $L(\theta, a)$. However, he may observe random variables $X_1, X_2, \ldots$ sequentially for as long as he likes before choosing the action, at a cost of c for each X_i observed. The random variables $X_1, X_2, \ldots$ are assumed to be i.i.d. given θ with a distribution of known form, $F(x|\theta)$. If he decides to stop taking observations after observing $X_1, \ldots, X_n$, then he would choose $a \in \mathcal{A}$ to minimize his conditional expected loss, and thus be expected to lose

$$\rho_n(X_1, \ldots, X_n) = \inf_{a \in \mathcal{A}} \mathrm{E}\{L(\theta, a)|X_1, \ldots, X_n\} \qquad \text{for} \quad n = 0, 1, \ldots.$$

The rule that chooses $a \in \mathcal{A}$ after observing $X_1, \ldots, X_n$ is called the terminal decision rule. It may be chosen independently of the stopping rule. (For a discussion, see section 7.2 of Ferguson (1967).) In this problem there is a loss plus cost, so in line with remarks 1 and 2 above, we let y_n denote the conditional minimum Bayes expected loss plus cost of stopping at n,

$$\begin{aligned} &y_n(x_1, \ldots, x_n) = \rho_n(x_1, \ldots, x_n) + nc \quad \text{for} \quad n = 0, 1, \ldots \\ &y_\infty(x_1, x_2, \ldots) = +\infty. \end{aligned}$$

The distribution of $X_1, X_2, \ldots$ is taken to be the marginal distribution derived from the joint distribution of $\theta, X_1, X_2, \ldots$ by integrating out the variable θ according to the given prior distribution. Thus even if the X_i are independent given θ, they become dependent when θ is integrated out.

4. THE ONE-ARMED BANDIT. (Bradt, Johnson and Karlin (1956)) There are two treatments available for the cure of a disease. The standard treatment, T2, has a known probability p_0 of cure, while treatment T1 has unknown probability p of cure, where the prior distribution of p is known. A group of n patients is to be treated sequentially, and you must decide which treatment to give each patient. If p were greater than p_0, you would prefer to give T1 to each patient. You may gain information on the value of p by observing the cure rate of patients assigned treatment T1. It is assumed that the patients respond independently and immediately to treatment. Assignment of a treatment to a patient may depend upon past outcomes. If T1 starts to look good because it is curing a proportion of patients greater than p_0, then one would like to keep assigning T1. Your objective is to cure as many of the patients as possible. Your payoff is the number of patients cured.

This is called the one-armed bandit problem. Bradt, Johnson and Karlin (1956) show that if it is ever optimal to use T2 on a patient, then it is optimal to continue to use T2 on all subsequent patients. Therefore, we need only consider rules that decide when, if ever, to start on treatment T2. In this way, the one-armed bandit problem is related to a stopping rule problem where stopping is identified with switching to treatment T2.

If treatment T1 is given to patient number j, we let X_j be 1 if the patient is cured and 0 if he is not. Thus, it is assumed that $X_1, \ldots, X_n$ given p are independent identically distributed Bernoulli random variables with $\mathrm{P}(X_j = 1) = p$, and that p has a known prior distribution, $G(p)$. This determines the distribution of the observations. If we decide to switch treatments after observing $X_1, \ldots, X_k$, then the number of patients cured is $Y_k = X_1 + \ldots + X_k + Z_{k+1} + \ldots + Z_n$, where Z_j is one or zero depending on whether the patient j is cured by treatment T2 or not. The values of the Z_j are not known when the decision to stop must be made, but we can, as pointed out in Remark 2, replace the Z_j by their expected values, p_0, without loss of generality. The reward for stopping at k becomes

$$Y_k = X_1 + \ldots + X_k + (n-k)p_0 \quad \text{for} \quad k = 0, 1, \ldots, n.$$

This problem has finite horizon, n. The problem is to choose a stopping rule, $N \le n$, to maximize $\mathrm{E}(Y_N)$. In this problem as in general bandit problems, we are not interested *per se* in estimating the unknown p. It is the sum of the observations that we are trying to maximize. General bandit problems are treated in Chapter 7.

5. DETECTING A CHANGE-POINT. (Shiryaev (1963)) You are monitoring a sequence of i.i.d. random variables, $X_1, X_2, \ldots$ with a known distribution, F_0. At some point T in time, unknown to you, the distribution will change to some other known distribution, F_1, and you want to sound an alarm as soon as possible after the change occurs. It is assumed that you know the distribution of T. If the cost of stopping after the change has occurred is the time since the change, and if the cost of a false alarm, that is, of stopping before the change has occurred, is taken to be a constant $c > 0$, then the total cost may be represented by

$$Y_n = cI\{n < T\} + (n-T)I\{n \ge T\} \qquad \text{for} \quad n = 0, 1, \ldots, \quad \text{and} \quad Y_\infty = \infty.$$

In this display, $I(A)$ represents the indicator function of a set A; so, for example, $I\{n < T\}$ is equal to 1 if $n < T$, and to zero otherwise. Since T is a random unobservable quantity, we may replace Y_n by its conditional expected value given $X_1, \ldots, X_n$,

$$y_n = c\mathrm{P}(T > n|\mathcal{F}_n) + \mathrm{E}((n-T)^+|\mathcal{F}_n) \qquad \text{for} \quad n = 0, 1, \ldots, \quad \text{and} \quad Y_\infty = \infty.$$

Applications include monitoring heart patients for a change in pulse rate, monitoring a production line for a change in quality, and monitoring missiles for a change of course.

§**1.3 Exercises.** Formulate the following problems as stopping rule problems; that is, give the distributions of the observations X_n, and give the payoffs Y_n or $y_n(X_1, ..., X_n)$.

1. *The Burglar Problem.* (Haggstrom (1966)) A burglar contemplates a series of burglaries. He may accumulate his larcenous earnings as long as he is not caught, but if he is caught during a burglary, he loses everything including his initial fortune, if any, and he is forced to retire. He wants to retire before he is cuaght. Assume that returns for each

burglary are i.i.d. and independent of the event that he is caught, which is, on each trial, equally probable. He wants to retire with a maximum expected fortune.

2. *Fishing.* (Starr and Woodroofe (1974)) You are fishing in a lake with n fish. Let T_j denote the time required to catch fish number j if you were to fish indefinitely. Assume the T_j are i.i.d. with known distribution F. You observe the order statistics of the T_j sequentially until you decide to stop, at which time you receive 1 for each fish you have caught and you pay c times the total time required. This is really a continuous time problem, but Starr and Woodroofe have shown that if F has non-decreasing failure rate (i.e. if $f(t)/(1-F(t))$ is non-decreasing, where f is the density), then it is optimal to stop only at the time of a catch.

3. *Search for a new species.* (Rasmussen and Starr (1979)) Individual wasps from the genus Zyzzyx are observed at unit time intervals. This genus is comprised of species $\mu_1, \mu_2, \ldots$ and the observations are assumed to be independently drawn from this genus with probability p_j for species j, assumed known. The cost of each observation is $c > 0$, and the reward when you stop is the number of different species observed.

4. *Proofreading.* (Yang, Wackerly and Rosalsky (1982)) A manuscript has just been typed. The number of misprints in the manuscript is a random variable, M, whose distribution is known. Misprints may be found and corrected through proofreading. Each proofreading costs an amount $c_1 > 0$. On the k^{th} proofreading, each undetected misprint is found independently with probability p_k independent of the number of misprints found on previous proofreadings. Each undetected misprint left in the manuscript when it is sent to the printer costs an amount $c_2 > 0$. The problem is to decide when to stop proofreading and send the manuscript to the printer. (This problem may also be stated in terms of deciding when one should stop testing software for bugs and send it to be marketed. See, for example, Dalal and Mallows (1988).)

5. *Success runs.* (Starr (1972)) Independent identically distributed Bernoulli trials with probability p of success are observed at a constant cost per observation until you decide to stop. When you stop, you receive a reward proportional to the number of successes in the current success run up to the time you stop.

Chapter 2. FINITE HORIZON PROBLEMS.

A stopping rule problem has a finite horizon if there is a known upper bound on the number of stages at which one may stop. If stopping is required after observing $X_1, \ldots, X_T$, we say the problem has horizon T. A finite horizon problem may be obtained as a special case of the general problem as presented in Chapter 1 by setting $y_{T+1} = \cdots = y_\infty = -\infty$. In principle, such problems may be solved by the method of **backward induction**. Since we must stop at stage T, we first find the optimal rule at stage $T-1$. Then, knowing the optimal rule at stage $T-1$, we find the optimal rule at stage $T-2$, and so on back to the initial stage (stage 0). We define $V_T^{(T)}(x_1, \ldots, x_T) = y_T(x_1, \ldots, x_T)$ and then inductively for $j = T-1$, backward to $j = 0$,

$$
\begin{aligned}
&V_j^{(T)}(x_1, \ldots, x_j) = \\
&\qquad \max\left\{ y_j(x_1, \ldots, x_j), \mathrm{E}(V_{j+1}^{(T)}(x_1, \ldots, x_j, X_{j+1}) | X_1 = x_1, \ldots, X_j = x_j) \right\}.
\end{aligned} \tag{1}
$$

Inductively, $V_j^{(T)}(x_1, \ldots, x_j)$ represents the maximum return one can obtain starting from stage j having observed $X_1 = x_1, \ldots, X_j = x_j$. At stage j, we compare the return for stopping, namely $y_j(x_1, \ldots, x_j)$, with the return we expect to be able to get by continuing and using the optimal rule for stages $j+1$ through T, which at stage j is $\mathrm{E}(V_{j+1}^{(T)}(x_1, \ldots, x_j, X_{j+1}) | X_1 = x_1, \ldots, X_j = x_j)$. Our optimal return is therefore the maximum of these two quantities, and it is optimal to stop at j if $V_j^{(T)}(x_1, \ldots, x_j) = y_j(x_1, \ldots, x_j)$, and to continue otherwise. The value of the stopping rule problem is then $V_0^{(T)}$.

In this chapter, we present a number of problems whose solutions may be effectively evaluated by this method. The most famous of these is the secretary problem. In the first three sections, we discuss this problem and some of its variations. In the fourth section, we treat the Cayley-Moser problem, a finite horizon version of the house selling problem mentioned in Chapter 1. In the last section, we present the parking problem of MacQueen and Miller. Other examples of finite horizon problems include the fishing problem and the one-armed bandit problem of the Exercises of Chapter 1.

§**2.1 The Classical Secretary Problem.** The secretary problem and its offshoots form an important class of finite horizon problems. There is a large literature on this problem, and one book, *Problems of Best Selection* (in Russian) by Berezovskiy and Gnedin

(1984) devoted solely to it. For an entertaining exposition of the secretary problem, see Ferguson (1989). The problem is usually described as that of choosing the best secretary (the secretary problem), but it is sometimes described as the problem of choosing the best spouse (the marriage problem) or the largest of an unknown set of numbers (Googol). First, we describe what is known as the classical secretary problem, or CSP.

1. There is one secretarial position available.

2. There are n applicants for the position; n is known.

3. It is assumed you can rank the applicants linearly from best to worst without ties.

4. The applicants are interviewed sequentially in a random order with each of the $n!$ orderings being equally likely.

5. As each applicant is being interviewed, you must either accept the applicant for the position and end the decision problem, or reject the applicant and interview the next one if any.

6. The decision to accept or reject an applicant must be based only on the relative ranks of the applicants interviewed so far.

7. An applicant once rejected cannot later be recalled.

8. Your objective is to select the best of the applicants; that is, you win 1 if you select the best, and 0 otherwise.

We place this problem into the guise of a stopping rule problem by identifying stopping with acceptance. We may take the observations to be the **relative ranks**, $X_1, X_2, \ldots, X_n$, where X_j is the rank of the jth applicant among the first j applicants, rank 1 being best. By assumption 4, these random variables are independent and X_j has a uniform distribution over the integers from 1 to j. Thus, $X_1 \equiv 1$, $\mathrm{P}(X_2 = 1) = \mathrm{P}(X_2 = 2) = 1/2$, etc.

Note that an applicant should be accepted only if it is relatively best among those already observed. A relatively best applicant is called a **candidate**, so the jth applicant is a candidate if and only if $X_j = 1$. If we accept a candidate at stage j, the probability we win is the same as the probability that the best among the first j applicants is best overall. This is just the probability that the best candidate overall appears among the first j applicants, namely j/n. Thus,

$$y_j(x_1, \ldots, x_j) = \begin{cases} j/n & \text{if applicant } j \text{ is a candidate,} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Note that $y_0 = 0$ and that for $j \geq 1$, y_j depends only on x_j.

This basic problem has a remarkably simple solution which we find directly without the use of (1). Let W_j denote the probability of win using an optimal rule among rules that pass up the first j applicants. Then $W_j \geq W_{j+1}$ since the rule best among those that pass up the first $j+1$ applicants is available among the rules that pass up only the first j applicants. It is optimal to stop with a candidate at stage j if $j/n \geq W_j$. This means that if it is optimal to stop with a candidate at j, then it is optimal to stop with a candidate at $j+1$, since $(j+1)/n > j/n \ge W_j \ge W_{j+1}$. Therefore, an optimal rule may be found among the rules of the following form, N_r for some $r \ge 1$:

N_r: Reject the first $r-1$ applicants and then accept the next relatively best applicant, if any.

Such a rule is called a **threshold rule** with threshold r. The probability of win using N_r is

$$\begin{aligned} P_r &= \sum_{k=r}^{n} \mathrm{P}(k^{th} \text{ applicant is best and is selected }) \\ &= \sum_{k=r}^{n} \mathrm{P}(k^{th} \text{ applicant is best })\mathrm{P}(k^{th} \text{ applicant is selected} \mid \text{it is best }) \\ &= \sum_{k=r}^{n} \frac{1}{n}\mathrm{P}(\text{best of first } k-1 \text{ appears before stage } r) \\ &= \sum_{k=r}^{n} \frac{1}{n}\,\frac{r-1}{k-1} = \frac{r-1}{n}\sum_{k=r}^{n}\frac{1}{k-1}, \end{aligned} \tag{3}$$

where $(r-1)/(r-1)$ represents 1 if $r=1$. The optimal r_1 is the value of r that maximizes P_r. Since

$$\begin{aligned} &P_{r+1} \le P_r \qquad \text{if and only if} \\ &\frac{r}{n}\sum_{r+1}^{n}\frac{1}{k-1} \le \frac{r-1}{n}\sum_{r}^{n}\frac{1}{k-1} \qquad \text{if and only if} \\ &\sum_{r+1}^{n}\frac{1}{k-1} \le 1, \end{aligned}$$

we see that the optimal rule is to select the first candidate that appears among applicants from stage r_1 on, where

$$r_1 = \min\{r \ge 1 : \sum_{r+1}^{n}\frac{1}{k-1} \le 1\}. \tag{4}$$

The following table is easily constructed.

$n =$	1	2	3	4	5	6	7	8
$r_1 =$	1	1	2	2	3	3	3	4
$P_{r_1} =$	1.0	.500	.500	.458	.433	.428	.414	.410

It is of interest to compute the approximate values of the optimal r_1 and the optimal P_{r_1} for large n. Since $\sum_{r+1}^{n} 1/(k-1) \sim \log(n/r)$, we have approximately $\log(n/r_1) = 1$, or $r_1/n = e^{-1}$. Hence, for large n it is approximately optimal to pass up a proportion

$e^{-1} = 36.8\%$ of the applicants and then select the next candidate. The probability of obtaining the best applicant is then approximately e^{-1}.

§**2.2 Arbitrary Monotonic Utility**. Problems in which the objective is to select the best of the applicants, as in assumption 8, are called **best-choice problems**. These are problems of the very particular person who will be satisfied with nothing but the very best. We extend the above result to more arbitrary payoff functions of the rank of the selected applicant. For example, we might be interested in getting one of the best two applicants; or we might be interested in minimizing the expected rank of the applicant selected, rank 1 being best. Let $U(j)$ be your payoff if the applicant you select has rank j among all applicants. It is assumed that $U(1) \geq U(2) \geq \cdots \geq U(n)$. The lower the ranking the more valuable. If no selection is made at all, the payoff is a fixed number denoted by Z_∞ (allowed to be greater than $U(n)$). In the best-choice problem, $U(1) = 1$, $U(2) = \cdots = U(n) = 0$, and $Z_\infty = 0$.

Let X_j be the relative rank of the jth applicant among the first j applicants observed. Then the X_j are independent, and X_j is uniformly distributed on the integers, $\{1, 2, ..., j\}$. The reward function $y_j(x_1, \ldots, x_j)$ is the expected payoff given that you have selected the jth applicant and it has relative rank x_j. The probability that an applicant of rank x among the first j applicants has eventual rank b among all applicants is the same as the probability that the applicant of rank b will be found in a sample of size j and have rank x there, namely

$$f(b|j,x) = \frac{\binom{b-1}{x-1}\binom{n-b}{j-x}}{\binom{n}{j}} \qquad \text{for} \quad b = x, \ldots, n-j+x,$$

(the negative hypergeometric distribution). Hence for $1 \leq j \leq n$,

$$y_j(x_1, ..., x_j) = \sum_{b=x}^{n-j+x} U(b) f(b|j,x)$$

where $x = x_j$. To force you to take at least one observation, we take $y_0 = -\infty$. We note that y_j depends only on x_j and may be written $y_j(x_j)$. As a practical matter, computation may be carried out conveniently using a backward recursion. The recursion for the probabilities,

$$f(b|j-1,x) = \frac{x}{j} f(b|j,x+1) + \frac{j-x}{j} f(b|j,x),$$

implies the backward recursion for the expected values,

$$y_{j-1}(x) = \frac{x}{j}\, y_j(x+1) + \frac{j-x}{j}\, y_j(x) \tag{5}$$

for $j > 1$, with initial conditions, $y_n(x) = U(x)$ for $1 \leq x \leq n$.

The horizon for the secretary problem is n. If you go beyond the horizon, you receive Z_∞, so the initial condition on the $V^{(n)}$ is: $V_n^{(n)}(x_n) = \max(U(x_n), Z_\infty)$. Since the X_i are independent, the conditional expectation in the right side of (1) reduces to an unconditional expectation. Since y_j depends on $x_1, \ldots, x_j$ only through the values of x_j, the same is true of $V_j^{(n)}$. Hence, for $j = n-1, \ldots, 1$,

$$V_j^{(n)}(x_j) = \max\{y_j(x_j), \frac{1}{j+1}\sum_{x=1}^{j+1} V_{j+1}^{(n)}(x)\}.$$

It is optimal to stop at j if

$$y_j(x_j) \ge \frac{1}{j+1}\sum_{x=1}^{j+1} V_{j+1}^{(n)}(x)$$

and to continue otherwise. The generalization of the result for the CSP that the optimal rule is a threshold rule, is contained in the following lemma.

Lemma. *If it is optimal to select an applicant of relative rank x at stage k, then*
(a) it is optimal to select an applicant of relative rank $x-1$ at stage k, and
(b) it is optimal to select an applicant of relative rank x at stage $k+1$.

Proof. Let $A(j) = (1/j)\sum_{i=1}^{j} V_j^{(n)}(i)$. The hypothesis is that $y_k(x) \ge A(k+1)$. We are to show that (a) $y_k(x-1) \ge A(k+1)$, and (b) $y_{k+1}(x) \ge A(k+2)$. (a) follows since $y_k(x-1) \ge y_k(x)$. To see (b), note first that $A(k+1) \ge A(k+2)$ since $A(k+1)$ is an average of quantities $V_{k+1}^{(n)}(i)$ each at least as large as $A(k+2)$. Thus, (b) will follow if we show $y_{k+1}(x) \ge y_k(x)$. To see this, use the recursion (5) to obtain,

$$\begin{aligned} y_k(x) &= (x/(k+1))y_{k+1}(x+1) + ((k+1-x)/(k+1))y_{k+1}(x) \\ &\le (x/(k+1))y_{k+1}(x) + ((k+1-x)/(k+1))y_{k+1}(x) \\ &= y_{k+1}(x). \quad \blacksquare \end{aligned}$$

This lemma implies that an optimal rule has the following form. Let r_x denote the first stage at which it is optimal to select an applicant of relative rank x. Then, $1 \le r_1 \le r_2 \le \ldots \le r_n \le n$ and if at stage j you see an applicant of relative rank x, you stop if $j \ge r_x$. For example, when $U(3) = U(4) = \ldots = U(n) = Z_\infty = 0$, the threshold rule takes the form $N_{r,s}$: continue through stage $r-1$; from stages r through $s-1$, select an applicant of relative rank 1; from stages s through n, select an applicant of relative rank 1 or 2. (See Exercise 3.)

As another application consider the problem of **minimizing the expected rank** of the applicant selected. In this case, $U(j) = j$, $y_j(x_j) = (n+1)x_j/(j+1)$ as may be found using (5), and we are trying to *minimize* $\mathrm{E}y_N(X_N)$. D. V. Lindley (1961) introduced this problem. For small values of n the optimal values of the r_x, call them $r_x(n)$, may be computed without difficulty. The problem of approximating these values when n is large has been solved by Chow, Moriguti, Robbins, and Samuels (1964). Their results may be summarized as follows. Let $V(n)$ denote the expected rank of the applicant selected by an optimal rule when there are n applicants. Then as $n \to \infty$,

$$V(n) \to V(\infty) = \prod_{1}^{\infty}(1+2/j)^{1/(j+1)} = 3.8695\cdots \quad \text{and}$$

$$r_x(n)/n \to (1/V(\infty)) \prod_{1}^{x-1}(1+2/j)^{1/(j+1)} = b(x).$$

Thus, $b(1) = .2584\cdots$, $b(2) = .4476\cdots$, $b(3) = .5640\cdots$, and for large x, $b(x) = 1 - 2/(x+1) + O(1/x^3)$. Therefore the optimal rule has the following description. Wait until 25.8% of the applicants have passed; then select any applicant of relative rank 1. After 44.8% of the applicants have passed, select any applicant of relative rank 1 or 2. After 56.4% have passed, select any candidate of relative rank 1, 2 or 3, etc.

The paper of Mucci (1973) contains an extension of these results to general non-decreasing payoff based on the ranks. An interesting possibility is that for some reward functions, y_j, the function $b(x) = \lim_{n\to\infty} r_x(n)/n$ may not tend to 1 as $x \to \infty$; that is, there may be an upper bound on the proportion of applicants it is optimal to view.

§**2.3 Variations.** There is a large literature dealing with many variations of the CSP. For example, there may be a probability q that an applicant will not accept an offered job (Exercise 2). Or, in the viewpoint of the marriage problem, it may be a worse outcome to marry the wrong person than not to get married at all (Exercise 1).

In Gilbert and Mosteller (1966), the intrinsic values of the objects are revealed, and you are allowed to use the information in your search for the best object. To be specific, there is a sequence of independent identically distributed continuous random variables, $X_1, \ldots, X_n$, that are shown to you one at a time, and you may stop the proceedings by selecting the number being shown to you. As before, the payoff is a function of the true rank of the number you select. For the best-choice problem, you win 1 if you select the largest X_i and you win 0 otherwise. Gilbert and Mosteller solve the best-choice problem when the distribution of the X_i's is fully known, and show that the limiting probability of choosing the best converges to $v* = .58016\cdots$ as $n \to \infty$. This is to be compared to the limiting value e^{-1} when the intrinsic values are not revealed.

The first description of the secretary problem in print occurs in Martin Gardner's Scientific American column of February 1960, in which he describes a game theoretic version called "Googol". In this version, player I chooses n numbers, $X_1, \ldots, X_n$, writes them on slips of paper and puts them in a hat. Player II, not knowing the numbers, pulls them out at random one at a time. Player II may stop at any time, and if he does he wins one if the last number he has drawn is the largest of all the n numbers. Clearly, player II can achieve at least probability P_{r_1} of winning by using only the relative ranks of the numbers drawn and the optimal stopping rule for the CSP. Can he do better using the actual values of the X's? Can player I choose the X's without giving away any extra information? (See Exercise 4.)

Problems such as Googol, in which the distribution of the X's is completely unknown, are called **no-information** problems. The classical secretary problem is included in this group. Problems such as the problem of Gilbert and Mosteller, in which the X's are i.i.d. with a known distribution, are called **full-information** problems. A number of intermediate versions have been proposed featuring some sort of **partial information** about the distribution of the X's. Petruccelli (1980) treats the normal distribution with unknown mean and shows that the best invariant rule achieves $v*$ as a limiting probability of selecting the best. Thus, asymptotically nothing is lost by not knowing the mean of the normal distribution. On the other hand, if the distribution is uniform on the interval (a, b) with a and b unknown, Stewart (1978) and Samuels (1981) show that the minimax stopping rule is based only on relative ranks, and so gives limiting probability e^{-1}, so that learning the distribution does you no good asymptotically. Campbell and Samuels (1981) consider the problem of selecting the best among the last n objects from a sequence of $m+n$ objects. The success probabilities converge as $m/(m+n) \to t$ to a quantity $p^*(t)$, where $p^*(t)$ increases from e^{-1} at $t=0$ (no-information) to $v*$ at $t=1$ (full-information).

Another variation lets the total number of objects that are going to be seen be unknown, violating assumption 2 of the CSP. It is assumed instead that the number of objects is random with some known distribution. These problems were introduced by Presman and Sonin (1972). (See Exercise 5.) Abdel-Hamid, Bather, and Trustrum (1982) consider admissibility of stopping rules for this problem, and J. D. Petruccelli (1983) gives conditions under which the optimal rule is a threshold rule.

Still another variation, introduced by Yang (1974) allows us to attempt to select an object we have already passed over. This is called backward solicitation. The probability of success of backward solicitation depends on how far in the past the object was observed. For example, one may be allowed to choose any of the last m objects that have been observed, as in Smith and Deely (1975). This problem has been extended to the full information case by Petruccelli (1982), and to the case where an option may be bought to be able to recall a desirable object subsequently by J. Rose (1984).

§**2.4 The Cayley-Moser Problem.** A finite horizon version of the problem of selling an asset was proposed by Arthur Cayley in 1875. A smooth version of Cayley's problem, due to Moser (1956), is presented in this section. Another important finite horizon problem, the parking problem of MacQueen and Miller (1960), is described in the next section.

In the Cayley problem, a population of m objects with known values, $x_1, x_2, \ldots, x_m$, is given. From this population, a simple random sample of size at most n, where $n \le m$, is drawn sequentially without replacement. You may stop the sampling at any time by selecting the last chosen object and you receive the value of that object as a reward, but you must stop and select the nth sample if you have not stopped before that stage. The problem is to find a stopping rule that maximizes the expected value of the selected object. Cayley discusses this problem and suggests solving it by the method of backward induction. He then applies the method to a simple example of a population of size $m = 4$

with values $\{1, 2, 3, 4\}$, and finds for $n = 1, 2, 3, 4$, the optimal rules and the maximal expected rewards. (Exercise 7.) The Cayley problem (sampling without replacement from a population of size m of known values $1, 2, \ldots, m$) has been extended to sampling with recall and an arbitrary payoff function by Chen and Starr (1980).

Moser reformulates Cayley's problem in a way that provides an approximation to the problem when m is large and the population values are $\{1, 2, \ldots, m\}$. He assumes that the observations, $X_1, X_2, \ldots, X_n$, are independent and identically distributed according to a uniform distribution on the interval, $(0, 1)$. The return for stopping at stage j is $Y_j = X_j$, at least one observation must be taken (so $y_0 = -\infty$), and stopping is required if stage n is reached.

Let us derive the optimal stopping rule by the method of equation (1), assuming an arbitrary common distribution, $F(x)$, for the X_j, with finite first moment. We have $V_n^{(n)} = Y_n = X_n$, and inductively for $j = n-1$ down to 1, $V_j^{(n)} = \max\{X_j, \mathrm{E}(V_{j+1}^{(n)}|\mathcal{F}_j)\}$. The independence of the X_j implies inductively that $V_j^{(n)}$ depends only on X_j and that $A_{n-j} = \mathrm{E}(V_{j+1}^{(n)}|\mathcal{F}_j)$ is a constant that depends only on $n - j$, the number of stages to go. Thus, the optimal stopping rule stops at j if $X_j \geq A_{n-j}$, where the A_j may be computed inductively as

$$
\begin{aligned}
A_0 &= -\infty \\
A_1 &= \mathrm{E}X_1 \qquad \text{and for } j \geq 1, \\
A_{j+1} &= \mathrm{E}\max\{X, A_j\} = \int_{-\infty}^{A_j} A_j\, dF(x) + \int_{A_j}^{\infty} x\, dF(x).
\end{aligned} \tag{6}
$$

When the X_j are uniformly distributed on $(0, 1)$, we may as well put $A_0 = 0$, and we find $A_1 = 1/2$, and for $j \geq 1$,

$$
A_{j+1} = \int_0^{A_j} A_j\, dx + \int_{A_j}^1 x\, dx = (A_j^2 + 1)/2 \tag{7}
$$

We have $A_2 = 5/8$, $A_3 = 89/128$, and so on. Moser finds an approximation to the A_n for large n as

$$
A_n \simeq 1 - \frac{2}{n + \log(n) + c} \tag{8}
$$

for some constant c.

This may be demonstrated as follows. If we let $B_n = 2/(1 - A_n) - n$, then $B_0 = 2$ and (7) reduces to

$$
B_{n+1} - B_n = \frac{1}{B_n + n - 1},
$$

and we are to show that $B_n - \log(n) \to c$. It is clear that the B_n are increasing. Then, $B_n \geq 2$ for all n so that the differences $B_{n+1} - B_n$ are bounded above by $1/(n+1)$ and

$$
B_n = B_0 + \sum_{j=0}^{n-1}(B_{j+1} - B_j) \leq 2 + \sum_{j=0}^{n-1} \frac{1}{j+1}. \tag{9}
$$

This upper bound on B_n gives us a lower bound on the differences,

$$B_{n+1} - B_n \geq \frac{1}{n+1+\sum_{j=0}^{n-1} \frac{1}{j+1}}$$

from which we may obtain a similar lower bound on the B_n:

$$B_n \geq 2 + \sum_{j=0}^{n-1} \frac{1}{j+1+\sum_{i=0}^{j-1} \frac{1}{i+1}}. \tag{10}$$

The difference of the bounds (9) and (10) is a sum of the form $\sum \log(n)/n^2$ and so is bounded. Thus we have $B_n = \log(n) + O(1)$. Finally, let $C_n = B_n - \sum_{j=1}^{n} 1/j$. Then,

$$C_{n+1} - C_n = \frac{1}{n-1+C_n+\sum_1^n 1/j} - \frac{1}{n+1} \sim -\frac{\log(n)}{(n+1)^2}$$

shows that C_n is a Cauchy sequence and hence converges. Thus, $B_n - \log(n)$ also converges to some constant c and (8) follows. Gilbert and Mosteller (1966) have approximated the constant c. To six figures, it is 1.76799.

Equations for the A_j are discussed in Guttman (1960) for the normal distribution, in Karlin (1962) for the exponential distribution (see Exercise 8), in Gilbert and Mosteller (1966) for the Pareto distribution and in Engen and Seim (1987) for the gamma distribution.

Various finite horizon extensions of the Moser problem have been suggested. Karlin (1962) and Saario (1986) treat the problem in which there are several objects to be selected. Hayes (1969) allows future returns to be discounted. Karlin (1962), Mastran and Thomas (1973) and Sakaguchi (1976) treat problems with random arrivals. But the main extension of Moser's problem is to allow an unbounded horizon, in which case a cost of observation or a discount factor must be included so that one is not inclined to continue forever. These problems are referred to as house-selling or selling an asset or search for the best offer and are treated in Chapter 4.

Another important generalization of the Moser problem may be found in the work of Derman, Lieberman and Ross (1972). In this problem, there are n workers of known values, $p_1 \leq \ldots \leq p_n$ and n jobs that appear sequentially with random values, $X_1, \ldots, X_n$, assumed to be independent and identically distributed according to a known distribution. The return for assigning worker i to job j is the product $p_i X_j$. The problem is to assign workers to jobs, immediately as the job values become known, in such a way to maximize the expectation of the sum of the returns. This contains the Moser problem when $p_1 = \cdots = p_{n-1} = 0$ and $p_n = 1$. It is a more complex decision problem than the stopping rule problems treated here since at each stage we must decide among a number of actions, not just two. Derman et al. show that there is an optimal policy of the following form: For each $n \geq 1$, there are numbers $-\infty = a_{0n} \leq a_{1n} \leq a_{2n} \leq \cdots \leq a_{nn} = +\infty$ such that if there are n workers remaining to be assigned and a job of value X_1 appears, the job is assigned to worker i if X_1 is contained in the interval $[a_{i-1,n}, a_{in}]$. A remarkable feature of this result is that the a_{ij}'s do not depend on the values of the p_i's, so long as they are arranged in increasing order. This result has been extended to random arrival infinite horizon problems with a discount by Albright (1974). See R. Righter (1990) for further extensions and recent developments.

§**2.5 The Parking Problem.** (MacQueen and Miller (1960)) You are driving along an infinite street toward your destination, the theater. There are parking places along the street but most of them are taken. You want to park as close to the theater as possible without turning around. If you see an empty parking place at a distance d before the theater, should you take it?

Here, we model this problem in a discrete setting. We assume that we start at the origin and that there are parking places at all integer points of the real line. Let $X_0, X_1, X_2, \ldots$ be independent Bernoulli random variables with common probability p of success, where $X_j = 1$ means that parking place j is filled and $X_j = 0$ means that it is available. Let $T > 0$ denote your target parking place. You may stop at parking place j if $X_j = 0$, and if you do you lose $|T - j|$. You cannot see parking place $j+1$ when you are at j, and if you once pass up a parking place you cannot return to it. If you ever reach T, you should choose the next available parking place. If T is filled when you reach it, your expected loss is then $(1-p) + 2p(1-p) + 3p^2(1-p) + \ldots = 1/(1-p)$, so that we may consider this as a stopping rule problem with finite horizon T and with loss

$$y_T = 0 \quad \text{if} \quad X_T = 0 \qquad \text{and} \qquad y_T = 1/(1-p) \quad \text{if} \quad X_T = 1$$

and for $j = 0, \ldots, T-1$,

$$y_j = T - j \quad \text{if} \quad X_j = 0 \qquad \text{and} \qquad y_j = \infty \quad \text{if} \quad X_j = 1.$$

The value $y_j = \infty$ forces you to continue if you reach a parking place j and it is filled. We seek a stopping rule, $N \le T$, to minimize $\mathrm{E}Y_N$.

First we show that if it is optimal to stop at stage j when $X_j = 0$, then it is optimal to stop at stage $j+1$ when $X_{j+1} = 0$. As in Moser's problem, $V_j^{(T)}$ depends only on X_j, and $A_{n-j} = \mathrm{E}(V_{j+1}^{(T)}|\mathcal{F}_j)$ is a constant that depends only on $n-j$. It is optimal to stop at stage $n-j$ if $y_{n-j} \le A_j$. We are to show that if $n - j \le A_j$, then $n - j - 1 \le A_{j-1}$. This follows from the inequalities, $n - j - 1 < n - j \le A_j \le A_{j-1}$.

Thus, there is an optimal rule of the threshold form, N_r for some $r \ge 0$: continue until r places from the destination and park at the first available place from then on. Let P_r denote the expected cost using this rule. Then, $P_0 = p/(1-p)$, and for $r \ge 1$, $P_r = (1-p)r + pP_{r-1}$. We will show by induction that

$$P_r = r + 1 + \frac{2p^{r+1} - 1}{1 - p}. \tag{11}$$

$P_0 = p/(1-p) = 1 + (2p-1)/(1-p)$, so it is true for $r = 0$. Suppose it is true for $r-1$; then $P_r = (1-p)r + pP_{r-1} = (1-p)r + pr + p(2p^r - 1)/(1-p) = (r+1) + (2p^{r+1} - 1)/(1-p)$, as was to be shown.

To find the value of r that minimizes (11), look at the differences, $P_{r+1} - P_r = 1 + (2p^{r+2} - 2p^{r+1})/(1-p) = 1 - 2p^{r+1}$. Since this is increasing in r, the optimal value is the first r for which this difference is nonnegative, namely, $\min\{r \geq 0 : p^{r+1} \leq 1/2\}$. For example, if $p \leq 1/2$, you should reach the destination before looking for a parking place. But if $p = .9$, say, we should start looking for a parking place $r = 6$ places before the destination.

There are various extensions to this problem. There may be a cost of time or gas. (See Exercise 10.) MacQueen and Miller (1960) allow you to drive around the block in search for a parking place, and Tamaki (1988) allows you to make a U-turn, at a cost, to return to a previously observed parking space. In Tamaki (1985), the problem is extended to allow the probability of finding parking space j free to depend on j. (See Exercises 11 and 12.) Other extensions are treated in Chapter 5.

§2.6 Exercises.

1. *The win-lose-or-draw marriage problem.* (Sakaguchi (1984)) Consider the marriage problem in which the payoff is 1 if you select the best, -1 if you select someone who is not the best, and 0 if you stay single.
(a) Find the optimal rule.
(b) Show that for large n the rule is approximately to pass up the first $1/\sqrt{e} = 60.6\cdots\%$ of the possibilities and to select the next relatively best, if any.

2. *Uncertain employment.* (Smith (1975)) Consider the classical secretary problem with the added possibility that an observed candidate may be unavailable (the applicant may refuse the offer). If this is the case, you are not allowed to select the applicant and the search goes on. Let ϵ_k represent the indicator of the event that applicant k is available; it is assumed that the ϵ_k are independent and independent of the X_j with $\mathrm{P}(\epsilon_k = 1) = p$ for all k. The outcome of availability or not is made known to you at the time of observation.
(a) Show that $y_k = (k/n)I(X_k = 1, \epsilon_k = 1)$.
(b) Show that there is a threshold rule that is optimal.
(c) Show that the probability of win using the threshold rule N_r is

$$P_r = \frac{p}{n}\sum_{k=r}^{n}\frac{\Gamma(r)\Gamma(k-p)}{\Gamma(k)\Gamma(r-p)}.$$

(d) Use $(n-p)^p\Gamma(n-p)/\Gamma(n) \to 1$ as $n \to \infty$ to show that the optimal threshold is approximately $r = np^{1/(1-p)}$.

3. *One of the best two.* (Gilbert and Mosteller (1966), Gusein-Zade (1966)) Consider the secretary problem with $U(3) = \ldots = U(n) = Z_\infty = 0$ and $U(1) = a \geq U(2) = b \geq 0$. An optimal rule is of the form, $N_{r,s}$: do not stop in stages 1 through $r-1$; in stages r through $s-1$, select an object if it has relative rank 1; in stages s through n, select an

object if it has relative rank 1 or 2.
(a) Show P(select the best$|N_{r,s}) = ((r-1)/n)[\sum_r^{s-1} 1/(k-1)+(s-2)\sum_s^n 1/((k-1)(k-2))]$.
(b) Show P(select second best$|N_{r,s}) = ((r-1)/n)[(1/(n-1))\sum_r^{s-1}(n-k)/(k-1) + (s-2)\sum_s^n 1/((k-1)(k-2))]$.
(c) Let $V(r,s) = \mathrm{E}(\text{return}|N_{r,s})$. Let $n \to \infty$, $r/n \to x$, and $s/n \to y$. Show that $V(r,s) \to (a+b)x(\log(y/x)+1-y) - bx(y-x)$.
(d) Show that this limit is maximized if $y = (a+b)/(a+2b)$ and x satisfies $\log(x) = (2b/(a+b))x - 1 + \log(y)$.
(e) Specialize to the case $a = b = 1$. (Ans. $x = .3475\cdots$, $y = 2/3$ and limiting probability of selecting one of the best two $= .574\cdots$.

