THE THEORY OF OPTIMAL STOPPING

RICHARD R. WEBER

DOWNING COLLEGE

CAMBRIDGE

Preface

The following essay is submitted as a partial fulfillment of the examination requirements for Part III of the Mathematical Tripos. No work was undertaken upon it before October, 1974, and I did not at any time discuss its specific content or form with anyone.

I would like to thank Dr. Doug Kennedy for undertaking to set the essay and for suggesting the main references. Dr. Bruce Brown receives my appreciation for an unwitting and extended loan of his copy of the book, "Great Expectations". I am also indebted to Dr. Peter Nash for providing copies of the papers on dynamic allocation indices.

Richard Weber

May, 1975

Downing College, Cambridge

Contents

Chapter 1. Introduction to Examples and the General Formulation

Optimal stopping problems are concerned with the control of random sequences in gambling and statistical decision. Often one desires to know the optimal instant to break off playing a game or to stop sampling in an inference problem.

A general theory that could give some insight to such problems was not seriously investigated until the last decade, beginning with E.B. Dynkin (1960). However several optimal stopping problems are quite famous for both their long history and attractive form. As an introduction to a subject which is firmly rooted in intuition we describe three problems. The first is an example in decision theory, the second in statistical sequential inference, and the third in the statistical design of experiments.

The chapter concludes with a formulation of the general optimal stopping problem on random sequences.

1.1 Three Problems of Optimal Stopping

1.1.1 The Secretary Problem

The Secretary, or Dowry, Problem has a long history, first appearing as a subject for discussion in a "Scientific American" article of 1960. Its solution was suggested there and then proved optimal by Dynkin in 1963.

The problem concerns that of an employer who must hire a secretary from among a group of n girls. At each interview he is only able to discern how the girl being interviewed compares with those whom he has seen previously. At the interview he must decide to hire the girl or reject her without any possibility of recall. His objective is to maximize, by some choice policy, the probability that he will select the best of the n girls.

1.1.2 The Sequential Probability Ratio Test

In 1947 A. Wald investigated the problem of hypothesis testing by sequential sampling. Suppose that $x_1, x_2, \ldots$ are independent, identically distributed samples from a distribution with density f. We wish to test the simple alternatives, H_0: $f = f_0$ against H_1: $f = f_1$. Costs are incurred for taking each sample as well as for ultimately taking an incorrect decision.

The desire is to take as few samples as possible while choosing between H_0 and H_1 with the best confidence possible. The search is essentially for a time, t, which tells us when to stop sampling and a decision rule, δ, which then tells us how to choose. Such a pair, (δ, t), is called a sequential decision procedure. The calculation of δ given t is only a standard hypothesis test. It is the choice of the stopping time t, which may depend on $x_1, \ldots, x_t$, that is an optimal stopping problem.

1.1.3 The Two-Armed Bandit Problem

Beyond considering the control of just one stochastic sequence, one might hope to control several simultaneously. The two arms, (1,2), of a two-armed bandit produce prizes or not when pulled. Arm i will produce a prize with probability p_i and will return nothing with probabilty $1 - p_i$, $i = 1,2$.

The interesting problems arise when one or both of p_1, p_2 are unknown and so must be infered from sampling on the arms. We are faced with trying to decide when to stop playing on one arm and play the other. Of course the desire is to maximise the total number of prizes obtained, either as an average number per play or as a total number when discounting operates with time.

The problem was first discussed by H. Robbins in 1952, but the form of the optimal rule for the case of unknown p_1 and p_2 was only described as recently as 1972 by J.C. Gittins and D.M. Jones.

These three problems will be discussed in the course of this essay as applications of the theory developed. We begin now by setting forth the general context and notation for optimal stopping problems.

1.2 General Formulation

We give definitions for random sequences, stopping times, and their associated rewards as considered in the essay.

1.2.1 Definition

A stochastic sequence $\{z_n, F_n\}$ is defined by:

(i) (Ω, F, P) is a probability space.

(ii) $\{F_n\}_1^\infty$ is an increasing sequence of sub σ- algebras.

(iii) $\{z_n\}_1^\infty$ is a sequence of random variables where z_n is F_n measurable, and takes values in $(-\infty, \infty]$.

(iv) $Ez_n^- < \infty$ for all n.

1.2.2 Definition

The non-negative, integer-valued random variable t is said to be an extended stopping time (variable or rule) if the event $[t = n]$ is in F_n all n. It is said to be a stopping time (variable or rule) if in addition $P(t < \infty) = 1$, ie. t takes the extended integer value ∞ with probabilty zero.

Given a stochastic sequence $\{z_n, F_n\}$, for the stopping problem on this sequence define:

$C = \{$ stopping times $t : Ez_t^- < \infty \}$

$\tilde{C} = \{$ extended stopping times $t : \tilde{E}z_t^- < \infty \}$ where $\tilde{E}$ is defined by

$$\tilde{E}z_t = E(z_t : t < \infty) + E(\lim_{n\to\infty} z_n : t = \infty)$$

Note that the restriction to times such that $Ez_t^- < \infty$ is only for convenience. For if t is any time, by letting $t' = \begin{cases} t \\ 1 \end{cases}$ as $E(z_t|F_1) \gtrless z_1$ then $Ez_{t'}^- \leqslant Ez_1^- < \infty$ and $Ez_{t'} \geqslant Ez_t$. So there is a $t' \varepsilon C$ for which the expected value of the stopped sequence is at least as large as that stopped by t.

That $\tilde{E}$ is the appropriate operator within $\tilde{C}$ will become clear in Chapter 3. But if $\{z_n\}$ were -1,-1,-1,... clearly we don't want to take $z_\infty = 0$ for then $\sup_{t\varepsilon\tilde{C}}\tilde{E}(z_t + 2) = 1$ does not equal $\sup_{t\varepsilon\tilde{C}}\tilde{E}z_t + 2 = 2$. In fact $\tilde{E}$ is precisely the operator that keeps $\sup_{t\varepsilon\tilde{C}}\tilde{E}(z_t + a) = \sup_{t\varepsilon\tilde{C}}\tilde{E}z_t + a$.

1.2.3 Definition

Given a stochastic sequence $\{z_n, F_n\}$ its value over C or $\tilde{C}$ is defined by $s = \sup_{t\varepsilon C} Ez_t$ or $\tilde{s} = \sup_{t\varepsilon\tilde{C}}\tilde{E}z_t$ respectively.

A time $t\varepsilon$ C or $\tilde{C}$ satisfying $Ez_t = s$ or $\tilde{E}z_t = \tilde{s}$ respectively is called a $(0,s)$ or $(0,\tilde{s})$-optimal rule.

A time $t\varepsilon C$ is called (ε,s)-optimal if $Ez_t \geqslant s - \varepsilon$.

Observe that without loss of generality the reward for stopping $\{z_n, F_n\}$ at n is taken as z_n. If it were actually some function, $f_n(z_1,\ldots,z_n)$, then a simple redefinition of z_n would cast the problem in the appropriate form.

1.3 General Problems of Optimal Stopping

Under the above formulation of the stopping problem on a stochastic sequence, the purpose of this essay will be to answer the following questions:

(a) What is s ($\tilde{s}$) ? Can it be computed given $\{z_n, F_n\}$?

(b) Do $(0,s)$, $(0,\tilde{s})$, (ε,s) or $(\varepsilon,\tilde{s})$-optimal times exist ?

(c) What is the form of an optimal stopping rule when it exists ?

We will answer these questions in two restricted contexts before stating the general results. Thereby it is hoped to make clear the way in which intuition might guide to develop the general theory from scratch.

Chapter 2. The Optimal Bounded Stopping of Random Sequences

In this chapter we formulate and solve the optimal stopping problem within the class of stopping times which are bounded by a fixed integer N. This inroad to the general problem proves to be a fruitful beginning. Not only does the bounded problem have a complete solution of interest in itself, but also, as we shall see in Chapter 3, its limiting form as $N \to \infty$ does in a sense describe the behavior of the general problem.

Many problems are of the bounded type in their own right. The Secretary problem of 1.1.1 is one such and its solution is derived.

2.1 Solution of the Bounded Problem

2.1.1 Definitions

Consider stopping times restricted to an interval and let $C_n^N = \{ t\varepsilon C : n < t < N \}$ $s_n^N = \sup_{t\varepsilon C_n^N} Ez_t$. For convenience write: $C^N = C_1^N$ and $s^N = s_1^N$.

Clearly the only stopping time in C_N^N is $t = N$. An intuitively likely construction of t_n^N optimal in C_n^N would take $t_n^N = n$ unless the expected reward of taking another step and applying the rule t_{n+1}^N were greater than the present reward, z_n. We show that this "backward construction" does produce the optimal rule.

2.1.2 Theorem [ref. Chow, et al. p. 50]

Define: $\gamma_N^N = z_N$ and $\gamma_n^N = \max\{ z_n , E(\gamma_{n+1}^N | F_n) \}$

Let $t_n^N = \min\{ i : i \geqslant n \text{ and } z_i = \gamma_i^N \}$.

