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:

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

$\psi_\infty(x_1, x_2, \ldots)$ 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 ϕ 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 $V(\phi)$*.

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 ϕ 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 ϕ 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

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

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

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

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

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

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

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

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,

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

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 $P(X_j = 1) = p$, 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 $E(Y_N)$. 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 $X_1, \ldots, X_n$,

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

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

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 kth 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.