4. *Googol.* (Berezovskiy and Gnedin (1984)) Let $X_1, \ldots, X_n$ be i.i.d. uniform on $(0,\theta)$ where θ has the Pareto distribution, $\mathcal{P}a(\alpha,1)$. The Pareto distribution, $\mathcal{P}a(\alpha,x)$ is defined to be the distribution with density

$$f(\theta|\alpha,x) = \alpha x^\alpha \theta^{-(1+\alpha)} I(\theta > x),$$

where $\alpha > 0$ and x is an arbitrary real number. Let $M_0 = X_0 = 1$, and for $j = 1, \ldots, n$ let $M_j = \max\{X_0, X_1, \ldots, X_j\} = \max\{M_{j-1}, X_j\}$. You observe the X's one at a time and you may stop at any time. If the most recently observed X_j when you stop is the largest of all the X's, including X_0, you win.
(a) Show the posterior distribution of θ given $X_1, \ldots, X_j$ is $\mathcal{P}a(j+\alpha, M_j)$.
(b) Show $y_j = \mathrm{P}\{X_j = M_n | X_1, \ldots, X_j\} = ((j+\alpha)/(n+\alpha))I(X_j = M_j)$.
(c) Show that if it is optimal to stop at stage j with $X_j = M_j$, then it is optimal to stop at stage $j+1$ if $X_{j+1} = M_{j+1}$.
(d) Show that the optimal stopping rule is to pass up $r-1$ numbers and then stop at the next j such that $X_j = M_j$, where r is that integer that maximizes $P_r = ((r-1+\alpha)/(n+\alpha))\sum_{j=r}^n 1/(j-1+\alpha)$.

5. *Random Number of Applicants.* (Presman and Sonin (1972), Rasmussen and Robbins (1975)) The number, K, of applicants is unknown, but it is assumed to be random with a uniform distribution on $\{1, \ldots, n\}$. The assumptions are as in the CSP, but if you pass up an applicant and it turns out that there are no more, you lose. You want to select the best of the K applicants.
(a) Find y_j for $j = 1, \ldots, n$.
(b) Show there is a threshold rule that is optimal.
(c) Find the probability of win using an arbitrary threshold rule.
(d) Find approximately, for large n, the optimal threshold rule and the optimal reward.

6. *The Duration Problem.* (Ferguson, Hardwick and Tamaki (1992)) The first 7 assumptions of the CSP are in force, but the payoff now is the proportion of time you are in possession of the relatively best applicant. Thus, if the jth applicant is relatively best and you select him, you win $(k-j)/n$ if the next relatively best applicant is the kth, and you win $(n+1-j)/n$ if the jth applicant is best overall.
(a) Find y_j for $j = 1, \ldots, n$.
(b) Show there is a threshold rule that is optimal.

(c) Find the expected winnings using an arbitrary threshold rule. Note that the answer is exactly the same as for Problem 5(c) so the asymptotic values of 5(d) hold for this problem as well.

7. *The Cayley Example.* Suppose in Cayley's problem that the population consists of $m = 4$ objects of values $1, 2, 3$ and 4. For $n = 1, 2, 3$ and 4, find the optimal stopping rule and its expected payoff if one may sample at most n times.

8. *Moser's problem with an exponential distribution.* (Karlin (1962)) In Moser's problem, suppose the common distribution of the X_i is exponential on $(0, \infty)$.
(a) Find the recurrence relation for the A_j.
(b) Show that $A_n = \log(n) + o(1)$; that is, show $A_n - \log(n) \to 0$.
(c) *Moser's problem with two stops.* (Karlin (1959) vol. 2, Chap. 9, Exercise 13.) In Moser's problem with $\mathcal{U}(0, 1)$ variables and finite horizon n, suppose you are allowed to choose two numbers from the sequence and are given the sum as a payoff. How do you proceed?

9. *Moser's problem with nonidentical distributions.* (Moser (1956)) Suppose, in Moser's problem, that $X_1, \ldots, X_n$ are independent and X_i has a uniform distribution on the interval $(0, n + 1 - i)$.
(a) Find the recurrence for the sequence of cutoff points A_n.
(b) Show that $A_n \simeq n - \sqrt{2n}$ in the sense that $(n - A_n)/\sqrt{2n} \to 1$.

10. *The parking problem with cost.* Solve the parking problem if the loss for stopping at parking place j is changed to $|T - j| + cj$, where $c > 0$ represents a cost of time or gas.

11. *The parking problem with arbitrary probabilities and distances between parking places.* Let $X_0, X_1, \ldots, X_T$ be independent Bernoulli random variables where $P(X_i = 1) = p_i$, $0 \le p_i \le 1$ for all i. Let $s_0 \le s_1 \le \ldots \le s_T$ be a nondecreasing sequence of numbers with $s_T > 0$. Consider the finite horizon stopping rule problem with observations $\{X_i\}$ and rewards Y_i for stopping at i, where $Y_i = s_i$ if $X_i = 1$, and $Y_i = 0$ if $X_i = 0$, for $i = 0, 1, \ldots, T$. (This contains the parking problem with non-constant distances between parking spaces, and probability of a free space dependent on its position.) Let W_j denote the optimal expected reward if it is decided to continue from stage j for $j = 0, 1, \ldots, T-1$, and $W_T = 0$. Note that $W_{j-1} = \mathrm{E}\{V_j^{(T)} | \mathcal{F}_{j-1}\}$ is a constant.
(a) Find the backward recursion equations for the W_j.
(b) Show there is an optimal rule of the form N_r: Stop at the first $j \ge r$ at which $X_j = 1$.

12. *The parking problem with random distances.* Generalize the parking problem to allow random nonnegative distances, Z_j, between parking places. (This extension also contains Exercise 11 when the ϵ_j below are identically one.) One can think of the Z_j as the travel time between parking places which is random due to traffic fluctuations. To be specific: Assume that the observations, $X_j = (Z_j, \epsilon_j)$ for $j = 1, \ldots, T$, are independent with the Z_j nonnegative, the ϵ_j Bernoulli with probability p_j of success, and with Z_j and ϵ_j independent for all j. Let $S_n = \sum_1^n Z_j$, and let the reward for stopping at n be $Y_n = S_n I(\epsilon_n = 1)$. Let $W_j(S_j) = \mathrm{E}\{V_{j+1}^{(T)} | \mathcal{F}_j\}$ denote the optimal expected reward if it is decided to continue from stage j.
(a) Find the backward recursion equations for the $W_j(s)$.

(b) Show that there exist numbers, $r_1 \geq r_2 \geq \cdots \geq r_T = 0$, such that it is optimal to stop at stage j if $S_j \geq r_j$ and $\epsilon_j = 1$.

13. *Fishing.* (Starr (1974)) Suppose that a lake contains n fish whose catch times, $T_1, \ldots, T_n$, are i.i.d. exponential, with density $f(t) = \exp\{-t\}I(t > 0)$. Let $X_1, \ldots, X_n$ denote the order statistics of the T_j. If you stop after catching j fish, you receive $Y_j = j - cX_j$, for $j = 0, 1, \ldots, n$ where $X_0 = 0$ and $c > 0$. Find an optimal stopping rule. Hint: Recall that $X_1, X_2 - X_1, X_3 - X_2, \ldots, X_n - X_{n-1}$ are independent and that $Z_j = X_j - X_{j-1}$ has density $f(z) = (n + 1 - j)\exp\{-(n + 1 - j)z\}I(z > 0)$.

14. *A Symmetric Random Walk Secretary Problem.* (Blackwell and MacQueen, (1996), personal communication.) Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables. Let $S_k = \sum_1^k X_i$ denote the partial sums with $S_0 = 0$, and let $M_k = \max\{S_0, S_1, \ldots, S_k\}$ be the maxima of the partial sums. If we stop at stage k, $0 \leq k \leq n$, we win if and only if $S_k = M_n$. This is a full-information secretary problem with the worths of the secretaries following a random walk. Note that ties among the S_j may occur, but you win if you tie for best. If you stop at a stage k with $S_k < M_k$ you can't possibly win. If you stop at stage k with $S_k = M_k$, then your probability of winning is $y_k = p_{n-k}$, where $p_j = \mathrm{P}(S_1 \leq 0, \ldots, S_j \leq 0)$. Thus this is a finite horizon stopping rule problem with payoff for stopping at k equal to $y_k \mathrm{I}(S_k = M_k)$.

The problem seems hard in general, but make the assumption that *the distribution of the X's is symmetric about* 0, and do the following.

(a) Suppose you are at a stage k with $S_k = M_k$ and you decide to continue until stage n in the hope that $S_n = M_n$. The probability you win is then q_{n-k}, where $q_j = \mathrm{P}(S_0 \leq S_j, \ldots, S_{j-1} \leq S_j)$. Show that $p_j = q_j$ for all j.

(b) Show that any stopping rule is optimal provided it does not stop at any $k < n$ for which $S_k < M_k$. In particular, the rule that continues to the last stage and stops is optimal.

(c) Suppose that the distribution of the X's is Bernoulli with $\mathrm{P}(X = 1) = \mathrm{P}(X = -1) = 1/2$. Show $p_n = \binom{n}{\lfloor n/2 \rfloor} 2^{-n}$.

(d) Suppose that the distribution of the X's is double exponential with density $\frac{1}{2}e^{-|x|}$. Show that

$$p_n = \frac{1}{2^{2n-1}} \binom{2n-1}{n} \qquad \text{for } n = 1, 2, \ldots$$

One way is to use the fundamental identity of Wald ($\mathrm{E}e^{tS_N} M(t)^{-N} \equiv 1$, valid for t for which $M(t)$ is finite, where $M(t) = \mathrm{E}e^{tX}$ is the moment generating function of X) to find the probability generating function of $N = \min\{n > 0 : S_n > 0\}$ is $\mathrm{E}u^N = 1 - \sqrt{1-u}$.

(e) Show that the answer to (d) holds for *any* continuous symmetric distribution of the X's. (Use Theorem 8.4.3 of Chung, (1974).)

15. *The Cayley-Moser problem with independent non-identically distributed distributions.* (Assaf, Goldstein and Samuel-Cahn (2000)) If, in the Cayley-Moser problem, the X_i have different distributions with different means etc., the decision maker should be able to improve his expected return by choosing the order in which he observes the X_i. Let $F_j(x)$ denote the distribution function of X_j and assume that $X_j \geq 0$ for all j. Suppose

we observe the X_i sequentially in reverse order, $X_n, X_{n-1}, \ldots, X_1$. Then corresponding to (6), the optimal rule stops at X_{j+1} if $X_{j+1} > A_j$, where $A_0 = 0$, $A_1 = \mathrm{E}X_1$ and for $j \geq 1$

$$A_{j+1} = \mathrm{E}\max\{X_{j+1}, A_j\} = \int_0^{A_j} A_j\, dF_{j+1}(x) + \int_{A_j}^{\infty} x\, dF_{j+1}(x).$$

Define the function $g_j(\alpha) = \mathrm{E}\max\{X_j, \alpha\}$; then $A_1 = g_1(0)$, $A_2 = g_2(g_1(0))$, and so forth.

Suppose the random variable X_j has distribution function $F(x|\theta_j)$, where

$$F(x|\theta) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{(1-\theta)^2}{(1-x\theta)^2} & \text{if } 0 \leq x < 1, \\ 1 & \text{if } x \geq 1. \end{cases}$$

for $0 < \theta < 1$. Note the $F(x|\theta)$ has a mass at 0 of size $(1-\theta)^2$. The θ_j are known and between 0 and 1.
(a) Show $\mathrm{E}X_j = \theta_j$.
(b) Show

$$g_j(\alpha) = 1 - \frac{(1-\theta_j)(1-\alpha)}{1-\theta_j\alpha} \qquad \text{for } 0 < \alpha < 1.$$

(c) Show $g_1(g_2(\alpha)) = g_2(g_1(\alpha))$ for all $0 \leq \alpha < 1$.
(d) Conclude that A_n is a symmetric function of the θ_i so that the decision maker is indifferent as to the order in which he observes the X_i!

Chapter 3. THE EXISTENCE OF OPTIMAL STOPPING RULES.

Consider the general stopping rule problem of Chapter 1 with observations $X_1, X_2, \ldots$ and rewards $Y_0, Y_1, \ldots, Y_\infty$ where $Y_n = y_n(X_1, \ldots, X_n)$. The following two assumptions are basic to the theory of this chapter.

A1. $\mathrm{E}\{\sup_n Y_n\} < \infty$.
A2. $\limsup_{n\to\infty} Y_n \le Y_\infty$ a.s.

Assumption A1 allows us to interchange expectation and summation in what follows. It implies that even a prophet who can foresee the future and stop at the time that Y_n assumes its maximum value, or comes close to the supremum if the maximum does not exist, can only obtain a finite expected return. Thus, certainly $\sup_N \mathrm{E}Y_N < \infty$, where the supremum is taken over all stopping rules, N.

In §3.1, we show that under these two assumptions an optimal stopping rule exists. The treatment follows the method of Chow and Robbins (1963) using the notion of a *regular* stopping rule. In §3.2, we discuss the principle of optimality and the optimality equation. We show under assumptions A1 and A2 that the rule given by the principle of optimality is optimal. In §3.3, we derive Wald's equation, and in §3.4, we examine prophet inequalities.

Here are two examples that show if either one of the assumptions is not satisfied, an optimal stopping rule may not exist.

EXAMPLE 1. Let $X_1, X_2, \ldots$ be independent Bernoulli trials with probability 1/2 of success, and let $Y_0 = 0$,

$$Y_n = (2^n - 1)\prod_1^n X_i, \tag{1}$$

and $Y_\infty = 0$. As long as only successes have occurred, you may stop at stage n and receive $2^n - 1$; after the first failure has occurred, you receive 0. Since $Y_n \to 0$ a.s., A2 is satisfied. On the other hand, $\sup_n Y_n = 2^k - 1$ with probability $1/2^{k+1}$ for $k = 0, 1, 2, \ldots$ so that $\mathrm{E}\{\sup_n Y_n\} = \sum_0^\infty (1 - 1/2^k)/2 = \infty$ and A1 is not satisfied. If you reach stage n without any failures, your return for stopping is $2^n - 1$, while if you continue one stage you can get an expected value of at least $(2^{n+1} - 1)/2 = 2^n - 1/2$, which is better. Thus, it can never be optimal to stop before a failure has occurred. Yet continuing forever gives you a zero payoff so there is no optimal stopping rule. In fact, $\sup_N \mathrm{E}Y_N = 1$, but the supremum is not attained. ■

EXAMPLE 2. Let $Y_0 = 0, Y_n = 1 - 1/n$ for $n = 1, 2, \ldots$ and $Y_\infty = 0$. (The X_n are immaterial.) Here A1 is satisfied and A2 is not. Yet, like the previous example, the longer you wait the better off you are, but if you wait forever you win nothing. There is no optimal rule. ■

We remark that for minimization problems, where Y_n represents a cost rather than a reward, conditions A1 and A2 should be replaced by

A1. $\mathrm{E}\{\inf_n Y_n\} > -\infty$.
A2. $\liminf_{n\to\infty} Y_n \geq Y_\infty$ a.s.

§**3.1. Regular Stopping Rules.** We precede the main theorem on the existence of optimal stopping rules by two lemmas involving the notion of a regular stopping rule, a concept due to Snell (1952).

Definition. *A stopping rule N is said to be* **regular**, *if for every n,*

$$\mathrm{E}\{Y_N|\mathcal{F}_n\} > Y_n \quad \text{a.s. on} \quad \{N > n\}. \tag{2}$$

In other words, N is regular if $\mathrm{E}\{Y_N|X_1 = x_1, \ldots, X_n = x_n\} > y_n(x_1, \ldots, x_n)$ for almost all $(x_1, \ldots, x_n) \in \{N > n\}$. In still other words, N is regular if it has the property that if N tells you to continue at a certain stage, then N gives you an improved conditional expected return compared to stopping at that stage. The first lemma shows how to replace a given stopping rule by a regular one having no worse expected payoff.

Lemma 1. *Assume A1. Given any stopping rule N, there is a regular stopping rule N' such that $\mathrm{E}Y_{N'} \geq \mathrm{E}Y_N$.*

Proof. Define $N' = \min\{n \geq 0 : \mathrm{E}(Y_N|\mathcal{F}_n) \leq Y_n\}$. That is, N' tells you to use N until N tells you to continue and stopping is at least as good, in which case you stop. It is clear that N' is a stopping rule and that $N' \leq N$. On $\{N' = n\}$, we have $\mathrm{E}(Y_N|\mathcal{F}_n) \leq Y_n$ a.s. for all n, while on $\{N' = \infty\}$, we have $Y_N = Y_{N'} = Y_\infty$ a.s. Hence,

$$\begin{aligned}\mathrm{E}Y_{N'} &= \sum_{n=0}^{=\infty} \mathrm{E}(I\{N' = n\}Y_n) \geq \sum_{n=0}^{\infty} \mathrm{E}(I\{N' = n\}\mathrm{E}(Y_N|\mathcal{F}_n)) + \mathrm{E}(I\{N' = \infty\}Y_\infty) \\ &= \sum_{n=0}^{=\infty} \mathrm{E}(I\{N' = n\}Y_N) = \mathrm{E}Y_N,\end{aligned} \tag{3}$$

where the summation is over n from 0 to ∞ inclusive. The interchange of expectation and summation is valid by A1. To see that N' is regular, we use the same argument as above conditional on $\mathcal{F}_n$ to find that on $\{N' > n\}$, we have $\mathrm{E}(Y_{N'}|\mathcal{F}_n) \geq \mathrm{E}(Y_N|\mathcal{F}_n)$ a.s. Since $\mathrm{E}(Y_N|\mathcal{F}_n) > Y_n$ a.s. on $\{N' > n\}$, we have $\mathrm{E}(Y_{N'}|\mathcal{F}_n) > Y_n$ on $\{N' > n\}$. ■

The next lemma shows that we may improve on two regular stopping rules by stopping when the one with the longer life tells us to stop.

Lemma 2. *Assume A1. If N and N' are regular stopping rules, then so is $N'' = \max\{N, N'\}$ and then*

$$\mathrm{E}Y_{N''} \geq \max\{\mathrm{E}Y_N, \mathrm{E}Y_{N'}\}. \tag{4}$$

Proof. N'' is equal to N except on sets of the form $\{N = n\} \cap \{N' > n\}$ in which case $\mathrm{E}(Y_{N''}|\mathcal{F}_n) = \mathrm{E}(Y_{N'}|\mathcal{F}_n) > Y_n$ a.s. Hence,

$$\begin{aligned} \mathrm{E}Y_{N''} &= \sum \mathrm{E}(I\{N = n\}Y_{N''}) = \sum \mathrm{E}(I\{N = n\}\mathrm{E}(Y_{N''}|\mathcal{F}_n)) \\ &\geq \sum \mathrm{E}(I\{N = n\}Y_n) = \mathrm{E}Y_N. \end{aligned} \tag{5}$$

By symmetry, $\mathrm{E}Y_{N''} \geq \mathrm{E}Y_{N'}$.

To see that N'' is regular, note that $\{N'' > n\} = \{N > n\} \cup \{N' > n\}$. On $\{N > n\}$, an argument similar to the above but conditional on $\mathcal{F}_n$ shows that $\mathrm{E}(Y_{N''}|\mathcal{F}_n) \geq \mathrm{E}(Y_N|\mathcal{F}_n)$ a.s., which is greater than Y_n from the regularity of N. By symmetry, $\mathrm{E}(Y_{N''}|\mathcal{F}_n) > Y_n$ a.s. on $\{N' > n\}$. Hence, on $\{N'' > n\}$, $\mathrm{E}(Y_{N''}|\mathcal{F}_n) > Y_n$ a.s. showing the regularity of N''. ∎

With these two lemmas in hand, the main theorem of this chapter is an application of the *Fatou-Lebesgue Lemma*, which states: If $Z, X_1, X_2, \ldots$ is a sequence of real-valued random variables such that $X_n \leq Z$ for all n and $\mathrm{E}Z < \infty$, then $\limsup_n \mathrm{E}X_n \leq \mathrm{E}\limsup_n X_n$.

Theorem 1. *Under A1 and A2, there exists a stopping rule N^* such that $\mathrm{E}Y_{N^*} = V^*$, where $V^* = \sup_N \mathrm{E}Y_N$.*

Proof. If $V^* = -\infty$, the result is trivial. So we may assume that $-\infty < V^* < \infty$. Let $N_1, N_2, \ldots$ be a sequence of stopping rules such that $\mathrm{E}Y_{N_j} \to V^*$. Let $N'_1, N'_2, \ldots$ be the regularized versions as in Lemma 1, so that $\mathrm{E}Y_{N'_j} \to V^*$. Let $N''_j = \max\{N'_1, \ldots, N'_j\}$ so that by Lemma 2, $\mathrm{E}Y_{N''_j} \geq \mathrm{E}Y_{N'_j}$ and consequently $\mathrm{E}Y_{N''_j} \to V^*$. Note that N''_j is a monotone nondecreasing sequence of stopping rules converging to the stopping rule $N^* = \sup\{N'_1, N'_2, \ldots\}$. Moreover, since N''_j is a nondecreasing sequence of integers, either $N''_j \to \infty$ or N''_j is a fixed integer from some j on. Thus, $\limsup_{j\to\infty} Y_{N''_j} \leq Y_{N^*}$ a.s from A2. From the Fatou-Lebesgue Theorem, since the $Y_{N''_j}$ are bounded above by $\sup_n Y_n$ which is integrable by A1,

$$V^* = \limsup \mathrm{E}Y_{N''_j} \leq \mathrm{E}\limsup Y_{N''_j} \leq \mathrm{E}Y_{N^*}. \tag{6}$$

Since $\mathrm{E}Y_{N^*} \leq V^*$ by definition of V^*, we have $\mathrm{E}Y_{N^*} = V^*$. ∎

§**3.2. The Principle of Optimality and the Optimality Equation.** At the initial stage, we can obtain y_0 without sampling, or we can obtain V^* by using an optimal rule. Therefore, it is optimal to stop without sampling if, and only if, $y_0 = V^*$. We expect to be able to apply this principle at later stages too. If we have observed $X_1 = x_1, \ldots, X_n = x_n$, we may obtain $y_n(x_1,\ldots,x_n)$ without sampling further, or we can, by using a rule optimal for the remaining stages, obtain

$$V_n^*(x_1,\ldots,x_n)=\sup_{N\geq n}\mathrm{E}\{Y_N|X_1=x_1,\ldots,X_n=x_n\}. \tag{7}$$

Note $V_0^*=V^*$. Here $\sup_{N\geq n}$ means supremum over the set of all stopping rules N such that $\mathrm{P}(N\geq n)=1$. We expect that it will be optimal to stop at stage n having observed $X_1=x_1,\ldots,X_n=x_n$ if and only if $y_n(x_1,\ldots,x_n)=V_n^*(x_1,\ldots,x_n)$. This is known as the **principle of optimality**. It is also of central importance in the more general dynamic programming problems.

This principle is valid here under A1 and A2 but it requires a modification. The trouble is that in general there are more than a countable number of stopping rules, $N\geq n$, and the supremum of an uncountable collection of random variables (here the supremum of $\mathrm{E}(Y_N|\mathcal{F}_n)$ over the set of those stopping rules $N\geq n$) may not be a random variable (i.e. measurable) and even if it is, it may not be what we want it to be. Thus, (7) is not well defined. We can get around this difficulty by using instead the essential supremum.

Definition. *Let X_t, for $t\in T$, be a collection of random variables. We say that a random variable Z is* **an essential supremum of** $(X_t)_{t\in T}$ *and write $Z=\operatorname{ess\,sup}_{t\in T}X_t$, if*

i. $\mathrm{P}(Z\geq X_t)=1$ for all $t\in T$, and

ii. if Z' is any other random variable such that $\mathrm{P}(Z'\geq X_t)=1$ for all $t\in T$, then $\mathrm{P}(Z'\geq Z)=1$.

As an example of a collection of random variables X_t for which $\operatorname{ess\,sup}_{t\in T}X_t\neq\sup_{t\in T}X_t$, let $T=[0,1]$ and let $X_t=I(t=U)$ where U is a random variable with a uniform distribution on $[0,1]$. Then $\sup_{t\in T}X_t=1$, yet $\operatorname{ess\,sup}_{t\in T}X_t=0$.

Lemma 3. *An essential supremum, $Z=\operatorname{ess\,sup}_{t\in T}X_t$, always exists, and there exists a countable subset $C\subset T$ such that $Z=\sup_{t\in C}X_t$ is an essential supremum.*

Proof. By taking arctan of the X_t, $(t\in T)$, if necessary, we may assume without loss of generality that the X_t are uniformly bounded. Let $\mathcal{C}$ be the class of all countable subsets of T, and let

$$\alpha=\sup_{S\in\mathcal{C}}\mathrm{E}\sup_{t\in S}X_t. \tag{8}$$

For every $n=1,2,\ldots$ find sets $S_n\in\mathcal{C}$ such that $\mathrm{E}\sup_{t\in S_n}X_t\geq\alpha-1/n$. Let $C=\bigcup_{n=1}^{\infty}S_n$ and $Z=\sup_{t\in C}X_t$ so that C is countable and $\mathrm{E}Z=\alpha$. We now show that $Z=\operatorname{ess\,sup}_{t\in T}X_t$.

i. Suppose for some $t\in T$, $\mathrm{P}(Z<X_t)>0$. Then $\mathrm{E}\max(Z,X_t)>\mathrm{E}Z\geq\alpha$, contradicting (8). Hence $\mathrm{P}(Z\geq X_t)=1)$ for all $t\in T$.

ii. Suppose for some Z' that $\mathrm{P}(Z'\geq X_t)=1$ for all $t\in T$. Then since C is countable, $\mathrm{P}(Z'\geq\sup_{t\in C}X_t)=1$. Hence $\mathrm{P}(Z'\geq Z)=1$. ∎

Before returning to the principle of optimality, we first indicate that the analogues of Lemmas 1 and 2 are valid conditionally on having reached stage n. For this purpose, we extend the notion of a regular stopping rule. When we write $X \geq Y$ or $X > Y$ etc., where X and Y are random variables, we take it as implied that the inequalities hold almost surely. Similarly, when we write $X > Y$ on a set A, we mean $\mathrm{P}(\{X > Y\} \cap A) = \mathrm{P}(A)$.

Definition. *A stopping rule $N \geq n$ is* **regular from** n **on**, *if for every $k \geq n$, $\mathrm{E}\{Y_N|\mathcal{F}_k\} > Y_k$ on $\{N > k\}$.*

Lemma 1′. *Under A1, for any stopping rule $N \geq n$ there exists a stopping rule $N' \geq n$, regular from n on, such that $\mathrm{E}\{Y_{N'}|\mathcal{F}_n\} \geq \mathrm{E}\{Y_N|\mathcal{F}_n\}$.*

Lemma 2′. *Under A1, if $N \geq n$ and $N' \geq n$ are both regular from n on, then so is $N'' = \max(N, N')$ and $\mathrm{E}(Y_{N''}|\mathcal{F}_n) \geq \max(\mathrm{E}\{Y_N|\mathcal{F}_n\}, \mathrm{E}\{Y_{N'}|\mathcal{F}_n\})$.*

The proofs of these lemmas are straightforward adaptations of the proofs of Lemmas 1 and 2, and so are omitted.

The next theorem is the **optimality equation** of dynamic programming. Let

$$V_n^* = \operatorname{ess\,sup}_{N \geq n} \mathrm{E}\{Y_N|\mathcal{F}_n\}. \tag{9}$$

Theorem 2. *Under A1, $V_n^* = \max(Y_n, \mathrm{E}\{V_{n+1}^*|\mathcal{F}_n\})$.*

Proof. Let $N \geq n$ be an arbitrary stopping rule. On $\{N > n\}$, $\mathrm{E}\{Y_N|\mathcal{F}_{n+1}\} \leq V_{n+1}^*$, so that on $\{N > n\}$, $\mathrm{E}\{Y_N|\mathcal{F}_n\} = \mathrm{E}\{\mathrm{E}(Y_N|\mathcal{F}_{n+1})|\mathcal{F}_n\} \leq \mathrm{E}\{V_{n+1}^*|\mathcal{F}_n\}$. Hence $\mathrm{E}\{Y_N|\mathcal{F}_n\} = I\{N = n\}Y_n + I\{N > n\}\mathrm{E}\{Y_N|\mathcal{F}_n\} \leq \max(Y_n, \mathrm{E}\{V_{n+1}^*|\mathcal{F}_n\})$ for all $N \geq n$. Therefore, $V_n^* = \operatorname{ess\,sup}_{N \geq n} \mathrm{E}\{Y_N|\mathcal{F}_n\} \leq \max(Y_n, \mathrm{E}\{V_{n+1}^*|\mathcal{F}_n\})$.

To show the reverse inequality, first note that $Y_n \leq V_n^*$ trivially. Now by Lemma 3, there exists a sequence $N_1, N_2, \ldots$ of stopping rules with each $N_k \geq n+1$ such that $V_{n+1}^* = \sup_k \mathrm{E}\{Y_{N_k}|\mathcal{F}_{n+1}\}$. By Lemma 1′, there exists for each k a stopping rule $N_k' \geq n+1$ regular from $n+1$ on such that $\mathrm{E}\{Y_{N_k'}|\mathcal{F}_{n+1}\} \geq \mathrm{E}\{Y_{N_k}|\mathcal{F}_{n+1}\}$. Let $N_k'' = \max(N_1', \ldots, N_k')$. Then,

$$\begin{aligned} V_n^* &\geq \mathrm{E}\{Y_{N_k''}|\mathcal{F}_n\} = \mathrm{E}(\mathrm{E}\{Y_{N_k''}|\mathcal{F}_{n+1}\}|\mathcal{F}_n) \\ &\geq \mathrm{E}(\max_{1 \leq j \leq k} \mathrm{E}\{Y_{N_j'}|\mathcal{F}_{n+1}\}|\mathcal{F}_n) \quad \text{by Lemma 2}', \\ &\geq \mathrm{E}(\max_{1 \leq j \leq k} \mathrm{E}\{Y_{N_j}|\mathcal{F}_{n+1}\}|\mathcal{F}_n) \\ &\to \mathrm{E}(V_{n+1}^*|\mathcal{F}_n) \end{aligned} \tag{10}$$

by monotone convergence. ∎

It is interesting to note that the optimality equation holds under A1 alone, even if an optimal rule does not exist.

The stopping rule given by the principle of optimality is the rule

$$N^* = \min\{n \geq 0 : Y_n = V_n^*\}. \tag{11}$$

It may be dangerous to use the rule give by the principle of optimality. In Example 2 of the introduction of this chapter, $V_n^* = 1$ for all finite n, and $Y_n < 1$ for all finite n so N^* tells you to continue forever. This gives a payoff of 0.

Using A1 only, N^* may not be optimal, but yet no rule is made worse by stopping when N^* tells you to stop, as the following lemma shows.

Lemma 4. *Assume A1. Let N be any stopping rule and let $N' = \min(N, N^*)$. Then, $\mathrm{E}Y_{N'} \geq \mathrm{E}Y_N$.*

Proof. On $\{N^* = n < N\}$, $V_n^* \geq \mathrm{E}(Y_N|\mathcal{F}_n)$. Hence,

$$\begin{aligned}
\mathrm{E}I\{N^* < N\}Y_{N^*} &= \sum_0^\infty \mathrm{E}I\{N^* = n < N\}Y_n \\
&= \sum_0^\infty \mathrm{E}I\{N^* = n < N\}V_n^* \\
&\geq \sum_0^\infty \mathrm{E}(I\{N^* = n < N\}\mathrm{E}\{Y_N|\mathcal{F}_n\}) \\
&= \sum_0^\infty \mathrm{E}I\{N^* = n < N\}Y_N \\
&= \mathrm{E}I\{N^* < N\}Y_N.
\end{aligned} \tag{12}$$

Hence,

$$\begin{aligned}
\mathrm{E}Y_{N'} &= \mathrm{E}I\{N^* < N\}Y_{N*} + \mathrm{E}I\{N^* \geq N\}Y_N \\
&\geq \mathrm{E}I\{N^* < N\}Y_N + \mathrm{E}I\{N^* \geq N\}Y_N = \mathrm{E}Y_N,
\end{aligned} \tag{13}$$

completing the proof. ■

Under A1 and A2, N^* is optimal. In fact, the proof of the following theorem shows that out of all optimal rules it stops the soonest (for any optimal rule N^o, $N^* \leq N^o$). A characterization of all optimal rules may be found in the paper of M. Klass (1973).

Theorem 3. *Under A1, if there exists an optimal rule, in particular if A2 holds, then N^* is optimal.*

Proof. Let N_0 be an optimal rule and let $N = \min\{N_0, N^*\}$. Then from Lemma 4, N is also optimal and $N \leq N^*$. We will complete the proof by showing that $N = N^*$. Suppose $\mathrm{P}\{N < N^*\} > 0$. Then for some n, $\mathrm{P}\{N = n < N^*\} > 0$, and on $\{N = n < N^*\}$, $Y_n < V_n^*$, so that we should be able to improve N by changing N on $\{N = n < N^*\}$ to something that gives return close to V_n^*. As in the proof of Theorem 2, find N_k regular from n on such that $\sup_k \mathrm{E}\{Y_{N_k}|\mathcal{F}_n\} = V_n^*$ and let $N_k' = \max\{N_1,\ldots,N_k\}$ so that

$$V_n^* \geq \mathrm{E}(Y_{N_k'}|\mathcal{F}_n) \geq \max_{1\leq j\leq k} \mathrm{E}(Y_{N_j}|\mathcal{F}_n) \to V_n^* \tag{14}$$

monotonically. Then, since $\mathrm{E}I\{N = n < N^*\}Y_{N_k'} \to \mathrm{E}I\{N = n < N^*\}V_n^*$, there exists a k such that $\mathrm{E}I\{N = n < N^*\}Y_n < \mathrm{E}I\{N = n < N^*\}Y_{N_k'}$. Letting $N' = N_k'$ on $\{N = n < N^*\}$ and $N' = N$ otherwise, we have $\mathrm{E}Y_{N'} > \mathrm{E}Y_N$, contradicting the optimality of N. Consequently, $\mathrm{P}(N < N^*) = 0$ which implies that $N = N^*$. ■

Often useful in applications is an alternate form of the rule given by the principle of optimality, based on the random variables

$$W_n^* = \mathrm{E}(V_{n+1}^*|\mathcal{F}_n).$$

By the optimality equation, $V_n^* = \max(Y_n, W_n^*)$, and the rule, N^*, becomes

$$N^* = \min\{n \geq 0 : Y_n \geq W_n^*\}.$$

One can show that $W_n^* = \operatorname{ess\,sup}_{N>n}\mathrm{E}(Y_N|\mathcal{F}_n)$, so W_n^* can be considered as the best return available at stage n among rules that continue at least one stage. The rule N^* calls for stopping when the return for stopping is at least as great as the best that can be obtained by continuing. The rule $N^{**} = \min\{n \geq 0 : Y_n > W_n^*\}$ is the optimal rule that stops last (for any optimal rule N^o, $N^{**} \geq N^o$).

§**3.3 The Wald Equation.** The following equation, due to Wald, is very useful in solving optimal stopping problems.

Theorem 4. *Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables such that $\mathrm{E}|X_1| < \infty$, let $\mu = \mathrm{E}X_1$, and let $S_n = X_1 + \cdots + X_n$. Let N be any stopping rule, adapted to $X_1, X_2, \ldots$. Then, if $\mathrm{E}N < \infty$,*

$$\mathrm{E}(S_N) = \mu\mathrm{E}(N). \tag{15}$$

Moreover, if $\mathrm{E}N = \infty$ and $\mu \neq 0$, then (15) holds provided $\mathrm{E}S_N$ exists (that is, provided not both $\mathrm{E}S_N^+ = \infty$ and $\mathrm{E}S_N^- = \infty$ where $x^+ = \max(0,x)$ and $x^- = -\min(0,x)$).

We give two proofs. The first does not cover the "moreover" part of the statement, but is more flexible for providing simple extensions.

Proof #1. Provided the change of summation can be justified, we have

$$\begin{aligned}\mathrm{E}S_N &= \sum_{n=1}^{\infty} \mathrm{E}\{\mathrm{I}(N = n)S_n\} \\ &= \sum_{n=1}^{\infty}\sum_{j=1}^{n} \mathrm{E}\{\mathrm{I}(N = n)X_j\} \\ &= \sum_{j=1}^{\infty}\sum_{n=j}^{\infty} \mathrm{E}\{\mathrm{I}(N = n)X_j\} \\ &= \sum_{j=1}^{\infty} \mathrm{E}\{\mathrm{I}(N \geq j)X_j\}.\end{aligned} \tag{16}$$

Since N is a stopping rule for the sequence $X_1, X_2, \ldots$, the event $\{N \geq j\}$ depends only on $X_1, \ldots, X_{j-1}$ and so is independent of X_j. The equation continues

$$= \sum_{j=1}^{\infty} \mathrm{P}\{N \geq j\}\mathrm{E}(X_j) = \mu \sum_{j=1}^{\infty} \mathrm{P}\{N \geq j\} = \mu \mathrm{E}N.$$

The interchange of summations is justified provided the double summation in (16) converges absolutely. This follows by replacing X_j with $|X_j|$ in (16) and obtaining as above, $\mathrm{E}(|S_N|) \leq \mathrm{E}(|X_1|)\mathrm{E}(N) < \infty$.

Proof #2. (This proof is due to Blackwell (1946). The 'Moreover' part is due Robbins and Samuel (1966).) Consider n stopping problems as follows. Let N_1 be the stopping rule N applied to the sequence $X_1, X_2, \ldots$, let N_2 be the stopping rule N applied to $X_{N_1+1}, X_{N_1+2}, \ldots$, etc., and let N_n be the stopping rule N applied to $X_{N_1+\cdots+N_{n-1}+1}$, $X_{N_1+\cdots+N_{n-1}+2}$, ... Let the returns for these problems be denoted by $Z_1, \ldots, Z_n$ where $Z_j = X_{N_1+\cdots+N_{j-1}+1} + \cdots + X_{N_1+\cdots+N_j}$. Then, the Z_j are independent with the same distribution as S_N, and we have

$$\frac{Z_1 + \cdots + Z_n}{n} = \frac{X_1 + \cdots + X_{N_1+\cdots+N_n}}{N_1 + \cdots + N_n} \cdot \frac{N_1 + \cdots + N_n}{n}. \tag{17}$$

From the strong law of large numbers, $(X_1 + \cdots + X_{N_1+\cdots+N_n})/(N_1 + \cdots + N_n) \to \mu$ a.s. If $\mathrm{E}N < \infty$, then $(N_1 + \cdots + N_n)/n \to \mathrm{E}N$ a.s., so that $(Z_1 + \cdots + Z_n)/n \to \mu\mathrm{E}N$ a.s., and hence by the converse to the strong law, $\mathrm{E}S_N = \mu\mathrm{E}N$. Similarly, if $\mathrm{E}S_N < \infty$ and $\mu \neq 0$, then $(N_1 + \cdots + N_n)/n \to \mathrm{E}S_N/\mu = \mathrm{E}N$ a.s. ∎

REMARK 1. One interpretation of this result for $\mu = 0$ is that it shows that no stopping strategy with $\mathrm{E}N < \infty$ in a sequence of identical fair games can yield a positive expected payoff. If each game has expectation $\mu = 0$, then the expectation of the sum S_N is zero for every stopping rule N with $\mathrm{E}N < \infty$. If $\mathrm{E}N = \infty$ is allowed, then we can obtain $\mathrm{E}S_N > 0$ by defining N as the first n such that $S_n > 0$, provided the distribution of X_1 is not degenerate at zero. The law of the iterated logarithm shows that N is finite with probability one.