Then t_n^N is optimal in C_n^N and $s_n^N = E\gamma_n^N = Ez_{t_n^N}$.

proof: (by backward induction on n)

True when $n = N$. Assume true for n and let $t \in C_{n-1}^N$, $A \in F_{n-1}$, $t' = \max(n,t)$. [we omit dP in the following integrals]

$$\int_A z_t = \int_{A\cap(t=n-1)} z_{n-1} + \int_{A\cap(t \geqslant n)} z_{t'}$$

$$= \int_{A\cap(t=n-1)} z_{n-1} + \int_{A\cap(t \geqslant n)} E(E(z_{t'}|F_n)|F_{n-1})$$

$$\leqslant \int_{A\cap(t=n-1)} z_{n-1} + \int_{A\cap(t \geqslant n)} E(\gamma_n^N|F_{n-1})$$

$$\leqslant \int_A \gamma_{n-1}^N$$

while $\int_A z_{t_{n-1}^N} =$

$$\int_{A\cap(z_{n-1} \geqslant E(\gamma_n^N|F_{n-1}))} z_{n-1} + \int_{A\cap(z_{n-1} < E(\gamma_n^N|F_{n-1}))} E(\gamma_n^N|F_{n-1})$$

$$= \int_A \max\{ z_{n-1} , E(\gamma_n^N|F_{n-1}) \} = \int_A \gamma_{n-1}^N$$

Unfortunately the computations $\gamma_n^N = \max\{ z_n , E(\gamma_{n+1}^N|F_n) \}$ are not going to be easy to carry out. We can make a simplification when γ_n^N is a random variable depending only on z_n, rather than on all the past history F_n. The optimal t_n^N will then choose to stop or not on the basis of only looking at the current state. Such memoryless or Markov nature is a feature of very many optimal stopping problems.

2.2 Solution of the Bounded Problem: Markov Case

2.2.1 Definitions

The stochastic sequence $\{z_n,F_n\}$ is said to have a stationary Markov representation if there exists a Markov sequence $\{x_n\}$ with state space E and transition probabilities P such that $F_n = B(x_n)$ and $z_n = g(x_n)$ where g is F_n- measurable, all n.

We write: $E_x g(x_1) = \int_E g(x_1)\, dP(x,x_1) = \int_E g(x_1)\, dP_x$.

Consider functions f mapping $E \to R$ and define:

$$L = \{ \text{ Borel-measurable } f : -\infty < f(x) \leqslant \infty \text{ and } E_x f^-(x_n) < \infty \text{ for all } n \text{ and } x \varepsilon E \}$$

To restrict attention to only those g which are in L does no more than simply ensure that the stopping time $t = n$ is in C. With this formulation Theorem 2.1.2 can be neatly restated.

2.2.2 Lemma

Define an operator $\overline{Q}:L \to L$ by $\overline{Q}f(x) = \max\{ g(x) , E_x f(x_1) \}$ If $z_n = g(x_n)$ in a Markov representation and $g\varepsilon L$, then with the notation of 2.1.2, $\gamma_n^N = \overline{Q}^{N-n} g(x_n)$.

proof: direct from the definitions

2.2.3 Lemma [ref. Shiryaev p. 23]

Define an operator $Q:L \to L$ by $Qf(x) = \max\{ f(x) , E_x f(x_1) \}$

Then $\overline{Q}^n g(x) = Q^n g(x)$ for all n and $x\varepsilon E$.

proof:

True for $n = 1$. Proceed by induction: $E_x Qf(x_1) \geqslant E_x f(x_1)$ hence $Q^2 g(x) = \max\{ Qg(x) , E_x Qg(x) \} = \max\{ g(x) , E_x Qg(x) \} = \overline{Q}^2 g(x)$ etc.

2.2.4 Theorem

Suppose $x_1, x_2, \ldots$ is a Markov random sequence and $g\varepsilon L$.

Then $s^N(x) = \sup_{t\varepsilon C^N} E_x g(x_t) = Q^N g(x) = \max\{ g(x) , E_x s^{N-1}(x_1) \}$

and $t^N = \min\{ i : i \leqslant N \text{ and } s^{N-n}(x_n) = g(x_n) \}$

proof: a consequence of the lemmas and definitons

This is just the statement that starting in state x and restricted to not more than N steps the optimal rule will choose the better of the two options:

(i) Stop now - receive $g(x)$

(ii) Take another step - receive on average $E_x s^{N-1}(x_1)$.

2.3 Solution of the Secretary Problem

Consider n objects indexed by 1,2,...,n permutated randomly with all permutations equally likely. Although observation of the objects does not reveal their true indices, comparison between two will disclose which is better. Examining the objects one by one we wish to stop at a t such that P(t^{th} object examined has index 1) is maximized.

2.3.1 Theorem [ref. Shiryaev pp. 46-48; Chow et al. pp. 51-52; Dynkin (1963) pp. 628-629]

The optimal rule for choosing the maximum of n objects as described above is to pass over the first k(n)-1 objects and then to choose the first to appear which is better than all the previous objects, where: $\frac{1}{n-1} + \dots + \frac{1}{k(n)} \leqslant 1 < \frac{1}{n-1} + \dots + \frac{1}{k(n)-1}$

and so $k(n) \sim n/e$

proof:

Let $x_0 = 1$ and x_{i+1} = the position in the observed sequence of the first object which is better than the object in position x_i.

(eg. if we were to see 10 objects as: 2,6,4,1,7,3,10,9,5,8, then $x_0=1$ $x_1=2$ $x_2=5$ $x_3=7$.)

Clearly the sequence x_i terminates at some $i' \leqslant n$, so let $x_i=0$ for all $i > i'$. (eg. $x_4=x_5=\dots= 0$ in the above)

Now suppose $x_i = b_i$. Then the first b_i-1 objects are simply arranged in one of the equally-likely random permutations of b_i-1 ordered objects. So we can deduce that they will have no effect on the distribution of x_{i+1} and can write:

$$P(x_{i+1}= b_{i+1} \mid x_i = b_i,\ x_{i-1}= b_{i-1}, \dots, x_1 = b_1) =$$

$$P(x_{i+1}= b_{i+1} \mid x_i = b_i) = \frac{P(x_{i+1}= b_{i+1} \text{ and } x_i = b_i)}{P(x_i = b_i)}$$

$$= \frac{P(\ b_{i+1}^{th} \text{ and } b_i^{th} \text{ objects are the 1st and 2nd best of the first } b_{i+1}\ )}{P(\ b_i^{th} \text{ object is the best of the first } b_i\ )}$$

$$= \frac{1/b_{i+1}(b_{i+1}-1)}{1/b_i} = \frac{b_i}{b_{i+1}(b_{i+1}-1)}$$

This shows that $x_0, x_1, ..$ is a Markov chain with transition probabilities: $P(0,0) = 1$ $\quad P(x,y) = 0$ when $x \geqslant y$

$$P(x,0) = \frac{x}{n} \qquad P(x,y) = \frac{x}{y(y-1)} \quad \text{when} \quad x < y$$

[note: $P(x,0) = 1 - \sum_{x+1}^{n} \frac{x}{y(y-1)} = \frac{x}{n}$]

Then $P(\ x_t$ is the position of the best object)

$$= \sum_{y=1}^{n} \frac{y}{n} P(x_t = y | x_0 = 1) = E_1 \frac{x_t}{n}$$

So in the formulation of the optimal stopping problem for a Markov random sequence, we are trying to maximize $E_1 g(x_t)$ where $g(x) = x/n$ $[\varepsilon L]$. Since $x_i = 0$ for all $i > n$ the optimal rule will lie in C^n. Hence Theorem 2.2.4 applies and

$$s^1(x) = \max\{ \frac{x}{n}, \sum_{y=x+1}^{n} \frac{x}{y(y-1)} \frac{y}{n} \} = \max\{ \frac{x}{n}, \frac{x}{n}[\frac{1}{x} + \dots + \frac{1}{n-1}] \}$$

$$\begin{aligned} &= x/n \quad \text{if} \quad x \geqslant k(n) \\ &> x/n \quad \text{if} \quad x < k(n) \end{aligned} \qquad \text{where } k(n) \text{ is defined as above.}$$

Continuing the construction it is clear that $s^i(x) \ \substack{= \\ >}\ x/n$ as $x \ \substack{\geqslant \\ <}\ k(n)$ $i = 1,2,3...$ and the optimal stopping time is:

$t = \min\{ i : s^{n-i}(x_i) = x_i/n \} = \min\{ i : x_i \geqslant k(n) \}$

Note: $1 \sim \frac{1}{n-1} + \dots + \frac{1}{k(n)} \sim \int_{k(n)}^{n-1} 1/x \, dx = \log_e(\frac{n-1}{k(n)})$

so $k(n) \sim n/e$ and the probability of success is

$$E \frac{x_t}{n} = \frac{1}{n} \sum_{k(n)}^{n} \frac{k(n)-1}{j-1} \frac{1}{j} j = \frac{k(n)-1}{n} \sum_{k(n)}^{n} \frac{1}{j-1} \sim \frac{k(n)}{n} \log(\frac{n-1}{k(n)-1})$$

$$\sim \frac{1}{e} \cong 0.368$$

Hence we have the rather remarkable fact that no matter how large the total number of objects it is always possible to choose the best with a probability greater than 0.368

Observe further that the probability that we are forced to take the last object unsuccessfully is P(best object is among the first k(n)-1 examined) = $\frac{k(n)-1}{n} \sim \frac{1}{e}$. So that if we were interested in say choosing the best wife, our chances of doing so would be about the same as our chances of never marrying.