REMARK 2. If the X_i are independent with mean μ but not identically distributed, then the result $\mathrm{E}S_N = \mu\mathrm{E}N$ may not hold. (See Exercise 4.) The natural generalization of this theorem to the dependent case is for sequences S_n that form a martingale. For this theory, see for example the books of Chung (1968), or Chow, Robbins and Siegmund (1971).

REMARK 3. If $\mathrm{E}X_1^- < \infty$, the argument involving (16) shows that when $\mathrm{E}N < \infty$, $\mathrm{E}S_N = \mathrm{E}X_1\mathrm{E}N$ even if $\mathrm{E}X_1 = \infty$, provided (when $N \equiv 0$) 0 times ∞ is taken to be 0.

EXAMPLE 3. As an example of the use of this equation, suppose that the X_i are i.i.d. taking integer values less than or equal to 1 with probabilities $\mathrm{P}(X_i = j) = p_j$, where $\sum_{-\infty}^{1} p_j = 1$, and assume that $\mathrm{E}X_i > 0$. For some integer $r > 0$, let $N = \min\{n > 0 : S_n = r\}$. Then $N < \infty$ with probability one, since $S_n \to \infty$ a.s. by the law of large numbers. Then, since $S_N \equiv r$, we have $\mathrm{E}N = \mathrm{E}S_N/\mathrm{E}X_1 = r/\mathrm{E}X_1$. In the particular case where $p_1 = p$ and $p_0 = 1-p$, we have the well-known result that the expectation of the number of trials until the rth success in a sequence of i.i.d. Bernoulli trials is r/p (a negative binomial random variable).

§**3.4 Prophet Inequalities.** Under A1 and A2, the decision maker can attain an expected payoff of $V^* = \sup_N \mathrm{E}Y_N$, using stopping rules, i.e. using rules for which stopping at stage n depends only on the observations up to that time. On the other hand, a prophet, who can foresee the future and who knows the values of all the Y_n, would stop at some Y_n which is close to the largest overall, and thus achieve an expected payoff as close as desired to $M^* = \mathrm{E}\sup_n Y_n$. We certainly have $V^* \le M^*$.

Obviously, the prophet has a big advantage over the decision maker. Therefore it is surprising that in many situations there are universal upper bounds on the advantage of the prophet over the decision maker. Inequalities bounding M^* above by some function of V^* are called prophet inequalities. The most basic and surprising of these is the first general prophet inequality discovered. It is due to Krengel, Sucheston and Garling, and found in Krengel and Sucheston (1977, 1978).

This inequality deals with a sequence of random variables $X_1, X_2, \ldots$ that are *independent* and *nonnegative*, and in which the payoff, Y_n, for stopping at stage n is X_n itself. We require that the decision maker use rules that stop at some finite time (i.e. we require $\mathrm{P}(N < \infty) = 1$), so there is no need to define Y_∞. We use the notation,

$$V^* = \sup_{N<\infty} \mathrm{E}(X_N) \qquad \text{and} \qquad M^* = \mathrm{E}(\sup_n X_n). \tag{18}$$

The basic theorem of Krengel, Sucheston and Garling is the following: *For independent nonnegative random variables,* $X_1, X_2, \ldots,$

$$M^* \le 2V^*. \tag{19}$$

Thus, a prophet cannot win more on the average than twice that of a real-time decision maker, when dealing with a sequence of independent nonnegative random variables.

This inequality follows from the finite horizon version, where the decision maker is restricted to using rules that stop by some prespecified time n. In this case we have

$$V_n^* = \sup_{N\le n} \mathrm{E}(X_N) \qquad \text{and} \qquad M_n^* = \mathrm{E}(\max_{1\le j\le n} X_j). \tag{20}$$

and the inequality becomes

$$M_n^* \le 2V_n^*. \tag{21}$$

To see that (21) implies (19), simply note, since $\max_{1\le j\le n} X_j$ is a.s. nondecreasing to $\sup_n X_n$, that $M_n^* \to M^*$ as $n \to \infty$. Then $V^* \ge V_n^* \ge (1/2)M_n^* \to (1/2)M^*$ as $n \to \infty$.

Examples. One cannot replace the condition that the X_i be nonnegative with the condition that the expectations be nonnegative. Here is an example with $n = 2$. Let the $X_1 = 1$ and let X_2 be B with probability $1/2$ and $-B$ with probability $1/2$, for some large number B. The best one can do with stopping rules is to stop with the first observation and receive $V_2^* = 1$, since continuing gives an expected value of 0. But the prophet can get $M_2^* = \mathrm{E}\max\{1, X_2\} = (1/2) + (1/2)B = (B + 1)/2$.

A similar example shows that the inequalities (19) and (21) are sharp. Let $X_1 = 1$, $X_3 = X_4 = \cdots = 0$, and for some number $B > 1$ let $\mathrm{P}(X_2 = B) = 1/B$ and $\mathrm{P}(X_2 = 0) = 1 - (1/B)$. Then again the best the real-time decision maker can do is $V^* = 1$. Whereas the prophet can obtain $M^* = \mathrm{E}\max\{1, X_2\} = (1 - (1/B)) + (1/B)B = 2 - (1/B)$. This can be made as close to 2 as desired by making B large.

If in (21) we rule out the trivial cases where $M^* = 0$ and $M^* = \infty$, then the inequality may be taken to be strict.

The inequality (21) is interesting because of its generality. Of course, in particular situations the bound may not be very good. For example, in the Cayley-Moser problem of Chapter 2, when $X_1, \ldots, X_n$ are i.i.d. uniform, the distribution of $\max\{X_1, \ldots, X_n\}$ has the beta distribution function, $F(y) = y^n$ on $[0, 1]$. This distribution has mean $M_n^* = n/(n+1) = 1 - (1/(n+1))$. This is just slightly bigger than $V_n^* = A_n \simeq 1 - (2/(n + \log(n) + 1.768))$ when n is large. ∎

For the proof of the prophet inequality (21), we use a method due to Ester Samuel-Cahn (1984). This proof has the advantage of being constructive. It exhibits a stopping rule that achieves at least half of the profit of a prophet. Moreover, it shows that the advantage of the prophet does not increase if the decision maker is restricted to using pure threshold rules. (A threshold rule is pure if the cutoff point does not depend on n.)

Let $M_n = \max_{1\le j\le n} X_j$. We define a **pure threshold stopping rule** with threshold c to be a rule of the form

$$s(c) = \begin{cases} \min\{1 \le j \le n : X_j > c\} & \text{if } M_n > c \\ n & \text{if } M_n \le c \end{cases} \tag{22}$$

or

$$t(c) = \begin{cases} \min\{1 \le j \le n : X_j \ge c\} & \text{if } M_n \ge c \\ n & \text{if } M_n < c. \end{cases} \tag{23}$$

Let m denote a median of the distribution of M_n, i.e.

$$\mathrm{P}(M_n < m) = q \le 1/2 \qquad \text{and} \qquad \mathrm{P}(M_n > m) = p \le 1/2. \tag{24}$$

and let

$$\beta = \sum_1^n \mathrm{E}(X_i - m)^+. \tag{25}$$

Then one of the two stopping rules, $s(m)$ or $t(m)$, will achieve at least half of $\mathrm{E}M_n$. More precisely,

Theorem 5. *If* $m \le \beta$, *then* $\mathrm{E}(M_n) \le 2\mathrm{E}(X_{s(m)})$. *If* $m \ge \beta$, *then* $\mathrm{E}(M_n) \le 2\mathrm{E}(X_{t(m)})$.

Proof. First note that $(M_n - m)^+ = \max_{1\le i\le n}(X_i - m)^+ \le \sum_1^n (X_i - m)^+$ a.s., so that $\mathrm{E}(M_n - m)^+ \le \beta$. Hence,

$$\mathrm{E}(M_n) = m + \mathrm{E}(M_n - m) \le m + \mathrm{E}(M_n - m)^+ \le m + \beta. \tag{26}$$

Suppose first that $m \le \beta$. Then

$$\begin{aligned} \mathrm{E}(X_{s(m)} - m)^+ &= \mathrm{E}\sum_1^n (X_i - m)^+ \mathrm{I}(s(m) = i) \\ &= \sum_1^n \mathrm{E}(X_i - m)^+ \mathrm{I}(s(m) > i - 1) \\ &= \sum_1^n \mathrm{E}(X_i - m)^+ \mathrm{P}(s(m) > i - 1). \end{aligned} \tag{27}$$

The last equality uses the fact that the event $\{s(m) > i-1\}$ depends only on $X_1, \ldots, X_{i-1}$ and so is independent of X_i. Moreover,

$$\mathrm{P}(s(m) > i - 1) \ge \mathrm{P}(s(m) = n) \ge \mathrm{P}(M_n \le m) = 1 - p \tag{28}$$

for all i. Together (27) and (28) show that $\mathrm{E}(X_{s(m)} - m)^+ \ge \beta(1 - p)$. Finally,

$$\mathrm{E}(X_{s(m)}) = \mathrm{E}(X_{s(m)}\mathrm{I}(X_{s(m)} > m)) + \mathrm{E}(X_n\mathrm{I}(X_{s(m)} \le m) \ge \mathrm{E}(X_{s(m)}\mathrm{I}(X_{s(m)} > m)) \tag{29}$$

since $X_n \ge 0$. Continuing this inequality,

$$\begin{aligned} \mathrm{E}(X_{s(m)}\mathrm{I}(X_{s(m)} > m)) &= \mathrm{E}((X_{s(m)} - m)\mathrm{I}(X_{s(m)} > m)) + m\mathrm{P}(X_{s(m)} > m) \\ &= \mathrm{E}(X_{s(m)} - m)^+ + m\mathrm{P}(M_n > m) = \mathrm{E}(X_{s(m)} - m)^+ + mp \\ &\ge \beta(1 - p) + mp = \beta - (\beta - m)p \\ &\ge \beta - (\beta - m)/2 = (m + \beta)/2 \\ &\ge \mathrm{E}(M_n)/2 \end{aligned} \tag{30}$$

using (26) and our assumption that $m \le \beta$. This proves the first statement of the theorem. The second statement is proved in a completely analogous fashion. ∎

It is interesting to note that this proof shows that for inequality (21) the assumption that all the X_i be nonnegative may be replaced by the assumption that just X_n be nonnegative. For the corresponding improvement in the conditions for inequality (19), see Exercise 6.

There are many other prophet inequalities. An important class of prophet inequalities concerns a sequence of independent uniformly bounded random variables. The basic theorem in this class is due to Hill and Kertz (1981): *For independent random variables,* $X_1, X_2, \ldots$ *such that* $0 \le X_i \le 1$ *for all* i,

$$M^* \le V^* + \frac{1}{4} \tag{31}$$

or better, as proved in Hill (1983)

$$M^* \le 2V^* - (V^*)^2. \tag{32}$$

The inequality here is sharp and may be attained. (See Exercise 7.)

Other classes of prophet inequalities include those where the variables are allowed to be dependent, or the variables are restricted to being i.i.d., or where the decision maker is given the freedom to choose the order in which the X_i are observed. For a review of these and other prophet inequalities, see the survey paper of Hill and Kertz (1992).

§3.5 Exercises.

1. **The one-stage look-ahead rule.** Let N_1 denote the one-stage look-ahead rule, sometimes called **the myopic rule**,

$$N_1 = \min\{n \ge 0 : Y_n \ge \mathrm{E}(Y_{n+1}|\mathcal{F}_n)\},$$

and let $N_1^{(J)}$ denote this rule truncated at J, $N_1^{(J)} = \min\{N_1, J\}$.
(a) Show that $N_1^{(J)}$ is regular if A1 is satisfied.
(b) Show that N_1 is regular if A1 and A2 are satisfied.
(c) Show by counterexample that if A1 or A2 is not satisfied, N_1 may not be regular.

2. **The hypermetropic rule.** Let N_∞ denote the hypermetropic rule,

$$N_\infty = \min\{n \ge 0 : Y_n \ge \mathrm{E}(Y_\infty|\mathcal{F}_n)\}.$$

Show that under condition A1, N_∞ is regular.

3. **The one-time look-ahead rule.** Let T_1 denote the one-time look-ahead rule,

$$T_1 = \min\{n \ge 0 : Y_n \ge \sup_{j>n, j\le\infty} \mathrm{E}(Y_j|\mathcal{F}_n)\}.$$

(a) Show that T_1 is regular under conditions A1 and A2.
(b) Assuming A1 and A2, conclude that the one-time look-ahead rule is at least as good as the one-stage look-ahead rule, and the hypermetropic rule.

4. **Wald's equation without the assumption of identical distributions.**
(a) Find an example of independent $X_1, X_2, \ldots$ such that $\mathrm{E}X_j = 0$ for all j, and a stopping rule N adapted to $X_1, X_2, \ldots$ with $\mathrm{E}N < \infty$ such that $\mathrm{E}(X_1 + \cdots + X_N) > 0$.
(b) Let $X_1, X_2, \ldots$ be independent with finite means $\mu_j = \mathrm{E}X_j$ for $j = 1, 2, \ldots$. Let N

be any stopping rule such that $\mathrm{E}N < \infty$. Find the extra conditions on the distributions of the X_j that are needed so that equation (16) goes through to show

$$\mathrm{E}(S_N) = \mathrm{E}\sum_{j=1}^{N} \mu_j. \tag{33}$$

5. Give an example of independent nonnegative random variables, $X_1, \ldots, X_n$, for which $\mathrm{E}(M_n) > 2\mathrm{E}(X_{s(m)})$, and another example for which $\mathrm{E}(M_n) > 2\mathrm{E}(X_{t(m)})$. (For this, it suffices to take $n = 2$.)

6. Prove the following extension of basic prophet inequality: If $X_1, X_2, \ldots$, are independent, and if $\sum_{i=1}^{\infty} \mathrm{P}(X_i \geq 0) = \infty$, then (19) holds.

7. For each value of V^* in (0,1), give an example of independent random variables $X_1, X_2, \ldots$ with $0 \leq X_i \leq 1$ for all i such that $M^* = 2V^* + (V^*)^2$.

Chapter 4. APPLICATIONS.

MARKOV MODELS.

In this chapter, we look at some stopping rule problems for which the principle of optimality provides an effective method of obtaining the solution. Each of these problems has a structure that reduces the problem to a Markov decision problem with a one-dimensional state space. This allows us to show that the stopping rule given by the principle of optimality has a simple form. In fact, when the returns are functions of a Markov chain, then we may restrict attention to stopping rules that are functions of the chain.

A typical form of these problems may be described as follows. Let $\{Z_n\}_{n=1}^{\infty}$ be a sequence of random variables such that Z_n is $\mathcal{F}_n$-measurable, where $\mathcal{F}_n$ denotes the σ-field generated by the observations $X_1,\ldots,X_n$. We assume that $\{Z_n\}$ is a Markov chain in the sense that the distribution of Z_{n+1} given $\mathcal{F}_n$ is the same as the distribution of Z_{n+1} given Z_n, and we suppose that the payoffs, Y_n, are functions of the chain, say $Y_n = u_n(Z_n)$ for some functions u_n. Then $V_n = \operatorname{ess\,sup}_{N\geq n}\mathrm{E}(Y_N|\mathcal{F}_n)$ is a function of Z_n, say $V_n(Z_n)$, and the rule given by the principle of optimality has the form

$$N^* = \min\{n\geq 0 : u_n(Z_n)\geq V_n(Z_n)\}. \tag{1}$$

Thus, stopping at time n can be taken to depend only on Z_n. In the first five problems treated in this chapter, there is a time invariance that reduces this rule to a simpler form, $N^* = \min\{n\geq 0 : Z_n\geq c\}$ where c is independent of n. For the general theory of Markov stopping rule problems, see Siegmund (1967), or Chow, Robbins and Siegmund (1972), or Shiryaev (1973).

In the first section, we solve the problem of selling an asset with and without recall. This is the house-selling problem described in Example 1 of Section 1.2. It is an extension to an infinite horizon of the Cayley-Moser problem treated in Section 2.4. In the second section, we look at the problem of stopping a discounted sum. This is a version of the burglar problem of Exercise 1 of Chapter 1. In the third section, we solve the problem due to Darling, Liggett and Taylor (1976) of stopping a sum with a negative drift. In the fourth section, we treat an extension of the problem of stopping a success run mentioned in Exercise 5 of Chapter 1. The fifth section is devoted to the problem of testing simple statistical hypotheses. This is a special case of the general Bayesian statistical problem described in Example 3 of Section 1.2. In the final section, we discuss the problem of maximizing the average, presented as Example 2 of Section 1.2. It is an example of a

Markov stopping rule problem without a time invariance, and so its solution is of a higher order of difficulty.

§4.1. Selling an asset with and without recall. The house-selling problem or problem of selling an asset, as described in Chapter 1, is defined by the sequence of observations and payoffs as follows. The observations are

$$X_1, X_2, \ldots \qquad \text{assumed to be i.i.d. with known distribution } F(x).$$

The reward sequence depends on whether or not recall of past observations is allowed. If recall is not allowed, then

$$Y_0 = -\infty, Y_1 = X_1 - c, \ldots, Y_n = X_n - nc, \ldots, Y_\infty = -\infty,$$

while if recall is allowed

$$Y_0 = -\infty, Y_1 = M_1 - c, \ldots, Y_n = M_n - nc, \ldots, Y_\infty = -\infty,$$

where $c > 0$ is the cost per observation, and $M_n = \max\{X_1, \ldots, X_n\}$. We arbitrarily put $Y_0 = -\infty$ to force you to take at least one observation. Putting $Y_\infty = -\infty$ is natural as the cost of an infinite number of observations is infinite.

The following theorem states that A1 and A2 are satisfied in problems of selling an asset with and without recall provided $X_1, X_2, \ldots$ are identically distributed with finite second moment. It is not assumed that the X_j are independent. This allows for the case where $X_1, X_2, \ldots$ are independent identically distributed given some parameter θ where θ is random with a known prior distribution. When θ is integrated out, the marginal process $X_1, X_2, \ldots$ is exchangeable (every permutation of $X_1, X_2, \ldots$ has the same distribution). The hypothesis of identical marginal distributions of the X_j, used in this theorem, is weaker than that of exchangeability. Proofs of this theorem and of a converse are contained in the appendix to this chapter.

Theorem 1. *Let $X, X_1, X_2, \ldots$ be identically distributed, let $c > 0$, and let $Y_n = X_n - nc$ or $Y_n = \max\{X_1, \ldots, X_n\} - nc$.*
If $\mathrm{E}X^+ < \infty$, then $\sup Y_n < \infty$ a.s. and $Y_n \to -\infty$ a.s.
If $\mathrm{E}(X^+)^2 < \infty$, then $\mathrm{E}\sup Y_n < \infty$.

We first consider the problem of selling an asset without recall: $Y_n = X_n - nc$. Suppose that $X_1, X_2, \ldots$ are independent identically distributed with finite second moment. Since A1 and A2 are satisfied, we know from Theorems 3.1 and 3.3 that an optimal stopping rule exists and is given by the principle of optimality.

Let V^* denote the expected return from an optimal stopping rule. Suppose you pay c and observe $X_1 = x_1$. Note that if you continue from this point then the x_1 is lost and the cost c has already been paid, so it is just like starting the problem over again; that is, *the problem is invariant in time.* So if you continue from this point, you can obtain an expected return of V^* but no more. Therefore, the principle of optimality says that if $x_1 < V^*$ you should continue, and if $x_1 > V^*$ you should stop. For $x_1 = V^*$ it is immaterial what you do, but let us say you stop. This argument can be made at any stage, so the rule given by the principle of optimality is

$$N^* = \min\{n \geq 1 : X_n \geq V^*\}. \tag{2}$$

The problem now is to compute V^*. This may be done through the optimality equation,

$$V^* = \mathrm{E}\max\{X_1, V^*\} - c = \int_{-\infty}^{V^*} V^*\,dF(x) + \int_{V^*}^{\infty} x\,dF(x) - c,$$

where F is the common distribution function of the X_i. Rearranging terms, we find

$$\int_{V^*}^{\infty} (x - V^*)\,dF(x) = c \qquad \text{or} \qquad \mathrm{E}(X - V^*)^+ = c. \tag{3}$$

The left side is continuous in V^* and decreasing from $+\infty$ to zero. Hence, there is a unique solution for V^* for any $c > 0$.

As an example, suppose F is $\mathcal{U}(0,1)$, the uniform distribution on the interval $(0,1)$. For $0 \leq v \leq 1$,

$$\int_v^1 (x - v)\,dF(x) = (1 - v)^2/2,$$

while for $v < 0$,

$$\int_0^1 (x - v)\,dF(x) = 1/2 - v.$$

Equating to c, we find

$$\begin{aligned} V^* &= 1 - (2c)^{1/2} \qquad &&\text{if} \quad c \leq 1/2 \\ V^* &= -c + 1/2 \qquad &&\text{if} \quad c > 1/2. \end{aligned}$$

The optimal rule N^* calls for accepting the first offer greater than or equal to V^*.

Now consider the problem with recall, $Y_0 = Y_\infty = -\infty$, and $Y_n = M_n - nc$ where $M_n = \max\{X_1, \ldots, X_n\}$. Again the rule given by the principle of optimality is optimal. Suppose at some stage you have observed $M_n = m$ and it is optimal to continue. Then at the next stage, if M_{n+1} is still m (because $X_{n+1} \leq m$), it is still optimal to continue due to the invariance of the problem in time. Thus the principle of optimality will never require you to recall an observation from an earlier stage. The best we can do among such rules is found above for the problem without recall. Thus, the same rule is optimal for both problems.

The Cayley-Moser problem has been extended to infinite horizon problems in a number of ways. Above, a cost is assessed for each observation. In another method, due to Karlin (1961), the future is discounted so that a positive return at the first stage is worth more than the same return at later stages. These problems are treated in Exercises 2 and 3. Another possibility is to have forced stopping at a random future time. In job processing, this forced stopping time is modeled as a deadline. See Exercise 4 for an example.

Is A1 needed? An interesting feature of this problem was pointed out in a paper of Robbins (1970). By an elegant direct argument based on Wald's equation and using only the assumption that $\mathrm{E}X^+ < \infty$, Robbins shows that the rule N^* given by (2) with V^* given by (3) is optimal within the class of stopping rules, N, such that $\mathrm{E}Y_N^- > -\infty$. This provides an extension of the optimal property of the rule N^* that is valid even if $\mathrm{E}(X^+)^2 = \infty$. Since, as is shown in the appendix, $\mathrm{E}(X^+)^2 < \infty$ is a necessary and sufficient condition for A1 to hold in this problem, this raises the possibility of similarly extending the general theory of Chapter 3.

However, there are difficulties of interpretation that arise because of the restriction to stopping rules that satisfy $\mathrm{E}Y_N^- > -\infty$. If $\mathrm{E}X^+ < \infty$ and $\mathrm{E}(X^+)^2 = \infty$, there will exist rules, N, such that $\mathrm{E}Y_N^- = -\infty$ and $\mathrm{E}Y_N^+ = +\infty$ (see the appendix). Restricting attention to rules such that $\mathrm{E}Y_N^- > -\infty$ seems to say that any rule, N, with $\mathrm{E}|Y_N| < \infty$, no matter how bad, is better than a rule whose expected payoff does not exist because $\mathrm{E}Y_N^- = -\infty$ and $\mathrm{E}Y_N^+ = +\infty$. Do you prefer a payoff of $-\$100$ (or $+\$100$) or a gamble giving you $\$X$, where X is chosen from a standard Cauchy distribution? Rather than attempting to answer such questions, we prefer the approach assuming condition A1.

§**4.2. Application to stopping a discounted sum.** (Dubins and Teicher (1967)) Let $X_1, X_2, \ldots$ be independent identically distributed random variables with $\mathrm{E}X_i^+ < \infty$, and let $S_n = \sum_1^n X_i$ $(S_0 = 0)$. The random variable X_n represents your return at stage n. Returns accumulate and are paid to you in a lump sum when you stop, but the future is discounted by a factor $0 < \beta < 1$, so your return for stopping at n is

$$Y_n = \beta^n S_n \quad \text{for} \quad n = 0, 1, \ldots .$$

If you never stop, you return is zero, $Y_\infty = 0$, so you will never stop at n with $S_n \le 0$. The problem is to decide how large to let S_n get before you stop.

We show that assumptions A1 and A2 are satisfied for this problem. From the strong law of large numbers $(1/n)\sum_1^n X_j^+ \to \mathrm{E}(X^+)$ a.s., so that

$$Y_n = n\beta^n (S_n/n) \le n\beta^n (1/n) \sum_1^n X_j^+ \simeq n\beta^n \mathrm{E}(X^+) \to 0 \quad \text{a.s.,}$$

and A2 is satisfied. To check A1, note that

$$\begin{aligned} \sup_n Y_n = \sup_n \beta^n S_n &\le \sup_n \beta^n \sum_1^n X_i^+ \\ &\le \sup_n \sum_1^n \beta^i X_i^+ \le \sum_1^\infty \beta^i X_i^+ , \end{aligned}$$

so that

$$\mathrm{E}\sup_n Y_n \leq \sum_1^\infty \beta^i \mathrm{E}X_i^+ = \mathrm{E}X_1^+ \beta/(1-\beta) < \infty.$$

Thus, an optimal stopping rule exists and is given by the principle of optimality.

Suppose $S_n = s$ and it is optimal to stop. Then the present return of $\beta^n s$ is at least as large as any expected future return $\mathrm{E}\beta^{n+N}(s+S_N)$. That is to say $s(1-\mathrm{E}\beta^N) \geq \mathrm{E}\beta^N S_N$ for all stopping rules N. The same must be true for all $s' \geq s$ so that the optimal rule N^* must be of the form for some number s_0,

$$N^* = \min\{n \geq 0 : S_n \geq s_0\}. \tag{4}$$

That is, stop at the first n for which $S_n \geq s_0$. To find s_0, note that if $S_n = s_0$, then we must be indifferent between stopping and continuing. The payoff for stopping, namely s_0, must be the same as the payoff for continuing using the rule that stops the first time the sum of the future observations is positive. That is, s_0 must satisfy the equation $s_0 = \mathrm{E}\beta^T(s_0 + S_T)$ or

$$s_0 = \mathrm{E}\beta^T S_T/(1-\mathrm{E}\beta^T), \tag{5}$$

where $T = \min\{n \geq 0 : S_n > 0\}$ is the rule: stop at the first n, if any, for which the sum of the next n observations is positive.

When X is positive with probability 1, then $T \equiv 1$ and $S_T \equiv X_1$, so that s_0 has the simple form $s_0 = \beta\mathrm{E}X/(1-\beta)$. This gives the optimal rule N^* of (4) in explicit form. The burglar problem of Exercise 1.1 may be put in this form where β represents the probability of not getting caught on any burglary. We return to this problem in Chapters 5 and 6.

When X may take in negative values with positive probability, the right side of (5) may difficult to evaluate in general, but in special cases the computational problems can be reduced. Dubins and Teicher call the distribution of the X_i **elementary** if X_i takes on only integer values less than or equal to one. In this special case, S_T in equation (5) becomes $+1$ for $T < \infty$ and is immaterial for $T = \infty$ since β^T is zero; hence,

$$s_0 = \mathrm{E}\beta^T/(1-\mathrm{E}\beta^T). \tag{6}$$

To find s_0, it suffices to find $\mathrm{E}\beta^T$. This may be computed from knowledge of the generating function of X, $G(\theta) = \mathrm{E}\theta^{-X}$ as follows. The distribution of T given $X_1 = x$ is the same as the distribution of $1 + T_1 + \ldots + T_{1-x}$, where $T_1, T_2, \ldots$ are independent with the same distribution as T. Hence, letting $\phi(\beta) = \mathrm{E}\beta^T$,

$$\begin{aligned}\phi(\beta) &= \mathrm{E}(\mathrm{E}\{\beta^T|X\}) = \mathrm{E}(\mathrm{E}\{\beta^{1+T_1+\ldots+T_{1-X}}|X\}) \\ &= \beta\mathrm{E}\phi(\beta)^{1-X} = \beta\phi(\beta)G(\phi(\beta)).\end{aligned} \tag{7}$$

Thus, $\phi(\beta)$ can be found by solving the equation $G(\phi(\beta)) = 1/\beta$. (See Exercise 5.)

When $\mathrm{E}X < 0$, then $S_n \to -\infty$ a.s. and the problem is still interesting even if $\beta = 1$, provided we keep $Y_\infty = 0$. However, we need a stronger assumption on the distribution of the X_i in order that A1 still hold. This problem is treated next.

§**4.3. Stopping a sum with negative drift.** (Darling, Liggett and Taylor (1972)) Let $X_1, X_2, \ldots$ be i.i.d. with negative mean $\mu = \mathrm{E}X < 0$, and let $S_n = X_1 + \ldots + X_n$, ($S_0 = 0$). The payoff for stopping at n is taken to be

$$Y_n = (S_n)^+ \qquad \text{for } n = 0, 1, \ldots, \text{ and } \quad Y_\infty = 0.$$

The law of large numbers implies that $S_n \to -\infty$ a.s. Hence, A2 is satisfied. We expect that an optimal rule may sometimes continue forever. In fact, it is clear that one need not stop at any n for which $S_n \le 0$.

This model has application to the problem of exercising options in stock market transactions (the American version without a deadline). The owner of an option has the privilege of purchasing a fixed quantity of a stock at a fixed price, here normalized to zero, and then reselling the stock at the market price which fluctuates as a classical random walk. Continuing forever may be interpreted as not exercising the option. If a cost of waiting is taken into account, the walk can easily have negative drift. Darling, Liggett and Taylor also treat the more realistic problem in which the logarithm of the prices form a random walk and there is a discount $0 < \beta \le 1$, so that the return function is $Y_n = \beta^n(e^{S_n} - 1)^+$, assuming $\mathrm{E}e^X < \beta^{-1}$.

We assume that $\mathrm{E}(X^+)^2 < \infty$. Then the following theorem of Kiefer and Wolfowitz (1956, Theorem 5), shows that A1 is satisfied and there exists an optimal rule.

Theorem 2. *Let $X, X_1, X_2, \ldots$ be i.i.d. with finite mean $\mu < 0$ and let $M = \sup_{n \ge 0} S_n$. Then,*

$$\mathrm{E}M < \infty \quad \textit{if, and only if,} \quad \mathrm{E}(X^+)^2 < \infty.$$

A proof is given in the appendix.

We show below that the optimal stopping rule is

$$N^* = \min\{n \ge 0 : S_n \ge \mathrm{E}M\}, \tag{8}$$

and its expected return is $\mathrm{E}(M - \mathrm{E}M)^+$.

Note that V_n^* depends on past observations only through S_n. In fact, $V_n^* = V_0^*(S_n)$, where $V_0^*(s) = \sup_{N \ge 0} \mathrm{E}(s + S_N)^+$ is the optimal return starting with initial fortune $S_0 = s$. Suppose $S_n = s > 0$ and it is optimal to stop. Then by the principle of optimality, the present return s is as least as large as future expected return $\mathrm{E}(s + S_N)^+$. That is to say, $\mathrm{E}\{\max(S_N, -s)\} \le 0$ for all stopping rules N. Since this expectation is nonincreasing in s and independent of n, it must also be optimal to stop when $S_m = s'$ for any m and any $s' \ge s$. Thus, N^* must be of the form $N^* = \min\{n \ge 0 : S_n \ge s_0\}$ for some s_0.

Let $N = N(s) = \min\{n \geq 0 : S_n \geq s\}$. We find the optimal value of s by computing the expected return $\mathrm{E}Y_N$. First note that the distribution of $M - S_N$ given S_N, where $S_N \geq s$, is the same as the distribution of M, because $M - S_N = \sup_n S'_n$, where $S'_n = X_{N+1} + \ldots + X_{N+n}$. Hence,

$$\begin{aligned}\mathrm{E}Y_N &= \mathrm{E}\{S_N \mathrm{I}(S_N \geq s)\} = \mathrm{E}\{M\,\mathrm{I}(S_N \geq s)\} - \mathrm{E}\{(M - S_N)\mathrm{I}(S_N \geq s)\} \\ &= \mathrm{E}\{M\,\mathrm{I}(S_N \geq s)\} - \mathrm{E}M \cdot \mathrm{P}\{S_N \geq s\} = \mathrm{E}\{(M - \mathrm{E}M)\mathrm{I}(S_N \geq s)\} \\ &= \mathrm{E}\{(M - \mathrm{E}M)\mathrm{I}(M \geq s)\}.\end{aligned} \tag{9}$$

This is clearly nondecreasing in s if $s \leq \mathrm{E}M$ and nonincreasing in s if $s \geq \mathrm{E}M$. The optimal value of s is therefore $s = \mathrm{E}M$ and the optimal expected return is $\mathrm{E}(M - \mathrm{E}M)^+$.

§**4.4. Rebounding from failures.** (Ferguson (1976)) Let $Z_1, Z_2, \ldots$ be i.i.d. random variables, and let $\epsilon_1, \epsilon_2, \ldots$ be i.i.d. Bernoulli random variables, independent of $Z_1, Z_2, \ldots$, with probability p of success, $\mathrm{P}(\epsilon = 1) = p = 1 - \mathrm{P}(\epsilon = 0)$, $0 < p < 1$. In a given economic system, Z_i represents your return and ϵ_i represents the indicator function of success in time period i. As long as successes occur consecutively, your returns accumulate, but when there is a failure, your accumulated return drops to zero. Failure represents a failure of your enterprise due to mismanagement, or a general calamity on the stock market, or a revolution in the country, etc. A failure does not remove you from the system as it does the burglar. You are allowed to accumulate future returns until the next failure drops you to zero again, and so on. If $X_0 = x$ denotes your initial fortune, then your fortune at the end of the nth time period is X_n, where

$$X_n = \epsilon_n(X_{n-1} + Z_n), \quad \text{for} \quad n = 1, 2, \ldots. \tag{10}$$

You want to retire to a safer, more comfortable environment while you are still young enough to enjoy it. Thus, if $c > 0$ represents the cost of the passing of time, your return if you stop at stage n is

$$Y_n = X_n - nc \tag{11}$$

for finite n. We take $Y_\infty = -\infty$ and note that A2 is satisfied. The problem is to choose a stopping rule N to maximize your expected net return, $\mathrm{E}Y_N$.

In the special case in which the Z_i are identically one, this is the problem of stopping a success run described in Exercise 5.

The following theorem states that A1 is satisfied if the distribution of the Z's has finite second moment. A proof is given in the appendix.

Theorem 3. $\mathrm{E}\sup_n(X_n - nc) < \infty$ *if, and only if,* $\mathrm{E}(Z^+)^2 < \infty$.

We note that the invariance of the problem in time implies that V_n^* depends on $X_1, \ldots, X_n$ only through the value of X_n, and that if $V_0^*(x)$ represents the optimal return with initial fortune x, then $V_n^*(X_n) = V_0^*(X_n) - nc$.

Suppose it is optimal to stop with $X_n = x$; then $x - nc \geq V_0^*(x) - nc$, or equivalently $x \geq \mathrm{E}Y_N(x)$ for all stopping rules N, where $Y_n(x)$ represents the payoff as a function of the initial fortune, x. For any $x' > x$, the difference $Y_n(x') - Y_n(x)$ is 0 or $x' - x$ according to whether the first failure has or has not occurred by time n. Thus, $\mathrm{E}Y_N(x') - \mathrm{E}Y_N(x) = (x' - x)\mathrm{P}(N < K) \leq x' - x$, where K is the time of the first failure. This implies that for all stopping rules N,

$$x' \geq \mathrm{E}Y_N(x') - \mathrm{E}Y_N(x) + x \geq \mathrm{E}Y_N(x'),$$

so that it is optimal to stop with $X_m = x'$ for any m and any $x' \geq x$. Hence, the optimal rule N^* given by the principle of optimality has the form

$$N(s) = \min\{n \geq 0 : X_n \geq s\} \tag{12}$$

for some s.

As an illustration, we take the distribution of the Z's to be exponential with density $f(z) = (1/\mu)\exp\{-z/\mu\}\mathrm{I}(z > 0)$. To find the optimal value of s, let us compute the expected payoff for $N = N(s)$, for $s > 0$ and for $x = 0$, namely,

$$\mathrm{E}Y_N = \mathrm{E}(X_N - Nc) = \mathrm{E}X_N - c\mathrm{E}N.$$

The lack of memory property of the exponential distribution implies that $X_N - s$ has the same distribution as Z. Hence, $\mathrm{E}X_N = s + \mu$. To find $\mathrm{E}N$, let

$$N' = N'(s) = \min\{n \geq 0 : S_n \geq s\}$$

where $S_n = Z_1 + \ldots + Z_n$. The sequence of points, $S_1, S_2, \ldots$, forms a Poisson point process with points occurring at rate $1/\mu$, so that $N' - 1$, the number of points in $(0, s)$, has a Poisson distribution with mean $\lambda = s/\mu$. Let K represent the time of the first failure, so that K has the geometric distribution, $\mathrm{P}(K = k) = (1 - p)p^{k-1}$ for $k = 1, 2, \ldots$. Then,

$$\begin{aligned} \mathrm{E}N &= \mathrm{E}N\mathrm{I}(N' < K) + \mathrm{E}N\mathrm{I}(N' \geq K) \\ &= \mathrm{E}N'\mathrm{I}(N' < K) + \mathrm{E}(K + \mathrm{E}N)\mathrm{I}(N' \geq K) \\ &= \mathrm{E}\min\{N', K\} + (\mathrm{E}N)\mathrm{P}(N' \geq K) \end{aligned}$$

so that $\mathrm{E}N = \mathrm{E}\min\{N', K\}/\mathrm{P}(N' < K)$. This may be computed from

$$\mathrm{P}(N' < K) = \mathrm{E}\{\mathrm{P}(N' < K|N')\} = \mathrm{E}p^{N'} = p\exp\{-\lambda(1 - p)\},$$

and

$$\begin{aligned} \mathrm{E}\min\{n, K - 1\} &= \sum_{k=1}^{n} k(1 - p)p^k + n\sum_{k=n+1}^{\infty}(1 - p)p^k \\ &= \sum_{k=1}^{n} kp^k - \sum_{k=1}^{n} kp^{k+1} + np^{n+1} \\ &= \sum_{k=1}^{n} p^k \\ &= p(1 - p^n)/(1 - p), \end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}\min\{N', K\} &= \mathrm{E}\min\{N'-1, K-1\} + 1 \\
&= (p/(1-p))\mathrm{E}(1-p^{N'-1}) + 1 \\
&= (p/(1-p))(1-e^{-\lambda(1-p)}) + 1 \\
&= (1/(1-p))[1-pe^{-\lambda(1-p)})].
\end{aligned}$$

We search for the value of s that maximizes

$$\mathrm{E}Y_N = s + \mu - (c/(p(1-p)))[e^{s(1-p)/\mu} - p].$$

Taking a derivative with respect to s, setting the result to zero and solving, gives the optimal value of s as

$$s = (\mu/(1-p))\log(p\mu/c).$$

provided $p\mu \ge c$. If $p\mu < c$, then $\mathrm{E}Y_N$ is a decreasing function of $s \ge 0$, and so the optimal value of s is $s = 0$, that is, stop without taking any observations.