Suppose that potential mates appear uniformly between the ages of 18 and 40. Then n= 22 and k(22)= 9. So we should marry when, for the first time after our 26th birthday, we meet a girl who is better than any other we have met before. [ref. Gilbert and Mosteller (1969) for tabulations of k(n)]

Of course it is unrealistic to assume that choosing the second best has no value whatsoever. If instead, the reward for choosing the object with order index i is n-i, then by a similar treatment to the preceding, the expected reward under optimal choice $\sim$ n - 3.8695 for large n. [ref. Chow, Mortiguti, Robbins and Samuels (1964)]

Chapter 3. The Optimal Stopping of Markov Random Sequences

The optimal stopping problem has been solved for stopping times in the class C^N. In the Markov case it has been observed that the value, $s^N(x)$, has the simple construction $Q^N g(x)$. In this chapter we now exploit this form by examining its limit as $N \to \infty$ to deduce results for the optimal stopping of Markov random sequences in C and $\tilde{C}$. Theorems are proved to show under what conditions $s^N(x) \to s(x)$ and $t^N \to$ a $(0,\tilde{s})$-optimal t.

The main technical lemma 3.1.5 appears in Shiryaev (pp. 29-31), as do the most of the proofs in this chapter. But I have rearranged the arguement leading up to the fundamental theorem 3.2.1 so as to use the results of Chapter 2. Not only does this treatment obtain 3.2.1 with substantially less bother, but it also demonstrates the significance of first treating bounded stopping in C^N. Having discarded many of Shiryaev's lemmas, I am forced to an independent proof of theorem 3.3.3.

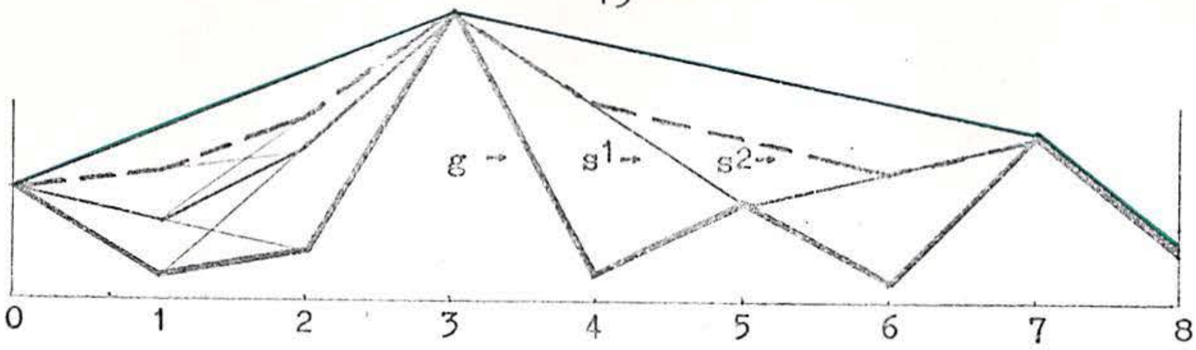
The chapter concludes in showing that the Sequential Probability Ratio Test has a Markov representation and its optimal character is proved.

3.1 Excessive Functions

3.1.1 Properties of $\lim\limits_{N\to\infty} s^N(x)$

We begin with an example:

Suppose $x_1, x_2, \ldots$ is a symmetric random walk on the integers $0,1,\ldots,8$, where 0 and 8 are absorbing. With $g(x)$ as shown, $s^1(x)$, $s^2(x)$ are constructed as:



It would appear that as $N \to \infty$, $s^N(x) \to$ the smallest concave function lieing above g (green line). More precisely we note the following:

(i) $Q^{N+1}g(x) \geq Q^N g(x)$ monotonic increasing. So $\lim_{N\to\infty} Q^N g(x)$ exists and equals, say, $s^*(x) = \lim_{N\to\infty} s^N(x)$.

(ii) $Q^N g(x) \geq -g^-(x)$ and $Q^N g(x) \geq E_x Q^{N-1} g(x_1)$. Assuming that $g\varepsilon L$, $E_x g^-(x_1) < \infty$, by monotone convergence:
$s^*(x) \geq E_x s^*(x_1)$ and $s^*(x) \geq g(x)$.

(iii) Suppose $f\varepsilon L$, $f(x) \geq g(x)$ and $f(x) \geq E_x f(x_1)$ for all $x\varepsilon E$. Then $Qf(x) = \max\{ f(x) , E_x f(x_1) \} = f(x)$ so that
$f(x) = Q^N f(x) \geq Q^N g(x) \to s^*(x)$ ie. $f(x) \geq s^*(x)$.

The existence of $\lim_{N\to\infty} s^N(x)$ and its properties (ii) and (iii) motivate the definitions given below.

3.1.2 Definitions

f is said to be an excessive function (write $f\varepsilon \mathcal{E}$) if $f\varepsilon L$ and $f(x) \geq E_x f(x_1)$ for all $x\varepsilon E$.

Given a function g, the funtion f is said to be an excessive majorant of g if $f\varepsilon\mathcal{E}$ and $f \geq g$.

(n.b. The excessive nature of a function is defined in terms of a particular Markov chain and transition probabilities. It is always assumed that this is the chain of the optimal stopping problem.)

From 3.1.1 (ii) and (iii) it is clear that s^* is an excessive majorant of g and that if f is any other excessive majorant of g then $f \geqslant s^*$. We call s^* the smallest excessive majorant of g (s.e.m.).

The basic properties of excessive funtions are included in the following lemmas. [ref. Dynkin (1963) p. 627 ; Shiryaev pp. 22,29]

3.1.3 Lemma

Let $f,g \in \mathcal{E}$. Then:

(i) constant funtions are excessive functions.

(ii) $\alpha f + \beta g$ is excessive for all $\alpha,\beta \geqslant 0$.

(iii) $E_x(f(x_{n+1})|x_n) \leqslant f(x_n)$ ie. $\{f(x_n),B(x_n)\}$ is a super-martingale.

(iv) the exact lower bound of non-negative excessive functions is a non-negative excessive function.

(v) if $\sup_n E_x f^-(x_n) < \infty$ then $\lim_{n\to\infty} f(x_n)$ exists P_x- a.s. (possibly $+\infty$).

proof:

(i) - (iv) are immediate consequences of the definitions.

(v) is the super-martingale convergence theorem.

3.1.4 Lemma

Suppose $t,s \in C^N$ with $t \geqslant s$ P_x- a.s.; then $E_x f(x_s) \geqslant E_x f(x_t)$.

proof:

To begin, suppose $t-s$ is just 0 or 1. Then:

$$E_x[\ f(x_s) - f(x_t)\ ] = \sum_0^N \int_{(s=n)\cap(t>n)} (f(x_n) - f(x_{n+1}))\, dP_x$$

$\geqslant 0$ since $(s=n)\cap(t>n) \in F_n$ and $E_x f(x_{n+1}) \leqslant E_x f(x_n)$ in a super-martingale. Now let $t_n = \min(t,s+n)$ for $n = 1,2,\ldots,N$. The t_n is a valid stopping time and $t_{n+1} - t_n$ is just 0 or 1. So $E_x f(x_s) \geqslant E_x f(x_{t_1}) \geqslant \cdots \geqslant E_x f(x_{t_N}) \geqslant E_x f(x_t)$.

As a summary of the state of knowledge so far, we know that $s^N(x) \to s^*(x) =$ s.e.m. of g and that $s^N(x) \leqslant s(x)$ (this is the key use of the link to Chapter 2). Therefore
$s^*(x) \leqslant s(x) \leqslant \tilde{s}(x)$.

also, $\tilde{s}(x) = \sup_{t \in \tilde{C}} \tilde{E}_x g(x_t) \leqslant \sup_{t \in \tilde{C}} \tilde{E}_x s^*(x_t)$. So if it were possible to show that $s^*(x) \geqslant \sup_{t \in \tilde{C}} \tilde{E}_x s^*(x_t)$ then we would have that $s^*(x) = s(x) = \tilde{s}(x)$. This follows from a final lemma.

3.1.5 Lemma [ref. Shiryaev pp. 29-31]

Let $f \in \mathcal{E}$ such that $E_x[\sup_n f^-(x_n)] < \infty$. Let $t,s \in \tilde{C}$ with $t \geqslant s$. Then $\tilde{E}_x f(x_s) \geqslant \tilde{E}_x f(x_t)$.
So in particular, if $E_x[\sup_n g^-(x_n)] < \infty$, then $s^*(x) \geqslant \tilde{E}_x s^*(x_t)$ for all $t \in \tilde{C}$.

proof:
By 3.1.3 (v): $\overline{\lim}_{n\to\infty} f(x_n) = \underline{\lim}_{n\to\infty} f(x_n)$. Assume that $f \leqslant K$ (f bounded) and let $s_n = \min(s,n)$; $t_n = \min(t,n)$. Omitting dP_x throughout,
by 3.1.4: $\int f(x_{s_n}) \geqslant \int f(x_{t_n})$ or

$$\int_{(s<n)} f(x_s) + \int_{(s \geqslant n)} f(x_n) \geqslant \int_{(t<n)} f(x_t) + \int_{(t \geqslant n)} f(x_n)$$

so:

$$\int_{(s<\infty)} f(x_s) + \int_{(s=\infty)} f(x_n) + \int_{(n \leqslant s < \infty)} [f(x_n) - f(x_s)] \geqslant$$
$$\int_{(t<\infty)} f(x_t) + \int_{(t=\infty)} f(x_n) + \int_{(n \leqslant t < \infty)} [f(x_n) - f(x_t)]$$

but since f is bounded the 3rd term on both sides $\to 0$ as $n \to \infty$, and so

$$\int_{(s<\infty)} f(x_s) \geqslant \int_{(t<\infty)} f(x_t) + \underline{\lim} \int_{(t=\infty)\setminus(s=\infty)} f(x_n)$$
$$\geqslant \int_{(t<\infty)} f(x_t) + \int_{(t=\infty)\setminus(s=\infty)} \lim f(x_n)$$

by Fatou's lemma and the remark that $\underline{\lim} f(x_n) = \lim f(x_n)$.
This last line is just $\tilde{E}_x f(x_s) \geqslant \tilde{E}_x f(x_t)$.

For a general f, let $f^m = \min(f,m)$ which is clearly excessive.

$f^m(x) \to f(x)$ and $\lim_{m\to\infty}(\lim_{n\to\infty} f^m(x_n)) = \lim_{n\to\infty} f(x_n)$ since if $\lim_{n\to\infty} f(x_n) = \alpha < \infty$ then for large m $\lim_{n\to\infty}(\min(m, f(x_n))) = \alpha$, and if $\lim_{n\to\infty} f(x_n) = \infty$ then $\lim_{n\to\infty}(\min(m, f(x_n))) = m$. Look at:

$$\int_{(s<\infty)} f(x_s) > \int_{(t<\infty)} f^m(x_t) + \int_{(t=\infty)\setminus(s=\infty)} \lim_{n\to\infty} f^m(x_n)$$

and let $m \to \infty$ to complete the proof.

In the particular case of $s = 1$, $f(x) \geqslant \tilde{E}_x f(x_1) \geqslant \tilde{E}_x f(x_t)$.

3.2 The Characterization of Value

The fundamental theorem about the value of the optimal stopping problem on a Markov random sequence may now be stated.

3.2.1 Theorem

If $E_x[\sup g^-(x_n)] < \infty$ then $s^*(x) = s(x) = \tilde{s}(x)$.

proof: the direct consequence of lemma 3.1.5.

The result is that there is no reduction in the value of the sequence when attention is restricted from extended stopping rules in C to those in C or even $\bigcup_1^\infty C^N$.

Some comment should be made about the condition $E_x[\sup g^-(x_n)] < \infty$ (which we shall write henceforth as $g \in L(A^-)$). It was used in the proof of 3.1.5 to ensure the conditions of 3.1.3 (v), Fatou's lemma and $\int f^-(x_n)$ bounded so that with $f^+ \leqslant K$ we could get $\int_{(n\leqslant t<\infty)} [f(x_n) - f(x_t)] \to 0$.

An example shows that it cannot be dropped. For let $\{x_n\}$ be a symmetric random walk on the integers with $g(x) = x$. Then the martingale theory of gambling systems or simply the recurrence $s^N(x) = \min\{ x , \frac{1}{2}[s^{N-1}(x+1) + s^{N-1}(x-1)] \}$ gives $s^N(x) = x$. While if $t^{(N)} = \min\{ n : x_n = N \}$, then $t^{(N)} \varepsilon$ C and $E_x g(x_{t^{(N)}}) = N$. Hence $\infty = s(x) \neq s^*(x) = x$.

At best we might hope for 3.2.1 to hold when 3.1.5 is true. To make this precise one can cook up the following definition.

3.2.2 Definition

A function f is said to be a regular excessive majorant of g if $f \geqslant g$ and $f(x) \geqslant \tilde{E}_x f(x_t)$ for all $x \varepsilon E$ and $t \varepsilon \tilde{C}$ (ie. for all t such that $\tilde{E}_x g^-(x_t) < \infty$).

Lemma 3.1.5 established that if $g \varepsilon L(A^-)$, then the excessive majorants of g are regular. Let s_r^* be the smallest regular excessive majorant of g. Then argueing as before from the definitions: $s_r^*(x) \geqslant \sup_{t \varepsilon \tilde{C}} \tilde{E}_x s_r^*(x_t) \geqslant \sup_{t \varepsilon \tilde{C}} \tilde{E}_x g(x_t) = \tilde{s}(x)$. The reverse inequality is also true even though it can no longer be obtained by appealing to $\lim s^N(x) = s^*(x)$, since in general $s^*(x) < s_r^*(x)$.

The actual proof is lengthy. So for completeness we conclude this section by simply stating the result.

3.2.3 Theorem [ref. Shiryaev pp. 50-56]

If (as always) $g \varepsilon L$ then $s_r^*(x) = s(x) = \tilde{s}(x)$.

eg. In the example above, $s_r^* > \acute{s} = \infty$. So $s_r^* = s$.

3.3 The Characterization of Optimal Times

Thus far we have been concerned with determining the value of a Markov random sequence, essentially through looking at the limit of $s^N(x)$ as $N \to \infty$. To do this, the neat recurrence construction of Theorem 2.2.4 was exploited.

However, 2.2.4 also gave an explicit construction for the optimal times, t^N. It is natural to ask whether there are stopping rules in C or $\tilde{C}$ which actually attain the value, s(x), and whether these can be related to the limit of t^N.

It will now be shown that within $\tilde{C}$ this is the case.

3.3.1 Lemma [ref. Shiryaev p. 34]

Suppose that the value, $s(x)$, is such that $s(x) < \infty$ for all $x \varepsilon E$. Define $t_o = \min\{ n : g(x_n) = s(x_n) \}$. Then,

$$s(x) = \int_{(t_o < N)} s(x_{t_o})\, dP_x + \int_{(t_o \geq N)} s(x_N)\, dP_x \quad \text{for all } N.$$

proof:

Clearly, $s(x) = \sup_{t \varepsilon C} E_x s(x_t) = E_x s(x_1)$ since $s(x)$ is the smallest regular excessive majorant of itself. Hence:

$$s(x) = \int_{(t_o = 1)} s(x_1)\, dP_x + \int_{(t_o > 1)} s(x_1)\, dP_x.$$

But on the set $(t_o > 1)$, $s(x_1) > g(x_1)$, so that $s(x_1) = E_{x_1} s(x_2)$.

Thus, $$s(x) = \int_{(t_o = 1)} s(x_{t_o})\, dP_x + \int_{(t_o \geq 2)} s(x_2)\, dP_x$$

(etc.) $$= \int_{(t_o < N)} s(x_{t_o})\, dP_x + \int_{(t_o \geq N)} s(x_N)\, dP_x$$

The stopping rule t_o seems to be the obvious candidate for the limit of t^N as given in theorem 2.2.4. Note that in fact, $t^{N+1} \geq t^N$. So $\lim t^N = t^*$, say, exists.

As previously, define the conditions

$g \varepsilon L(A^-)$ to mean $E_x[\sup_n g^-(x_n)] < \infty$ for all $x \varepsilon E$, and

$g \varepsilon L(A^+)$ to mean $E_x[\sup_n g^+(x_n)] < \infty$ for all $x \varepsilon E$.

The next theorem relates t^* and t_o to the optimal rule.

3.3.2 Theorem [ref. Shiryaev pp. 57,58,62]

(i) if $g \varepsilon L(A^+)$, then t_o is $(0,\tilde{s})$-optimal.

(ii) if $g \varepsilon L(A^+) \cap L(A^-)$, then $t_o = t^* = \lim_{N \to \infty} t^N$.

proof:

By 3.3.1 (i) $$s(x) = \int_{(t_o < N)} s(x_{t_o})\, dP_x + \int_{(t_o \geq N)} s(x_N)\, dP_x$$

$$\leq \int_{(t_o < N)} g(x_{t_o})\, dP_x + \int_{(t_o \geq N)} \sup_{n \geq N} g(x_n)\, dP_x$$

from the definitions of t_o and $s(x_N)$. Now apply $\overline{\lim}_{N\to\infty}$ to the above using Fatou's lemma and the condition $g\varepsilon L(A^+)$ to deduce:

$$s(x) \leqslant \int_{(t_o<\infty)} g(x_{t_o})\, dP_x + \int_{(t_o=\infty)} \overline{\lim}_{n\to\infty} g(x_n)\, dP_x$$
$$= \tilde{E}_x g(x_{t_o}).$$

(ii) For the first inequality we still only assume $g\varepsilon L(A^+)$.

<u>$t^* \leq t_o$</u>: If $t^*= n$ then there exists an N such that $t^N= n$. So $g(x_i) < s^{N-i}(x_i)$ for $i = 1,\dots,n-1$. This then implies that $g(x_i) < s(x_i)$ for $i = 1,\dots,n-1$, and hence that $t_o \geqslant n$.

If $t^*= \infty$ then for a given k there exists an N such that $g(x_i) < s^{N-i}(x_i)$ for $i = 1,\dots,k$. But $s^{N-i}(x_i) \leqslant s(x_i)$ then gives $g(x_i) < s(x_i)$ for $i = 1,\dots,k$. True all k. So t_o must equal ∞.

To prove the reverse inequality we need $g\varepsilon L(A^-)$.