§**4.5 Testing Simple Statistical Hypotheses.** The Bayes approach to general sequential statistical problems was discussed in Example 3 of Chapter 1. Here we specialize to the problem of testing simple statistical hypotheses. In this problem, there are two hypotheses, H_0 and H_1, and one distribution corresponding to each hypothesis. In the notation of Example 1.3, the parameter space is a two-point set, $\Theta = \{H_0, H_1\}$, and the observations, $X_1, X_2, \ldots$, are assumed to be i.i.d. according to a density $f_0(x)$ if H_0 is true, and density $f_1(x)$ if H_1 is true, where $f_0(x)$ and $f_1(x)$ are distinct as distributions. We must decide which hypothesis to accept. Thus, the action space is also a two-point set, $\mathcal{A} = \{a_0, a_1\}$, where a_0 (resp. a_1) represents the action "accept H_0" (resp. "accept H_1"). We lose nothing if we accept the true hypothesis , but if we accept the wrong hypothesis we lose an amount depending on which hypothesis is true; thus,

$$L(H_i, a_j) = \begin{cases} 0 & \text{if} \quad i = j, \\ L_i & \text{if} \quad i \ne j, \end{cases}$$

where L_0 and L_1 are given positive numbers.

We are given the prior probability, τ_0, that H_1 is the true hypothesis. After observing $X_1, \ldots, X_n$, the posterior probability that H_1 is the true hypothesis becomes, according to Bayes rule,

$$\tau_n(X_1, \ldots, X_n) = \frac{\tau_0 \prod_1^n f_1(X_i)}{\tau_0 \prod_1^n f_1(X_i) + (1-\tau_0)\prod_1^n f_0(X_i)}. \tag{13}$$

The likelihood ratio is $\lambda(x) = f_1(x)/f_0(x)$ with the understanding that $\lambda(x) = 0$ if $f_1(x) = 0$ and $f_0(x) > 0$, $\lambda(x) = \infty$ if $f_1 > 0$ and $f_0(x) = 0$, and $\lambda(x)$ is undefined if $f_1(x) = 0$ and $f_0(x) = 0$. Using this, we may rewrite τ_n as

$$\tau_n(X_1, \ldots, X_n) = \frac{\tau_0 \prod_1^n \lambda(X_i)}{\tau_0 \prod_1^n \lambda(X_i) + (1-\tau_0)} = \frac{\tau_0 \lambda_n}{\tau_0 \lambda_n + (1-\tau_0)} \tag{14}$$

where $\lambda_n = \lambda_n(X_1, \ldots, X_n)$ denotes the likelihood ratio, or probability ratio of the first n observations, $\lambda_n = \prod_{i=1}^n f_1(X_i)/f_0(X_i)$ and $\lambda_0 \equiv 1$.

Suppose it is decided to stop and the probability of H_1 is τ. If H_1 is accepted, the expected loss is $(1-\tau)L_0$, while if H_0 is accepted, the expected loss is τL_1. Thus, it is optimal to accept H_1 if $(1-\tau)L_0 < \tau L_1$ and to accept H_0 otherwise, incurring an expected loss of

$$\rho(\tau) = \min\{\tau L_1, (1-\tau)L_0\} \tag{15}$$

Therefore, if we stop at stage n having observed $X_1, \ldots, X_n$, we would accept H_1 if $\tau_n(X_1, \ldots, X_n)L_1 < (1 - \tau_n(X_1, \ldots, X_n))L_0$, and the expected terminal loss would then be $\rho(\tau_n(X_1, \ldots, X_n))$.

There is a cost of $c > 0$ for each observation taken, so the total loss plus cost of stopping at stage n after observing $X_1, \ldots, X_n$ is

$$Y_n = \rho(\tau_n(X_1, \ldots, X_n)) + nc \quad \text{for} \quad n = 0, 1, 2, \ldots, \tag{16}$$

and $Y_\infty = +\infty$ if we never stop. The problem is to find a stopping rule N to minimize $\mathrm{E}Y_N$.

Now that the stopping rule problem has been defined, let us check conditions A1 and A2. Since this is a minimization problem, A1 and A2 must be replaced by

A1. $\mathrm{E}\{\inf_n Y_n\} > -\infty$.

A2. $\liminf_{n\to\infty} Y_n \geq Y_\infty$ a.s.

We note that A1 follows from $Y_n \geq 0$, and A2 follows since $Y_n \geq nc \to \infty = Y_\infty$. Thus there is an optimal rule, N^*, given by the principle of optimality. Let $V_0^*(\tau_0)$ denote the expected loss plus cost using this rule, as a function of the prior probability τ_0. There is an invariance in time; at stage n after observing $X_1, \ldots, X_n$, the probability distribution of future payoffs and costs is the same as it was at stage 0 except that the prior has been changed to $\tau_n(X_1, \ldots, X_n)$ and the cost of the first n observations must be paid. Thus,

$$V_n^*(X_1, \ldots, X_n) = V_0^*(\tau_n(X_1, \ldots, X_n)) + nc, \tag{17}$$

where $V_n^*(X_1, \ldots, X_n)$ is the conditional minimum expected loss plus cost having observed $X_1, \ldots, X_n$. The rule given by the principle of optimality reduces to

$$\begin{aligned} N^* &= \min\{n \geq 0 : Y_n = V_n^*(\tau_n(X_1, \ldots, X_n))\} \\ &= \min\{n \geq 0 : \rho(\tau_n(X_1, \ldots, X_n)) = V_0^*(\tau_n(X_1, \ldots, X_n))\} \end{aligned} \tag{18}$$

From this it follows that the optimal decision to stop may be based on the value of $\tau_n(X_1, \ldots, X_n)$.

We now note that $V_0^*(\tau)$ is a concave function of $\tau \in [0,1]$. Let $\alpha, \tau, \tau' \in [0,1]$; we are to show

$$\alpha V_0^*(\tau) + (1-\alpha)V_0^*(\tau') \leq V_0^*(\alpha\tau + (1-\alpha)\tau'). \tag{19}$$

Suppose before stage 0 in the above decision problem, a coin with probability α of heads is tossed and if the coin comes up heads the prior $\tau_0 = \tau$ is used, while if the coin comes up tails the prior $\tau_0 = \tau'$ is used. The left side of (19) is the minimum value of this stopping rule problem when the information on the outcome of the toss may be used. The right side is the minimum value when this information may not be used. Since the class of stopping rules that ignore this information is a subset of the class that may use it, the inequality follows.

In addition, we note that $V_0^*(0) = 0 = \rho(0)$ and $V_0^*(1) = 0 = \rho(1)$. This together with concavity and $V_0^*(\tau) \le \rho(\tau)$ implies that there are numbers a and b with $0 \le a \le L^* \le b \le 1$ such that the optimal rule (18) has the form,

$$N^* = \min\{n \ge 0 : \tau_n(X_1,\ldots,X_n) \le a \quad \text{or} \quad \tau_n(X_1,\ldots,X_n) \ge b\}.$$

where $L^* = L_0/(L_0 + L_1)$. Writing the inequalities in this expression in terms of the likelihood ratio, $\lambda_n = \prod_1^n f_1(X_i)/f_0(X_i)$, we find

$$N^* = \min\{n \ge 0 : \lambda_n \le \frac{(1-\tau_0)a}{\tau_0(1-a)} \quad \text{or} \quad \lambda_n \ge \frac{(1-\tau_0)b}{\tau_0(1-b)}\}.$$

This is the Wald sequential probability ratio test: Sample as long as the probability ratio, λ_n lies between two preassigned numbers, $\alpha < \lambda_n < \beta$; if λ_n falls below α, stop and accept H_0, if L_n rises above β, stop and accept H_1.

The problem of finding the values of a and b of the optimal rule N^* usually requires approximation. There are standard methods of approximation that originate with Wald. See the book of Siegmund (1985) for general methods and applications to statistical problems. Numerical approximation on a computer may also be used. This involves evaluating EY_N for an arbitrary rule of the form $N_{\alpha,\beta} = \min\{n \ge 0 : \lambda_n \le \alpha \text{ or } \lambda_n \ge \beta\}$, and then searching for α and β to minimize this quantity. In a few cases, this may be done explicitly.

§4.6. Application to maximizing the average. Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean $\mu = \mathrm{E}X$. For $n \ge 1$, let $Y_n = S_n/n$, and let $Y_0 = -\infty$ and $Y_\infty = \mu$. By the law of large numbers, $Y_n \to \mu$ a.s. so A2 is satisfied. Therefore an optimal stopping rule will exist if A1 is satisfied; that is, if $\mathrm{E}\sup_n(S_n/n) < \infty$. The following theorem states that A1 holds if $\mathrm{E}X\log^+(X) < \infty$, where $\log^+(x)$ is defined to be $\log(x)$ if $x > 1$, and 0 otherwise.

Theorem 4. *Let $X_1, X_2, \ldots$ be i.i.d. with finite mean μ. Then*

$$\mathrm{E}\sup_n(S_n/n) < \infty \quad \textit{if, and only if,} \quad \mathrm{E}X\log^+(X) < \infty.$$

The sufficiency of this condition is due essentially to Marcinkiewitz and Zygmund (1937). The necessity is due to Burkholder (1962). See the appendix for a proof.

The lack of a time invariance in this problem makes it more difficult than the other problems in this chapter. The problem still has a Markov structure so the decision to stop at stage n can be made on the basis of S_n alone. The principle of optimality may be used to show that the optimal rule is given by a monotone sequence of constants $a_1 \geq a_2 \geq \cdots$ with stopping at stage n if $S_n/n \geq a_n$. However, even in specific simple cases, such as the Bernoulli case, the values of the optimal a_n are quite difficult to evaluate.

An interesting question arises: Does the optimal rule for maximizing the average stop with probability one? Chow and Robbins (1965) show that it does in the Bernoulli case, and Dvoretzky (1967) and Teicher and Wolfowitz (1966) show that it does provided $\mathrm{E}X^2 < \infty$. The general problem is treated in Klass (1973) where it is seen that the optimal rule stops with probability one provided $\mathrm{E}(X^+)^\alpha < \infty$ for some $\alpha > 1$, and examples are given for which the optimal rule does not stop with probability one.

§4.7 Exercises.

1. Let $X_1, X_2, \ldots$ be a sample from the negative exponential distribution with density

$$f(x|\sigma) = \sigma \exp\{-\sigma x\}\mathrm{I}\{x > 0\},$$

where $\sigma > 0$ is known. Find the optimal stopping rule for the problem $Y_n = X_n - nc$ (or $Y_n = \max\{X_1, \ldots, X_n\} - nc$).

2. *Selling an asset without recall and without cost, but with discounted future.* Let $X_1, X_2, \ldots$ be independent identically distributed with finite first moment, let $0 < \beta < 1$, and let $Y_n = \beta^n X_n$ with $Y_0 = Y_\infty = 0$.
(a) Show that A1 and A2 are satisfied. (Hint: $\sup_n Y_n \leq \sum_n \beta^n |X_n|$.)
(b) Show that it is optimal to stop after the first n for which $X_n \geq V^*$, where V^* is the optimal expected return and the unique solution of the equation $V^* = \beta \mathrm{E} \max\{X, V^*\}$.
(c) Specialize to the cases $\mathcal{U}(0,1)$ and $\mathcal{U}(-1,1)$.

3. Solve the discounted version of the house-selling problem for sampling with recall, where $Y_n = \beta^n \max\{X_1, \ldots, X_n\}$. (Hint for (a): $Y_n \leq \max\{\beta|X_1|, \ldots, \beta^n|X_n|\}$.)

4. *Job Processing.* A given job must be assigned to a processor before a (continuous time) deadline, D, which is unknown (until it occurs) but has an exponential distribution at rate $\alpha > 0$: $\mathrm{P}(D > d) = \exp\{-\alpha d\}$ for $d > 0$. Job processors arrive at times, $T_1, T_2, \ldots$, given by a renewal process independent of D; that is, the interarrival times, $T_1, T_2 - T_1, T_3 - T_2, \ldots$ are i.i.d. according to a known distribution G on the positive real line and independent of D. As each processor arrives, its value, that is to say the return it provides if the job is assigned to it, becomes known. These values, denoted by $Z_1, Z_2, \ldots$, are assumed to be i.i.d. according to a known distribution, F, and independent of the arrival times, $T_1, T_2, \ldots$, and deadline, D. If the deadline occurs before the job is assigned to a processor, the return is some constant, say μ. Therefore, the payoff, Y_n, for assigning the job to the nth processor to arrive is taken to be $Y_0 = \mu$, $Y_\infty = \mu$, and

$$Y_n = Z_n \mathrm{I}(T_n < D) + \mu \mathrm{I}(T_n \geq D).$$

Let $K = \min\{n \geq 1 : T_n > D\}$ denote the index of the first processor to arrive after the deadline, and let $\beta = \mathrm{P}(D > T_1) = \mathrm{E}(\exp\{-\alpha T_1\})$. Then K is independent of $Z_1, Z_2, \ldots$ and has a geometric distribution, $\mathrm{P}(K \geq k) = \beta^{k-1}$ for $k = 1, 2, \ldots$. The payoff may be written in an equivalent form,

$$Y_n = Z_n \mathrm{I}(n < K) + \mu \mathrm{I}(n \geq K).$$

In this form, we see that the problem depends on α and the distribution G only through the constant β.
(a) Assume that $\mathrm{E}Z^+ < \infty$. Show that A1 and A2 are satisfied.
(b) Show that it is optimal to stop at the first n for which $Z_n \geq V^*$, where V^* is the optimal expected return and the unique solution of the equation

$$V^* = \mu + (\beta/(1-\beta))\mathrm{E}((Z - V^*)^+).$$

(c) Take $\mu = 0$ and specialize to the cases $\mathcal{U}(0,1)$ and $\mathcal{U}(-1,1)$.

5. In the problem of Section 4.2, suppose the distribution of X gives probability p to $+1$ and probability $1-p$ to -1.
(a) Find the optimal stopping rule, N^*.
(b) Find $\mathrm{P}(N^* < \infty)$.
(c) For what values of $p < 1/2$ is it true that N^* requires you to stop at the first time that $S_n = 1$ no matter what β is?

6. In the problem of stopping a sum with negative drift, assume that the distribution of the X_i is elementary.
(a) Show that the distribution of M is geometric with probability of success $\mathrm{P}(T < \infty)$ where $T = \min\{n \geq 0 : S_n > 0\}$.
(b) Show how to find $\mathrm{P}(T < \infty)$ and hence $\mathrm{E}M$ from knowledge of the generating function, $G(\theta) = \mathrm{E}\theta^{-X}$.
(c) Specialize to the case $\mathrm{P}(X = 1) = p$ and $\mathrm{P}(X = -1) = 1 - p$, where $p < 1/2$.

7. *Setting a record* (Ferguson and MacQueen (1992).) Let $X_1, X_2, \ldots$ be i.i.d. with nonpositive mean, let $S_n = X_1 + \cdots + X_n$, and let $M_n = \max\{M_0, S_1, \ldots, S_n\}$ with $M_0 = 0$. Let $c > 0$ and let $Y_n = M_n - nc$, with $Y_\infty = -\infty$. The problem of choosing a stopping rule to maximize $\mathrm{E}Y_N$ is the problem of deciding when to give up trying to set a new record, the return being the value of the record.
(a) Show that A1 and A2 are satisfied if $\mathrm{E}(X^+)^2 < \infty$.
(b) Show that the optimal rule has the form, $N = \min\{n \geq 0 : M_n - S_n \geq \gamma\}$ for some $\gamma \geq 0$.
(c) Suppose $X_1, X_2, \ldots$ are i.i.d. with $\mathrm{P}(X = 1) = \mathrm{P}(X = -1) = 1/2$. Show the stopping rule of (b) for γ an integer has expected return, $\mathrm{E}Y_N = \gamma - \gamma(\gamma+1)c$. Find the optimal rule.

8. *Attaining a goal* (Ferguson and MacQueen) Let $X_1, X_2, \ldots$ be i.i.d. and let $S_n = X_0 + X_1 + \ldots + X_n$, where X_0 is a given number. Let $a > 0$, $c > 0$, let $Y_n = \mathrm{I}(S_n \geq a) - nc$ and let $Y_\infty = -\infty$. This is the problem of choosing a stopping rule to maximize the

probability of attaining a goal when there is a cost of time.
(a) Show that A1 and A2 are satisfied.
(b) Show that the optimal rule has the form, $N = \min\{n \ge 0 : S_n < \gamma \quad \text{or} \quad S_n \ge a\}$ for some γ.
(c) Suppose that $X_1, X_2, \ldots$ are i.i.d. with $\mathrm{P}(X = 1) = \mathrm{P}(X = -1) = 1/2$. Show the stopping rule of (b) for γ an integer has expected return, for $\gamma < X_0 < a$, of $\mathrm{E}Y_N = (X_0 - \gamma)/(a - \gamma) - c(X_0 - \gamma)(a - X_0)$. Find the optimal rule.

9. (Ferguson and MacQueen) Let $X_1, X_2, \ldots$ be i.i.d. with a distribution that is symmetric about zero. Let $S_n = X_1 + \ldots + X_n$, let $c > 0$ and let $Y_n = |S_n| - nc$ for n finite and $Y_\infty = -\infty$.
(a) Show that A1 and A2 are satisfied if $\mathrm{E}X^2 < \infty$.
(b) Show that an optimal rule has the form $N = \min\{n \ge 0 : |S_n| \ge \gamma\}$ for some $\gamma \ge 0$.
(c) Suppose that $X_1, X_2, \ldots$ are i.i.d. with $\mathrm{P}(X = 1) = \mathrm{P}(X = -1) = 1/2$. Show that the return of the stopping rule of (b) is $\mathrm{E}Y_N = \gamma - \gamma^2 c$. Find an optimal rule.

10. *A change-point repair model.* (This model is due essentially to Girshick and Rubin (1952).) Let T denote an unobservable change-point, and assume the distribution of T is geometric with known parameter π, $\mathrm{P}(T = t) = (1 - \pi)\pi^{t-1}$ for $t = 1, 2, \ldots$. Given $T = t$, the observations, $X_1, X_2, \ldots$, are independent with $X_1, \ldots, X_{t-1}$ i.i.d. having density $f_0(x)$ with mean $\mu_0 \ge 0$, and $X_t, X_{t+1}, \ldots$ i.i.d. having density $f_1(x)$ with mean $\mu_1 < 0$ and finite variance. The observations represent the daily returns from operating a machine. Let the return for stopping at n be $Y_n = S_n - a\mathrm{I}(T \le n) - c$, where $S_n = \sum_1^n X_j$, $c > 0$ represents the cost of shutdown for repair, and $a > 0$ is the excess cost of repair when the machine is in the poor state. Since T is unobservable, it is preferable to work with $Y_n = S_n - aQ_n - c$, where $Q_n = \mathrm{P}\{T \le n | X_1, \ldots, X_n\}$ for finite n ($Q_0 = 0$) and $Y_\infty = -\infty$.
(a) Show that A1 and A2 are satisfied.
(b) Find Q_{n+1} as a function of Q_n and $\lambda(X_{n+1})$, where $\lambda(x)$ is the likelihood ratio, $\lambda(x) = f_1(x)/f_0(x)$.
(c) Show there is an optimal rule of the form, $N = \min\{n \ge 0 : Q_n \ge \gamma\}$ for some constant $\gamma \ge 0$.

11. *Selling two assets.* (Bruss and Ferguson, (1997).) You want to buy Christmas presents for your two children. After deciding which two presents to buy, you go to various stores. With two presents to buy, you can be a little more choosy. If the price of one of the gifts is clearly too high, you know you will have to go to another store anyway, so you will reject a borderline price for the other gift.

We restate this problem as a selling problem to be able to use the formulas of §4.1. You have two objects to sell, x and y. Offers come in daily for these objects, $(X_1, Y_1), (X_2, Y_2), \ldots$, assumed i.i.d. with finite second moments. There is a cost of $c > 0$ to observe each vector. At each stage you may sell none, one or both of the objects. If just one of the objects is sold, you must continue until the other object is sold. Your payoff is the sum of the selling prices minus c times the number of vectors observed.

Once one object is sold, the problem reduces to the standard problem of §4.1. Let V_x and V_y denote the optimal values of selling the x-object and y-object separately. These values are the unique solutions of the Optimality Equations

$$\mathrm{E}(X - V_x)^+ = c \qquad \text{and} \qquad \mathrm{E}(Y - V_y)^+ = c.$$

Therefore, one can consider the two-asset selling problem as a stopping rule problem in which stopping at stage n with offers (X_n, Y_n) yields payoff $W_n = \max\{X_n + Y_n, X_n + V_y, Y_n + V_x\}$.

(a) Find the Optimality Equation for V_{xy}, the value of the two-asset problem, and describe the optimal stopping rule.

(b) Suppose the (X_n, Y_n) are a sample from (X, Y) with X and Y independent having uniform distributions on the interval $(0, 1)$. Take $c = 1/8$ (so that $V_x = V_y = 1/2$), and find the optimal rule.

(c) In (b), suppose $c = 1/2$ (so that $V_x = V_y = 0$), and find the optimal rule.

APPENDIX TO CHAPTER 4

The following proof of Theorem 1 is taken partly from DeGroot (1970) pp. 350-352, where it is attributed to Bramblett (1965). For this theorem and the others in this appendix, the following inequalities are basic.

$$\sum_{n=1}^{\infty} \mathrm{P}(Z > n) \leq \mathrm{E}(Z^+) = \int_0^{\infty} \mathrm{P}(Z > z)\, dz \leq \sum_{n=0}^{\infty} \mathrm{P}(Z > n)$$

so that

$$\mathrm{E}(Z^+) < \infty \quad \text{if and only if} \quad \sum_n \mathrm{P}(Z > n) < \infty.$$

Similarly,

$$\mathrm{E}(Z^+)^2 = 2\int_0^{\infty} z\mathrm{P}(Z > z)\, dz < \infty \quad \text{if and only if} \quad \sum_n n\mathrm{P}(Z > n) < \infty.$$

and

$$\mathrm{E}(Z^+)^2 = 2\int_0^{\infty}\int_0^{\infty} \mathrm{P}(Z > z+u)\, du\, dz < \infty \quad \text{if and only if} \quad \sum_k \sum_n \mathrm{P}(Z > n+k) < \infty.$$

Theorem 1. *Let $X, X_1, X_2, \ldots$ be identically distributed, let $c > 0$, and let $Y_n = X_n - nc$ or $Y_n = M_n - nc$, where $M_n = \max\{X_1, \ldots, X_n\}$. If $\mathrm{E}(X^+) < \infty$, then $\sup Y_n < \infty$ a.s. and $Y_n \to -\infty$ a.s. If $\mathrm{E}(X^+)^2 < \infty$, then $\mathrm{E}\sup_{n\geq 1} Y_n \leq \mathrm{E}(X^+)^2/(2c)$.*

Proof. Since $M_n - nc = \max(X_1 - nc, \ldots, X_n - nc) \leq \max(X_1 - c, \ldots, X_n - nc)$, one sees that $\sup(X_n - nc) = \sup(M_n - nc)$. This implies that in the statements about $\sup Y_n$ it does not matter which definition of Y_n we take; so let us take $Y_n = X_n - nc$. Suppose $\mathrm{E}(X^+) < \infty$. Then,

$$\begin{aligned}\mathrm{P}(\sup_{n\geq 1} Y_n > z) &\leq \sum_{n=1}^{\infty} \mathrm{P}(Y_n > z) = \sum_{n=1}^{\infty} \mathrm{P}(X > z + nc) \\ &= \sum_{n=1}^{\infty} \mathrm{P}((X - z)/c > n) \leq \mathrm{E}((X - z)^+/c) \to 0\end{aligned}$$

as $z \to \infty$. Thus, $\sup Y_n < \infty$ a.s. Moreover, $Y_n \leq M_n - nc = (M_n - nc/2) - nc/2 \leq U - nc/2 \to -\infty$, where $U = \sup(M_n - nc/2)$.

Now assume that $\mathrm{E}(X^+)^2 < \infty$.

$$\begin{aligned}
\mathrm{E}\sup_{n\geq 1} Y_n &\leq \int_0^\infty \mathrm{P}(\sup_{n\geq 1} Y_n > z)dz \\
&\leq \int_0^\infty \mathrm{E}((X-z)^+/c)\,dz \\
&= \int_0^\infty \int_z^\infty (x-z)\,dF(x)\,dz/c \\
&= \int_0^\infty \int_0^x (x-z)\,dz\,dF(x)/c \\
&= \int_0^\infty (x^2/2)\,dF(x)/c = \mathrm{E}(X^+)^2/(2c). \ \blacksquare
\end{aligned}$$

For the converse, the variables are required to be independent.

Theorem 1′. *If $X, X_1, X_2, \ldots$ are i.i.d. and if $\mathrm{E}\sup(X_n - nc) < \infty$, then $\mathrm{E}(X^+)^2 < \infty$.*

Proof. Take $c = 1$ without loss of generality and suppose $\mathrm{E}\sup_{n\geq 0}(X_n - n) < \infty$. Then,

$$\begin{aligned}
\mathrm{P}(\sup_n (X_n - n) > z) &= 1 - \prod_{n=1}^\infty \mathrm{P}(X_n - n \leq z) \\
&= 1 - \prod_{n=1}^\infty (1 - \mathrm{P}(X > z+n)) \geq 1 - \exp\{-\sum_{n=1}^\infty \mathrm{P}(X > z+n)\}.
\end{aligned}$$

Since $\sum_z \mathrm{P}(\sup_n(X_n - n) > z) < \infty$, we have $\mathrm{P}(\sup_n(X_n - n) > z) \to 0$ as $z \to \infty$, which in turn implies that $\sum_{n=1}^\infty \mathrm{P}(X > z+n) \to 0$, so that for sufficiently large z,

$$\mathrm{P}(\sup_n(X_n - n) > z) \geq \sum_{n=1}^\infty \mathrm{P}(X > z+n)/2$$

(using $1 - \exp\{-x\} \geq x/2$ for x sufficiently small.) Hence,

$$\begin{gathered}
\sum_z \mathrm{P}(\sup_n(X_n - n) > z) < \infty \quad \text{implies} \\
\sum_z \sum_n \mathrm{P}(X > z+n) < \infty \quad \text{which implies} \\
\mathrm{E}(X^+)^2 < \infty. \ \blacksquare
\end{gathered}$$

As an alternate proof of Theorem $1'$, we show that the rule $N = \min\{n \geq 1 : X_n \geq 2cn\}$ gives $\mathrm{E}(X_N - Nc)^+ = \infty$ when $\mathrm{E}X^+ < \infty$ and $\mathrm{E}(X^+)^2 = \infty$.

$$\begin{aligned}
\mathrm{E}(X_N - Nc)^+ &= \sum_{n=1}^{\infty} \mathrm{E}(X_n - nc)I(N = n) \\
&\geq \sum_{n=1}^{\infty} cn\mathrm{P}(N = n) \\
&= \sum_{n=1}^{\infty} cn\mathrm{P}(N > n-1)\mathrm{P}(X_n > 2cn) \\
&\geq \sum_{n=1}^{\infty} cn\mathrm{P}(N = \infty)\mathrm{P}(X_n > 2cn) = \infty.
\end{aligned}$$

since $\mathrm{E}(X^+)^2 = \infty$ implies $\sum_{n=1}^{\infty} n\mathrm{P}(X_n > 2cn) = \infty$, and

$$\begin{aligned}
\mathrm{P}(N = \infty) &= \mathrm{P}(X_n < 2cn \quad \text{for all} \quad n) \\
&= \prod_{n=1}^{\infty} \mathrm{P}(X_n < 2nc) = \prod_{n=1}^{\infty}(1 - \mathrm{P}(X_n \geq 2nc)) \\
&\sim \exp\{-\sum \mathrm{P}(X_n > 2cn)\} \sim \exp\{-\mathrm{E}X^+/2c\} > 0.
\end{aligned}$$

The following proof of the result of Kiefer and Wolfowitz (1956) is a modification, due to Thomas Liggett, of a computation of Kingman (1962). We assume a finite variance for X and derive an upper bound for $\mathrm{E}M$. The theorem of Section 4.3, assuming only that $\mathrm{E}(X^+)^2 < \infty$, may be deduced from Theorem 2 and 2' below by truncating the distribution of X below at $-B$ where B is chosen large enough so that if $X' = \max\{X, -B\}$, then $\mathrm{E}X'$ is still negative. Let $X'_j = \max\{X_j, -B\}$, $S'_n = \sum_1^n X'_j$, and $M' = \sup_{n\geq 0} S'_n$. Then $M \leq M'$, so that from Theorem 2 below, $\mathrm{E}M \leq \mathrm{E}M' \leq \mathrm{Var}\,(X')/(2|\mathrm{E}X'|) < \infty$.

Theorem 2. *Let $X, X_1, X_2, \ldots$ be i.i.d. with $\mu = \mathrm{E}X < 0$ and $\sigma^2 = \mathrm{Var}\,X < \infty$. Let $S_n = \sum_1^n X_j$, $S_0 = 0$ and $M = \sup_{n\geq 0} S_n$. Then*

$$\mathrm{E}M \leq \sigma^2/(2|\mu|).$$

Proof. Let $M_n = \max_{0\leq j\leq n} S_j$. Note that $M_{n+1} = \max(0, X_1 + \max_{1\leq j\leq n+1}(S_j - X_1))$ so that the distribution of M_{n+1} is the same as the distribution of $(M_n + X)^+$. Then, writing $M_n + X = (M_n + X)^+ - (M_n + X)^-$, and noting that $(M_n + X)^2 = ((M_n + X)^+)^2 + ((M_n + X)^-)^2$, we find,

$$\begin{aligned}
\mathrm{E}(M_n + X)^- &= \mathrm{E}(M_{n+1}) - \mathrm{E}(M_n) - \mathrm{E}(X) \quad \text{and} \\
\mathrm{E}((M_n + X)^-)^2 &= -\mathrm{E}(M_{n+1}^2) + \mathrm{E}(M_n^2) + 2\mu\mathrm{E}(M_n) + \mathrm{E}(X^2)
\end{aligned}$$

so that

$$\begin{aligned} 0 &\le \mathrm{Var}\left((M_n + X)^-\right) \\ &= 2\mu \mathrm{E}(M_{n+1}) + \sigma^2 - (\mathrm{E}M_{n+1} - \mathrm{E}M_n)^2 - (\mathrm{E}(M_{n+1}^2) - \mathrm{E}(M_n^2)) \\ &\le 2\mu \mathrm{E}(M_{n+1}) + \sigma^2. \end{aligned}$$

Hence, $\mathrm{E}(M_{n+1}) \le \sigma^2/(2|\mu|)$ for all n and the result now follows by passing to the limit using monotone convergence. ■

The proof of the following converse to Theorem 2 is due to Michael Klass. In the proof it is seen that if $\mathrm{E}(X^+)^2 = \infty$, then the simple rule $T = \min\{n \ge 0 : S_n > 0\}$ has infinite expected return, $\mathrm{E}S_T^+ = \infty$.

Theorem 2′. *Let $X, X_1, X_2, \ldots$ be i.i.d. with finite first moment $\mu < 0$. Then, $\mathrm{E}M < \infty$ implies $\mathrm{E}(X^+)^2 < \infty$.*

Proof. Let $T = \min\{n \ge 0 : S_n > 0\}$ with $T = \infty$ if $S_n \le 0$ for all n. Then $\mathrm{E}M < \infty$ implies that $\mathrm{E}S_T \mathrm{I}(T < \infty) < \infty$. But,

$$\begin{aligned} \mathrm{E}S_T \mathrm{I}(T < \infty) &= \sum_n \mathrm{E}S_T \mathrm{I}(T = n) \\ &\ge \sum_n \mathrm{E}S_T \mathrm{I}(T = n, S_{n-1} > 2n\mu, X_n > -3n\mu) \\ &\ge \sum_n n|\mu| \mathrm{P}(T = n, S_{n-1} > 2n\mu, X_n > -3n\mu) \\ &= \sum_n n|\mu| \mathrm{P}(n \le T \le \infty, S_{n-1} > 2n\mu, X_n > -3n\mu) \\ &= \sum_n n|\mu| \mathrm{P}(n \le T \le \infty, S_{n-1} > 2n\mu) \mathrm{P}(X_n > -3n\mu). \end{aligned}$$

But $\mathrm{P}(n \le T \le \infty, S_{n-1} > 2n\mu) \to \mathrm{P}(T = \infty) > 0$ as $n \to \infty$, so that

$$\sum_n n\mathrm{P}(X_n > -3n\mu) < \infty,$$

which implies that $\mathrm{E}(X^+)^2 < \infty$. ■

In the restatement of Theorem 3, we put $c = 1$ without loss of generality. The proof of the "only if" part of the theorem was suggested by Thomas Liggett.

Theorem 3. *Let $Z, Z_1, Z_2, \ldots$ be i.i.d., let $\epsilon, \epsilon_1, \epsilon_2, \ldots$ be i.i.d. Bernoulli with $p = \mathrm{P}(\epsilon = 1) = 1 - \mathrm{P}(\epsilon = 0)$, with $0 < p < 1$. Let the $\{Z_j\}$ and $\{\epsilon_j\}$ be independent, and let $X_0 = 0$ and $X_n = \epsilon_n(X_{n-1} + Z_n)$ for $n = 1, 2, \ldots$. Then,*

$$\mathrm{E}\sup_n (X_n - n) < \infty \quad \textit{if and only if} \quad \mathrm{E}(Z^+)^2 < \infty.$$

Proof. First, suppose that $\mathrm{E}(Z^+)^2 < \infty$. Since X_n is the sum of the Z_j since the last $\epsilon = 0$, it has the same distribution as $\sum_1^{\min(K,n)} Z_j$, where K represents the distance back from n to the most recent failure if any. We may take $K \geq 0$ to have a geometric distribution with success probability p and to be independent of the Z_j. This latter sum is less than $Q = \sum_1^K Z_j^+$.

$$\begin{aligned}\mathrm{E}\sup_n(X_n - n) &\leq \sum_{x=0}^{\infty} \mathrm{P}(\sup_n(X_n - n) \geq x) \\ &\leq \sum_{x=0}^{\infty}\sum_{n=0}^{\infty} \mathrm{P}(X_n - n \geq x) \\ &\leq \sum_{x=0}^{\infty}\sum_{n=0}^{\infty} \mathrm{P}(Q \geq x + n).\end{aligned}$$

This last sum is finite since the variance of Q is finite:

$$\begin{aligned}\mathrm{E}(Q^2) = \mathrm{E}(\mathrm{E}(Q^2|K)) &= \mathrm{E}(\mathrm{E}(\sum_{i=1}^{K}\sum_{j=1}^{K} Z_i^+ Z_j^+ |K)) \\ &= \mathrm{E}K\mathrm{E}(Z^+)^2 + \mathrm{E}K(K-1)(\mathrm{E}Z^+)^2 < \infty.\end{aligned}$$

Conversely, note that

$$\sup_n(X_n - n) \geq \sup_n(\mathrm{I}(\epsilon_n = 1, \epsilon_{n-1} = 0)(Z_n - n)).$$

This latter sup is equal in distribution to $\sup_n(Z_n - K_n)$, where K_n is the time of the nth appearance of the pattern $\epsilon_{j-1} = 0, \epsilon_j = 1, K_n$ starting with $\epsilon_0 = 0$. Therefore, the differences $K_2 - K_1, K_3 - K_2, \ldots$ are i.i.d. with finite second moment. For an arbitrary positive number c', we have

$$\sup(Z_n - c'n) \leq \sup(Z_n - K_n) + \sup(K_n - c'n).$$

The expectation of the first term on the right is finite since it is bounded above by $\mathrm{E}\sup(X_n - n) < \infty$. The expectation of the second term is finite from Theorem 2, provided c' is chosen large enough, say $c' \geq 2\mu$, where $\mu = \mathrm{E}(K_2 - K_1)$. Hence, $\mathrm{E}\sup(Z_n - c'n) < \infty$ so that from the converse part of Theorem 1, $\mathrm{E}(Z^+)^2 < \infty$. ∎

We precede the proof of Theorem 4 by two lemmas. Let $X_1, X_2, \ldots$ be i.i.d. with finite mean and let $S_n = \sum_1^n X_j$.

Lemma 1. *For* $j \leq n$, $\mathrm{E}\{S_j^+/j | S_n, S_{n+1}, \ldots\} \geq S_n^+/n$.

Proof. By symmetry, $\mathrm{E}\{X_j | S_n, S_{n+1}, \ldots\}$ is the same for all $j \leq n$, and since the sum is S_n, we must have $\mathrm{E}\{X_j | S_n, S_{n+1}, \ldots\} = S_n/n$. Hence, $\mathrm{E}\{S_j/j | S_n, S_{n+1}, \ldots\} = S_n/n$, and hence $S_n^+/n \leq \mathrm{E}\{S_j^+/j | S_n, S_{n+1}, \ldots\}$ (since $(\mathrm{E}X)^+ \leq \mathrm{E}X^+$). ∎

Lemma 2. *Let* $A_n = \{\sup_{j\le n} S_j^+/j \ge \lambda\}$ *where* $\lambda > 0$. *Then* $\mathrm{E}\{X_1^+\mathrm{I}(A_n)\} \ge \lambda \mathrm{P}\{A_n\}$.

Proof. Let $A_{nj} = \{S_j^+/j \ge \lambda \quad \text{and} \quad S_k^+/k < \lambda \quad \text{for} \quad k = j+1,\ldots,n\}$. Then $A_n = \bigcup_{j=1}^n A_{nj}$ and $X_1^+ = S_1^+$, so

$$\begin{aligned}
\mathrm{E}\{X_1^+\mathrm{I}(A_n)\} &= \sum_{j=1}^n \mathrm{E}\{S_1^+\mathrm{I}(A_{nj})\} \\
&= \sum_{j=1}^n \mathrm{E}[\mathrm{E}\{S_1^+\mathrm{I}(A_{nj})|S_j, S_{j+1},\ldots\}] \\
&= \sum_{j=1}^n \mathrm{E}[\mathrm{I}(A_{nj})\mathrm{E}\{S_1^+|S_j, S_{j+1},\ldots\}] \\
&\ge \sum_{j=1}^n \mathrm{E}[\mathrm{I}(A_{nj})S_j^+/j] \\
&\ge \lambda \sum_{j=1}^n \mathrm{E}\{\mathrm{I}(A_{nj})\} = \lambda \mathrm{P}(A_n). \ \blacksquare
\end{aligned}$$

Theorem 4. *If* $\mathrm{E}\{X_1 \log^+ X_1\} < \infty$, *then* $\mathrm{E}\sup_n S_n/n < \infty$.

Proof. (Doob (1953) p. 317) Let $Z_n = \sup_{j\le n} S_j^+/j$. Then $\mathrm{E}Z_n < \infty$ (since $Z_n \le \sum_1^n S_j^+/j$).

$$\begin{aligned}
\mathrm{E}Z_n &= \int_0^\infty \mathrm{P}(Z_n \ge z)\,dz \le 1 + \int_1^\infty \mathrm{P}(Z_n \ge z)\,dz \\
&\le 1 + \int_1^\infty (1/z)\mathrm{E}\{X_1^+\mathrm{I}(Z_n \ge z)\}\,dz \quad \text{(Lemma 2)} \\
&= 1 + \mathrm{E}\{X_1^+ \int_1^{Z_n} (1/z)\,dz \mathrm{I}(Z_n \ge 1)\} = 1 + \mathrm{E}X_1^+ \log^+ Z_n.
\end{aligned}$$

Now, note that $\log x \le x/e$ for all $x > 0$. (There is equality at $x = 1/e$, the slopes are equal there, and $\log x$ is concave.) Replacing x by Z_n/X_1^+, we find $X_1^+ \log^+ Z_n \le X_1^+ \log^+ X_1 + Z_n/e$. Therefore,

$$\mathrm{E}Z_n \le 1 + \mathrm{E}X_1^+ \log^+ X_1 + \mathrm{E}Z_n/e,$$

so that

$$\mathrm{E}Z_n \le (e/(1-e))(1 + \mathrm{E}X_1 \log^+ X_1).$$

But Z_n converges monotonically to $\sup_j S_j^+/j$, so that $\mathrm{E}\sup_j S_j/j \le \mathrm{E}\sup_j S_j^+/j < \infty$. $\blacksquare$

The above proof uses the fact that S_n is a sum of i.i.d. random variables only through Lemma 1. Thus, the theorem is valid for sequences S_n^+/n satisfying that lemma (a non-negative reverse supermartingale).