<u>$t^* \geq t_o$</u>: If $t_o= n$ then $g(x_i) < s(x_i)$ for $i = 1,\dots,n-1$. Under $g\varepsilon L(A^-)$ $s^N \to s$ and so $g(x_i) < s^{N-i}(x_i)$ for $i = 1,\dots,n-1$ and large enough N. Hence $t^N \geqslant n$ and so $t^* \geqslant n$.

If $t_o= \infty$ then $g(x_i) < s(x_i)$ for all $i = 1,\dots$ Therefore, given an n, $t^N \geqslant n$ for large enough N. Hence $t^*= \infty$.

3.3.3 Corollary [ref. Shiryaev p. 58]

If $g\varepsilon L(A^+)$ and $\lim_{n\to\infty} g(x_n) = -\infty$ P_x- a.s., then t_o is $(0,s)$-optimal.

proof:

If $P_x(t_o= \infty) > 0$ then $s(x) = -\infty$. But $s(x) \geqslant g(x) > -\infty$. So $P_x(t_o= \infty) = 0$ and we have the existence of a $(0,s)$-optimal rule.

In general, a $(0,s)$-optimal rule may not exist. For example, if $E = \{0,1,2,\dots\}$, $P(x_n= i+1 | x_{n-1}= i) = 1$, and $g(x) = x/(1+x)$, then $s(x) = 1$. The time $t^*= \infty$ is $(0,\tilde{s})$- optimal, but there is clearly no $(0,s)$-optimal rule.

The best that can be achieved in C is often just a rule arbitrarily close to a (0,s)-optimal one. The following theorem states the conditions for a (ε,s)-optimal rule.

3.3.4 Theorem

If $g \varepsilon L(A^+) \cap L(A^-)$, then the stopping time defined as $t_\varepsilon = \min\{ n : g(x_n) \geqslant s(x_n) - \varepsilon \}$ is (ε,s)-optimal.

proof:

Clearly $t_o \geqslant t_\varepsilon$, so that by 3.1.5:

$$\tilde{E}_x g(x_{t_\varepsilon}) \geqslant \tilde{E}_x s(x_{t_\varepsilon}) - \varepsilon \geqslant \tilde{E}_x s(x_{t_o}) - \varepsilon \geqslant \tilde{E}_x g(x_{t_o}) - \varepsilon = \tilde{s}(x) - \varepsilon.$$

This shows that t_ε is $(\varepsilon,\tilde{s})$-optimal. Also, $s(x) \geqslant \tilde{E}_x s(x_{t_o}) \geqslant \tilde{E}_x g(x_{t_o}) = s(x)$ gives:

$$E_x[\, g(x_{t_o}) : t_o < \infty \,] + E_x[\, \overline{\lim}\, g(x_n) : t_o = \infty \,] =$$
$$E_x[\, s(x_{t_o}) : t_o < \infty \,] + E_x[\, \overline{\lim}\, s(x_n) : t_o = \infty \,].$$

Here the first terms on either side are equal and certainly $\overline{\lim}\, s(x_n) \geqslant \overline{\lim}\, g(x_n)$. The conditions imply that $P_x(\overline{\lim}\, s(x_n) = \pm\infty) = 0$. Hence we deduce that $P_x(\, t_o = \infty$ and $\overline{\lim}\, s(x_n) > \overline{\lim}\, g(x_n)\,) = 0$.

Now $t_\varepsilon = \infty$ implies that $t_o = \infty$ and that $g(x_n) < s(x_n) - \varepsilon$ for all n, ie. that $\overline{\lim}\, g(x_n) < \overline{\lim}\, s(x_n)$. Therefore, $P_x(\, t_\varepsilon = \infty\,) = 0$ and t_ε is (ε,s)-optimal.

Note that in the line above showing t_ε to be $(\varepsilon,\tilde{s})$-optimal, we could deduce $E_x s(x_{t_\varepsilon}) - \varepsilon = s(x) - \varepsilon$ without assuming $g \varepsilon L(A^-)$. This can be done by proving lemma 3.3.1 and theorem 3.3.2 (i) for t_ε in exactly the same way as they were proved for t_o.

This now concludes the characterization of the solution of the optimal stopping problem on a Markov random sequence.

3.4 The Optimal Character of the Sequential Probability Ratio Test

Assume that $x_1, x_2, \ldots$ are independent, identically distributed samples from some density f. We wish to decide between the hypotheses, $H_0: f = f_0$ and $H_1: f = f_1$, using a sequential decision procedure as outlined in 1.2.2. The cost of taking each sample is 1 and the cost of an incorrest decision is:

a when H_0 is true and we choose H_1 and

b when H_1 is true and we choose H_0.

Using the procedure (δ,t), let $\alpha_i(\delta,t) = P_i(\text{ reject } H_i)$ $i = 0,1$.

Assume further that we know H_0 to be true with prior probability π so that $r(\pi,\delta,t) = \pi[a\alpha_0 + E_0 t] + (1-\pi)[b\alpha_1 + E_1 t]$ is the expected loss which we desire to minimize.

3.4.1 Theorem [ref. Chow pp. 46-49,105; Shiryaev pp. 124,125]

Let $\pi_n = \dfrac{\pi f_0(x_1)\cdots f_0(x_n)}{\pi f_0(x_1)\cdots f_0(x_n) + (1-\pi)f_1(x_1)\cdots f_1(x_n)} =$ the posterior probability of H_0 given $x_1, x_2, \ldots, x_n$. Then the sequential decision procedure (δ,t) which minimizes the risk, $r(\pi,\delta,t)$ is described by:

$t = \min\{ n : \pi_n \varepsilon [0,\underline{\pi}] \cup [\overline{\pi},1] \}$ where $0 < \underline{\pi} < \overline{\pi} < 1$ and

$$\delta = \begin{cases} \text{accept } H_0 \text{ if } \pi_t a \geqslant (1-\pi_t)b \\ \text{accept } H_1 \text{ if } \pi_t a < (1-\pi_t)b. \end{cases}$$

The procedure is to continue sampling until the posterior probability of H_0 is sufficiently close to 0 or 1 and then to choose the hypothesis whose rejection risks the greater loss under this probability.

proof:

For fixed t the form of δ is easily found. The part of the loss depending on δ is $\pi a\alpha_0 + (1-\pi)b\alpha_1 =$

(omit $dx_1 \cdots dx_n$ in the following integrals)

$$= \sum_{n=1}^{\infty} \left\{ \pi a \int_{\{t = n;\ \text{accept } H_1\}} f_0(x_1)\cdots f\ (x_n) + (1-\pi) b \int_{\{t = n;\ \text{accept } H_0\}} f_1(x_1)\cdots f_1(x_n) \right\}$$

$$> \sum_{n=1}^{\infty} \int_{\{t = n\}} \min\{ \pi a f_0(x_1)\cdots f_0(x_n) \ ,\ (1-\pi) b f_1(x_1)\cdots f_1(x_n) \}$$

$$= \sum_{n=1}^{\infty} \int_{\{t = n\}} [\min\{ \pi_n a \ ,\ (1-\pi_n) b \}][\pi f_0(x_1)\cdots f_0(x_n) + (1-\pi) f_1(x_1)\cdots f_1(x_n)]$$

$= \pi a \alpha_0 + (1-\pi) b \alpha_1$ for the δ which accepts $\begin{matrix} H_0 \\ H_1 \end{matrix}$ as $\pi_t a \gtrless (1-\pi_t) b$.

Now $$\pi_n = \frac{\pi_{n-1} f_0(x_n)}{\pi_{n-1} f_0(x_n) + (1-\pi_{n-1}) f_1(x_n)} ,$$

so $\pi, \pi_1, \pi_2, \ldots$ is a stationary Markov sequence; $F_n = B(x_1, \ldots, x_n)$. The loss is $E_\pi[\min\{ a\pi_t \ ,\ (1-\pi_t) b\} - t]$ and so in the notation of the chapter we can take:

$(\pi,0), (\pi_1,1), (\pi_2,2), \ldots$ a stationary Markov sequence with state space $E = \{ (\pi,n) : 0 < \pi < 1 \text{ and } n = 0,1,\ldots \}$. Then, $g(\pi,n) = -h(\pi) - n$ where $h(\pi) = \min\{ a\pi, (1-\pi) b \}$. And we are interested in finding $\sup_{t \varepsilon C} E_{(\pi,0)} g(\pi_n, n) = s(\pi,0)$. [note that $\lim g(\pi_n,n) = -\infty$ precludes t which take the value ∞ .]

Since $g \varepsilon L(A^+)$, theorem 3.3.2 states that a $(0,s)$-optimal t exists and is given by $t = \min\{ n : g(\pi_n,n) = s(\pi_n,n) \}$. This optimal t is the least n such that:

$$-h(\pi_n) - n \geqslant \sup_{t \varepsilon C} E_{(\pi_n,n)} g(\pi_{n+t}, n+t) =$$

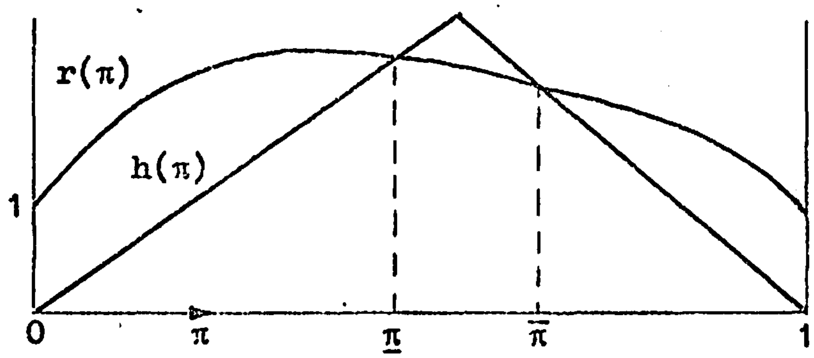
$$\sup_{t \varepsilon C} \sup_{\delta} [-\pi_n(\alpha_0 a + E_0(t+n)) - (1-\pi_n)(\alpha_1 b + E_1(t+n))], \text{ or}$$

the least n such that:

$$h(\pi_n) \leqslant \inf_{t \varepsilon C} \inf_{\delta} [\pi_n(\alpha_0 a + E_0 t) + (1-\pi_n)(\alpha_1 b + E_1 t)] = r(\pi_n), \text{ say.}$$

Not surprisingly, this is just to say that the optimal t stops sampling when for the first time the expected loss of stopping now is less than the expected loss of all procedures which continue.

The quantities, $\alpha_0, \alpha_1, E_0 t, E_1 t$, are determined for fixed δ, t, and are therefore independent of π. We deduce that $r(\pi)$ is the infimum of linear functions of π and hence is concave on $[0,1]$. Concave functions are continuous. The graphs of $h(\pi)$ and $r(\pi)$ appear as:



Clearly, $t = \min\{\ n : h(\pi_n) \leqslant r(\pi_n)\ \}$ is equivalent to

$$t = \min\{\ n : \pi_n \varepsilon\ [0,\underline{\pi}] \cup [\bar{\pi},1]\ \}.$$

Chapter 4. The Optimal Stopping of Random Sequences

Having characterized the solution to the optimal stopping problem on Markov random sequences, we have a good idea of the type of theorems which are likely to prove true when considering the problem on general stochastic sequences.

Given a stochastic sequence, $\{z_n, F_n\}$, we might define $x_n = (z_1, z_2, \ldots, z_n)$ and $g(x_n) = z_n$. So that in a sense, every stochastic sequence has a Markov representation, even though the state space and transition probabilities may be far too complex for direct treatment. This thought suggests that the results of Chapter 3, such as " $s = \tilde{s}$ " or " a $(0, \tilde{s})$-optimal time exists ", should carry across to general sequence results, for they contain no statements about the form of the Markov sequence involved. This chapter describes the form taken by the theorems of Chapter 3 when extended to the general context.

The Two-armed Bandit Problem of 1.1.3 is an example in the sequential design of experiments which can be solved through the use of optimal stopping times. As in many of the more complex problems, general theory lends only the first insights, while deeper investigation proceeds with reference to the specific features of the problem. The solution to the Bandit Problem is a nice example of the way stopping times feature in one area of contemporary research.

4.1 The Characterization of Value

4.1.1 Properties of $\lim_{N \to \infty} \gamma_n^N$

Just as in 3.1 we examined the limit of $s^N(x)$ as $N \to \infty$, so here we look at the limit of the random variable, γ_n^N, defined in Theorem 2.1.2. We note the following:

(i) γ_n^N is the pointwise supremum of the set of random variables, $\{ E(z_t|F_n) : t\varepsilon C_n^N \}$, except possibly for a set of point with probability zero. This is contained in the proof of 2.1.2. We write: $\gamma_n^N = \text{ess} \sup_{t\varepsilon C_n^N} E(z_t|F_n)$ [essential supremum].

(ii) $\gamma_n^N \geqslant \gamma_n^{N-1}$ monotonic incresing. So $\lim_{N\to\infty} \gamma_n^N$ exists and equals, say, γ_n^*.

(iii) $\gamma_n^N = \max\{ z_n , E(\gamma_{n+1}^N|F_n) \}$ so that by monotone convergence, $\gamma^* = \max\{ z_n , E(\gamma_{n+1}^*|F_n) \}$, ie. $\gamma_n^* \geqslant z_n$ and $\{\gamma_n^*, F_n\}$ is a supermartingale.

(iv) Suppose that $\{\beta_n, F_n\}$ is a supermartingale such that $\beta_n \geqslant z_n$ for all n. Then, $\beta_N \geqslant z_N$ $\gamma_{N-1}^N = \max\{ z_{N-1} , E(\gamma_N^N|F_{N-1}) \} \leqslant \max\{ \beta_{N-1} , E(\beta_N^N|F_{N-1}) \} = \beta_{N-1}$. Continuing the induction, we find $\gamma_n^N \leqslant \beta_n$ for all N. So $\gamma_n^* \leqslant \beta_n$.

It is now clear what should take the place of excessive functions considered in Chapter 3.

4.1.2 Definition

The super martingale $\{\beta_n, F_n\}$ is said to dominate the stochastic sequence $\{z_n, F_n\}$ if $\beta_n \geqslant z_n$ a.s. for all n. If all other supermartingales which dominate $\{z_n, F_n\}$ also dominate $\{\beta_n, F_n\}$, then $\{\beta_n, F_n\}$ is said to be the smallest supermartingale dominating $\{z_n, F_n\}$.

By (iii) and (iv) above, γ_n^* is the smallest supermartingale dominating z_n. We also define:

$\gamma_n = \text{ess} \sup_{t\varepsilon C_n} E(z_t|F_n)$ and $\tilde{\gamma}_n = \text{ess} \sup_{t\varepsilon \tilde{C}_n} \tilde{E}(z_t|F_n)$ where it is assumed that z_t takes the value $\overline{\lim_{n\to\infty}}\, z_n$ on the set $(t = \infty)$.

Clearly $\gamma_n^* \leqslant \gamma_n \leqslant \tilde{\gamma}_n$. If $E[\sup_n z_n^-] < \infty$ then we will show that $\tilde{\gamma}_n \leqslant \operatorname{ess}\sup_{t\in\tilde{C}_n} \tilde{E}(\gamma_t^*|F_n) \leqslant \gamma_n^*$ where the last inequality follows from the lemma below, exactly paralleling 3.1.5.

4.1.3 Lemma

Let $\{\beta_n, F_n\}$ be a supermartingale satisfying the condition A^-: $E[\sup_n \beta_n^-] < \infty$. Let $t, s \in \tilde{C}_n$ with $t \geqslant s$. Then $\tilde{E}(\beta_t|F_n) \leqslant \tilde{E}(\beta_s|F_n)$. So in particular, if $E[\sup_n z_n^-] < \infty$ then $\gamma_n^* \geqslant \tilde{E}(\gamma_t^*|F_n)$ for all $t\in\tilde{C}_n$.

proof:

Lemma 3.1.4 becomes: if $t, s \in C_n^N$ with $t \geqslant s$, then $E(\beta_t|F_N) < E(\beta_s|F_N)$ by an exactly analogous proof.

Assume β_n is a.s. bounded above. The supermartingale convergence theorem shows that $\lim_{n\to\infty} \beta_n$ exists a.s. if $\sup E\beta_n^- < \infty$, which is implied by A^-.

Let $A\in F_n$. Then as in 3.1.5 we can get:

$$\int_{(s<\infty)\cap A} \beta_s \geqslant \int_{(t<\infty)\cap A} \beta_t + \int_{[(t=\infty)\smallsetminus(s=\infty)]\cap A} \beta_N$$
$$+ \int_{(N\leqslant t<\infty)\cap A} (\beta_N - \beta_t) - \int_{(N\leqslant s<\infty)\cap A} (\beta_N - \beta_s)$$

Apply $\lim_{N\to\infty}$ to both sides. Look at the integrals on the right hand side.

The second is $\geqslant \underline{\lim} \int_{[\ ]\cap A} \beta_N \geqslant \int_{[\ ]\cap A} \underline{\lim}\, \beta_N$ since A^- implies (β_N^-) uniformly integrable, which implies Fatou's lemma.

The third is $\geqslant -\lim \int_{(\ )\cap A} \beta_N^- - \lim \int_{(\ )\cap A} \beta_t^+$. Both limits here are zero since β_N^- and β_t^+ are both bounded by variables with finite expectation (ie. $\beta_N^- < \sup_n \beta_n^-$) and $P(N \leqslant t < \infty) \to 0$.

The fourth is $\geqslant -\lim \int_{(\ )\cap A} \beta_N^+ - \lim \int_{(\ )\cap A} \beta_s^-$. Again both limits are zero.

As these results hold for all $A\in F_n$, $\tilde{E}(\beta_s|F_n) \geqslant \tilde{E}(\beta_t|F_n)$. The boundedness assumption is relaxed just as in 3.1.5.

The result to parallel 3.2.1 follows immediately: if $E[\sup_n z_n^-] < \infty$ then $\gamma_n^* = \gamma_n = \tilde{\gamma}_n$ and hence $s = \tilde{s}$.

Note that for $t,s \in C_n$ the lemma may be proved under the weaker condition $\underline{\lim} \int_{(t \geqslant N)} \beta_N^- = 0$.

In order to include all stocastic sequences in one theorem the appropriate class of supermartingales is defined. The general characterization of value then follows.