The following converse, due to McCabe and Shepp (1970) and Davis (1971), explicitly exhibits a simple stopping rule with $\mathrm{E}S_N/N = \infty$.

Theorem $4'$**.** *Let* $X_1, X_2, \ldots$ *be i.i.d. with finite first moment and suppose* $\mathrm{E}X_1 \log^+ X_1 = \infty$*. Let* $c > 0$ *be such that* $\mathrm{P}(X < c) > 0$*, and let* N *denote the stopping rule* $N = \min\{n \geq 1 : X_n \geq nc\}$*. Then* $\mathrm{E}S_N/N = \infty$*.*

Proof. Without loss of generality, we take $c = 1$ since we could work as well with the sequence $X_1/c, X_2/c, \ldots$. First note that $\mathrm{P}(N = \infty) > 0$, since $\mathrm{P}(X_1 < 1) > 0$ and

$$\begin{aligned}\mathrm{P}(N = \infty) &= \mathrm{P}(X_n < n \quad \text{for all} \quad n) = \prod_{n=1}^{\infty} \mathrm{P}(X_n < n) \\ &\geq [\prod_{n=1}^{m-1} \mathrm{P}(X_n < n)](1 - \sum_{n=m}^{\infty} \mathrm{P}(X_n \geq n)),\end{aligned}$$

and

$$\begin{aligned}\sum_{n=m}^{\infty} \mathrm{P}(X_n \geq n) &= \sum_{n=m}^{\infty} \int_n^{\infty} dF(x) \\ &\leq \int_m^{\infty} x\, dF(x) \to 0 \quad \text{as} \quad m \to \infty,\end{aligned}$$

where F denotes the distribution function of X_1. Now, choose m so that the latter sum is less than 1.

Second, note that $\mathrm{E}X_N/N = \infty$, since

$$\begin{aligned}\mathrm{E}((X_N/N)\mathrm{I}\{N < \infty\}) &= \sum_1^{\infty} \mathrm{P}(N = n) n^{-1} \mathrm{E}\{X_n | N = n\} \\ &= \sum_1^{\infty} \mathrm{P}(N = n) n^{-1} \mathrm{E}\{X_n | X_n \geq n\}\end{aligned}$$

(since the X_j are independent)

$$= \sum_1^{\infty} \mathrm{P}(N \geq n) n^{-1} \int_n^{\infty} x\, dF(x)$$

(since $\mathrm{P}(N = n) = \mathrm{P}(N \geq n)\mathrm{P}(X_n \geq n)$)

$$\begin{aligned}&\geq \mathrm{P}(N = \infty) \sum_1^{\infty} n^{-1} \int_n^{\infty} x\, dF(x) \\ &= \mathrm{P}(N = \infty) \int_1^{\infty} \sum_1^{[x]} n^{-1} x\, dF(x) \\ &\geq \mathrm{P}(N = \infty) \int_1^{\infty} x \log(x)\, dF(x) = \infty.\end{aligned}$$

Finally, note that $\mathrm{E}(S_N/N) = \mathrm{E}(X_N/N) + \mathrm{E}(\sum_{k=1}^{N-1} X_k/N)$. But,

$$\begin{aligned}\mathrm{E}(\sum_{k=1}^{N-1} X_k/N) &= \sum_{n=1}^{\infty} \mathrm{P}(N=n)n^{-1}\sum_{k=1}^{n-1}\mathrm{E}(X_k|N=n)\\ &= \sum_{n=1}^{\infty} \mathrm{P}(N=n)n^{-1}\sum_{k=1}^{n-1}\mathrm{E}(X_k|X_k<k)\\ &\geq \sum_{n=1}^{\infty} \mathrm{P}(N=n)n^{-1}\mathrm{E}(X_1|X_1<1)(n-1) > -\infty,\end{aligned}$$

since $\mathrm{E}(X|X<k)$ is nondecreasing in k. Thus, $\mathrm{E}(S_N/N) = \infty$. ■

Chapter 5. MONOTONE STOPPING RULE PROBLEMS.

The ease with which the various problems of Chapter 4 were solved may be misleading. In general, stopping rule problems do not have closed form solutions and methods of finding approximate solutions must be used. Indeed, most problems without some form of Markovian structure are essentially intractable due to the high dimension of the observations involved.

In principle, it is possible to approximate a stopping rule problem by considering a truncated version of the problem. We choose a large truncation integer, T, and require stopping by stage T. More generally, for those problems for which continuing forever is a useful possibility, we would require that at stage T, the decision maker must choose between stopping at T and continuing forever, that is, between receiving Y_T and receiving $\mathrm{E}(Y_\infty|\mathcal{F}_T)$. Essentially, we replace the payoff for stopping at T by $Y_T^{(T)} = \max\{Y_T, \mathrm{E}(Y_\infty|\mathcal{F}_T)\}$, and we require stopping if stage T is reached. This is a finite horizon problem which in principle can be solved by the method of backward induction of Chapter 2. In this chapter, we find conditions under which the infinite problem may be approximated to any desired degree of accuracy by truncated problems with sufficiently large truncation points.

From a practical point of view, this method of approximating a solution by truncation is not a very good one. Solving the problem truncated at T requires computation and storage of the truncated values, $V_j^{(T)}(x_1,\ldots,x_j)$ for $j = T, T-1,\ldots,0$ as defined by equation (1) of Chapter 2. If each x_i is allowed to assume 10 values and $T = 20$ say, this is already too large a problem for today's computers. Even if the problem can be reduced by sufficient statistics or by taking advantage of a Markovian structure, there are better methods of approximation. We consider in Section 5.1 the k-stage look-ahead rules as a simple but powerful improvement on the method of truncation. We also consider the k-time look-ahead rules.

The main topic of this chapter appears in Section 5.2. The 1- stage look-ahead rules are generally quite good; sometimes they are optimal. Under the condition that the problem be *monotone*, the 1-stage look-ahead rule is optimal for finite horizon problems. The extension of this result to infinite horizon problems requires some new conditions, detailed in Section 5.3. Essentially, these conditions are needed so that the infinite problem can be approximated by truncated problems in the sense that the limiting value of the truncated problems is the value of the original problem, $\lim_{T\to\infty} V_0^{(T)} = V^*$. Then, if the problem is monotone, the 1-stage look-ahead rule is optimal for the truncated problems, and by extension for the infinite problem as well. These ideas provide a second general method for finding simple solutions to many complex problems. (The first is the method of Chapter 4.) There are numerous applications contained in Section 5.4 and the Exercises.

§**5.1 The k-stage look-ahead rules.** For stopping rule problems, the k-stage look-ahead rule (k-sla) is described by the stopping time,

$$N_k = \min\{n \geq 0 : Y_n \geq V_n^{(n+k)}\} = \min\{n \geq 0 : Y_n \geq \mathrm{E}(V_{n+1}^{(n+k)}|\mathcal{F}_n)\}. \tag{1}$$

The k-sla is the rule which at each stage stops or continues according to whether the rule optimal among those truncated k stages ahead stops or continues. Thus at stage n, if the optimal rule among those truncated at $n+k$ continues, the k-sla continues; otherwise, the k-sla stops.

The simplest of these rules is the 1-stage look-ahead rule,

$$N_1 = \min\{n \geq 0 : Y_n \geq \mathrm{E}(Y_{n+1}|\mathcal{F}_n)\}, \tag{2}$$

In words, N_1 calls for stopping at the first n for which the return for stopping is at least as great as the expected return of continuing one stage and then stopping.

The one-stage look-ahead rule, sometimes called the myopic rule, is reasonably good, and the two- and three-stage look-ahead rules are often quite good. An important property of these rules is that *if an optimal rule exists, and if the k-sla tells you to continue, then it is optimal to continue*, for then there is at least one rule that continues and gives you at least as great an expected return as stopping at once. This property suggests a simplification of the 2-sla. Use the 1-sla until it tells you to stop, and then use the 2-sla. Similarly, the 3-sla is equivalent to : Use the 1-sla until it tells you to stop, then the 2-sla until it tells you to stop, and then use the 3-sla.

On the other hand, sometimes the 1-sla will tell you to stop, while the 2-sla, and hence the optimal rule, will tell you to continue, as examples given later will show. Therefore, it would be good to know how close to optimal the 1-sla is when it calls for stopping. Theorem 2 in Section 5.2 gives a sufficient condition for the one-stage look-ahead rule to be optimal. This theorem may be described as follows. *Suppose $V_0^{(T)} \to V^*$ as $T \to \infty$. If at some stage the 1-sla calls for stopping, and if no matter what happens in the future the 1-sla will continue to call for stopping at all future stages, then stopping immediately is optimal.* This result is true also if "1-sla" is replaced by "k-sla".

SEQUENTIAL STATISTICAL ESTIMATION. As an illustration of the computation of the 1-sla and the 2-sla, we specialize the Bayes sequential decision problems of Example 3 of Chapter 1 to the problem of statistical estimation of an unknown parameter with squared error loss. The problem is to estimate a parameter θ based on a sequentially observed sequence of random variables, $X_1, X_2, \ldots$, known to be independent and identically distributed from a distribution $F(x|\theta)$. It is assumed that θ is a real parameter and that the loss incurred when θ is estimated by a real number a is the square of the error,

$L(\theta, a) = (\theta - a)^2$. One may observe $X_1, X_2, \ldots$ for as long as desired before estimating θ at a cost of c per observation. In the Bayes approach to this problem, the prior distribution, τ, of θ is assumed known. If after observing $X_1, \ldots, X_n$, it is decided to stop and estimate θ, the estimate that minimizes the conditional expected loss, $\mathrm{E}\{(\theta-a)^2|X_1, \ldots, X_n\}$, i.e. the Bayes estimate, is the mean of the posterior distribution of θ given $X_1, \ldots, X_n$, namely, $a = \hat{\theta}_n = \mathrm{E}(\theta|X_1, \ldots, X_n)$. The minimum expected loss is then just the conditional variance of θ, $\rho_n(X_1, \ldots, X_n) = \mathrm{Var}\,(\theta|X_1, \ldots, X_n) = \mathrm{E}\{(\theta - \hat{\theta}_n)^2|X_1, \ldots, X_n\}$. Therefore, the total loss plus cost of stopping at stage n is

$$Y_n = \mathrm{Var}\,(\theta|X_1, \ldots, X_n) + nc. \tag{3}$$

As n increases, the posterior variance ordinarily decreases almost surely to zero, while the cost of sampling increases to ∞. The problem is to choose a stopping rule N to minimize $\mathrm{E}Y_N$.

As an example, consider estimating the mean θ of a Poisson distribution

$$f(x|\theta) = e^{-\theta}\theta^x/x! \qquad \text{for } x = 0, 1, 2, \ldots,$$

based on a sequential sample, $X_1, X_2, \ldots$, with constant cost c per observation and squared error loss. Let the prior distribution of θ be gamma, $\mathcal{G}(\alpha, 1/\lambda)$, with density

$$g(\theta) = \frac{\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda\theta}\theta^{\alpha-1} \qquad \text{on } (0, \infty).$$

It is assumed that the prior parameters, α and λ, are known. The joint density of $X_1, \ldots, X_n$ and θ is the product $g(\theta)\prod_1^n f(x_i|\theta)$, and the posterior distribution of θ given $X_1, \ldots, X_n$ is proportional to this, namely, $f(\theta|x_1, \ldots, x_n) \propto e^{-(\lambda+n)\theta}\theta^{\alpha+x_1+\cdots+x_n-1}$, which is $\mathcal{G}(\alpha + S_n, 1/(\lambda + n))$ where $S_n = \sum_1^n X_i$. Thus, past information furnished by the observations may be summarized by "up-dating" the parameters of the prior from (α, λ) to $(\alpha + S_n, \lambda + n)$. Since the mean of $\mathcal{G}(\alpha, 1/\lambda)$ is α/λ, the Bayes estimate of θ based on $X_1, \ldots, X_n$ is $\hat{\theta}_n = (\alpha + S_n)/(\lambda + n)$. Since the variance of $\mathcal{G}(\alpha, 1/\lambda)$ is α/λ^2, equation (3) becomes

$$Y_n = \frac{\alpha + S_n}{(\lambda + n)^2} + nc.$$

At stage 0, the 1-sla compares the expected loss of stopping now, $Y_0 = \alpha/\lambda^2$, with the expected loss plus cost of taking one observation and stopping, $\mathrm{E}Y_1$, where, since $\mathrm{E}X_1 = \mathrm{E}(\mathrm{E}(X_1|\theta)) = \mathrm{E}(\theta) = \alpha/\lambda$,

$$\begin{aligned} \mathrm{E}Y_1 &= \mathrm{E}(\alpha + X_1)/(\lambda + 1)^2 + c \\ &= \alpha/(\lambda(\lambda + 1)) + c. \end{aligned}$$

The 1-sla calls for stopping without taking any observations if $\alpha/\lambda^2 \le \alpha/(\lambda(\lambda+1)) + c$ or equivalently if

$$\frac{\alpha}{\lambda^2(\lambda + 1)} \le c. \tag{4}$$

The general 1-sla may be obtained from this by replacing (α, λ) by the the updated parameters $(\alpha + S_n, \lambda + n)$. Hence, the 1-sla of equation (2) with the inequality reversed since we are dealing with a cost rather than a return becomes is

$$N_1 = \min\{n \geq 0 : \frac{\alpha + S_n}{(\lambda + n)^2(\lambda + n + 1)} \leq c\}.$$

To compute the 2-sla at stage 0, we compare the expected loss of stopping without any observations, $Y_0 = \alpha/\lambda^2$, with the expected loss plus cost of observing X_1 and then using the optimal 1-stage procedure, namely,

$$c + \mathrm{E}\min\{(\alpha + X_1)/(\lambda + 1)^2, (\alpha + X_1)/((\lambda + 1)(\lambda + 2)) + c\}, \tag{5}$$

The 2-sla calls for stopping without taking any observations if the former is no greater than the latter. We can simplify this expression and make the comparison of the 1-sla and the 2-sla easier by subtracting $(\alpha + X_1)/(\lambda + 1)^2$ from both terms in the minimum of (5) to rewrite it as

$$\begin{aligned} &c + \mathrm{E}\min\{0, -\frac{\alpha + X_1}{(\lambda + 1)^2(\lambda + 2)} + c\} + \mathrm{E}(\alpha + X_1)/(\lambda + 1)^2 \\ &\qquad = c - \mathrm{E}(\frac{\alpha + X_1}{(\lambda + 1)^2(\lambda + 2)} - c)^+ + \alpha/(\lambda(\lambda + 1)). \end{aligned} \tag{6}$$

from this, it follows that the 2-sla, N_2, calls for stopping at stage 0 if

$$\frac{\alpha}{\lambda^2(\lambda + 1)} \leq c - \mathrm{E}(\frac{\alpha + X_1}{(\lambda + 1)^2(\lambda + 2)} - c)^+ \tag{7}$$

The 2-sla at stage n can be obtained from this by replacing (α, λ) by $(\alpha + S_n, \lambda + n)$ where the expectation must also be taken with the updated parameters. For a problem in which the 2-sla can be found in closed form, see Exercise 1.

THE TWO-TIMER. A simple improvement on the k-stage look-ahead rule, called the **k-time look-ahead rule**, has been suggested by A. Biesterfeld (1996). The one-time look-ahead rule, which is an improvement over both the myopic rule (the 1-sla) and the hypermetropic rule of Exercise 3.2, is the rule that calls for stopping at stage n if $Y_n \geq \sup_{t>n} \mathrm{E}(Y_t|\mathcal{F}_n)$. In other words, if at stage n there is some fixed future time $t > n$ (possibly $t = \infty$) such that the conditional expected return of continuing to stage t and stopping is greater than the return of stopping immediately, then the one-time look-ahead rule continues at least to stage $n + 1$. Otherwise, it stops immediately. We denote the one-time look-ahead rule by T_1. Thus,

$$T_1 = \min\{n \geq 0 : Y_n \geq \sup_{t>n} \mathrm{E}(Y_t|\mathcal{F}_n)\}. \tag{8}$$

Like the one-stage look-ahead rule, N_1, if T_1 calls for continuing it is optimal to continue (provided an optimal rule exists). In addition, if N_1 calls for continuing, so does T_1. From Exercise 3.3, T_1 is at least as good as N_1. It is easy to see it can be better. For example, when the Y_n are degenerate with $Y_0 = 1$, $Y_1 = 0$, $Y_2 = 2$, and $Y_3 = Y_4 = \cdots = Y_\infty = 0$, then $N_1 = 0$ with a return of 1, and $T_1 = 2$ with a return of 2.

Let us see how well T_1 does in the example of statistical sequential estimation of the mean of a Poisson distribution. The observations, $X_1, X_2, \ldots$, are i.i.d. Poisson with mean θ, and the prior distribution of θ is $\mathcal{G}(\alpha, 1/\lambda)$. To find T_1, let us find the conditions under which T_1 stops at stage zero. We compute

$$\mathrm{E}(Y_t) = \mathrm{E}(\frac{\alpha + S_t}{(\lambda+t)^2} + tc) = \frac{\alpha}{\lambda(\lambda+t)} + tc$$

since $\mathrm{E}(S_t) = \mathrm{E}(\mathrm{E}(S_t|\theta)) = \mathrm{E}(t\theta) = t\mathrm{E}(\theta) = t\alpha/\lambda$. To find T_1, we find the t at which $\mathrm{E}(Y_t)$ is a minimum. (You may check that this occurs at $t = \{\sqrt{\frac{\alpha}{c\lambda} + \frac{1}{2}} - \lambda\}$, where $\{x\}$ represents the integer closest to x.) Clearly, for some α, c and λ, this can be greater than one. So is T_1 different from N_1?

Surprisingly, the answer is no. Here is why. As noted before, if N_1 says continue, then T_1 also says continue. Suppose N_1 says stop. Then $\alpha \le c\lambda^2(\lambda+1)$ from (4). This implies that for $t \ge 1$, $t\alpha \le tc\lambda^2(\lambda+t)$, or equivalently,

$$\frac{\alpha}{\lambda^2} \le \frac{\alpha}{\lambda(\lambda+t)} + tc.$$

Thus T_1 calls for stopping also.

It is somewhat disappointing that the one-time look-ahead rule does not improve upon the one-stage look-ahead rule for this example. Moreover, this holds true for rather general distributions. It holds whenever $\mathrm{E}(\mathrm{Var}\,(\theta|X_1, \ldots, X_n))$ is a convex function of n.

Therefore, to get an improvement we must look at the k-time look-ahead rule for $k > 1$. This rule may be described as follows. At stage n, choose k fixed times, $n < t_1 < \cdots < t_k$, and consider the best sequential rule among those that stop only at these times. If there exists a set of $t_1, \ldots, t_k$ for which this rule gives smaller expected loss than stopping at n then continue; otherwise stop.

Consider the two-time rule (called simply *the two-timer*) at stage $n = 0$. We choose two times, $0 < t_1 < t_2$. Then we consider the sequential rule that looks first at $X_1, \ldots, X_{t_1}$ and decides whether to stop or to continue to t_2 and stop. If we stop, we pay Y_{t_1}, and if we continue we expect to pay $\mathrm{E}(Y_{t_2}|\mathcal{F}_{t_1})$. We stop at t_1 if the former is less than the latter. The expected loss using such a rule is $\mathrm{E}(\min\{Y_{t_1}, \mathrm{E}(Y_{t_2}|\mathcal{F}_{t_1})\})$. Therefore, the two-timer stops without taking any observations if

$$Y_0 \le \mathrm{E}(\min\{Y_{t_1}, \mathrm{E}(Y_{t_2}|\mathcal{F}_{t_1})\}) \qquad \text{for all } 0 < t_1 < t_2. \tag{9}$$

Let us compute the expectation on the right side of (9) in our example. From

$$\mathrm{E}(Y_{t_2}|\mathcal{F}_{t_1}) = \frac{\alpha + S_{t_1}}{(\lambda+t_1)(\lambda+t_2)} + t_2 c,$$

we find

$$\begin{aligned}\min\{Y_{t_1}, \mathrm{E}(Y_{t_2}|\mathcal{F}_{t_1})\} &= \min\{\frac{\alpha+S_{t_1}}{(\lambda+t_1)^2}+t_1c, \frac{\alpha+S_{t_1}}{(\lambda+t_1)(\lambda+t_2)}+t_2c\} \\ &= \frac{\alpha+S_{t_1}}{(\lambda+t_1)^2}+t_1c+\min\{0, -\frac{(t_2-t_1)(\alpha+S_{t_1})}{(\lambda+t_1)^2(\lambda+t_2)}+(t_2-t_1)c\} \\ &= \frac{\alpha+S_{t_1}}{(\lambda+t_1)^2}+t_1c-(t_2-t_1)\left(\frac{\alpha+S_{t_1}}{(\lambda+t_1)^2(\lambda+t_2)}-c\right)^+,\end{aligned}$$

from which we may compute the expectation in (9) as

$$\mathrm{E}(\min\{Y_{t_1}, \mathrm{E}(Y_{t_2}|\mathcal{F}_{t_1})\}) = \frac{\alpha}{\lambda(\lambda+t_1)}+t_1c-(t_2-t_1)\mathrm{E}\left(\frac{\alpha+S_{t_1}}{(\lambda+t_1)^2(\lambda+t_2)}-c\right)^+.$$

Thus the two-timer stops without taking observations if for all $0<t_1<t_2$, $Y_0=\alpha/\lambda^2$ is less than or equal to this, that is, if

$$\frac{\alpha t_1}{\lambda^2(\lambda+t_1)} \le t_1c-(t_2-t_1)\mathrm{E}\left(\frac{\alpha+S_{t_1}}{(\lambda+t_1)^2(\lambda+t_2)}-c\right)^+$$

or, equivalently,

$$\frac{\alpha}{\lambda^2} \le c(\lambda+t_1)-\frac{(t_2-t_1)}{t_1(\lambda+t_1)(\lambda+t_2)}\mathrm{E}\left(\alpha+S_{t_1}-(\lambda+t_1)^2(\lambda+t_2)c\right)^+. \qquad (10)$$

for all $0<t_1<t_2$. To compute the expectation on the right, note that the marginal distribution of S_{t_1} is negative binomial:

$$\begin{aligned}\mathrm{P}(S_t=x) &= \mathrm{E}(\mathrm{P}(S_t=x|\theta)) = \mathrm{E}(\frac{1}{x!}e^{-t\theta}(t\theta)^x) \\ &= \frac{t^x}{x!}\frac{\lambda^\alpha}{\Gamma(\alpha)}\int_0^\infty e^{-(\lambda+t)\theta}\theta^{\alpha+x-1}\,d\theta \\ &= \frac{t^x}{x!}\frac{\lambda^\alpha}{\Gamma(\alpha)}\frac{\Gamma(\alpha+x)}{(\lambda+t)^{\alpha+x}}\end{aligned}$$

for $x=0,1,2,\ldots$.

For $\alpha=1$, this is the geometric distribution, $\mathcal{G}(t/(\lambda+t))$. If $X\in\mathcal{G}(p)$, it is easy to compute $\mathrm{E}(X-k)^+$ for $k>-1$ as follows.

$$\begin{aligned}\mathrm{E}(X-k)^+ &= \sum_{x=\lceil k\rceil}^\infty (x-k)p^x(1-p) = p^{\lceil k\rceil}\sum_{x=\lceil k\rceil}^\infty (x-k)p^{x-\lceil k\rceil}(1-p) \\ &= p^{\lceil k\rceil}\sum_{y=0}^\infty (y+\lceil k\rceil-k)p^y(1-p) = p^{\lceil k\rceil}(\frac{p}{1-p}+\lceil k\rceil-k).\end{aligned} \qquad (11)$$

Now assume $\alpha = 1$ in (10) and put $p = t_1/(\lambda + t_1)$ and $k = (\lambda + t_1)^2(\lambda + t_2)c - 1$ into (11). We find that the two-timer calls for stopping at stage 0 if for all $0 < t_1 < t_2$,

$$\frac{1}{\lambda^2} \le c(\lambda + t_1) - \frac{(t_2 - t_1)}{t_1(\lambda + t_1)(\lambda + t_2)} \left(\frac{t_1}{\lambda + t_1}\right)^{\lceil k \rceil} (\frac{t_1}{\lambda} + \lceil k \rceil - k).$$

Now suppose λ is large, say $\lambda = 100$, and c is small, say $c = .000001$. Then $1/(\lambda^2(\lambda + 1)) < c$, so the 1-sla and the one-timer call for stopping. Then for small values of t_1 and t_2 such that $\lceil k \rceil = 1$ (i.e., $(1 + .01t_1)^2(1 + .01t_2) < 2$), the above inequality becomes

$$\begin{aligned} \frac{1}{\lambda^2} &\le c(\lambda + t_1) - \frac{(t_2 - t_1)}{(\lambda + t_1)^2(\lambda + t_2)}(\frac{t_1}{\lambda} + 1 - k) \\ &= c(\lambda + t_2) - \frac{(t_2 - t_1)}{(\lambda + t_1)^2(\lambda + t_2)}(\frac{t_1}{\lambda} + 2). \end{aligned}$$

For fixed t_2, the right side is increasing in t_1, and so the inequality is sharpest when t_1 is smallest, namely $t_1 = 1$. At $t_1 = 1$, the inequality becomes

$$1 \le c(\lambda + t_2)\lambda^2 - \frac{(t_2 - 1)\lambda}{(\lambda + 1)^2(\lambda + t_2)}(1 + 2\lambda).$$

At $t_2 = 2$, this inequality is $1 \le 1.00068$, so the 2-sla also calls for stopping. Yet the right side of this inequality decreases in t_2 until $t_2 = 41$ (value there $= .851023$), and then increases. Therefore the two-timer calls for taking the first observation.

In this example, the two-timer provides an improvement over the 2-stage look-ahead rule with no real increase in the cost of computation. But in continuous-time problems the k-time look-ahead rule plays a more fundamental role. In such problems, stopping can occur at any real time, not just at integer times; stopping does not occur in stages so the k-stage look-ahead approximations to the optimal rule are not available. The analog of the one-stage look-ahead rule in continuous-time problems is the infinitesimal look-ahead rule of Ross (1971). (See Example 5 in Section 5.4.) Although analogs to the k-stage look-ahead rules are not meaningful, the analogs to the k-time look-ahead rules exist and generally provide improvements over the infinitesimal look-ahead rule. The two-timer will generally look infinitesimally ahead as well as at a second time a positive amount into the future. See Biesterfeld (1996) for an example.

§**5.2 Monotone Stopping Rule Problems.** We seek conditions under which it is optimal to stop when the 1-sla calls for stopping. The basic condition is that the problem be monotone, a notion due to Chow and Robbins (1961). In finite horizon monotone stopping rule problems, the 1-sla is optimal.

Definition. *Let A_n denote the event $\{Y_n \ge \mathrm{E}(Y_{n+1}|\mathcal{F}_n)\}$. We say that the stopping rule problem is* **monotone** *if*

$$A_0 \subset A_1 \subset A_2 \subset \ldots \quad a.s. \tag{12}$$

Monotone problems may be described as follows. The set A_n is the set on which the 1-sla calls for stopping at stage n (given that stage is reached). The condition $A_n \subset A_{n+1}$ means that if the 1-sla calls for stopping at stage n, then it will also call for stopping at stage $n+1$ no matter what X_{n+1} happens to be (a.s.). Similarly, $A_n \subset A_{n+1} \subset A_{n+2} \subset \ldots$ means that if the 1-sla calls for stopping at stage n, then it will call for stopping at all future stages no matter what the future observations turn out to be (a.s.).

Theorem 1. *In a finite horizon monotone stopping rule problem, the one-stage look-ahead rule is optimal.*

Proof. Suppose the horizon is J. One optimal rule is

$$N^* = \min\{n \geq 0 : Y_n \geq \mathrm{E}(V_{n+1}^{(J)}|\mathcal{F}_n)\}$$

where $V_{J+1}^{(J)} = -\infty, V_J^{(J)} = Y_J$, and by backward induction,

$$V_n^{(J)} = \max\{Y_n, \mathrm{E}(V_{n+1}^{(J)}|\mathcal{F}_n)\} \quad \text{for} \quad n = 0, 1, \ldots, J-1.$$

Fix $n < J$. If N_1 calls for continuing at n, then, since $V_{n+1}^{(J)} \geq Y_{n+1}$ a.s., N^* calls for continuing at n also. Suppose N_1 calls for stopping at n, that is, suppose A_n holds. Then, since the problem is monotone $A_{n+1}, \ldots, A_{J-1}$ also hold. Thus

$$\begin{aligned}
Y_{J-1} &\geq \mathrm{E}(Y_J|\mathcal{F}_{J-1}) = \mathrm{E}(V_J^{(J)}|\mathcal{F}_{J-1}). \quad \text{Hence} \quad V_{J-1}^{(J)} = Y_{J-1}.\\
Y_{J-2} &\geq \mathrm{E}(Y_{J-1}|\mathcal{F}_{J-2}) = \mathrm{E}(V_{J-1}^{(J)}|\mathcal{F}_{J-2}). \quad \text{Hence} \quad V_{J-2}^{(J)} = Y_{J-2}.\\
&\vdots\\
Y_n &\geq \mathrm{E}(Y_{n+1}|\mathcal{F}_n) = \mathrm{E}(V_{n+1}^{(J)}|\mathcal{F}_n). \quad \text{Hence} \quad V_n^{(J)} = Y_n.
\end{aligned}$$

Thus, N^* also calls for stopping. ■

In particular, for a monotone stopping rule problem, the k-sla is no better than the 1-sla for any $k > 1$.

Corollary 1. *For a monotone stopping rule problem, for all $k > 1$, the k-stage look-ahead rule is equivalent to the 1-stage look-ahead rule.*

There are various problems associated with the extension of Theorem 1 to the infinite horizon case. Consider Examples 1 and 2 of Chapter 3. In both of these examples the problem is monotone because the 1-sla always tells you to continue. Yet the 1-sla is not optimal; in fact it is the worst of all stopping rules since it has you continue forever and receive nothing. Even if we assume A1 and A2, the 1-sla still might not be optimal for a monotone problem as the following two counterexamples show.

Counterexample 1. If $Y_n = 1/(n+1)$ for $n = 0, 1, \ldots$, and $Y_\infty = 2$, then the unique optimal stopping rule is $N = \infty$ with return $V = 2$. The 1-sla always calls for stopping so the problem is monotone. But stopping at stage 0 has return $V = 1$ so the 1-sla is not optimal. ■

Clearly, it is suboptimal to stop if continuing forever gives a greater expected return, that is, if the hypermetropic rule calls for continuing. In other words, any stopping rule is improved by replacing any decision to stop by the decision to continue forever if that gives a greater expected payoff. Hence, we may replace Y_n by $\max\{Y_n, \mathrm{E}(Y_\infty|\mathcal{F}_n)\}$ without changing the problem; this merely rules out some suboptimal stopping rules. Thus, we may assume

$$Y_n \geq \mathrm{E}(Y_\infty|\mathcal{F}_n) \quad \text{a.s.}$$

without loss of generality. Given that A1 and A2 hold, we then would have the following strengthened form of A2,

$$\text{A3: } \lim_{n\to\infty} Y_n = Y_\infty \text{ a.s.} \tag{13}$$

since by the martingale convergence theorem $\mathrm{E}(Y_\infty|\mathcal{F}_n) \to \mathrm{E}(Y_\infty|\mathcal{F}_\infty) = Y_\infty$ a.s. (See, for example, Chow Robbins and Siegmund (1971), p. 18.) We have assumed that Y_∞ is $\mathcal{F}_\infty$-measurable; that is, we have replaced Y_∞ by its expectation given $\mathcal{F}_\infty$.

Counterexample 2. Let K have a geometric distribution $\mathrm{P}(K = k) = 1/2^k$ for $k = 1, 2, \ldots$ and define $X_n = I(n \neq K)$ for $n = 1, 2, \ldots$. For a fixed ϵ, $0 < \epsilon < 1$, let $Y_0 = -1 + \epsilon$, $Y_n = (-2^n + \epsilon)X_n$ for $n = 1, 2, \ldots$ and $Y_\infty = -\infty$. Then A1 and A3 are satisfied and the optimal rule is obviously $N = \min\{n \geq 0 : X_n = 0\}$ having return 0. However, at stage n with $K > n$, if we continue one stage and stop, we expect $(1/2)0 + (1/2)(-2^{n+1} + \epsilon) = -2^n + \epsilon/2$ compared with $-2^n + \epsilon$ for stopping immediately. Hence, the 1-sla always calls for stopping so the problem is monotone. The 1-sla stops without taking any observations and has the return $-1 + \epsilon$. ■

In spite of these counterexamples, it is usually true that the 1-sla is optimal for monotone infinite horizon problems. What is needed is a condition to ensure that the infinite horizon problem can be approximated well by finite horizon problems in the sense that $V_0^{(J)} \to V_0^{(\infty)}$ as $J \to \infty$, where $V_0^{(J)}$ denotes the optimal return for the problem truncated at J, and $V_0^{(\infty)}$ denotes V^*, the optimal return for the infinite horizon problem. The following theorem states this formally. A similar approach is used in Bayes sequential statistical problems. (See, for example, Theorem 7.2.5 in Ferguson (1967).)

Theorem 2. *Suppose A1 and A2 are satisfied and suppose the problem is monotone. If $V_0^{(J)} \to V_0^{(\infty)}$ as $J \to \infty$, then the one-stage look-ahead rule is optimal.*

Proof. Let N^* denote the 1-sla and let N_j be the 1-sla truncated at j, $N_j = \min\{N^*, j\}$. Then N_j is the 1-sla for the problem truncated at j and so by Theorem 1, N_j is optimal for the problem truncated at j, so that $\mathrm{E}Y_{N_j} = V_0^{(j)}$. Note that N_j is an increasing sequence of stopping rules converging to N^*. Thus, as in the proof of Theorem 3.1,

$$V_0^{(\infty)} = \lim \mathrm{E}Y_{N_j} \leq \mathrm{E}\limsup Y_{N_j} \leq \mathrm{E}Y_{N^*},$$

showing that N^* is optimal. ■

§5.3 Approximation of the Infinite Problem by Finite Horizon Problems. This brings up the problem of the approximation of optimal rules by truncated rules. Counterexample 2 indicates that we need some sort of lower bounds for the Y_n. Under the extra condition that $T_n = \sup_{j\ge n}(Y_j - Y_n)$ be uniformly integrable, the infinite horizon problem can be approximated by the finite horizon problems.

Definition. *A set of random variables* $\{T_n\}$ *is said to be* **uniformly integrable** *(u.i.) if*

$$\sup_n \mathrm{E}\{|T_n|I(|T_n|>a)\} \to 0 \quad \textit{as} \quad a\to\infty. \tag{14}$$

Note 1. If $\mathrm{E}(|T_n|)\to 0$ as $n\to\infty$, then T_n is u.i.
(*Proof.* Let $\epsilon>0$. Find N such that $n>N$ implies $\mathrm{E}(|T_n|)<\epsilon$. For each n, find a_n such that $\mathrm{E}\{|T_n|I(|T_n|>a_n)\}<\epsilon$. Let $A=\max_{1\le n\le N} a_n$. Then $a>A$ implies that $\mathrm{E}\{|T_n|I(|T_n|>a)\}<\epsilon$, for all n, so that $\sup_n \mathrm{E}\{|T_n|I(|T_n|>a)\}\le\epsilon$.)

Note 2. If $\limsup \mathrm{E}(|T_n|)=\infty$, then T_n is not u.i.
(*Proof.* Fix a. $\mathrm{E}(|T_n|)=\mathrm{E}\{|T_n|I(|T_n|\le a)\}+\mathrm{E}\{|T_n|I(|T_n|>a)\}\le a+\mathrm{E}\{|T_n|I(|T_n|>a)\}$, so $\sup_n \mathrm{E}\{|T_n|I(|T_n|>a)\}\ge\sup_n \mathrm{E}(|T_n|)-a=\infty$.)

Note 3. If $\mathrm{E}(|T_n|)$ stays bounded away from zero and infinity, then T_n may or may not be u.i. For example, if T_n is i.i.d. with $\mathrm{E}(|T_n|)=1$, then T_n is u.i., but if $T_n=n$ w.p. $1/n$ and $T_n=0$ otherwise, then $\mathrm{E}(|T_n|)=1$ for all n, yet T_n is not u.i.

Lemma 1. *If* $\{T_n\}$ *is uniformly integrable and if* A_n *are any events such that* $\mathrm{P}(A_n)\to 0$ *as* $n\to\infty$, *then* $\mathrm{E}I(A_n)|T_n|\to 0$ *as* $n\to\infty$.

Proof.

$$\begin{aligned}\mathrm{E}I(A_n)|T_n| &= \mathrm{E}I(A_n)I(|T_n|\le a)|T_n| + \mathrm{E}I(A_n)I(|T_n|>a)|T_n|\\ &\le a\mathrm{P}(A_n)+\mathrm{E}I(|T_n|>a)|T_n|.\end{aligned}$$

From uniform integrability, for fixed $\epsilon>0$ there exists an a such that the second term is less than ϵ for all n. Now let $n\to\infty$ and the first term tends to zero. ■

Theorem 3. *Assume A1 and A3 and let* $T_n=\sup_{j\ge n}\{Y_j-Y_n\}$. *If the* T_n *are uniformly integrable, then* $V_0^{(n)}\to V_0^{(\infty)}$ *as* $n\to\infty$.

Proof. Let N denote an optimal stopping rule, and let $N(n)=\min\{N,n\}$ for $n=1,2,\ldots$. Then,

$$\begin{aligned}0\le V_0^{(\infty)}-V_0^{(n)} &\le \mathrm{E}Y_N-\mathrm{E}Y_{N(n)}\\ &=\mathrm{E}I(n<N<\infty)(Y_N-Y_n)+\mathrm{E}I(N=\infty)(Y_\infty-Y_n)\\ &\le \mathrm{E}I(n<N<\infty)T_n+\mathrm{E}(Y_\infty-Y_n)^+.\end{aligned}$$

The first term tends to zero by Lemma 1. If we let ϵ be an arbitrary positive number, the second term may be written as

$$\begin{aligned}&\mathrm{E}I((Y_\infty-Y_n)\le\epsilon)(Y_\infty-Y_n)^+ + \mathrm{E}I((Y_\infty-Y_n)>\epsilon)(Y_\infty-Y_n)^+\\ &\qquad\le \epsilon+\mathrm{E}I((Y_\infty-Y_n)>\epsilon)T_n.\end{aligned}$$

Since $\mathrm{P}((Y_\infty - Y_n)^+ > \epsilon) \to 0$ as $n \to \infty$ by A3, the final expectation tends to zero from Lemma 1, so the whole can be made less than 2ϵ for n sufficiently large. ∎

The proof of this theorem shows more. Namely, under the conditions of that theorem, we have $\mathrm{E}Y_{N(n)} \to V_0^{(\infty)}$, i.e., one loses very little in truncating an optimal rule at sufficiently large n. In addition, it follows from this theorem that $\mathrm{E}Y_{N_k} \to V_0^{(\infty)}$, where N_k is the k-stage look-ahead rule.

The following corollary gives a simple sufficient condition for the uniform integrability of $\{T_n\}$ that is easily checked in many cases in which the payoff is a reward depending on the observations minus a constant cost per observation.