4.1.4 Definition

The supermartingale $\{\beta_n, F_n\}$ is said to be regular if for all $t \in \tilde{C}$ $\tilde{E}\beta_t$ exists and $\tilde{E}(\beta_t | F_n) \leqslant \beta_n$ on the set $(t \geqslant n)$.

4.1.5 Theorem [ref. Chow pp. 66,75-76,81]

If $Ez_n^- < \infty$ all n (as assumed) then:

(i) $\gamma_n = \tilde{\gamma}_n =$ the smallest regular supermartingale dominating z_n.

(ii) $\gamma_n = \max\{z_n, E(\gamma_{n+1} | F_n)\}$

(iii) $s = \tilde{s} = E\gamma_1$

proof:

(i) see Chow pg. 81 (Theorem 4.7)

(ii) Let $t \in C_n$. By definition $E(z_t | F_{n+1}) \leqslant \gamma_{n+1}$ on $(t > n)$. So

$$E(z_t | F_n) \leqslant E(\gamma_{n+1} | F_n) \text{ on } (t > n)$$
$$= z_n \text{ on } (t = n).$$

Thus $\gamma_n \leqslant \max\{z_n, E(\gamma_{n+1} | F_n)\}$

Conversely, $\gamma_n \geqslant E(z_t | F_n)$ for all $t \in C_{n+1}$. Take t_i such that $E(z_{t_i} | F_{n+1}) \nearrow \gamma_{n+1}$. Then $\gamma_n \geqslant E(E(z_{t_i} | F_{n+1}) | F_n) \nearrow E(\gamma_{n+1} | F_n)$ by monotone convergence. Clearly $\gamma_n \geqslant z_n$.

(iii) immediate

4.2 The Characterization of Optimal Times

The two conditions which again play an important role are: A^+: $E[\sup_n z_n^+] < \infty$ and A^-: $E[\sup_n z_n^-] < \infty$. The results which can be proved mimic 3.3.2 and 3.3.3. In summary:

4.2.1 Theorem [ref. Chow p. 82]

Let $t_o = \min\{ n : z_n \geqslant \gamma_n \}$ then

(i) if A^+ holds then t_o is $(0,\tilde{s})$-optimal

(ii) if A^+ and A^- hold then $t^N = \min\{ n : z_n \geqslant \gamma_n^N \} \to t_o$ a.s.

proof:

(i) $$E\gamma_1 = \int_{(t_o=1)} \gamma_1 + \int_{(t_o>1)} \gamma_1 =$$

$$\int_{(t_o=1)} z_1 + \int_{(t_o\geqslant 2)} E(\gamma_2|F_1) = \int_{(t_o=1)} z_1 + \int_{(t_o\geqslant 2)} \gamma_2 =$$

$$\ldots = \int_{(t_o<N)} z_{t_o} + \int_{(t_o\geqslant N)} \gamma_N$$

$$\leqslant \varlimsup_{N\to\infty} \int_{(t_o<N)} z_{t_o} + \int_{(t_o\geqslant N)} \sup_{n\geqslant N} z_n = Ez_{t_o}.$$

(ii) $t^N \to t^*$ say

If $t^* = n$ then there exists N such that $t^N = n$. $z_i < \gamma_i$ $i = 1\ldots n-1$ so $z_i < \gamma_i$ $i = 1\ldots n-1$. Thus $t_o \geqslant n$. (similarly $t^* = \infty$ implies $t_o = \infty$)

If $t_o = n$ then $z_i < \gamma_i$ $i = 1\ldots n-1$. By A^- there exists large N so $z_i < \gamma_i^N$ $i = 1\ldots n-1$. Thus $t^* \geqslant n$. (similarly $t_o = \infty$ implies $t^* = \infty$)

4.2.2 Corollary

If A^+ holds and $\lim z_n = -\infty$ then t_o is $(0,s)$-optimal.

As a parallel to 3.3.4 we state a theorem on (ε,s)-optimal times. For variety it can be put in a form where the conditions do not explicitly mention A^+ or A^-.

4.2.3 Theorem

It t_o is $(0,\tilde{s})$-optimal and $s < \infty$ then $t_\varepsilon = \min\{ n : z_n > \gamma_n - \varepsilon \}$ is (ε,s)-optimal.

proof:

$$E\gamma_1 \leqslant \int_{(t_\varepsilon=1)} (z_1 + \varepsilon) + \int_{(t_\varepsilon>1)} \gamma_1 \leqslant$$

$$\int_{(t_\varepsilon < 1)} (z_1 + \varepsilon) + \int_{(t_\varepsilon \geq 2)} E(\gamma_2 | F_1) \leqslant \int_{(t_\varepsilon < 1)} (z_1 + \varepsilon) + \int_{(t_\varepsilon \geq 2)} \gamma_2$$

$$\leqslant \ldots \leqslant \int_{(t_\varepsilon < N)} (z_{t_\varepsilon} + \varepsilon) + \int_{(t_\varepsilon \geq N)} \gamma_N$$

$\leqslant E(z_{t_\varepsilon} : t_\varepsilon < \infty) + \varepsilon P(t_\varepsilon < \infty) + E(\overline{\lim}_{N\to\infty} \gamma_N : t_\varepsilon = \infty)$ Putting $\varepsilon = 0$ we get $\tilde{E}\gamma_1 \leqslant \tilde{E}\gamma_{t_o}$ as $z_{t_o} = \gamma_{t_o}$ on $(t_o < \infty)$. But $\tilde{E}\gamma_1 \geqslant \tilde{E}\gamma_{t_o}$ since $\{\gamma_n\}$ is its own smallest dominating regular supermartingale. Hence $\tilde{E}z_{t_o} = E\gamma_1 = \tilde{E}\gamma_{t_o} < \infty$. This implies that $\overline{\lim}\ \gamma_n > \overline{\lim}\ z_n$ only on a set of probability zero. However $t_o = \infty$ whenever $t_\varepsilon = \infty$ and $z_n < \gamma_n - \varepsilon$ all n implies $\overline{\lim}\ z_n < \overline{\lim}\ \gamma_n$. Thus $t_\varepsilon = \infty$ with probability zero and the above line gives $s = E\gamma_1 \leqslant Ez_{t_\varepsilon} + \varepsilon$.

The characterization of the solution to the optimal stopping problem on general stochastic sequences is now complete. The problem differs on Markov sequences and general sequences only in that the latter requires knowledge of the entire past. We summarize the answers to the questions posed in 1.3:

(a) s always satisfies $s = \max\{ z_1 , \sup_{t\varepsilon C_2} Ez_t \}$ and under $E[\sup_n z_n^-] < \infty$ can be computed as the limit of the value on the bounded problem, s^N.

(b) $(0,\tilde{s})$ and (ε,s) times exist when $s < \infty$ and $E[\sup_n z_n^+] < \infty$ and a $(0,s)$-optimal time exists if in addition $z_n \to -\infty$.

(c) The nature of $(0,\tilde{s})$-optimal rules are always " stop when for the first time the reward attained by stopping is greater than the best that could be expected to be obtained from going on ".

4.3 Solution of the Two-Armed Bandit Problem

Suppose we have the option of playing one of two bandit arms at each time instant. Arms 1 and 2 pay 1 unit with probabilities p_1 and p_2 or 0 units with probabilites $1-p_1$ and $1-p_2$ respectively.

In order to keep the expected payoff finite we discount at a rate α where $0 < \alpha < 1$. This may be thought as equivalent to the situation where there is a probability $1-\alpha$ that an arm will at any time instant become inoperable, never again available for play. The expected reward we desire to maximize is then $E\{\sum_1^\infty \alpha^{i-1} x_i\}$, where x_i is the reward received from the arm that is chosen and played at time i.

Of interest is the optimal design of play when one or both of p_1, p_2 are only known to have been chosen from some prior distribution. We examine the two cases in turn.

4.3.1 Theorem

If p_2 is known and p_1 has prior density f_0 on $[0,1]$ then

(i) There is an extended stopping time t^* such that the optimal play is: pull arm 1 for $1,\ldots,t^*-1$ and then pull arm 2 at all times t^* and beyond.

(ii) The expected reward is $E\{\sum_1^{t^*-1} \alpha^{i-1} x_i + \frac{\alpha^{t^*-1}}{1-\alpha} p_2\}$.

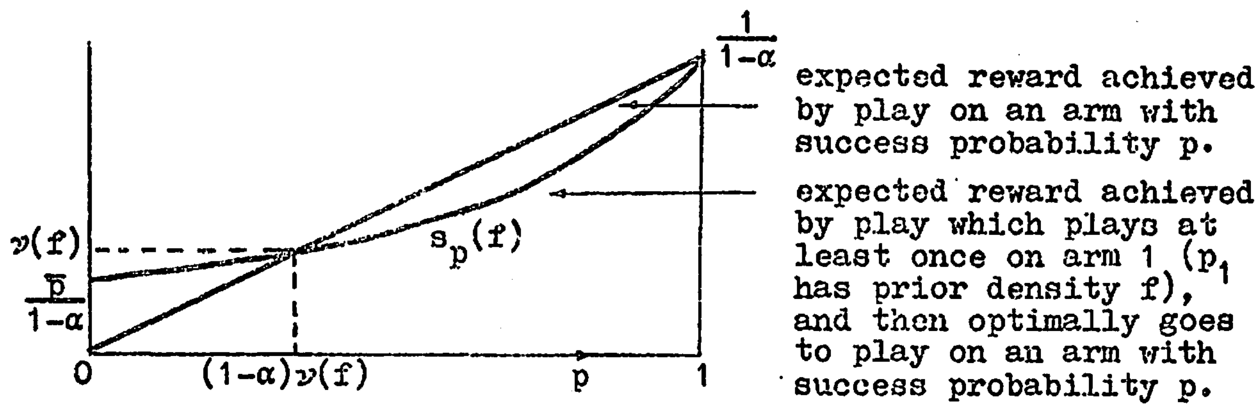
(iii) t^* can be written as $t^* = \min\{ n : \nu(f_n) \leq \frac{p_2}{1-\alpha} \}$ where f_n is the posterior density of p_1 after n plays on arm 1 and is a function satisfying $\nu(f) = \sup_{t \in \tilde{C}} E_f\{ \sum_1^t \alpha^{i-1} x_i + \alpha^t \nu(f)\}$.

proof:

(i),(ii) As in the discussion of the sequential probability ratio test it is easily seen that $\{f_n\}$ is a Markov chain. If the optimal policy ever recommends playing arm 2 it must continue to do so ever after since that decision is taken by looking at f_n and play of arm 2 leaves f_n fixed. Hence we wish to maximize $E\{\sum_1^{t-1} \alpha^{i-1} x_i + \sum_t^\infty \alpha^{i-1} p_2\} = E\{\sum_1^{t-1} \alpha^{i-1} x_i + \frac{\alpha^{t-1}}{1-\alpha} p\}$ in $\tilde{C}$.

Theorem 4.2.1 (i) applies so that an optimal t^* does exist.

(iii) Let $s_p(f) = \sup_{t\varepsilon C} E_f\{\sum_1^t \alpha^{i-1}x_i^1 + \frac{\alpha^t}{1-\alpha}p\}$, where the x_i^1's are taken from play on arm 1 (with prior density f for p_1). As the supremum of linear increasing functions of p, $s_p(f)$ is convex, continuous and increasing in p on $[0,1]$. Clearly $s_o(f) = E_f\{\sum_1^\infty \alpha^{i-1}x_i\} = \frac{\bar{p}_1}{1-\alpha}$ where $\bar{p}_1 = \int_0^1 u\, f(u)\, du$. Also, $s_1(f) = \frac{1}{1-\alpha}$. Pictorially this looks like:



Hence there is a unique (f) such that $s_{(1-\alpha)\nu(f)}(f) = \sup_{t\varepsilon C} E_f\{\sum_1^t \alpha^{i-1}x_i^1 + \alpha^t\ (f)\} = \nu(f)$. It is also clear from 4.2.1 and the picture that it is optimal in the two-armed bandit problem to stop play on arm 1 if $\frac{p_2}{1-\alpha} \geqslant \nu(f_o)$ and to go on for at least one more play on 1 if $\frac{p_2}{1-\alpha} < \nu(f_o)$. Then $t^* = \min\{n: \nu(f_n) \leqslant \frac{p_2}{1-\alpha}\}$ is $(0,\tilde{s})$- optimal.

4.3.2 Definition

The function defined by being the unique solution of $\nu(f_o) = E_{f_o}\{\sum_1^{t^*} \alpha^{i-1}x_i + \alpha^{t^*}\nu(f_o)\}$ and $t^* = \min\{n:\nu(f_n) \leqslant \nu(f_o)\}$, is called the dynamic allocation index (DAI) of the bandit arm (from which the x_i are obtained). The DAI, $\nu(f)$, of an arm whose success probability has prior density f may be thought to be the success probability of a second arm against which optimal play would give no preference as to which of the two arms to play.

It is rather an amazing fact that the optimal policy for playing two arms for which neither p_1 or p_2 is known can also be described in terms of the DAIs of the arms.

4.3.3 Theorem [ref. Gittins and Jones; Gittins and Nash]

Suppose that p_1 and p_2 are known to have prior densities f_o^1 and f_o^2 respectively. Let f_n^i be the posterior density of p_i after n plays have been made on arm i. Let $\nu_n^i = \nu(f_n^i)$. If then at time n there have n_1 and n_2 plays on arms 1 and 2 respectively (where $n_1 + n_2 = n-1$), then it is unquely optimal to make the nth pull on the arm for which $\nu_{n_i}^i$ is greatest.

proof:

We will refer to the above described playing policy as the "DAI strategy". The proof follows several stages:

(1) Given $\varepsilon > 0$ N_o such that for all $N \geq N_o$

$$E\begin{pmatrix}\text{reward of the strategy that plays according to the DAI strategy}\\ \text{for } n = N+1, \ldots \text{ and optimally subject to this constraint for}\\ n = 1, \ldots, N.\end{pmatrix}$$

$>$ E(reward of any other strategy) $- \varepsilon$. This is because $\sum_{N_o}^{\infty} \alpha^{i-1} < \varepsilon$ for large enough N_o.

(2) Let $t = \min\{n > 1 : \nu_n < \mu\}$ then by 4.3.1 $\mu \lesseqqgtr E_{f_o}\{\sum_1^t \alpha^{i-1} x_i + \alpha^t \mu\}$ as $\mu \lesseqqgtr \nu_o$.

(3) Let $\mu = \max_i\{\nu_o^i\}$ and let $t^i = \min\{n > 0 : \nu_n^i < \mu\}$. Let E_{ij} be the expected reward of the strategy which plays arm i for times $1, \ldots, t^i$, then arm j for times $t^i + 1, \ldots, t^i + t^j$, and the DAI strategy thereafter. Call this strategy S_{ij}. Then:

$$E_{12} = E\{\sum_1^{t^1} \alpha^{r-1} x_r^1 + \alpha^{t^1} \sum_1^{t^2} \alpha^{s-1} x_s^2\} + E(\text{reward beyond } t^1 + t^2 + 1)$$

$$E_{21} = E\{\sum_1^{t^2} \alpha^{s-1} x_s^2 + \alpha^{t^2} \sum_1^{t^1} \alpha^{r-1} x_r^1\} + E(\text{reward beyond } t^2 + t^1 + 1)$$

Since the values of ν^1 and ν^2 at $t^1 + t^2$ are the same when S_{12} has been played as when S_{21} has been played, the final terms in the two expressions above are equal. Hence $E_{12} - E_{21} =$

$$E\{(1-\alpha^{t^2})\sum_1^{t^1}\alpha^{r-1}x_r^1\} - E\{(1-\alpha^{t^1})\sum_1^{t^2}\alpha^{s-1}x_s^2\} \begin{matrix} < \\ = \\ < \end{matrix}$$

$$E\{(1-\alpha^{t^2})(1-\alpha^{t^1})\mu\} - E\{(1-\alpha^{t^1})(1-\alpha^{t^2})\mu\} = 0 \text{ as}$$

$\nu_0^1 \begin{matrix} = \\ = \\ < \end{matrix} \mu \begin{matrix} > \\ = \\ = \end{matrix} \nu_0^2$ by (2) above.

(4) When $\nu_0^1 = \nu_0^2$, S_{12} and S_{21} are both simply the DAI strategy. Hence (3) implies that it doesn't matter which arm is played first.

Otherwise, interpretation of S_{12} and S_{21} tells us that the strategy which plays the arm with smaller DAI once and then the DAI strategy thereafter is strictly bettered by the strategy which plays the arm with larger DAI first until its DAI is less than its intial value, then the other arm once, and the DAI strategy thereafter.

From not more than N_0 applications of this observation linked one after another in the obvious fashion we deduce that E(reward of the DAI strategy) > E(reward of any other strategy) − ε. The DAI strategy is optimal because ε is arbitrary. It is uniquely optimal because the inequalities which hold in the above arguement are always strict.

Note that the proof is easily generalized to the case of finitely many arms (the Multi-Armed Bandit Problem).

5. References

Chow, Y.S., Robbins, H., and Siegmund, D.
"Great Expectations", Houghton Miffin, 1971

Chow, Y.S., Mortigui, S., Robbins, H., and Samuels, S.M.
"Optimal selection based on relative rank",
Israel Journal of Mathematics 2 (1964), 81-90 **

Dynkin, E.B.
"The optimal choice of the instant for stopping a Markov process", Doklady Akad. Nauk. 4 (1963), 627-629 **

Dynkin, E.B., and Yushkevitch, A.A.
"Markov Processes, Theorems and Problems", Plenum Press, 1969 *

Gardner, M.
"Mathmatical games", Scientific American 202 (1960), 150 **

Gilbert, J., and Mosteller, F.
"Recognizing the maximum of a sequence",
Amer. Stat. Assoc. 61 (1966), 35-73 *

Gittins, J.C., and Jones, D.M.
"A dynamic allocation index for the sequential design of experiments", Colloquia Mathematica Societatis János Bolyai, Budapest, 1972 ***

Gittins, J.C., and Nash, P.
"Scheduling, queues and dynamic allocation indices" (unpub.) ***

Shiryaev, A.N.
"Statistical Sequential Analysis", Translations of Mathematical Monographs, Vol. 38, Amer. Math. Soc., 1973 *

Wald, A.
"Sequential Analysis", J. Wiley and Sons, 1947 *

These may be located in Cambridge at:

* Wishart Library (or Pure Mathematics)

** Scientific Periodicals Library

*** Engineering Department