Corollary 2. *Assume A3 and suppose that $Y_n = Z_n - W_n$ where $\mathrm{E}\sup_n |Z_n| < \infty$ and W_n is nonnegative and nondecreasing a.s. Then, A1 holds and $V_0^{(J)} \to V_0^{(\infty)}$ as $J \to \infty$.*

Proof. A1 is satisfied since $Y_n \le Z_n$ is bounded above by an integrable function. Moreover, for $j > n$, $Y_j - Y_n \le Z_j - Z_n$ so that $0 \le T_n = \sup_{j \ge n}(Y_j - Y_n) \le 2\sup_n |Z_n| = Z'$, say, which has finite expectation. This implies uniform integrability since for all n $\mathrm{E}I(|T_n| > a)|T_n| \le \mathrm{E}I(|Z'| > a)Z' \to 0$ as $a \to \infty$. ∎

In particular, putting $W_n \equiv 0$, we see that the strengthened form of A1, $\mathrm{E}(\sup_n |Y_n|) < \infty$, together with A3 implies that $V_0^{(J)} \to V_0^{(\infty)}$. It is worth remarking that for a cost problem, where we are trying to minimize $\mathrm{E}Y_N$, the only change required in Corollary 2 is to replace $Y_n = Z_n - W_n$ by $Y_n = Z_n + W_n$.

Consider Counterexample 2. One finds that $T_n = (2^n - \epsilon)I(K > n)$. The T_n cannot be u.i. since otherwise the 1-sla would be optimal. Indeed, one can check directly that the T_n are not u.i. For fixed a, just find n such that $2^n - \epsilon > a$; then $\mathrm{E}|T_n|I(|T_n| > a) = (2^n - \epsilon)\mathrm{P}(K > n) = 1 - (\epsilon/2^n)$.

If ϵ in Counterexample 2 were negative, then the 1-sla continues until $n = K$ and so is optimal. However, its optimality is not implied by the theorems of this section. First of all, the problem is not monotone since if you continue beyond $n = K$, the 1-sla will again have you continue. But even if we modify the problem to make it monotone by changing X_n to be $I(n < K)$, the $\{T_n\}$ are still as in the previous paragraph and so are not u.i. This provides an example of a monotone problem in which the 1-sla is optimal even though $V_0^{(J)}$ does not converge to $V_0^{(\infty)}$; that is, it provides a counterexample to the converse of Theorem 2.

§**5.4 Examples.** 1. THE BURGLAR PROBLEM. Suppose that the returns from the burglaries are i.i.d. non-negative random variables, $X_1, X_2, \ldots$ with known distribution having finite mean, μ. Let $Z_1, Z_2, \ldots$ be the random variables that indicate whether the burglar is caught, with $Z_n = 1$ denoting that the nth burglary is successful and $Z_n = 0$ indicating that he is caught during the nth burglary, and assume that the Z_n are i.i.d. with $\mathrm{P}(Z_n = 1) = \beta$ and $\mathrm{P}(Z_n = 0) = 1 - \beta$ where β is known and $0 < \beta < 1$. The payoff for stopping after the nth burglary is

$$Y_n = (\prod_1^n Z_i) \sum_1^n X_i,$$

for $n = 0, 1, \ldots$, and we take $Y_\infty = 0$ so that A3 is satisfied. Letting $\mathcal{F}_n$ denote the σ-field generated by both $X_1, \ldots, X_n$ and $Z_1, \ldots, Z_n$, we compute on $\{\prod_1^n Z_i = 1\}$,

$$\mathrm{E}(Y_{n+1}|\mathcal{F}_n) = \mathrm{E}(Z_{n+1} \sum_1^{n+1} X_i|\mathcal{F}_n) = \beta(\sum_1^n X_i + \mu).$$

Hence, the one-stage look-ahead rule is

$$\begin{aligned} N_1 &= \min\{n \geq 0 : \sum_1^n X_i \geq \beta(\sum_1^n X_i + \mu)\} \\ &= \min\{n \geq 0 : \sum_1^n X_i \geq \beta\mu/(1-\beta)\}, \end{aligned}$$

that is, stop at the first n at which your accumulated gain is at least $\beta\mu/(1-\beta)$. Because the X_i are assumed to be non-negative, one sees that the problem is monotone. To check optimality, we note that the Y_n are bounded below by 0 and above by $\sup_n Y_n = \sum_1^M X_i$, where M is the number of successful burglaries, $M = \max\{n \geq 0 : \prod_1^n Z_i = 1\}$. Since $\mathrm{E}(\sum_1^M X_i) = \mu\mathrm{E}(M) = \mu/(1-\beta)$, the conditions of Corollary 2, with $W_n \equiv 0$, are satisfied and so N_1 is optimal.

This problem is just a version of the discounted stopping of a sum problem of Dubins and Teicher where the returns are nonnegative. Instead of a discount β, there is a probability β that your fortune is set forever to zero. Since you have no control over the Z_n and they are independent of everything else in the problem, you may as well replace them by their expected values. When you do, you find that $Y_n = \beta^n \sum_1^n X_i$, as in the problem of Dubins and Teicher.

In §4.2, this problem was seen to have an optimal rule of the simple form, $N = \min\{n > 0 : S_n \geq s_0\}$ for some s_0, whether or not the X_i are assumed to be non-negative. If the X_i may assume negative values, then the problem is not monotone and the methods of the present chapter are not capable of finding the solution directly. However, if the X_i are non-negative, then these methods may sometimes be used to solve more general problems where the X_i are dependent. (See Exercises 2 and 3.)

2. SELLING AN ASSET. Let $X_1, X_2, \ldots$ (with X_n representing the n-th offer) be i.i.d. with known distribution F having finite variance, and consider the problem of selling an asset with constant cost $c > 0$ and with recall so that $Y_n = M_n - nc$, where $M_n = \max\{X_1, \ldots, X_n\}$, $Y_0 = -\infty$ and $Y_\infty = -\infty$. Since

$$\begin{aligned} \mathrm{E}(Y_{n+1}|\mathcal{F}_n) - Y_n &= \mathrm{E}(\max\{M_n, X_{n+1}\}|\mathcal{F}_n) - M_n - c \\ &= \mathrm{E}((X_{n+1} - M_n)^+|\mathcal{F}_n) - c \\ &= \int (x - M_n)^+ \, dF(x) - c, \end{aligned}$$

the 1-sla may be written

$$N_1 = \min\{n \geq 1 : \int (x - M_n)^+ dF(x) \leq c\}$$

Since $\int (x-v)^+ \, dF(x)$ is a nonincreasing function of v, this reduces to $N_1 = \min\{n \geq 1 : M_n \geq v\}$, where v satisfies $\int (x-v)^+ \, dF(x) = c$, the stopping rule already seen to be optimal in §4.1. Let us check optimality using the theorems of this section. The problem is clearly monotone. Note that Corollary 2 does not apply, since $\sup_n M_n = \infty$ if the distribution of X is unbounded. We check the condition of Theorem 3. Write

$$T_n = \sup_{j \geq n}(Y_j - Y_n) = \sup_{j \geq n}(M_j - M_n - (j-n)c)$$

and note that conditional on M_n, the right side of this equality has the same distribution as $\sup_{j>0}(M'_j - jc)$ where M'_j is the maximum of a sample of size j from the distribution of $(X - M_n)^+$. Thus, the $\mathrm{E}(T_n|M_n)$ are a.s. decreasing to zero, and since they have finite mean by Theorem 4.1, $\mathrm{E}(T_n) \to 0$ by monotone convergence, and thus the T_n are uniformly integrable.

We point out that the problem of selling an asset without recall, $Y_n = X_n - nc$, is not monotone. This is because the 1-sla can call for stopping at a very high value of X_n, but if you continue and observe a very low value of X_{n+1}, the 1-sla would certainly call for continuing. The 1-sla is

$$N'_1 = \min\{n \geq 0 : X_n > \mathrm{E}X - c\},$$

quite different from N_1 above. In fact, as noted in Chapter 4, N_1 is optimal for sampling without recall as well, since for all n, $X_n \leq M_n$, and this implies that for all stopping rules, N,

$$\mathrm{E}(X_N - Nc) \leq \mathrm{E}(M_N - Nc) \leq \mathrm{E}(M_{N_1} - N_1 c) = \mathrm{E}(X_{N_1} - N_1 c).$$

3. THE PARKING PROBLEM. The formulation of the parking problem given in Chapter 2 does not produce a monotone stopping rule problem. This is because if the 1-sla calls for stopping at some open parking place before your destination and you continue one step, the next parking place may be filled so the 1-sla would call for continuing. This is somewhat artificial since you are not allowed to stop if the parking place is filled; a different formulation does lead to a monotone problem. We describe this formulation and at the same time give an extension due to Tamaki (1982) to the case where the destination is not known precisely. This problem does not have a finite horizon and so is not amenable to the methods of Chapter 2.

You are at the origin driving to a target destination $T > 0$. The distribution of T is known and has finite mean but the value of T is not known until you reach it. Parking places occur at random along the street at points $Z_1, Z_2, \ldots$ chosen according to a Poisson process independent of T with intensity $\lambda > 0$; that is, $Z_1, Z_2 - Z_1, Z_3 - Z_2, \ldots$ are i.i.d. with an exponential distribution of mean $1/\lambda$. Your loss is the distance you have to walk to your destination,

$$Y_n = |T - Z_n| \quad \text{for} \quad n = 0, 1, 2, \ldots \quad \text{and} \quad Y_\infty = \infty,$$

where Z_0 is defined to be 0 and $Y_0 = |T|$ represents the loss if you walk all the way.

The information available at stage n is the set of values of $Z_1, \ldots, Z_n$ and the information as to whether $Z_n < T$ or $Z_n \geq T$, and in the latter case the exact value of T. On $\{Z_n < T\}$, Y_n may be replaced by

$$\mathrm{E}\{Y_n | \mathcal{F}_n\} = \mathrm{E}\{T | Z_n, Z_n < T\} - Z_n.$$

If $Z_n \geq T$, it is clearly optimal to stop; so let us assume $Z_n < T$ and look at the 1-sla. To compute the 1-sla, we find on $\{Z_n < T\}$

$$\mathrm{E}\{Y_{n+1} | \mathcal{F}_n\} = \mathrm{E}\{|T - Z_n - Z| | Z_n, Z_n < T\},$$

where Z is exponential with mean $1/\lambda$. For arbitrary $u > 0$,

$$\begin{aligned}
\mathrm{E}|u - Z| &= \mathrm{E}(u - Z) + 2\mathrm{E}(Z - u)^+ \\
&= (u - 1/\lambda) + 2P(Z > u)\mathrm{E}\{Z - u | Z > u\} \\
&= (u - 1/\lambda) + 2e^{-\lambda u}(1/\lambda),
\end{aligned}$$

so that on $\{Z_n < T\}$

$$\mathrm{E}\{|T - Z_n - Z| | \mathcal{F}_n\} = \mathrm{E}\{(T - Z_n) - 1/\lambda + (2/\lambda)e^{-\lambda(T - Z_n)} | \mathcal{F}_n\}.$$

Comparing with Y_n and letting $g(x) = \mathrm{E}\{e^{-\lambda(T-x)} | T > x\}$, we see that the 1-sla is

$$N_1 = \min\{n \geq 0 : g(Z_n) \geq 1/2\}.$$

It is evident from this that the problem is monotone if $g(x)$ is monotonically nondecreasing, since the Z_n are a.s. increasing. To check the conditions of Corollary 2, modified to reflect the fact that this is a minimum problem, write $Y_n = Z'_n + W'_n$ where $Z'_n = (T - Z_n)^+$ is bounded below by zero and above by T and $W'_n = (Z_n - T)^+$ is non-decreasing a.s. Thus, the conditions of Corollary 2 are satisfied, so that the 1-sla is optimal if $g(x)$ is nondecreasing.

Consider the case where T has an exponential distribution. The lack of memory property of the exponential distribution implies that $g(x)$ is a constant, and hence nondecreasing. Thus, the exponential distribution is the borderline case. For distributions with thinner tails than the exponential, i.e. distributions with increasing failure rate, $g(x)$ will be nondecreasing; for distributions with decreasing failure rates, $g(x)$ will be nonincreasing.

4. PROOFREADING. Let us find the 1-stage look-ahead rule for Example 1.4, the proofreading problem. For this problem, the number of misprints M and the numbers of misprints detected on subsequent consecutive proofreadings $X_1, X_2, \ldots$ have a known joint distribution such that $X_j \geq 0$ and $\sum X_j \leq M$ a.s. and $\mathrm{E}M < \infty$. The cost for stopping after n proofreadings is

$$Y_n = nc_1 + (M - \sum_1^n X_j)c_2 \quad \text{and} \quad Y_\infty = \infty,$$

where $c_1 > 0$ is the cost of each proofreading, and $c_2 > 0$ is the cost of each undetected misprint.

If you stop at stage n you expect to lose

$$\mathrm{E}\{Y_n | X_1, \ldots, X_n\} = nc_1 + [\mathrm{E}\{M | X_1, \ldots, X_n\} - \sum_1^n X_j]c_2.$$

If you continue one stage and stop, you expect to lose

$$\mathrm{E}\{Y_{n+1} | X_1, \ldots, X_n\} = (n+1)c_1 + [\mathrm{E}\{M | X_1, \ldots, X_n\} - \sum_1^n X_j - \mathrm{E}\{X_{n+1} | X_1, \ldots, X_n\}]c_2.$$

The 1-sla calls for stopping if the former is not greater than the latter,

$$N_1 = \min(n \geq 0 : \mathrm{E}\{X_{n+1} | X_1, \ldots, X_n\} \leq c_1/c_2).$$

Clearly, the problem is monotone if and only if $\mathrm{E}\{X_{n+1} | X_1, \ldots, X_n\}$ is nonincreasing a.s. Assuming the problem is monotone let us check Corollary 2 for optimality of the 1-sla. Y_∞ was chosen so that A3 is satisfied. Moreover, $Y_n = Z_n + W_n$, where $W_n = nc_1$ is nondecreasing and $Z_n = (M - \sum_1^n X_j)c_2$ is bounded in absolute value by Mc_2 which has finite expectation by assumption. Hence, the 1-sla is optimal if the problem is monotone.

One case which leads to a monotone problem is mentioned in Yang, Wackerly and Rosalsky (1982). (See also the correction in Chow and Schechner (1985).) Let M have a Poisson distribution with known mean $\lambda > 0$ and for $n = 0, 1, \ldots$ let X_{n+1} have the binomial distribution with sample size $M - \sum_1^n X_j$ and known success probability p, $0 < p < 1$. Let $M_n = M - \sum_1^n X_j$ denote the number of misprints remaining after n proofreadings. Then the posterior distribution of M_n given $X_1, \ldots, X_n$ is Poisson with mean $\lambda(1-p)^n$ independent of $X_1, \ldots, X_n$ as is easily checked. Since

$$\begin{aligned} \mathrm{E}\{X_{n+1} | X_1, \ldots, X_n\} &= \mathrm{E}\{M_n p | X_1, \ldots, X_n\} \\ &= \lambda(1-p)^n p \end{aligned}$$

is a decreasing function of n, the problem is monotone. The corresponding optimal rule,

$$N_1 = \min\{n \geq 0 : \lambda p (1-p)^n \leq c_1/c_2\},$$

is a fixed sample size rule, i.e. a rule that stops at a fixed predetermined number of observations.

5. FISHING. (Starr, Wardrop and Woodroofe (1976)) Consider a lake with n fish and let $T_1, \ldots, T_n$ denote the capture times, assumed to be i.i.d. according to a given distribution function $F(t)$ on $(0, t)$. For $t > 0$, let $X(t)$ denote the number of fish caught by time t, i.e. $X(t)$ is the number of T_j less than or equal to t. The payoff if you stop at time t is $Y(t) = X(t) - ct$, where $c > 0$ is the cost of time. If stopping is allowed at all times, it may be optimal to stop between catches. However, as mentioned in Exercise 2 of Chapter 1, Starr and Woodroofe (1974) have shown that if F has increasing failure rate, then there is an optimal rule that stops only at catch times. In spite of this, the easy case turns out to be the case of decreasing failure rate! If F has decreasing failure rate, then it may be optimal to stop between catch times, but there is an analogue of the 1-sla for continuous time problems that is easy to compute and is optimal for this problem, namely, the infinitesimal look-ahead rule. The theory for this rule is developed by Ross (1971) based on the theory of Markov processes. Here, we find an approximation by discretizing the time axis, finding the ordinary 1-sla and passing to the limit.

Let $F(t)$ be an absolutely continuous distribution function with density $f(t)$ on the interval $(0, \infty)$. The *failure rate* (or *hazard rate*) of F at t is defined to be $h(t) = f(t)/(1 - F(t))$, and may be interpreted as the instantaneous death rate of those individuals who have reached age t. Conversely, from the failure rate, $h(t)$, one can obtain the distribution function by the formula, $1 - F(t) = \exp\{-\int_0^t h(s)\,ds\}$. The distribution with a constant failure rate equal to λ on $(0, \infty)$ is the exponential distribution, $F(t) = 1 - \exp\{-\lambda t\}$.

We say that F has *decreasing failure rate* (DFR) if the failure rate, $h(t)$, is nonincreasing, and *increasing failure rate* (IFR) if $h(t)$ is nondecreasing. Monotone failure rate distributions may be characterized in terms of the stochastic order of the residual lifetimes at t. Specifically, if F has DFR (resp. IFR), then the distribution of the residual life time at t,

$$P(T < t + \epsilon | T > t) = 1 - \exp\{-\int_t^{t+\epsilon} h(s)\,ds\},$$

is a nonincreasing (resp. nondecreasing) function of t, for every $\epsilon > 0$.

Let ϵ be a small fixed positive number, and consider the return of stopping at time t, $Y(t) = X(t) - tc$, compared to the conditional expected return of continuing to time $t + \epsilon$ and stopping, namely

$$\mathrm{E}(Y(t+\epsilon)|\mathcal{F}_t) = X(t) + (n - X(t))\mathrm{P}(T < t + \epsilon | T > t) - (t + \epsilon)c.$$

The former is greater than or equal to the latter if and only if

$$(n - X(t))P(T < t + \epsilon | T > t) \le c\epsilon.$$

The first factor on the left, $n - X(t)$, is nonnegative and nonincreasing a.s., and the second is nonincreasing in t, provided F has DFR. Thus, in the DFR case, once this inequality becomes satisfied at some time t, it stays satisfied at all future times. This is a version of the monotonicity property.

Suppose that stopping is allowed only at times $t = k\epsilon$, for $k = 0, 1, 2, \ldots$. The 1-sla would be

$$N_1 = \min\{k\epsilon \geq 0 : (n - X(k\epsilon))\mathrm{P}(T < (k+1)\epsilon | T > k\epsilon) \leq c\epsilon\}.$$

As argued above, if the distribution of T has DFR, then the problem is monotone, and so from Corollary 2, with $Z_k = X(k\epsilon)$ bounded and $W_k = k\epsilon c$ nondecreasing, the 1-sla is optimal. This is true for all $\epsilon > 0$.

Now, noting that $\mathrm{P}(T < t+\epsilon | T > t)/\epsilon \to h(t)$ as $\epsilon \to 0$, we find as an approximation to the 1-sla when ϵ is small,

$$N^* = \min\{t \geq 0 : (n - X(t))h(t) \leq c\}.$$

This is the infinitesimal look-ahead rule. This problem is treated in Starr, Wardrop and Woodroofe (1976), where the payoff is extended to be of the form $Y(t) = g(X(t)) - c(t)$, where $g(k+1) - g(k)$ is nonincreasing in k, and c is a convex function. The corresponding optimal rule is $N^* = \min\{t \geq 0 : (n - X(t))(g(X(t)+1) - g(X(t)))h(t) \leq c'(t)\}$.

As a simple numerical example, consider the Pareto distribution, $F(t) = 1 - \lambda/t$, with decreasing hazard function, $h(t) = \lambda/t$. It is optimal to stop as soon as the number of fish left, $n - X(t)$, is less than ct/λ. For exponential distributions, $F(x) = 1 - \exp\{-\lambda x\}$, the only distributions that have both DFR and IFR, $h(t)$ is the constant λ so that the optimal rule is to stop as soon as $n - X(t) \leq c/\lambda$. This is a fixed sample size rule; we stop as soon as we catch $n - c/\lambda$ fish. For the Rayleigh distribution, $F(t) = 1 - \exp\{-t^2/2\}$, with increasing hazard function, $h(t) = t$, the problem becomes harder because the 1-sla is not optimal. But since we will only stop at catch times, the problem has a finite horizon and can be solved in principle by the method of backward induction of Chapter 2.

A related model, allowing an unknown number of fish of differing sizes with the catch time of a fish, T, being dependent on its size, Z, is due to Cozzolino (1972). The number of fish, M, has a prior Poisson distribution, $\mathrm{P}(\lambda)$, for a known $\lambda > 0$. Given $M = m$, the catch times and sizes of the fish, $(T_1, Z_1), \ldots, (T_m, Z_m)$, are i.i.d. with common known distribution function, $F(t, z)$, independent of m. (Cozzolino takes Z to be have a gamma distribution and T given $Z = z$ to have an exponential distribution with rate γz.) At time $s > 0$, the known data are the catch times and sizes of those fish, if any, whose catch times are less than or equal to s. The payoff for stopping at time s is the sum of the sizes of the fish caught by time s minus a constant cost per unit time: $Y_s = \sum_j Z_j \mathrm{I}(T_j \leq s) - cs$. See Exercise 13.

The most interesting feature of this model is that if by time s, k fish of sizes $z_1, \ldots, z_k$ have been caught at times $t_1, \ldots, t_k$ respectively, then the posterior distribution of future catch times and sizes is independent of this information and may be described as follows. The number of fish remaining, $N - k$, at time s is $\mathcal{P}(\lambda S(s))$, where $S(s) = \mathrm{P}(T > s)$ is the survival function. Given the number of fish remaining, their catch times and sizes are i.i.d. with distribution function, $(F(t,z)-F(s,z))/S(s)$, on the half-plane $\{(t,z) : t > s\}$. The conclusion to be drawn from this is that if there is an optimal stopping rule, then there is an optimal fixed time stopping rule. This conclusion is independent of the payoff function.

6. A BEST-CHOICE PROBLEM — SUM-THE-ODDS. Here we consider a far-reaching generalization of the secretary problem due to Hill and Krengel (1992) and Thomas Bruss (2000). In this generalization, the observations are independent random variables, $X_1, X_2, \ldots$, taking values 0 or 1, with 0 representing failure and 1 representing success. Our goal is to stop on the last success. Let the success probabilities be denoted by $p_n = P(X_n = 1) = 1 - P(X_n = 0)$ for $n = 1, 2, \ldots$. Since we would never stop at stage n if there is an $i > n$ such that $p_i = 1$, we assume that all p_i are strictly less than 1 for all $i > 1$, but allow an initial success with probability 1. If we stop at stage n, our payoff, the probability of stopping at the last success, is for $n = 1, 2, \ldots$,

$$Y_n = X_n \prod_{i=n+1}^{\infty} (1-p_i) = \begin{cases} \prod_{i=n+1}^{\infty}(1-p_i) & \text{if } X_n = 1 \\ 0 & \text{if } X_n = 0 \end{cases} \tag{15}$$

and we take $Y_0 = Y_\infty = 0$. Of course, if there are an infinite number of successes, then no stopping rule can achieve the goal of stopping on the last success. By the Borel-Cantelli Lemma, there will be a finite number of successes almost surely if and only if $\sum_1^\infty p_i < \infty$, or equivalently, $\prod_{i=2}^\infty (1-p_i) > 0$.

Therefore, to avoid the trivial case where every rule is optimal and gives payoff 0, we assume $\sum_1^\infty p_i < \infty$.

The secretary problem is a finite horizon problem. In the formulation above, a problem is said to have finite horizon n if $p_i = 0$ for all $i > n$. The secretary problem is obtained from this by taking X_i as the indicator function of the event that the ith object is relatively best, for $i = 1, \ldots, n$. A result of Rényi (1962) states that these events are independent and the probability that the ith object is best out of the first i is $1/i$. (For a generalization of this result, see Exercise 17.) Thus the classical secretary problem is obtained in this formulation if we put

$$p_i = \begin{cases} 1/i & \text{for } i = 1, \ldots, n \\ 0 & \text{for } i > n. \end{cases} \tag{16}$$

The secretary problem is not monotone and the optimal rule is not the one-stage look-ahead rule. But this is only because if one continues from a relatively best option, the next option may not be relatively best, in which case stopping at the next stage is obviously bad. In one of the early treatments of the secretary problem, Dynkin (1963) shows that the problem can be interpreted so that it is monotone simply by not allowing stopping on an observation that is not relatively best. With this restriction, the problem becomes monotone and the optimal rule is the one-stage look-ahead rule.

In the structure for the theory of optimal stopping as treated in Chapter 3, stopping is allowed after each observation. So to change the above problem so that stopping is allowed only after a success is observed, we change the definition of an observation. We pretend that the observations are the times at which successes occur, say $T_1, T_2, \ldots$, where T_k is the time at which the kth success occurs. Let K denote the time at which the last success occurs, with $K = \infty$ if no successes occur. We put $T_j = \infty$ if $T_n = K$ and $j > n$. With this change, the 1-sla may be computed as follows. The payoff if we stop at $T_n = t$ is 1 if $K = T_n$ and 0 otherwise. So the expected payoff at the time of the nth success is

$$Y_n = P(K = t | T_n = t) = \begin{cases} \prod_{i=t+1}^{\infty}(1 - p_i) & \text{if } t < \infty \\ 0 & \text{if } t = \infty, \end{cases} \tag{17}$$

and we take $Y_\infty = 0$. If we continue from $T_n = t < \infty$ and stop at T_{n+1}, we expect to receive

$$\begin{aligned} P(K = T_{n+1} | T_1, \ldots, T_n = t) &= p_{t+1} \prod_{t+2}^{\infty}(1 - p_i) + (1 - p_{t+1}) p_{t+2} \prod_{t+3}^{\infty}(1 - p_i) + \cdots \\ &= [\prod_{i=t+1}^{\infty}(1 - p_i)] \sum_{i=t+1}^{\infty} \frac{p_i}{1 - p_i} \end{aligned} \tag{18}$$

Therefore, the one-stage look-ahead rule is

$$\begin{aligned} N_1 &= \min\{n \geq 0 : \sum_{i=T_n+1}^{\infty} \frac{p_i}{1 - p_i} \leq 1\} \\ &= \min\{t \geq 1 : X_t = 1 \quad \text{and} \quad \sum_{i=t+1}^{\infty} \frac{p_i}{1 - p_i} \leq 1\}. \end{aligned} \tag{19}$$

This is the rule that stops on a success at time t if the sum of the odds, $p_i/(1 - p_i)$, from $i = t + 1$ to infinity is is less than or equal to 1.

We now show that the rule N_1 is optimal. Let $r_i = p_i/(1-p_i)$ be odds of success on the ith trial. First, the problem is monotone because $\sum_{i=T_n+1}^{\infty} r_i \leq 1$ implies $\sum_{i=T_{n+1}+1}^{\infty} r_i \leq 1$. A_3 is satisfied because $Y_n \to 0 = Y_\infty$ a.s. as $n \to \infty$ and we are assuming that $\sum_1^{\infty} p_i < \infty$. Finally, the conditions of Corollary 2 hold with $W_n = 0$ since $|Y_n| \leq 1$.

As an example, consider the secretary problem with the p_i given by (16). The stopping rule (19) reduces to

$$N_1 = \min\{t \geq 1 : X_t = 1 \quad \text{and} \quad \sum_{i=t+1}^{n} \frac{(1/i)}{1 - (1/i)} \leq 1\}.$$

This is the rule that stops at time t if the tth object is relatively best, and if $\sum_{i=t+1}^{n} 1/(i - 1) \leq 1$. This agrees with the rule found in equation (4) of Chapter 2.

§5.5 Exercises.

1. *Bayesian Estimation of the mean of an exponential distribution.* Consider the problem of sequential statistical estimation of a real parameter θ with squared error loss and constant cost $c > 0$ per observation, so that the loss for stopping at n is given by (3). For given $\theta > 0$, let $X_1, X_2, \ldots$ be i.i.d. according to an exponential distribution with mean θ and density $f(x|\theta) = (1/\theta)\exp\{-x/\theta\}$ for $x > 0$. Let the prior density of θ be the reciprocal gamma distribution with density

$$g(\theta) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \exp\{-\lambda/\theta\}\theta^{-(\alpha+1)}, \quad \theta > 0,$$

denoted by $\mathcal{G}^{-1}(\alpha, \lambda)$. Suppose that $\alpha > 2$.
(a) Show that the posterior distribution of θ given $X_1, \ldots, X_n$, is reciprocal gamma, $\mathcal{G}^{-1}(\alpha + n, \lambda + S_n)$, where $S_n = \sum_1^n X_i$.
(b) Show $Y_0 = \mathrm{Var}\,(\theta) = \lambda^2/((\alpha-1)^2(\alpha-2))$.
(c) Show that the 1-sla calls for stopping without taking any observations if and only if

$$\lambda^2/(\alpha(\alpha-1)^2(\alpha-2)) \le c.$$

(d) Show that the 2-sla calls for stopping without taking any observations if and only if

$$\lambda^2/(\alpha(\alpha-1)^2(\alpha-2)) \le c - (2c/(\alpha-2))(\lambda^2/(c(\alpha+1)\alpha^2(\alpha-1)))^{\alpha/2}.$$

2. *The adaptive burglar.* A burglar moves to a new city where he does not know precisely the distribution of the rewards he might expect from his burglaries, $X_1, X_2, \ldots$, but he is willing to assume that they are i.i.d. with an exponential density $f(x|\theta) = \theta e^{-\theta x} I(x > 0)$ for some unknown θ, whose distribution in turn he is willing to approximate by the gamma, $\mathcal{G}(1, 1/\lambda)$, with density $g(\theta) = \lambda e^{-\lambda\theta} I(\theta > 0)$ for some known λ. Let β denote the probability he is caught, $0 < \beta < 1$, and write his reward for stopping at n in the Dubins/Teicher form, $Y_n = \beta^n S_n$ for $n = 0, 1, 2, \ldots$, where $S_n = X_1 + \ldots + X_n$, and $Y_k = 0$.
(a) Show that the problem is monotone.
(b) Note that $\mathrm{E}X_1 = \infty$ so that A1 is not satisfied. But since the burglar should be especially eager to take the first observation anyway, let us suppose that he has done so and that $X_1 = x_1$ is known. Show that the 1-sla is optimal in this conditional problem.

3. *Vingt-et-un.* The burglar problem may also be generalized to allow capture times to be dependent on the rewards. (See Taylor (1975).) We also generalize to a concave utility of the accumulated payoff. Let the returns for the burglaries, $X_1, X_2, \ldots$, be nonnegative i.i.d. random variables with known distribution function, F, and finite mean, $\mu > 0$, and let $S_n = X_1 + \cdots + X_n$. Let T denote the random integer-valued time of capture and assume that for $n = 0, 1, \ldots$,

$$\mathrm{P}(T = n+1 | X_1, X_2, \ldots, T > n) = 1 - r(S_n),$$

where $0 \le r(z) \le 1$ is a known nonincreasing function on $[0, \infty)$ not identically 1. (In the casino games vingt-et-un and blackjack, one loses when one's total card count exceeds 21. Such games may be modeled as this burglar problem with $r(z) = \mathrm{P}(z + X \le 21)$.)
(a) *Maximizing utility of reward.* Let $u(z)$ denote the burglar's utility of retiring with an accumulated gain of z, where u is assumed to be a concave nondecreasing function on $[0, \infty)$, normalized so that $u(0) = 0$. Thus,

$$\begin{aligned} Y_n &= u(S_n)I(T > n) \quad \text{for} \quad n = 0, 1, \ldots \\ Y_\infty &= 0. \end{aligned}$$

Find the 1-sla and show it is optimal.
(b) *Maximizing the duration of operation.* Suppose instead that

$$\begin{aligned} Y_n &= u(n)I(T > n) \quad \text{for} \quad n = 0, 1, \ldots \\ Y_\infty &= 0, \end{aligned}$$

where $u(n)$ is such that $u(0) = 0$ and $u(n)/u(n+1)$ is nondecreasing. We also need a condition to make $\mathrm{E}(u(T)) < \infty$, for example, $u(n)/u(n+1) \to 1$. (The function $u(n) = n$, where we are trying to maximize the duration of operation, satisfies these conditions.) Find the 1-sla and show it is optimal.

4. *Adaptively selling an asset with recall.* (Ferguson (1974) and Rothschild (1974)) Offers $X_1, X_2, \ldots$ for an asset you own come in, independently drawn from a distribution F that you do not know exactly, but for which you have a Dirichlet process prior $\mathcal{D}(\alpha)$ with some parameter $\alpha = MF_0$. (All you need to know about the Dirichlet process to, solve this problem is that the expectation of $F(x)$ is $F_0(x)$ and that the posterior distribution of F given $X_1, \ldots, X_n$ is also Dirichlet but with parameter $\alpha_n = (M+n)F_n$ where $F_n = p_n F_0 + (1 - p_n)F_n^*$ and $p_n = M/(M+n)$ and F_n^* is the sample distribution function.) Assume that F_0 has a finite second moment.
(a) Assume a cost model $Y_n = M_n - nc$, find the 1-sla and show it is optimal.
(b) Assume a discount model $Y_n = \beta^n M_n$, find the 1-sla and show it is optimal.

5. *Selling several assets.* (MacQueen and Miller (1960).) A company desires to hire r persons. Interviews at a cost of $c > 0$ each yield applicants with expected worths to the company of $X_1, X_2, \ldots$, an i.i.d. sequence with known distribution function F with finite variance. Hiring is done with recall so that $Y_n = X_1^{(n)} + X_2^{(n)} + \ldots + X_r^{(n)} - nc$ for $n = r, r+1, \ldots$ and $Y_0 = \ldots = Y_{r-1} = Y_\infty = -\infty$, where $X_1^{(n)} \ge X_2^{(n)} \ge \ldots \ge X_n^{(n)}$ are the order statistics of $X_1, \ldots, X_n$. Find the 1-sla and show it is optimal.

6. *The parking problem.* Suppose the target destination $t > 0$ is known, and that the points $X_1, X_2, \ldots$ at which parking is available form a random walk; more precisely, we assume that the differences $X_1, X_2 - X_1, X_3 - X_2, \ldots$ are independent identically distributed positive random variables. The cost of stopping at stage n is taken to be

$$Y_n = g(X_n) \quad \text{for} \quad n = 1, 2, \ldots \quad (Y_0 = g(0), Y_\infty = \infty),$$

where g is a nonnegative convex function with minimum 0 at t: $g(t) = 0$. Assume that $g(x) \to \infty$ as $x \to \infty$, so that A3 is satisfied.
(a) Show that the problem is monotone.
(b) Show that the 1-sla is optimal.

7. *Finding the parking place closest to the destination.* (Boneh (1989)) Consider the parking problem with parking places occurring at sites $S_1, S_2, \ldots$ chosen according to a Poisson process with intensity λ. This time you want to find the parking place closest to the target t, assumed to be known. That is, you win if and only if you stop at the S_n closest to t. Thus,

$$Y_n = \begin{cases} 1 & \text{if } S_n > t \text{ and } S_n - t < |S_{n-1} - t| \\ 0 & \text{if } S_n > t \text{ and } S_n - t > |S_{n-1} - t| \\ \exp\{-2\lambda(t - S_n)\} & \text{if } S_n < t. \end{cases}$$

(a) Find the 1-sla.
(b) Show it is optimal.

8. *Proofreading.* (Ferguson and Hardwick (1989)) (a) Consider the proofreading problem of the text with $M \in \mathcal{P}(\lambda)$, but with the change that the success probability on the nth proofreading depends on n and is denoted by p_n. Find conditions on the p_n, assumed known, under which the problem is still monotone. Note that if the p_n are decreasing, as they well might be in practice, the 1-sla is still optimal, but that even if the 1-sla is not optimal, it is still easy to find the optimal rule because it is a fixed sample size rule.
(b) Consider a manuscript with a known number of words, W, each of which has a known probability π of being a misprint, so that M, the number of misprints, has a binomial distribution with sample size W and success probability π. Assume that the conditional distribution of the number of misprints found is as in part (a). Show that the posterior distribution of M_n given $X_1, \ldots, X_n$ is binomial. Find conditions on the p_n under which the problem is monotone. Note that the 1-sla is optimal if the p_n are nonincreasing.

9. *A search problem.* (Chew (1967)) An object is placed in a box with known probability π, $0 < \pi < 1$. You may search for it there as many times as you like. If it is there, the probability that you find it in any given search is p, $0 < p < 1$, independent of how many times you have already searched. Each search costs $c > 0$ and you win 1 if and only if you find the object.
(a) Find the 1-sla and show it is optimal.
(b) Generalize to the case where π and p are unknown with an arbitrary prior distribution.
(c) Extend to k objects placed in the box independently with probabilities π_j, $j = 1, \ldots, k$. Searches may find many objects, the objects being found independently with probabilities p_j, $j = 1, \ldots, k$, independent of the number of times you have searched and the number of objects found so far. Object j is worth x_j, $j = 1, \ldots, k$.

10. *Search for new species.* (Rasmussen and Starr (1979)) Independent observations may be drawn from a population Π consisting of finitely many species, $\Pi_1, \ldots, \Pi_s$, with probability p_j of observing a member of Π_j for $j = 1, \ldots, s$ where $p_j \geq 0$ and $\sum p_j = 1$.

There is a cost $c > 0$ for each observation, and a reward for each new species discovered. If we stop after n observations for some $n = 0, 1, \ldots$, our payoff is $Y_n = h(K_n) - nc$, where K_n is the number of distinct species observed by time n, and $h(k) \geq 0$ is the reward for having observed k distinct species. ($Y_\infty = -\infty$.) Assume that $h(k+1) - h(k)$ is nonincreasing in k.
(a) Suppose the p_j are known numbers. Find the 1-sla and show it is optimal.
(b) Suppose the p_j are unknown, but have a Dirichlet prior distribution, $\mathcal{D}(\alpha_1, \ldots, \alpha_s)$, where $\alpha_j > 0$ for all j. Find the 1-sla and show it is optimal.
(c) Show by an example that for a general prior distribution for the p_j, the 1-sla may not be optimal.

11. *Dispatching.* (See Ross (1969)) Passengers arrive randomly in the time interval (0,1) at a bus depot and wait for the next bus to leave. One bus is scheduled to leave at time 1; the problem is to choose the time t, $0 < t < 1$, at which an unscheduled bus should be dispatched in order to minimize the total expected waiting time of all the passengers. Discretize the interval (0,1) into T intervals of length $\delta = 1/T$, and let $X_j \geq 0$ denote the number of passengers who arrive in the time interval $(j-1)\delta$ to $j\delta$, for $j = 1, \ldots, T$. The total waiting time if the unscheduled bus leaves at time $t = n\delta$ is $W_n = \delta[\sum_{j=1}^{n}(n-j)X_j + \sum_{j=n+1}^{T}(T-j)X_j]$. Assume that the X_j are independent with means λ_j (e.g. $X_j \in \mathcal{P}(\lambda_j)$) for $j = 1, \ldots, T$.
(a) Find the 1-sla stopping rule N_1 for minimizing $\mathrm{E}(W_N)$ and show it is optimal provided that $\lambda_j(T-j)$ is nonincreasing in j.
(b) Suppose the λ_j are constant, equal to λ, unknown, but that λ has a known gamma prior distribution. Find the 1-sla and show it is optimal.

12. *Detecting a change-point.* (Shiryaev (1963)) The change-point K is unobservable but has known distribution $\mathrm{P}(K = k) = \pi_k$, $k = 1, 2, \ldots$, with finite mean. The observable quantities, $X_1, X_2, \ldots$ are independent given $K = k$, with $X_1, \ldots, X_{k-1}$ having a distribution with density $f_0(x)$ and $X_k, X_{k+1}, \ldots$ having a distribution with density $f_1(x)$. If you stop at n your loss is an inspection cost $c > 0$ if $n < K$, and the length of time since the change, $n - K$, if $n \geq K$. Thus, $Y_n = cI(n < K) + (n-K)I(n \geq K)$ or, conditioned on $\mathcal{F}_n$,

$$Y_n = c\mathrm{P}\{K > n|\mathcal{F}_n\} + \mathrm{E}\{(n-K)^+|\mathcal{F}_n\} \quad \text{for} \quad n = 1, 2, \ldots,$$

and we take $Y_\infty = \infty$.
(a) Find the 1-sla in the form $N_1 = \min\{n \geq 0 : U_n \geq c\}$.
(b) Find a recurrence to compute the U_n.
(c) Take K to be geometric with parameter π, $\mathrm{P}(K = n) = (1-\pi)\pi^{n-1}$ for $n = 1, 2, \ldots$. Take $f_0(x)$ to be exponential with mean 1, and $f_1(x)$ to be exponential with mean $\mu > 1$. Show the 1-sla is optimal if $c\pi\mu \leq 1 + c$.

13. *A Poisson fishing model.* (Cozzolino (1972) and Ferguson (1997)) Let $M \in \mathcal{P}(\lambda)$ and suppose that $X_1, \ldots, X_m$ given $M = m$ are i.i.d. according to a probability distribution P on some space. Let A be any measurable set in the space, and let $N(A)$ be the random variable denoting the number of $X_i \in A$. Then, $N(A)$ and $N(A^c)$ are independent with $\mathcal{P}(\mathrm{P}(A)\lambda)$ and $\mathcal{P}(\mathrm{P}(A^c)\lambda)$ distributions respectively. Moreover, given $N(A)$

and $N(A^c)$, the $X_i \in A$ (resp. the $X_i \in A^c$) are i.i.d. with distribution $\mathrm{P}/\mathrm{P}(A)$ on A (resp. $\mathrm{P}/\mathrm{P}(A^c)$ on A^c).
(a) Using this result, establish the following. Suppose the number, M, of fish in the lake is Poisson, $\mathcal{P}(\lambda)$, and given $M = m$ let the catch times and sizes of the fish, $(T_1, Z_1), \ldots, (T_m, Z_m)$, be i.i.d. with known distribution $F(t, z)$. Then at time t, independent of the number of fish caught and their catch times and sizes, the number of fish remaining in the lake is $\mathcal{P}(\lambda S(t))$, where $S(t) = \mathrm{P}(T > t)$ and given the number of fish remaining, their catch times and sizes are i.i.d. with distribution function, $\mathrm{P}(T \le s, Z \le z | T > t) = (F(s, z) - F(t, z))/S(t)$, on the set $\{s > t\}$.
(b) Suppose the joint distribution of T and Z has density $f(t, z)$ with finite $\mathrm{E}|Z|$. Let $Y_t = \sum_{j=1}^{M} Z_j I(T_j \le t) - ct$. Find $\mathrm{E}Y_t$, the expected return of the fixed time stopping rule, $N \equiv t$.
(c) Suppose T has the inverse power distribution with density, $f(t) = \theta/(1+t)^{\theta+1}$ on $(0, \infty)$, and let the distribution of Z given $T = t$ be the gamma, $\mathcal{G}(\alpha, \gamma(t+1))$, where $\theta > 0$, $\alpha > 0$, and $\gamma > 0$ are given constants. (Smaller fish are easier to catch.) Find the optimal stopping rule.
(d) Suppose the distribution of the size Z is gamma, $\mathcal{G}(\alpha, \beta)$ and that given $Z = z$ the catch time is exponential $\mathcal{G}(1, 1/(z\gamma))$. (The larger fish are easier to catch.) Find the optimal rule.

14. *Additive Damage Model.* (See Derman and Sacks (1960)) Consider a machine that accumulates observable damage. Let X_n denote the damage accrued on day n, where $X_1, X_2, \ldots$ are i.i.d. with known distribution function $F(x)$ on $[0, \infty)$. The accumulated damage on day n is $S_n = X_1 + \cdots + X_n$. When the accumulated damage exceeds a known threshold L, the machine breaks down. If the machine has not broken down by day n, it will produce a random return with mean $r_n(S_n)$ on that day. If it breaks down on that day, a penalty of b_n is assessed. Let Z denote the time of breakdown, $Z = \min\{n \ge 1 : S_n > L\}$. Then, the return for stopping on day n is

$$Y_n = \begin{cases} \sum_1^n r_j(S_j) & \text{if } n < Z \\ \sum_1^Z r_j(S_j) - b_Z & \text{if } n \ge Z. \end{cases}$$

(a) Find the 1-sla.
(b) Find some reasonable conditions on the r_n, b_n and F so that the problem is monotone.
(c) What if L is random?

15. *A Win-Lose-or-Draw Sum-the-Odds Problem.* Extend the Bruss sum-the-odds result of Example 6 Section 5.4, to the win-lose-or-draw problem of Sakaguchi. Suppose you win 1 if you stop on the last success, win nothing if you stop on a success that is not the last, and win an amount θ if you don't stop, where $0 < \theta < 1$. Assume that successes are independent events, that the probability of success on trial n is p_n and that $\sum_1^\infty p_i < \infty$. Find the 1-sla, note that it is a sum-the-odds rule, and show it is optimal.

16. *The Group Interview Secretary Problem.* (Hsiau and Yang (2000)) Consider the secretary problem in which groups of applicants are interviewed together. It is possible to select any applicant of the present group, but one may not recall applicants of previous

groups. Suppose there are n rankable applicants arranged in a completely random order, that are to be interviewed sequentially in m groups of sizes $k_1, k_2, \ldots, k_m$, where $\sum_1^m k_i = n$. You will stop only if the present group has a relatively best applicant, and if you stop you will select that applicant. Find the optimal rule.

17. *Another Sum-the-Odds Stopping Rule.* Let $Z_1, Z_2, \ldots,$ be independent random variables, and suppose Z_n has an exponential distribution, $\mathcal{E}(\theta_n)$, with density $f(z|\theta_n) = \theta_n e^{-\theta_n z}$ for $z > 0$. The θ_n are given numbers that satisfy the restriction, $\sum_1^\infty \theta_i < \infty$. The Z_n are observed sequentially, and it is desired to choose a stopping rule that maximizes the probability of stopping on the smallest Z_n. Let X_n denote the indicator of the event that $Z_n = \min_{1 \le i \le n} Z_i$, that Z_n is a relatively best observation. Thus, we want to stop on the last X_n that is equal to 1.
(a) Show $P(X_n = 1) = \theta_n/(\theta_1 + \cdots + \theta_n)$.
(b) Show that the X_i are independent.
(c) Find the optimal stopping rule.

18. *A lower bound for the value of a sum-the-odds problem.* (Hill and Krengel (1992) and Bruss (2003)) When the secretary problem first appeared, it was found surprising that as the number of applicants tends to infinity, the probability of selecting the very best does not tend to zero, but instead is always greater than e^{-1}. It is even more surprising that this same bound, e^{-1}, holds for these more general problems, provided only that the sum of the odds is at least 1. In fact, there is a simple bound for finite horizon problems as well.
(a) Let $q_i = 1 - p_i$ and $r_i = p_i/(1-p_i)$. Consider the problem with finite horizon, n, and suppose that $\sum_1^n r_i \ge 1$. Let s denote an integer such that $\sum_{i=s}^n r_i \ge 1$ and $\sum_{i=s+1}^n r_i \le 1$. Show that the optimal probability of success is

$$V_n^* = \begin{cases} [\prod_{i=s}^n q_i] \sum_{i=s}^n r_i & \text{if } p_s < 1 \\ \prod_{i=s+1}^n q_i & \text{if } p_s = 1. \end{cases}$$

(b) Show that when $q_{s+1}, \ldots, q_n$ are fixed, V_n^* is increasing in p_s and so is smallest when p_s is as small as it can be made, which is just small enough to make $\sum_{i=s}^n r_i = 1$.
(c) Show that $\prod_{i=s}^n q_i$ is minimized subject to the constraint $\sum_{i=s}^n r_i = 1$ when all the q_i for $i \ge s$ are equal.
(d) Conclude that when $\sum_1^n r_i \ge 1$,

$$V_n^* \ge \left(1 + \frac{1}{n}\right)^{-n}$$

with equality if and only if $p_i = 1/(n+1)$ for $i = 1, \ldots, n$.
(e) Conclude that when $\sum_1^\infty r_i \ge 1$ and $\sum_1^\infty p_i < \infty$,

$$V_\infty^* \ge e^{-1}.$$

Chapter 6. MAXIMIZING THE RATE OF RETURN.

In stopping rule problems that are repeated in time, it is often appropriate to maximize the average return per unit of time. This leads to the problem of choosing a stopping rule N to maximize the ratio $\mathrm{E}Y_N/\mathrm{E}N$. The reason we wish to maximize this ratio rather than the true expected average per stage, $\mathrm{E}(Y_N/N)$, is that if the problem is repeated independently n times with a fixed stopping rule leading to i.i.d. stopping times $N_1, \ldots, N_n$ and i.i.d. returns $Y_{N_1}, \ldots, Y_{N_n}$, the total return is $Y_{N_1}+\cdots+Y_{N_n}$ and the total time is $N_1+\cdots+N_n$, so that the average return per unit time is the ratio $(Y_{N_1}+\cdots+Y_{N_n})/(N_1+\cdots+N_n)$. If both sides of this ratio are divided by n and if the corresponding expectations exist, then this ratio converges to $\mathrm{E}Y_N/\mathrm{E}N$ by the law of large numbers. We call this ratio the rate of return. We wish to maximize the rate of return.

In the first section of this chapter, we describe a method of solving the problem of maximizing the rate of return by solving a sequence of related stopping rule problems as developed in the earlier chapters. There are a number of applications that are treated in subsequent sections and in the exercises. In Section 6.2, the main ideas are illustrated using the house-selling problem. In Section 6.3, application is made to problems where the payoff is a sum of discounted returns. This provides a background for the treatment of bandit problems in Chapter 7. In Section 6.4, a simple maintenance model is considered to illustrate the general method of computation. Finally in Section 6.5, a simple inventory model is treated.

§**6.1 Relation to Stopping Rule Problems.** We set up the problem more generally by allowing different stages to take different amounts of time. There are observations $X_1, X_2, \ldots$ as before, but now there are two sequences of payoffs, $Y_1, Y_2, \ldots$ and $T_1, T_2, \ldots$ with both Y_n and T_n assumed to be $\mathcal{F}_n$-measurable, where $\mathcal{F}_n$ is the sigma-field generated by $X_1, \ldots, X_n$. In this formulation, Y_n represents the return for stopping at stage n and T_n represents the total time spent to reach stage n. Throughout this chapter, we assume that the T_n are positive and nondecreasing almost surely,

$$(1) \qquad 0 < T_1 \le T_2 \le \ldots \quad \text{a.s.}$$

We restrict attention to stopping rules that take at least one observation and note that $\mathrm{E}(T_N) \ge \mathrm{E}(T_1) > 0$ for every stopping rule $N \ge 1$. Thus, in forming the ratio $\mathrm{E}Y_N/\mathrm{E}T_N$, we avoid the problem of dealing with $0/0$. To avoid the troublesome $\pm\infty/+\infty$, we restrict attention to stopping rules such that $\mathrm{E}T_N < \infty$. Thus, we let $\mathcal{C}$ denote the class of stopping rules,

$$\mathcal{C} = \{N : N \geq 1, \mathrm{E}T_N < \infty\} \tag{2}$$

and we seek a stopping rule $N \in \mathcal{C}$ to maximize the rate of return, $\mathrm{E}Y_N/\mathrm{E}T_N$.

Without entering into the question of the existence of a stopping rule that attains a finite supremum of the above ratio, we can relate the solution of the problem of maximizing the rate of return to the solution of an ordinary stopping rule problem with return $Y_n - \lambda T_n$ for some λ.

Theorem 1. *(a) If for some* λ, $\sup_{N\in\mathcal{C}} \mathrm{E}(Y_N - \lambda T_N) = 0$ *, then* $\sup_{N\in\mathcal{C}} \mathrm{E}(Y_N)/\mathrm{E}(T_N) = \lambda$. *Moreover, if* $\sup_{N\in\mathcal{C}} \mathrm{E}(Y_N - \lambda T_N) = 0$ *is attained at* $N^* \in \mathcal{C}$, *then* N^* *is optimal for maximizing* $\sup_{N\in\mathcal{C}} \mathrm{E}(Y_N)/\mathrm{E}(T_N)$.
(b) Conversely, if $\sup_{N\in\mathcal{C}} \mathrm{E}(Y_N)/\mathrm{E}(T_N) = \lambda$ *and if the supremum is attained at* $N^* \in \mathcal{C}$, *then* $\sup_{N\in\mathcal{C}} \mathrm{E}(Y_N - \lambda T_N) = 0$ *and the supremum is attained at* N^*.

Proof. If $\sup_{N\in\mathcal{C}} \mathrm{E}(Y_N - \lambda T_N) = 0$, then for all stopping rules $N \in \mathcal{C}$, $\mathrm{E}(Y_N - \lambda T_N) \leq 0$ so that $\mathrm{E}(Y_N)/\mathrm{E}(T_N) \leq \lambda$. If, for some $\epsilon \geq 0$, the rule $N^* \in \mathcal{C}$ is ϵ-optimal, so that $\mathrm{E}(Y_{N^*} - \lambda T_{N^*}) \geq -\epsilon$, then

$$\mathrm{E}(Y_{N^*})/\mathrm{E}(T_{N^*}) \geq \lambda - \epsilon/\mathrm{E}(T_{N^*}) \geq \lambda - \epsilon/\mathrm{E}(T_1),$$

so that N^* is $(\epsilon/\mathrm{E}(T_1))$-optimal for maximizing $\mathrm{E}(Y_N)/\mathrm{E}(T_N)$.

Conversely, suppose $\sup_{N\in\mathcal{C}} \mathrm{E}(Y_N)/\mathrm{E}(T_N) = \lambda$, and suppose the supremum is attained at $N^* \in \mathcal{C}$. Then, $\mathrm{E}Y_{N^*} - \lambda \mathrm{E}T_{N^*} = 0$ and for all stopping rules $N \in \mathcal{C}$, $\mathrm{E}Y_N - \lambda \mathrm{E}T_N \leq 0$. ∎

The optimal rate of return, λ, may also be considered as the "shadow" cost of time measured in the same units as the payoffs. This is because, when λ is the optimal rate of return, we search for the stopping rule that maximizes $\mathrm{E}(Y_N - \lambda T_N)$. It is as if we are being charged λ for each time unit. This is the mathematical analog of the aphorism, "Time is money".

Sometimes an extra argument may be provided to show that the limiting average payoff cannot be improved using rules for which $\mathrm{E}T_N = \infty$. (See §6.2.)

COMPUTATION. Many of the good applications require heavy computation to reach the solution and so we mention a fairly effective method suggested by G. Michaelides. We use part (a) of Theorem 1 to approximate the solution to the problem of computing the optimal rate of return. To use this theorem, we first solve the ordinary stopping rule problem for stopping $Y_n - \lambda T_n$ with arbitrary λ, and find the value. Ordinarily, this value will be a decreasing function of λ going from $+\infty$ at $\lambda = -\infty$ to $-\infty$ at $\lambda = \infty$. We then search for that λ that makes the value equal to zero.

To be more specific, let us make the assumption that for each λ there exists a rule $N(\lambda) \in \mathcal{C}$ that maximizes $\mathrm{E}(Y_N - \lambda T_N)$, and let $V(\lambda)$ denote the optimal return,

$$V(\lambda) = \sup_{N \in \mathcal{C}} [\mathrm{E}(Y_N) - \lambda \mathrm{E} T_N] = \mathrm{E}(Y_{N(\lambda)}) - \lambda \mathrm{E}(T_{N(\lambda)}).$$

Lemma 1. *$V(\lambda)$ is decreasing and convex.*

Proof. Let $\lambda_1 < \lambda_2$. Then

$$V(\lambda_2) = \mathrm{E}Y_{N(\lambda_2)} - \lambda_2 \mathrm{E}T_{N(\lambda_2)} < \mathrm{E}Y_{N(\lambda_2)} - \lambda_1 \mathrm{E}T_{N(\lambda_2)} \le \mathrm{E}Y_{N(\lambda_1)} - \lambda_1 \mathrm{E}T_{N(\lambda_1)} = V(\lambda_1),$$

so $V(\lambda)$ is decreasing in λ. To show convexity, let $0 < \theta < 1$, fix λ_1 and λ_2 in $\mathcal{C}$, and let $\lambda = \theta\lambda_1 + (1-\theta)\lambda_2$. Then,

$$\begin{aligned} V(\lambda) &= \mathrm{E}Y_{N(\lambda)} - (\theta\lambda_1 + (1-\theta)\lambda_2)\mathrm{E}T_{N(\lambda)} \\ &= \theta(\mathrm{E}(Y_N) - \lambda_1 \mathrm{E}T_N) + (1-\theta)(\mathrm{E}(Y_N) - \lambda_2 \mathrm{E}T_N) \\ &\le \theta V(\lambda_1) + (1-\theta)V(\lambda_2). \quad \blacksquare \end{aligned}$$

With this result, we may describe a simple iterative method of approximating the optimal rate of return and the optimal stopping rule. This method is a variation of Newton's method and so converges quadratically. Let λ_0 be an initial guess at the optimal value. At λ_0, the line $y = V(\lambda_0) - \mathrm{E}T_{N(\lambda_0)}(\lambda - \lambda_0)$ is a supporting hyperplane. This follows because $V(\lambda_0) - \mathrm{E}T_{N(\lambda_0)}(\lambda - \lambda_0) = \mathrm{E}Y_{N(\lambda_0)} - \lambda \mathrm{E}T_{N(\lambda_0)} \le V(\lambda)$. Therefore, in Newton's method, $\lambda_{n+1} = \lambda_n - V(\lambda_n)/V'(\lambda_n)$, we may replace the derivative of $V(\lambda)$ at λ_n with $-\mathrm{E}T_{N(\lambda_n)}$. This gives the iteration for $n = 0, 1, 2, \ldots$,

$$\lambda_{n+1} = \lambda_n + \frac{V(\lambda_n)}{\mathrm{E}T_{N(\lambda_n)}} = \frac{\mathrm{E}Y_{N(\lambda_n)}}{\mathrm{E}T_{N(\lambda_n)}}. \tag{3}$$

For any initial value, λ_0, this sequence will converge quadratically to the optimal rate of return. It is interesting to note that the convergence is quadratic even if the derivative of $V(\lambda)$ does not exist everywhere. See §6.4 for an example.

§**6.2. House-Selling.** Consider the problem of selling a house without recall and with i.i.d. sequentially arriving offers of $X_1, X_2, \ldots$ dollars, constant cost $c \ge 0$ dollars per observation and with return $X_n - cn$ for stopping at n. When the house is sold, you may construct a new house to sell. Construction cost is $a \ge 0$ dollars and construction time is $b \ge 0$ time units, measured in units of time between offers. Thus your return for one cycle is $Y_n = X_n - a - cn$, and the time of a cycle is $T_n = n + b$. Note that in this formulation, the cost of living, c, is not assessed while the house is being built. We assume the cost of living while building is included in the cost a.

To solve the problem of maximizing the rate of return, $\mathrm{E}(Y_N)/\mathrm{E}(T_N)$, we solve the related stopping rule problem with return for stopping at n taken to be $Y_n - \lambda T_n =$

$X_n - a - cn - \lambda n - \lambda b$, and then choose λ so that the optimal return is zero. If we assume that the X_n have finite second moment, $\mathrm{E}(X^2) < \infty$, this is the problem solved in §4.1 with return X_n replaced by $X_n - a - \lambda b$ and cost c replaced by $c + \lambda$. The solution found there requires $c + \lambda > 0$. The optimal rule is to accept the first offer $X_n \geq V^* + a + \lambda b$, where V^* satisfies

$$\mathrm{E}(X - a - \lambda b - V^*)^+ = c + \lambda. \tag{4}$$

The value of λ that gives $V^* = 0$ is then simply the solution of (4) with $V^* = 0$:

$$\mathrm{E}(X - a - \lambda b)^+ = c + \lambda. \tag{5}$$

If $b > 0$, the left side is a continuous decreasing function of λ from $\mathrm{E}(X - a + bc)^+$ at $\lambda = -c$, to zero at $\lambda = \infty$, and the right side is continuous increasing from 0 at $\lambda = -c$, to ∞. If $b = 0$, the left side is constant. In either case, if $\mathrm{E}(X - a + bc)^+ > 0$, there is a unique root, λ^*, of (5) such that $\lambda^* > -c$. The optimal rule is to accept the first offer of $a + \lambda^* b$ or greater:

$$N^* = \min\{n \geq 1 : X_n \geq a + \lambda^* b\}. \tag{6}$$

This rule is optimal for maximizing the limiting average payoff out of all rules N such that $\mathrm{E}N < \infty$, provided $\mathrm{E}(X^+)^2 < \infty$ and $\mathrm{E}(X - a + bc)^+ > 0$.

If $\mathrm{E}(X - a + bc)^+ = 0$, then $\lambda^* = -c$ and $N^* \equiv \infty$. In other words, if $\mathrm{P}(X > a - bc) = 0$, we never sell the house and our expected rate of return is $-c$. This makes sense since stopping can only make the rate of return less than $-c$, but since we have not defined a limiting average payoff for continuing forever, we make the assumption that $\mathrm{E}(X - a + bc)^+ > 0$.

If $b = 0$, then (5) has a simple solution. The optimal rule for maximizing the rate of return is to *accept the first offer greater than the construction cost*, and the optimal rate of return becomes

$$\lambda^* = \mathrm{E}(X - a)^+ - c.$$

If the offers are a.s. greater than a, this means that we accept the first offer that comes in, so that N is identically equal to 1.

Can we do better with rules N such that $\mathrm{E}T_N = \infty$? From §4.1, it follows that N^* is optimal for maximizing $\mathrm{E}(Y_N - \lambda^* T_N)$ out of all stopping rules provided we define the payoff for not stopping to be $-\infty$. We have not defined Y_∞ or T_∞ for this problem but we can extend the optimality of the rule N^* for maximizing the limiting average payoff to the class of rules N such that $\mathrm{P}(N < \infty) = 1$ as follows. Let N be such a rule. As in the first paragraph of this chapter, we consider the problem repeated independently n times using the same stopping rule each time. Let the i.i.d. stopping times be denoted by $N_1, \ldots, N_n$, the corresponding i.i.d. returns by $Y_{N_1}, \ldots, Y_{N_n}$ and the corresponding i.i.d. reward times by $T_{N_1}, \ldots, T_{N_n}$. From §4.1, it follows that for any rule N with $\mathrm{P}(N < \infty) = 1$, we have $\mathrm{E}(Y_N - \lambda^* T_N) \leq 0$, possibly $-\infty$, so that from the strong law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n}(Y_{N_i} - \lambda^* T_{N_i}) \xrightarrow{a.s.} \mathrm{E}(Y_N - \lambda^* T_N) \leq 0.$$

Also from the strong law of large numbers,

$$\frac{1}{n}\sum_{i=1}^{n} T_{N_i} \xrightarrow{a.s.} \mathrm{E}T_N > \mathrm{E}T_1 > 0,$$

so that

$$\limsup_{n\to\infty} \frac{\sum_{i=1}^n Y_{N_i}}{\sum_{i=1}^n T_{N_i}} - \lambda^* \le 0. \tag{7}$$

From the Fatou-Lebesgue Lemma, the expected value of the lim sup of the average return is also nonpositive. This shows that N^* achieves the optimal rate of return out of all stopping rules for which $\mathrm{P}(N<\infty)=1$, provided $\mathrm{E}(X^+)^2<\infty$ and $\mathrm{E}(X-a+bc)^+>0$.

It is interesting to note that we can now make use of the observation of Robbins (1970) to weaken the condition $\mathrm{E}(X^+)^2<\infty$ to requiring only $\mathrm{E}X^+<\infty$. Under this weaker condition, Robbins shows that the rule N^* is optimal for stopping $Y_n-\lambda T_n$ out of all rules N such that $\mathrm{E}(Y_N-\lambda T_N)^- > -\infty$. But one can show that when $\mathrm{E}X^+<\infty$ and $\mathrm{E}(Y_N-\lambda^*T_N)^- = -\infty$, we still have $(1/n)\sum_1^n(Y_{N_i}-\lambda^*T_{N_i}) \xrightarrow{a.s.} -\infty$ even though $\mathrm{E}(Y_N-\lambda T_N)^+ = \infty$. One may conclude that if $\mathrm{E}X^+<\infty$, then N^* is optimal for maximizing the rate of return in the sense that if N is any stopping rule with $\mathrm{P}(N<\infty)=1$ and $N_1,N_2,\ldots$ are i.i.d. with the distribution of N, then (7) holds, and equality is achieved if $N=N^*$.

§**6.3. Sum of Discounted Returns.** Let $X_1,X_2,\ldots$ represent your returns for working on days $1,2,\ldots$. It is assumed that the X_j have some known joint distribution. For example, it might be assumed that $X_1,X_2,\ldots$ are daily returns from some mining operation or from studying some new mathematical problem. It may be that the returns indicate that the mine or problem is not likely to be very profitable, and so you should switch to a different mine or problem. The future is discounted by $0<\beta<1$ so that your total return for working n days has present value $Y_n=\sum_1^n \beta^{j-1}X_j$. In considering time, we should also discount, so the total time used earning this reward has present value $T_n=\sum_1^n\beta^{j-1}$. The problem of maximizing the rate of return is the problem of finding a stopping rule N to achieve the supremum in

$$V^* = \sup_{N\ge 1} \frac{\mathrm{E}\sum_1^N \beta^{j-1}X_j}{\mathrm{E}\sum_1^N \beta^{j-1}}. \tag{8}$$

We assume that the expectations of the X_j exist and are uniformly bounded above, $\sup_n \mathrm{E}X_n<\infty$. It may be noted that in this problem, we may allow N to assume the value $+\infty$ with positive probability; both sums in (8) will still be finite.

The problem given by (8) can be justified by a method similar to that used in the first paragraph of this chapter. We assume the original problem can be repeated independently as many times as desired. The kth repetition yields the sequence $X_{k1},X_{k2},\ldots$, each sequence being an independent sample from the original joint distribution of $X_1,X_2,\ldots$.

At any time n, after observing $X_{11}, \ldots, X_{1n}$ and earning $\sum_1^n \beta^{j-1} X_{1j}$, you may ask to start the problem over using the second sequence, but these returns will be discounted by an extra β^n because they start at time $n+1$. Similarly for any k, while observing sequence k, you may call for a restart and begin to observe the sequence $k+1$, etc. This is called the **restart problem** in Katehakis and Vienott (1987). Suppose the same stopping rule N is used in each restarted problem, yielding i.i.d. random variables $N_1, N_2, \ldots$. Then the total discounted return is

$$V = \sum_1^{N_1} \beta^{j-1} X_{1j} + \beta^{N_1} \sum_1^{N_2} \beta^{j-1} X_{2j} + \beta^{N_1+N_2} \sum_1^{N_3} \beta^{j-1} X_{3j} + \cdots.$$

Its expected return is

$$\begin{aligned} \mathrm{E}V &= \mathrm{E} \sum_1^{N_1} \beta^{j-1} X_{1j} + \mathrm{E}\beta^{N_1} \mathrm{E} \left[\sum_1^{N_2} \beta^{j-1} X_{2j} + \beta^{N_2} \sum_1^{N_3} \beta^{j-1} X_{3j} + \cdots \right] \\ &= \mathrm{E} \sum_1^{N_1} \beta^{j-1} X_{1j} + \mathrm{E}\beta^{N_1} \mathrm{E}V. \end{aligned}$$

Solving for $\mathrm{E}V$, we find

$$\mathrm{E}V = \frac{\mathrm{E}\sum_1^N \beta^{j-1} X_j}{1 - \mathrm{E}\beta^N} = \frac{\mathrm{E}\sum_1^N \beta^{j-1} X_j}{(1-\beta)\mathrm{E}\sum_1^N \beta^{j-1}}. \tag{9}$$

Thus, the optimal rate of return given in (8) is equal to $1-\beta$ times the optimal value of the restart problem.

To take an example of the computation of (8), assume that $X_1, X_2, \ldots$ are i.i.d. given a parameter $\theta > 0$ with distribution

$$\begin{aligned} \mathrm{P}(X = 0|\theta) &= 1/2 \\ \mathrm{P}(X = \theta|\theta) &= 1/2 \end{aligned}$$

for all $\theta > 0$. We assume that the prior distributin of θ on $(0, \infty)$ is such that $\mathrm{E}\theta < \infty$. To find the supremum in (8), we first solve the associated stopping rule problem of finding a stopping rule to maximize $\mathrm{E}Y_N - \lambda \mathrm{E}T_N = \mathrm{E}(\sum_1^N \beta^{j-1}(X_j - \lambda))$. Let $V^* = V^*(\lambda)$ denote this maximum value. We must take at least one observation. With probability 1/2, $X_1 = 0$, we lose λ and gain no information. In this case, the future looks as it did at the initial stage (except that it is now discounted by β), so we would continue if $V^* > 0$ and stop otherwise. With probability 1/2, $X_1 = \theta$, we receive $\theta - \lambda$ and we would have complete information. In this case, if $\theta/2 > \lambda$, we would continue forever and expect to receive $\sum_2^\infty \beta^{j-1}(\theta/2 - \lambda) = (\beta/(1-\beta))(\theta/2 - \lambda)$, while if $\theta/2 \le \lambda$, we would stop now and receive nothing further. Combining this, we arrive at the following equation for V^*.

$$V^* = (1/2)(-\lambda + \beta \max(0, V^*)) + (1/2)(\mathrm{E}(\theta - \lambda) + \mathrm{E}((\theta/2 - \lambda)^+)(\beta/(1-\beta))).$$

Therefore,

$$\begin{aligned} V^* &= [-2\lambda + (\beta/(1-\beta))\mathrm{E}(\theta/2-\lambda)^+ + \mathrm{E}\theta]/(2-\beta) \quad \text{if} \quad V^* > 0 \\ &= [-2\lambda + (\beta/(1-\beta))\mathrm{E}(\theta/2-\lambda)^+ + \mathrm{E}\theta]/2 \qquad \text{if} \quad V^* \le 0. \end{aligned}$$

To find the maximal rate of return, we choose λ so that $V^* = 0$. This gives the optimal rate of return, λ^*, as the root of the equation,

$$2\lambda = \mathrm{E}(\theta) + (\beta/(1-\beta))\mathrm{E}(\theta/2-\lambda)^+.$$

The left side is increasing from $-\infty$ to $+\infty$, and the right side is nonincreasing, so there is a unique root. The optimal rule is: take one observation; if $X_1 > 2\lambda^*$, then continue forever; otherwise stop.

For a specific example, suppose θ has a uniform distribution on the interval $(0,1)$. Then $\mathrm{E}(\theta) = 1/2$ and $\mathrm{E}(\theta/2-\lambda)^+ = \lambda^2 - \lambda + 1/4$, so that λ^* is the root of $2\lambda = 1/2 + (\beta/(1-\beta))(\lambda^2-\lambda+1/4)$ between $1/4$ and $1/2$, namely

$$\lambda^* = [2-\beta-\sqrt{2(2-\beta)(1-\beta)}]/(2\beta).$$

§**6.4. Maintenance.** A machine used in production of some item will produce a random number of items each day. As time progresses, the performance of the machine deteriorates and it will eventually need to be overhauled entailing a cost for the service and a loss of time for use of the machine. Suppose that if the machine has just been overhauled it produces X_1 items where X_1 has a Poisson distribution with mean μ. Suppose also that deterioration is exponential in time so that the number of items produced on the nth day after overhaul, X_n, is Poisson with mean μq^{n-1}, where q is a given number, $0 < q < 1$. Let $c > 0$ denote the cost of the overhaul and suppose that the service takes one day. The problem of finding a time at which to stop production for overhaul in order to maximize the return per day is then the problem of finding a stopping rule $N \ge 1$ to maximize $\mathrm{E}(S_N - c)/\mathrm{E}(N+1)$, where $S_n = X_1 + \ldots + X_n$.

To solve this problem, we first consider the problem of finding a stopping rule to maximize $\mathrm{E}(Y_N - \lambda T_N)$, where $Y_n = S_n - c$ and $T_n = n+1$. Let us see if the 1-sla is optimal. If we stop at stage n, we gain $S_n - c - \lambda(n+1)$; if we continue one stage and stop, we expect to gain

$$S_n + \mathrm{E}X_{n+1} - c - \lambda(n+2) = S_n + \mu q^n - c - \lambda(n+2).$$

Therefore, the 1-sla is

$$N_1 = \min\{n > 0 : \lambda \ge \mu q^n\} = \min\{n > 0 : n \ge \log(\mu/\lambda)/\log(1/q)\}. \tag{10}$$

The problem is monotone and the 1-sla is an optimal stopping rule of a fixed sample size, $N_1 = m$, where $m = 1$ if $\lambda \ge \mu$ and $m = \lceil \log(\mu/\lambda)/\log(1/q) \rceil$ if $\lambda < \mu$. Its expected return is simply

$$\begin{aligned} V(\lambda) &= \mathrm{E}(S_m - c - \lambda(m+1)) = \mu(1 + q + \ldots + q^{m-1}) - c - \lambda(m+1) \\ &= \mu(1-q^m)/(1-q) - c - \lambda(m+1). \end{aligned} \tag{11}$$

We set this expression to zero and solve for λ, which looks easy until we remember that m depends on λ. We illustrate the general method of solving for λ suggested in §6.1 on a simple numerical example.

Suppose $\mu = 3$, $q = .5$, and $c = 1$. As an initial guess at the optimal rate of return, let us take $\lambda_0 = 1.5$. The iteration involved in (3) requires that we iterate the following two equations in order:

$$m = \left\lceil \frac{\log(\mu/\lambda)}{\log(1/q)} \right\rceil \qquad \text{and} \qquad \lambda = \frac{\mu\frac{1-q^m}{1-q} - c}{m+1}.$$

On the first iteration, we find $m = \lceil 6.579 \rceil = 7$ and $\lambda_1 = 1.813$. Applying the iteration again, we find $m = \lceil 4.686 \rceil = 5$ and $\lambda_2 = 1.881$. On the third iteration we find $m = \lceil 4.431 \rceil = 5$, and we must therefore have $\lambda_3 = \lambda_2$. The iteration has converged in a finite number of steps. We overhaul every sixth day ($m = 5$) and find as the average return per day, $\lambda^* = 1.881$.

In this problem, the iteration converges in a finite number of steps whatever be the values of μ, q and c, because the value function, $V(\lambda)$, is piecewise linear. It is only in very simple problems that this will be the case.

§**6.5. An Inventory Problem.** A warehouse can hold W items of a certain stock. Each day a random number of orders for the item are received. The items are sold up to the number of items in stock; orders not filled are lost. Each item sold yields a net profit of $C_1 > 0$ (selling price of the item minus its cost). The warehouse may be restocked to capacity at any time by paying a restocking fee of $C_2 > 0$. On day n, orders for X_n items are received, $n = 1, 2, \ldots$, where $X_1, X_2, \ldots$ are independent and all have the same distribution, $f(x) = P(X_n = x)$ for $x = 0, 1, 2, \ldots$. The problem is to find a restocking time N (a stopping rule) to maximize the rate of return,

$$\mathrm{E}(\min(S_N, W)C_1 - C_2)/\mathrm{E}(N), \tag{12}$$

where $S_n = X_1 + \ldots + X_n$.

If C_2 were zero, we would restock every day ($N = 1$) and have a rate of return equal to $\mathrm{E}(\min(X_1, W))C_1$. Since $C_2 > 0$, it may be worthwhile to wait until the number of items on stock gets low before reordering.

To find the optimal restocking time, we first solve the stopping rule problem for maximizing the return

$$\begin{aligned} Y_n &= \min(S_n, W)C_1 - C_2 - n\lambda \\ &= (W - (W - S_n)^+)C_1 - C_2 - n\lambda. \end{aligned}$$

The one-stage look-ahead rule is

$$\begin{aligned} N_1 &= \min\{n \geq 1 | Y_n \geq \mathrm{E}(Y_{n+1}|\mathcal{F}_n)\} \\ &= \min\{n \geq 1 | (W - S_n)^+ - \mathrm{E}((W - S_n - X_{n+1})^+|\mathcal{F}_n) \leq \lambda/C_1\} \\ &= \min\{n \geq 1 | W - S_n \leq z\}, \end{aligned} \tag{13}$$

where,

$$z = \max\{u | u - \mathrm{E}(u - X)^+ \le \lambda / C_1\}. \tag{14}$$

The 1-sla is monotone and the theorems of Chapter 5 show that it is optimal. Thus, the optimal restocking rule has the form: *restock the warehouse as soon as the inventory has z items or less.* The optimal value of z may be found from (14) when λ is chosen to make $\mathrm{E}(Y_{N_1}) = 0$. In this problem, it may be simpler to calculate the ratio (12) for all stopping rules $N(z)$ of the form $N(z) = \min\{n \ge 1 | W - S_n \le z\}$ and find z that makes this ratio largest.

We carry out the computations when f is the geometric distribution,

$$f(x) = (1-p)p^x \quad \text{for} \quad x = 0, 1, \ldots.$$

First, we compute the numerator of (12) using $N(z)$. The distribution of $S_{N(z)} - (W - z)$ is the same geometric distribution, f , so that

$$\begin{aligned} \mathrm{E}\min(S_{N(z)}, W) &= W - z + \mathrm{E}\min(X, z) \\ &= W - z + (1 - p^z)p/(1-p). \end{aligned}$$

Second, to compute the denominator of (12), note that the geometric random variable X_n may be considered as the number of heads in a sequence of tosses of a coin with probability p of heads tossed until the first tail occurs. Therefore, $N(z) - 1$ represents the number of tails observed before $W - z$ heads, and so $N(z) - 1$ has a negative binomial distribution with probability of success $1 - p$ and $W - z$ fixed failures. Hence,

$$\mathrm{E}(N(z)) = 1 + (W - z)(1-p)/p.$$

Combining these two expectations and letting $\lambda(z)$ represent the ratio (1) when $N = N(z)$, we find,

$$\lambda(z) = [W - z + (1 - p^z)p/(1-p) - C_2]/[1 + (W - z)(1-p)/p],$$

where without loss of generality we have taken $C_1 = 1$, since the optimal rule depends only on the ratio C_2/C_1. After some wonderfully exciting and beautiful algebraic manipulations, we find that

$$\lambda(z-1) > \lambda(z) \quad \text{if and only if} \quad C_2 > p^z(W - z).$$

Since the right side of this inequality is decreasing in z, we find that the maximum of $\lambda(z)$ occurs at $z = 0$ if $C_2 \ge p(W-1)$ and at $z = n$ if

$$p^{n+1}(W - n - 1) \le C_2 \le p^n(W - n) \quad \text{for} \quad n = 1, \ldots, W-1.$$

As a numerical example, suppose that $C_1 = 1$, $W = 10$, and $p = 2/3$. Then if $C_2 > 6$, we have $z = 0$; we wait until we run out completely before reordering. (If

$C_2 > 10$, we are operating at a loss.) If $3.555 < C_2 < 6$, we have $z = 1$; we reorder when there is at most one left. Similarly down to: if $0 < C_2 < .026$, we have $z = 9$; we reorder as soon as at least one item is sold.

§6.6 Exercises.

1. *Selling an asset.* You can buy an item at a constant cost $c > 0$ and sell it when you like. Bids for the item come in one each day, $X_1, X_2, \ldots$ i.i.d. $F(x)$ and although it does not cost anything to observe these, it takes d days to obtain a new item. The problem is to find a stopping rule to maximize $\mathrm{E}(X_N - c)/\mathrm{E}(N + d)$. Assume that $\mathrm{P}(X > c) > 0$, and that $\mathrm{E}(X^2) < \infty$. Find an optimal rule and the optimal rate of return. Specialize to the case where F is the uniform distribution on $(0, 1)$.

2. *Maintenance.* (Taylor (1975), Posner and Zuckerman (1986), and Aven and Gaardner (1987)) A machine accumulates observable damage at discrete times through a series of shocks. Shocks occur independently at times $t = 1, 2, \ldots$, with probability, q, $0 < q < 1$, independent of time. When a shock occurs, the machine will accrue a certain amount of damage, assumed to be exponentially distributed with mean μ. If X_n denotes the damage accrued at time n, it is thus assumed that the X_n are i.i.d. with distribution $\mathrm{P}(X_n > x) = q\exp\{-x/\mu\}$ for $x > 0$, and $\mathrm{P}(X_n = 0) = 1 - q$. The total damage accrued to the machine by time n is $S_n = \sum_1^n X_j$. The machine breaks down at time n if S_n exceeds a given number, $M > 0$. The time of breakdown is thus $T = \min\{n \ge 1 : S_n > M\}$. A machine overhaul costs an amount $C > 0$. If the machine breaks down, the machine must be overhauled and there is an additional cost of $K > 0$. The problem is to decide when to overhaul the machine. The cost of overhauling the machine at stage n is thus $Y_n = C + K\mathrm{I}(n = T)$. To enforce stopping at T, we may put $Y_n = \infty$ on $\{n > T\}$. We want to choose a stopping rule, N, to *minimize* the cost per unit time, $\mathrm{E}(Y_N)/\mathrm{E}(N)$. This reduces to seeking a stopping rule to minimize $\mathrm{E}(Y_N - \lambda N)$, for some $\lambda > 0$.
(a) Find the 1-sla for the latter problem and show it is optimal.
(b) Show how to solve the original problem. (Choose λ as the root of $\lambda[M - \mu\log(Kq/\lambda)] = Cq\mu$.)

3. *Foraging.* Consider an animal that forages for food in spacially separated patches of prey. He feeds at one patch for awhile and then moves on to another. The problem of when to move to a new patch in order to maximize the rate of energy intake is addressed in the papers of Allan Oaten (1977), and Richard Green (1984, 1987). As an example, take the fisherman who moves from waterhole to waterhole catching fish. Suppose that in each waterhole there are initially n fish, where n is known. Assume that each fish has an exponential catch time at rate 1, and captures are independent events. This problem is treated in Example 5 of §5.4. Suppose that the expected time it takes to move from one waterhole to another is a known constant, $\tau > 0$. The problem is to find a stopping rule N to maximize the rate of return, $\mathrm{E}(N)/\mathrm{E}(X_N + \tau)$, where X_j is the jth order statistic of a sample of size n from the exponential distribution. Find an optimal rule and the optimal rate of return. As a numerical example, take $n = 10$ and $\tau = 1$.

4. *Attaining a goal.* Let $X_1, X_2, \ldots$ be independent Bernoulli trials with probability $1/2$ of success, and let S_n denote the sum, $\sum_1^n X_j$. Your goal is to achieve $S_n = a$,

where a is a fixed positive integer. If you attain your goal you win $c_1 > 0$, but the cost is 1 per trial. You may give up at any time by paying an additional amount, c_2. The real problem, however, is to choose a stopping rule, N, to maximize the rate of return, $c_1\mathrm{P}(N \geq a)/\mathrm{E}(N + c_2)$. Find the optimal rule and the optimal rate of return. (Refer to Exercise 4.8.)

Chapter 7. BANDIT PROBLEMS.

Bandit problems are problems in the area of sequential selection of experiments, and they are related to stopping rule problems through the theorem of Gittins and Jones (1974). In the first section, we present a description of bandit problems and give some historical background. In the next section, we treat the one-armed bandit problems by the method of the previous chapter. In the final section, we discuss the Theorem of Gittins and Jones which shows that the k-armed bandit problems may be solved by solving k one-armed problems. An excellent reference to bandit problems is the book of Berry and Fristedt, (1985).

§**7.1 Introduction to the Problem.** Consider a sequential decision problem in which at each stage there are k possible actions or choices of experiment. Choice of action j results in an observation being taken from the jth experiment, and you receive the numerical value of this observation as a reward. The observations you make may give you information useful in future choices of actions. Your goal is to maximize the present value of the infinite stream of rewards you receive, discounted in some way.

The name "bandit" comes from modeling these problems as a k-armed bandit, which is a slot machine with k arms, each yielding an unknown, possibly different distribution of payoffs. You do not know which arm gives you the greatest average return, but by playing the various arms of the slot machine you can gain information on which arm is best. However, the observations you use to gain information are also your rewards. You must strike a balance between gaining rewards and gaining information. For example, it is not good to always pull the arm that has performed best in the past, because it may have been that you were just unlucky with the best arm. If you have many trials to go and it only takes a few trials to clarify the matter, you can stand to improve your average gain greatly with only a small investment. Typically in these problems, there is a period of gaining information, followed by a period of narrowing down the arms, followed by a period of "profit taking", playing the arm you feel to be the best.

A more important modeling of these problems comes from clinical trials in which there are k treatments for a given disease. Patients arrive sequentially at the clinic and must be treated immediately by one of the treatments. It is assumed that response from treatment is immediate so that the effectiveness of the treatment that the present patient receives is known when the next patient must be treated. It is not known precisely which one of the treatments is best, but you must decide which treatment to give each patient, keeping in mind that your goal is to cure as many patients as possible. This may require you to give a patient a treatment which is not the one that looks best at the present time in order to gain information that may be of use to future patients.

To describe bandit problems more precisely, let the k reward distributions be denoted by $F_1(x|\theta_1),\ldots,F_k(x|\theta_k)$ where $\theta_1,\ldots,\theta_k$ are parameters whose exact values are not known precisely, but whose joint prior distribution is known and denoted by $G(\theta_1,\ldots,\theta_k)$. Initially, an action a_1 is chosen from the set $\{1,\ldots,k\}$ and then an observation, Z_1, the reward for the first stage, is taken from the distribution F_{a_1}. Based on this information, an action a_2 is then taken from the same action space, and an observation, Z_2, is taken from F_{a_2} and so on. It is assumed that given a_n and the parameters $\theta_1,\ldots,\theta_k$, Z_n is chosen from F_{a_n} independent of the past. A decision rule for this problem is a sequence $A=(a_1,a_2,\ldots)$ of functions adapted to the observations; that is, a_n may depend on past actions and observations,

$$a_n(a_1,Z_1,a_2,Z_2,\ldots,a_{n-1},Z_{n-1}).$$

It is hoped that no confusion results from using one symbol, a_n, to denote both the function of past observations and the action taken at stage n.

There is a discount sequence, denoted by $B=(\beta_1,\beta_2,\ldots)$, such that the jth observation is discounted by β_j where $0\le\beta_j\le 1$ for $j=1,2,\ldots$. The total discounted return is then $\sum_1^\infty \beta_j Z_j$. The problem is to choose a decision rule A to maximize the expected reward, $\mathrm{E}\sum_1^\infty \beta_j Z_j$. This problem is called the k-armed bandit problem. The one-armed bandit problem, mentioned in Exercise 1.4, is defined as the 2-armed bandit problem in which one of the arms always returns the same known amount, that is, the distribution F associated with one of the arms is degenerate at a known constant.

To obtain a finite value for the expected reward, we assume

(1) each distribution, F_j for $j=1,\ldots,k$, has finite first moment, and

(2) $\sum_1^\infty \beta_j<\infty$.

Two important special cases of the discount sequence are

(1) the *n-horizon uniform discount* for which $\beta_1=\ldots=\beta_n=1$ and $\beta_{n+1}=\beta_{n+2}=\ldots=0$, and

(2) the *geometric discount* in which $B=(1,\beta,\beta^2,\beta^3,\ldots)$, that is, $\beta_j=\beta^{j-1}$ for $j=1,2,\ldots$.

In the former, the payoff is simply $\sum_1^n Z_j$, the sum of the first n observations. The problem becomes one with a finite horizon which can in principle be solved by backward induction. In the latter, there is a time invariance in which the future after n stages looks like it did at the start except for the change from the prior distribution to the posterior distribution. We will treat mainly problems with geometric discount and independent arms, that is, prior distributions G for which $\theta_1,\ldots,\theta_k$ are independent, so that an observation on one arm will not influence your knowledge of the distribution of any other arm.

First, we give a little historical background to add perspective to what follows. Early work on these problems centered mainly on the finite horizon problem with Bernoulli trials. These problems were introduced in the framework of clinical trials by Thompson (1933) for two treatments with outcomes forming Bernoulli trials with success probabilities having independent uniform prior distributions on (0,1). Robbins (1952) reintroduced the problem from a non-Bayes viewpoint, and suggested searching for the minimax decision rule. In the Bernoulli case, he proposed the play-the-winner/switch-from-a-loser strategy, and discussed the asymptotic behavior of rules.

The first paper to search for Bayes decision rules for this problem is the paper of Bradt, Johnson and Karlin (1956). One of their important results, mentioned in Exercise 1.4, is that for the one-armed bandit with finite horizon and Bernoulli trials, if the known arm is optimal at any stage, then it is optimal to use that arm at all subsequent stages. Another important result is that for the 2-armed bandit with finite horizon and Bernoulli trials with success probabilities p_1 and p_2, if the prior distribution gives all its weight to points (p_1, p_2) on the line $p_1 + p_2 = 1$, then the 1-stage look-ahead rule is optimal; that is, it is optimal to choose the arm with the higher present probability of success. In addition, they conjecture that if the prior distribution gives all its weight to two points (a, b) and (b, a) symmetrically placed about the line $p_1 = p_2$, then again the 1-stage look-ahead rule is optimal. This conjecture was proved to be true by Feldman (1962).

These are about the only cases in which the optimal rule is easy to evaluate. In particular, in the important practical case of independent arms, the difficulty of computation of the optimal rule hindered real progress. One important result for the 2-armed Bernoulli bandit with independent arms and finite horizon is the "stay-on-a winner" principle, proved in Berry (1972). This principle states that if an arm is optimal at some stage and if it proves to be successful at that stage, then it is optimal at the following stage also. This was proved for the one-armed Bernoulli bandit with finite horizon by Bradt, Johnson and Karlin (1956). This was the state of affairs when the theorem of Gittins and Jones showed that for the k-armed bandit with independent arms and geometric discount, the problem can be solved by solving k one-armed bandit problems.

We treat a problem of a somewhat more general structure than that indicated above by letting the returns for each arm be an arbitrary sequence of random variables, (not necessarily exchangeable). Thus for each arm, say arm j, there is assumed to be a sequence of returns, $X_{j1}, X_{j2}, \ldots$ with an arbitrary joint distribution, subject to the condition that $\sup_n \mathrm{E}X_{jn}^+ < \infty$ for all $j = 1, \ldots, k$. It is assumed that the arms are independent, that is, that the sets $\{X_{11}, X_{12}, \ldots\}, \ldots, \{X_{k1}, X_{k2}, \ldots\}$ are independent sets of random variables. When an arm, say j, is pulled, the first random variable of the sequence X_{j1} is received as the reward. The next time j is pulled, the reward X_{j2} is received, etc. That the theorem of Gittins and Jones applies to this more general probabilistic structure has been proved by Varaiya, Walrand and Buyukkoc (1985). See also Mandelbaum (1986).

§**7.2 The one-armed bandit.** As a preliminary to the solution of the k-armed bandit with geometric discount and independent arms, it is important to understand the one-armed bandit. The one-armed bandit is a really a bandit problem with two arms, but one of the arms has a known i.i.d. distribution of returns, and so plays only a minor role. We first show how the one-armed bandit can be related to a stopping rule problem. We assume that arm 1 has an associated sequence of random variables, $X_1, X_2, \ldots$ with known joint distribution satisfying $\sup_n \mathrm{E}X_n^+ < \infty$. For the other arm, arm 2, the returns are assumed to be i.i.d. from a known distribution with expectation λ. We take the discount sequence to be geometric, $B = (1, \beta, \beta^2, \ldots)$ where $0 < \beta < 1$, and seek to find a decision rule $A = (a_1, a_2, \ldots)$ to maximize

$$V(A) = \mathrm{E}\left(\sum_1^\infty \beta^{j-1} Z_j | A\right). \tag{1}$$

First we argue that we may assume without loss of generality that the returns from arm 2 are degenerate at λ. Note that in the expectation above, any Z_j from arm 2 may be replaced by λ. However, the rule A may allow the actual values of these Z_j to influence the choice of the future a_j . But the statistician may produce his own private sequence of i.i.d. random variables from the distribution of arm 2 and use them in A in place of the actual values he sees. This produces a randomized decision rule which may be denoted by A^*. Use of A^* in the problem where the random variables are given to be degenerate at λ produces the same expected payoff as A does in the original problem.

The advantage of making this observation is that we may now assume that the decision rule A does not depend on Z_j when $a_j = 2$, since Z_j is known to be λ. Thus we assume that arm 2 gives a constant return of λ each time it is pulled.

We now show that if at any stage it is optimal to pull arm 2, then it is optimal to keep pulling arm 2 thereafter. This implies that if there exists an optimal rule for this problem, there exists an optimal rule with the property that every pull of arm 2 is followed by another pull of arm 2. Thus one need only decide on the time to switch from arm 1 to arm 2. This relates this problem to a stopping rule problem in which the stopping time is identified with the time of switching from arm 1 to arm 2. Without loss of generality, we may assume that arm 2 is optimal at the initial stage, and state the theorem as follows.

Theorem 1. *If it is initially optimal to use arm 2 in the sense that* $\sup_A V(A) = V^* = \sup\{V(A) : A \text{ such that } a_1 = 2\}$, *then it is optimal to use arm 2 always and* $V^* = \lambda/(1-\beta)$.

Proof. For a given $\epsilon > 0$, find a decision rule A such that $a_1 = 2$ and $V(A) \geq V^* - \epsilon$. Then,

$$\begin{aligned} V(A) &= \lambda + \beta \mathrm{E}\left(\sum_2^\infty \beta^{j-2} Z_j | A\right) \\ &= \lambda + \beta \mathrm{E}\left(\sum_1^\infty \beta^{j-1} Z_{j+1} | A\right) \\ &= \lambda + \beta \mathrm{E}\left(\sum_1^\infty \beta^{j-1} Z'_j | A^1\right) \\ &\leq \lambda + \beta V^*, \end{aligned}$$

where $A^1 = (a_2, a_3, \ldots)$ is the rule A shifted by 1, and $Z'_j = Z_{j+1}$. Thus, we have $V^* - \epsilon \le \lambda + \beta V^*$, or equivalently, $V^* \le (\lambda + \epsilon)/(1-\beta)$. Since $\epsilon > 0$ is arbitrary, this implies $V^* \le \lambda/(1-\beta)$, but this value is achievable by using arm 2 at each stage. ■

This theorem is also valid for the n-uniform discount sequence. In fact, Berry and Fristedt (1985) show that if the sequence $\{X_1, X_2, \ldots\}$ is exchangeable, then an optimal strategy for the problem has the property that every pull of arm 2 is followed by a pull of arm 2 if, and essentially only if, the discount sequence is what they call regular. The discount sequence B is said to be *regular* if it has increasing failure rate, that is, if $\beta_n / \sum_n^\infty \beta_j$ is non-decreasing on its domain of definition. That the above theorem is not true for such discount sequences is easily seen: If $B = \{.1, 1, 0, 0, 0, \ldots\}$, then B is regular, yet if X_1 is degenerate at 10, X_2 is degenerate at 0, and $\lambda = 0$, then clearly the only optimal strategy is to follow an initial pull of arm 2 with a pull of arm 1. Exactly what property of B is required for the above theorem seems to be unknown.

As a corollary, we see that there exists an optimal rule for this problem. It is either the rule that uses arm 2 at all stages, or the rule corresponding to the stopping rule $N \ge 1$ that is optimal for the stopping rule problem with payoff,

$$Y_n = \sum_1^n \beta^{j-1} X_j + \lambda \sum_{n+1}^\infty \beta^{j-1}. \tag{2}$$

In fact, we can say more.

Theorem 2. *Let $\Lambda(\beta)$ denote the optimal rate of return for using arm 1 at discount β,*

$$\Lambda(\beta) = \sup_{N \ge 1} \frac{\mathrm{E}(\sum_1^N \beta^{j-1} X_j)}{\mathrm{E}(\sum_1^N \beta^{j-1})}. \tag{3}$$

Then arm 2 is optimal initially if, and only if, $\lambda \ge \Lambda(\beta)$.

Proof. By Theorem 1, we may restrict attention to decision rules A specified by a stopping time N which represents the last time that arm 1 is used. The payoff using N is $\mathrm{E}(\sum_1^N \beta^{j-1} X_j + \lambda \sum_{N+1}^\infty \beta^{j-1})$, which for $N = 0$ is $\lambda/(1-\beta)$. Therefore, arm 2 is optimal initially if, and only if, for all stopping rules $N \ge 1$,

$$\mathrm{E}(\sum_1^N \beta^{j-1} X_j + \lambda \sum_{N+1}^\infty \beta^{j-1}) \le \lambda/(1-\beta)$$

or, equivalently,

$$\mathrm{E}(\sum_1^N \beta^{j-1} X_j) \le \lambda \mathrm{E}(\sum_1^N \beta^{j-1})$$

or, equivalently,

$$\mathrm{E}(\sum_1^N \beta^{j-1} X_j)/\mathrm{E}(\sum_1^N \beta^{j-1}) \le \lambda.$$

This is equivalent to $\Lambda(\beta) \leq \lambda$. ∎

The value $\Lambda(\beta)$ depends only on β and on the distribution of the returns from arm 1, $X_1, X_2, \ldots$. It is called the **Gittins index for arm 1** and it represents the indifference point: that value of λ for arm 2 in the one-armed bandit at which you would be indifferent between starting off on arm 1 and choosing arm 2 all the time.

§**7.3 The Gittins Index Theorem**. We return to the k-armed bandit with geometric discount and independent arms having returns denoted by

arm 1: $X(1,1), X(1,2), \ldots$
arm 2: $X(2,1), X(2,2), \ldots$
$\ldots$
arm k: $X(k,1), X(k,2), \ldots$

where it is assumed that the variables are independent between rows and that the first absolute moments exist and are uniformly bounded, $\sup_{k\geq 1, t\geq 1} \mathrm{E}|X(k,t)| < \infty$. The discount is β, where $0 \leq \beta < 1$, and we seek a decision rule $A = (a_1, a_2, \ldots)$ to maximize the total discounted return,

$$V(A) = \mathrm{E}(\sum_{t=1}^{\infty} \beta^{t-1} Z_t | A). \tag{4}$$

For each arm we may compute a Gittins index,

$$\Lambda_j = \sup_{N\geq 1} \mathrm{E} \sum_{t=1}^{N} \beta^{t-1} X(j,t) / \mathrm{E} \sum_{t=1}^{N} \beta^{t-1} \quad \text{for} \quad j = 1, \ldots, k \tag{5}$$

where we suppress β in the notation for Λ since β is held constant throughout this section.

The celebrated theorem of Gittins and Jones (1974) states that for the k-armed bandit with geometric discount and independent arms, it is optimal at each stage to select the arm with the highest index. We give a proof of this theorem due to Varaiya, Walrand and Buyukkoc (1983) and adopt their didactic strategy of presenting the proof first in the special case in which there are just two arms ($k = 2$) and where all the random variables are degenerate. We denote the returns from arm 1 by $X(1), X(2), \ldots$ and from arm 2 by $Y(1), Y(2), \ldots$. Thus, it is assumed that $X(1), X(2), \ldots$ and $Y(1), Y(2), \ldots$ are bounded sequences of real numbers. Let the Gittins indices be denoted by

$$\Lambda_X = \sup_{j\geq 1} \sum_1^j \beta^{t-1} X(t) / \sum_1^j \beta^{t-1}$$
$$\Lambda_Y = \sup_{j\geq 1} \sum_1^j \beta^{t-1} Y(t) / \sum_1^j \beta^{t-1}.$$

Since $X(t)$ is assumed bounded, the series $\sum_1^n \beta^{t-1} X(t)$ converges, so that there exists a value of j, possibly ∞, at which the supremum in the definition of Λ_X is taken on. In the following lemma, we suppose that s is this value of j so that $1 \leq s \leq \infty$.

Lemma 1. *Suppose the sequence* $X(1), X(2), \ldots$ *is non-random and bounded. If* $\Lambda_X = \sum_1^s \beta^{t-1}X(t)/\sum_1^s \beta^{t-1}$*, then for all* $j \le s$*,*

$$\sum_{t=j}^{s} \beta^{t-1}X(t) \ge \Lambda_X \sum_{t=j}^{s} \beta^{t-1}, \tag{6}$$

and for s *finite and all* $j > s$*,*

$$\sum_{t=s+1}^{j} \beta^{t-1}X(t) \le \Lambda_X \sum_{t=s+1}^{j} \beta^{t-1}. \tag{7}$$

Proof. Since we have both

$$\sum_1^s \beta^{t-1}X(t) = \Lambda_X \sum_1^s \beta^{t-1}, \quad \text{and}$$
$$\sum_1^{j-1} \beta^{t-1}X(t) \le \Lambda_X \sum_1^{j-1} \beta^{t-1} \quad \text{for all } j,$$

subtracting the latter from the former gives (6) when j is less than or equal to s, and gives (7) when $j > s$. ■

Inequality (6) implies that after $j-1 < s$ stages have elapsed, the new Gittins index (which is at least as great as $\sum_j^s \beta^{t-1}X(t)/\sum_j^s \beta^{t-1}$) is at least as great as the original index, Λ_X. Similarly, (7) shows that after exactly s stages have elapsed, the new Gittins index can be no greater than the original one. In fact, the following proof shows that if $\Lambda_X \ge \Lambda_Y$ and $\Lambda_X = \sum_1^s \beta^{t-1}X(t)/\sum_1^s \beta^{t-1}$, then it is optimal to start with at least s pulls of arm 1.

Theorem 3. *If the sequences* $X(1), X(2), \ldots$ *and* $Y(1), Y(2), \ldots$ *are non-random and bounded, if* $\Lambda_X \ge \Lambda_Y$ *then it is optimal initially to use arm 1.*

Proof. Assume $\Lambda_X \ge \Lambda_Y$ and let A be any rule. We will show that there is a rule A' that begins with arm 1 and gives at least as great a value as A. Then it is clear that the supremum of $V(A)$ over all A is the same as the supremum of $V(A)$ over all A that begin with arm 1.

Find s such that $\Lambda_X = \sum_1^s \beta^{t-1}X(t)/\sum_1^s \beta^{t-1}$ where $1 \le s \le \infty$. Let $k(t)$ denote the number of times that rule A calls for the use of arm 2 just before arm 1 is used for the tth time, $t = 1, 2, \ldots$. It may be that arm 2 is not used between pulls $t-1$ and t of arm 1 so that $k(t) = k(t-1)$ and it is possible that A does not use arm 1 t times, in which case $k(t)$ is defined as $+\infty$. We have $0 \le k(1) \le k(2) \le \ldots \le \infty$. We define a new decision rule A' that starts out with s X's followed by $k(s)$ Y's and then, if $k(s) < \infty$, following this with the same sequence of X's and Y's as in A. Subject to s and the first s $k(t)$'s being finite, the sequence of observations occurs in the following order for A:

$$Y(1),\ldots,Y(k(1)),X(1),Y(k(1)+1),\ldots,Y(k(2)),X(2),\ldots,Y(k(s)),X(s),Z(T+1),\ldots$$

where T is the time that the sth X occurs using decision rule A and $Z(T+1),\ldots$ represents the rest of this sequence beyond T. The sequence of observations occurs in the following order using A':

$$X(1),X(2),\ldots,X(s),Y(1),\ldots,Y(k(1)),\ldots,Y(k(2)),\ldots,Y(k(s)),Z(T+1),\ldots$$

We are to show that $V(A')-V(A)\geq 0$. Write this difference as $V(A')-V(A)=\Delta(X)-\Delta(Y)$, where $\Delta(X)$ represents the improvement in the value caused by shifting the X's forward, and $\Delta(Y)$ represents the loss due to shifting the Y's backward. Thus,

$$\begin{aligned}
\Delta(X) &= \sum_{1}^{s}\beta^{t-1}X(t)-\sum_{1}^{s}\beta^{t+k(t)-1}X(t)\\
&= \sum_{1}^{s}\beta^{t-1}(1-\beta^{k(t)})X(t)\\
&= \sum_{t=1}^{s}\beta^{t-1}X(t)\sum_{j=1}^{t}(\beta^{k(j-1)}-\beta^{k(j)})\\
&= \sum_{j=1}^{s}(\beta^{k(j-1)}-\beta^{k(j)})\sum_{t=j}^{s}\beta^{t-1}X(t)\\
&\geq \Lambda_X\sum_{j=1}^{s}(\beta^{k(j-1)}-\beta^{k(j)})\sum_{t=j}^{s}\beta^{t-1}\\
&= \Lambda_X\sum_{1}^{s}\beta^{t-1}(1-\beta^{k(t)}),
\end{aligned}$$

where $k(0)$ represents 0, and the inequality follows from (6). It is important to note that this computation is valid even if some of the $k(t)=\infty$, that is, even if A did not contain s X's. It is also valid if $s=\infty$. Similarly, we have

$$\begin{aligned}
\Delta(Y) &= \sum_{j=1}^{s}\beta^{j-1}\sum_{t=k(j-1)+1}^{k(j)}\beta^{t-1}Y(t)-\beta^{s}\sum_{t=1}^{k(s)}\beta^{t-1}Y(t)\\
&= \sum_{j=1}^{s}(\beta^{j-1}-\beta^{s})\sum_{t=k(j-1)+1}^{k(j)}\beta^{t-1}Y(t)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{s} \sum_{m=j}^{s} (\beta^{m-1} - \beta^{m}) \sum_{t=k(j-1)+1}^{k(j)} \beta^{t-1} Y(t) \\
&= \sum_{m=1}^{s} (\beta^{m-1} - \beta^{m}) \sum_{j=1}^{m} \sum_{t=k(j-1)+1}^{k(j)} \beta^{t-1} Y(t) \\
&= (1-\beta) \sum_{m=1}^{s} \beta^{m-1} \sum_{t=1}^{k(m)} \beta^{t-1} Y(t) \\
&\le \Lambda_Y (1-\beta) \sum_{m=1}^{s} \beta^{m-1} \sum_{t=1}^{k(j)} \beta^{t-1} \\
&= \Lambda_Y \sum_{m=1}^{s} \beta^{m-1} (1 - \beta^{k(m)}).
\end{aligned}
$$

The inequality follows from the definition of Λ_Y. This computation is also valid if $s = \infty$ or if some of the $k(t) = \infty$. Now, using the assumption that $\Lambda_X \ge \Lambda_Y$, we find that $\Delta(X) - \Delta(Y) \ge 0$, as was to be shown. ■

The proof of the general theorem is similar, but at the crucial step involving the inequalities, we will not be able to interchange summation and expectation because the limits of summation are random. To circumvent this difficulty, the following two lemmas will be used. We deal with possibly randomized stopping rules by allowing the increasing sequence of σ-fields,

$$
\mathcal{F}(1) \subset \mathcal{F}(2) \subset \ldots \subset \mathcal{F}(\infty) \tag{8}
$$

to be such that $\mathcal{F}(t)$ is the σ-field generated by $X(1), \ldots, X(t)$ and any number of other random variables independent of $X(t+1), X(t+2), \ldots$. To say now that a random variable Z is $\mathcal{F}(t)$-measurable means essentially that Z and $\{X(t+1), X(t+2), \ldots\}$ are conditionally independent given $X(1), \ldots, X(t)$. In particular, we have

$$
\begin{aligned}
\mathrm{E}X(t+1)Z &= \mathrm{E}(\mathrm{E}\{X(t+1)Z | X(1), \ldots, X(t+1)\}) \\
&= \mathrm{E}(X(t+1)\mathrm{E}\{Z | X(1), \ldots, X(t)\})
\end{aligned} \tag{9}
$$

for any $\mathcal{F}(t)$-measurable Z. This observation is useful in the proof of the following lemma.

Lemma 2. *Let $X(t)$ be a sequence of random variables such that $\sup_t \mathrm{E}|X(t)| < \infty$, let $0 < \beta < 1$ and let Λ_X denote the Gittins index. Then, for every stopping rule N, and every sequence of random variables $\alpha(t)$, $t = 1, 2, \ldots$, such that $\alpha(t)$ is $\mathcal{F}(t-1)$-measurable and $1 \ge \alpha(1) \ge \alpha(2) \ge \ldots \ge 0$ a.s., we have*

$$
\mathrm{E} \sum_{t=1}^{N} \alpha(t) \beta^{t-1} X(t) \le \Lambda_X \mathrm{E} \sum_{t=1}^{N} \alpha(t) \beta^{t-1}. \tag{10}
$$

Proof. Let $W(t) = \beta^{t-1}(X(t) - \Lambda_X)$. Then, the definition of Λ_X implies that for every stopping rule $N \geq 1$,

$$\mathrm{E}\sum_{t=1}^{N} W(t) \leq 0.$$

For any stopping rule N, $\mathrm{I}(N \geq t)$ is $\mathcal{F}(t-1)$-measurable. Hence, from (9),

$$\begin{aligned}
\mathrm{E}\sum_{t=1}^{N} W(t) &= \mathrm{E}\sum_{n=1}^{=\infty} I(N = n)\sum_{t=1}^{n} W(t) \\
&= \mathrm{E}\sum_{t=1}^{\infty} W(t)\sum_{n=t}^{=\infty} I(N = n) \\
&= \mathrm{E}\sum_{t=1}^{\infty} W(t)\gamma(t) \leq 0,
\end{aligned}$$

where $\gamma(t) = \mathrm{P}(N \geq t | X(1), \ldots, X(t-1))$. Any sequence, $1 \geq \gamma(1) \geq \gamma(2) \geq \ldots \geq 0$ a.s. with $\gamma(t)$ $\mathcal{F}(t-1)$-measurable, determines a stopping rule N such that $\mathrm{P}(N \geq t | \mathcal{F}(t-1)) = \gamma(t)$. Hence, the hypothesis that $\mathrm{E}\sum_1^N W(t) \leq 0$ for all stopping rules N is thus equivalent to the hypothesis that $\mathrm{E}\sum_{t=1}^{\infty} W(t)\gamma(t) \leq 0$ for all sequences $\gamma(t)$ such that $\gamma(t)$ is $\mathcal{F}(t-1)$-measurable and $1 \geq \gamma(1) \geq \gamma(2) \geq \ldots \geq 0$ a.s. Now, since

$$\begin{aligned}
\mathrm{E}\sum_{t=1}^{N} \alpha(t)W(t) &= \mathrm{E}\sum_{n=1}^{=\infty} I(N = n)\sum_{t=1}^{n} \alpha(t)W(t) \\
&= \mathrm{E}\sum_{t=1}^{\infty} \alpha(t)W(t)\sum_{n=t}^{=\infty} I(N = n) \\
&= \mathrm{E}\sum_{t=1}^{\infty} W(t)\gamma(t) \leq 0,
\end{aligned}$$

where $\gamma(t) = \mathrm{E}\{\alpha(t)I(N \geq t) | X(1), \ldots, X(t-1)\}$, the conclusion follows. ■

The next lemma provides the required generalization of (6) of Lemma 1. Since $\mathrm{E}|X_n|$ is assumed to be bounded, conditions A1 and A2 are satisfied and there exists an optimal stopping rule for every λ. In particular, there exists a rule N^* that attains the Gittins index.

Lemma 3. *Let $X(t)$ be a sequence of random variables such that $\sup_t \mathrm{E}|X(t)| < \infty$, let $0 < \beta < 1$, and let N^* denote a stopping rule that attains the Gittins index, Λ_X. Then, for every sequence of random variables $\xi(t)$, $t = 1, 2, \ldots$ such that $\xi(t)$ is $\mathcal{F}(t-1)$-measurable and $0 \leq \xi(1) \leq \xi(2) \leq \ldots \leq 1$ a.s., we have*

$$\mathrm{E}\sum_{t=1}^{N^*} \xi(t)\beta^{t-1}X(t) \geq \Lambda_X \mathrm{E}\sum_{t=1}^{N^*} \xi(t)\beta^{t-1}. \tag{11}$$

Proof. Since the Gittins index is attained at N^*,

$$\mathrm{E}\sum_{t=1}^{N^*}\beta^{t-1}X(t) = \Lambda_X \mathrm{E}\sum_{t=1}^{N^*}\beta^{t-1}.$$

From Lemma 2 with $\alpha(t) = 1-\xi(t)$, we have

$$\mathrm{E}\sum_{t=1}^{N^*}(1-\xi(t))\beta^{t-1}X(t) \le \Lambda_X \mathrm{E}\sum_{t=1}^{N^*}(1-\xi(t))\beta^{t-1}.$$

Subtracting the latter from the former gives the result. ∎

We now turn to the general problem with k independent arms, and denote the sequence of returns from arm j by $X(j,1), X(j,2), \ldots$ for $j = 1,\ldots,k$. It is assumed that the sets $\{X(1,t)\}_{t=1}^{\infty},\ldots,\{X(k,t)\}_{t=1}^{\infty}$ are independent, and that $\sup_{j,t}\mathrm{E}|X(j,t)| < \infty$. We shall be dealing with random variables $k(j,t)$ that depend upon the sequence $X(j,1), X(j,2),\ldots$ only through the values of $X(j,1),\ldots,X(j,t-1)$, though possibly on some of the $X(m,n)$ for $m \ne j$. Such random variables are measurable with respect to the σ-field generated by $X(j,1),\ldots,X(j,t-1)$, and $X(m,n)$ for $m \ne j$ and $n = 1,2,\ldots$. We denote this σ-field by $\mathcal{F}(j,t-1)$.

Any decision rule that at each stage chooses an arm that has the highest Gittins index is called a **Gittins index rule**.

Theorem 4. *For a k-armed bandit problem with independent arms and geometric discount, any Gittins index rule is optimal.*

Proof. Suppose that $\Lambda_1 = \max \Lambda_j$. Let A be an arbitrary decision rule. We prove the theorem by showing that there is a rule A' that begins with arm 1 and gives at least as great a value as A.

Let $k(j,t)$ denote the (random) time that A uses arm j for the tth time, $j = 1,\ldots,k$, $t = 1,2,\ldots$, with the understanding that $k(j,t) = \infty$ if arm j is used less than t times. The value of A may then be written

$$\begin{aligned} V(A) &= \mathrm{E}\{\sum_{t=1}^{\infty}\beta^{t-1}Z(t)|A\} \\ &= \mathrm{E}\sum_{j=1}^{k}\sum_{t=1}^{\infty}\beta^{k(j,t)-1}X(j,t). \end{aligned}$$

Let N^* denote the stopping rule that achieves the supremum in

$$\Lambda_1 = \sup_{N\ge 1}\mathrm{E}(\sum_1^N \beta^{t-1}X(1,t))/\mathrm{E}(\sum_1^N \beta^{t-1}),$$

and let T denote the (random) time that A uses arm 1 for the N^*th time, $T = k(1, N^*)$. Define the decision rule A' as follows:
(a) use arm 1 at times $1, 2, \ldots, N^*$, and then if $N^* < \infty$,
(b) use the arms $j \neq 1$ at times $N^* + 1, \ldots, T$ in the same order as given by A, and then if $T < \infty$,
(c) continue according to A from time T on.

Let $k'(j, t)$ denote the time when A' uses arm j for the tth time, so that $k'(1, t) = t$ for $t = 1, \ldots, N^*$. Finally, let $m(j)$ denote the number of times that arm j is used by time T, so that $m(1) = N^*$. Then,

$$
\begin{aligned}
V(A') - V(A) &= \mathrm{E}\sum_{j=1}^{k}\sum_{t=1}^{m(j)}(\beta^{k'(j,t)-1} - \beta^{k(j,t)-1})X(j,t) \\
&= \mathrm{E}\sum_{t=1}^{N^*}(\beta^{t-1} - \beta^{k(1,t)-1})X(1,t) \\
&\quad - \mathrm{E}\sum_{j=2}^{k}\sum_{t=1}^{m(j)}\beta^{t-1}(\beta^{k(j,t)-t} - \beta^{k'(j,t)-t})X(j,t) \\
&= \mathrm{E}\sum_{t=1}^{N^*}\xi(t)\beta^{t-1}X(1,t) - \sum_{j=2}^{k}\mathrm{E}\sum_{t=1}^{m(j)}\alpha(j,t)\beta^{t-1}X(j,t)
\end{aligned} \tag{12}
$$

where

$$
\begin{aligned}
\xi(t) &= 1 - \beta^{k(1,t)-t}, \quad \text{and} \\
\alpha(j,t) &= \beta^{k(j,t)-t} - \beta^{k'(j,t)-t}.
\end{aligned}
$$

Since $k(1, t) - t$ represents the number of times that an arm other than arm 1 has been pulled by the time the tth pull of arm 1 occurs, we have that $k(1, t) - t$ is $\mathcal{F}(1, t-1)$-measurable and nondecreasing in t a.s. so that $\xi(t)$ is $\mathcal{F}(1, t-1)$-measurable and $0 \leq \xi(1) \leq \xi(2) \leq \ldots \leq 1$ a.s. Thus from Lemma 3, we have

$$
\mathrm{E}\sum_{t=1}^{N^*}\xi(t)\beta^{t-1}X(1,t) \geq \Lambda_1\mathrm{E}\sum_{t=1}^{N^*}\xi(t)\beta^{t-1}. \tag{13}
$$

For $j > 1$, $\alpha(j, t) = \beta^{k(j,t)-t}(1 - \beta^{k'(j,t)-k(j,t)})$. Since $k(j, t) - t$ is $\mathcal{F}(j, t-1)$-measurable and nondecreasing and since $k'(j, t) - k(j, t)$ is equal to N^* minus the number of times arm 1 is pulled before the tth pull of arm j, and hence is $\mathcal{F}(j, t-1)$-measurable and nonincreasing, we find that $\alpha(j, t)$ is $\mathcal{F}(j, t-1)$-measurable and $1 \geq \alpha(j, 1) \geq \alpha(j, 2) \geq \ldots \geq 0$. Hence from Lemma 2, we have for $j = 2, \ldots, k$,

$$
\mathrm{E}\sum_{t=1}^{m(j)}\alpha(j,t)\beta^{t-1}X(j,t) \leq \Lambda_j\mathrm{E}\sum_{t=1}^{m(j)}\alpha(j,t)\beta^{t-1}. \tag{14}
$$

Combining (13) and (14) into (12) and recalling that $\Lambda_j \leq \Lambda_1$ for all $j > 1$, we find

$$V(A') - V(A) \geq \Lambda_1 \mathrm{E} \sum_{j=1}^{k} \sum_{t=1}^{m(j)} (\beta^{k'(j,t)-1} - \beta^{k(j,t)-1}).$$

This last expectation is zero since it is just $V(A') - V(A)$ with all payoffs put equal to 1. ■