4. Optimal Stopping

4.1. **Discrete-time Framework: An Introduction.** Historically, the optimal stopping problem arose in the Bayesian approach to sequential analysis of problems on testing two statistical hypotheses. In the discrete-time framework the optimal stopping problem can be formulated as follows.

Problem Statement: Given

1. a sequence of random variables, $X_1, X_2, \cdots$, with known joint distribution;
2. a sequence of real-valued reward functions

$$g_0, g_1(x_1), g_2(x_1, x_2), \cdots, g_n(x_1, \cdots, x_n), \cdots, g_\infty(x_1, x_2, \cdots),$$

where g_0 is a real number and g_n maps $\mathbb{R}^n$ to $\mathbb{R}$, $n \geq 1$,

find a "stopping rule" Φ that maximizes the expected reward, i.e., find

$$\max_N \mathbb{E}\left[g_N(x_1, \cdots, x_N)\right],$$

where the max (assuming it is achieved) is over random variables N that are determined via a (randomized) stopping rule of the form

$$\Phi = \left(\phi_0, \phi_1(x_1), \phi_2(x_1, x_2), \cdots\right),$$

where, $\phi_n : \mathbb{R}^n \mapsto [0,1]$, $n \in \mathbb{Z}_+$, and

$$\phi_n(x_1, \ldots, x_n) = \mathbb{P}(N = n | (X_1, \ldots, X_n) = (x_1, \ldots, x_n), N \geq n).$$

Here, ϕ_0 represents the probability that you stop before making any observations at all and, informally speaking, Φ_n represents the decision "whether to stop at n". If each $\phi_n(x_1, \ldots, x_n)$ is either 0 or 1, then the stopping rule is said to be non-randomized.

A simpler, more widely used, notation to model stopping rule problems is as follows. Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a probability space.

Alternative Problem Statement: Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting

1. a filtration $\{\mathcal{F}_n\}_{n \geq 0}$, where

$$\mathcal{F}_0 = \{\emptyset, \Omega\} \subset \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_n \ldots \subseteq \mathcal{F}_\infty \subseteq \mathcal{F};$$

2. a sequence of random reward variables $Y_1, \cdots, Y_n, \cdots$, where each Y_n is $\mathcal{F}_n$-measurable,

the goal is to maximize the expected reward:

$$\max_N \mathbb{E}[Y_N],$$

where the maximum (assuming it is achieved) is over all $\mathcal{F}_n$-stopping times N (i.e., random variables N such that $\{N = n\} \in \mathcal{F}_n$ for every $n \in \mathbb{N}$).

Observe that when one chooses the filtration $\{\mathcal{F}_n\}$ to be

$$\mathcal{F}_0 = \{\emptyset, \Omega\}\,; \quad \mathcal{F}_n = \sigma\left(X_1, \cdots, X_n\right); \quad \mathcal{F}_\infty = \sigma\left(\cup_{n=1}^{\infty} \mathcal{F}_n\right);$$

then the second formulation essentially reduces to the first. However, the second formulation is more general than the first as there may be filtrations that are not generated by any sequence of random variables.

We now provide some examples of problems that can be formulated as optimal stopping problems.

Example 1: " House-selling problem".
Suppose we are given "offers" $X_i \sim$ i.i.d. from a known distribution. Each offer costs \$c to observe, and we need to decide when to sell the house. We consider two scenarios that lead to different reward structures:

Can't accept past offers	Can accept past offers
$y_0 = 0$	$y_0 = 0$
$y_n(x_1, \cdots, x_n) = x_n - nc$	$y_n(x_1, \cdots, x_n) = \max\limits_{i=1,\cdots,n} x_i - nc$

Example 2: "Bayes Sequential Decision Problems" (Wald 1947)
Suppose you are given a sequence $X_1, X_2, \cdots$ (possibly dependent); A parameter θ is chosen from a parameter space Θ according to some prior distribution τ. The statistician must eventually choose an action a in a given action space $\mathcal{A}$ (action space) and incur a loss $L(\theta, a)$. She may observe as many random variable as she likes before choosing the action, at a cost of c for each X_i observed. Moreover, the variables X_i are i.i.d *given* θ, with known distribution $F(x \mid \theta)$. If he decides to stop after taking n observations, she would choose an action $a \in \mathcal{A}$ to minimize her conditional expected loss, and thus be expected to lose

$$\rho_n(x_1, \cdots, x_n) \;=\; \inf_{a \in \mathcal{A}} \mathbb{E}\left[L(\theta, a) \mid X_1, \cdots, x_n\right].$$

It turns out that, the terminal decision of what action a to choose can be made independently of the stopping rule. Taking into account both the loss and cost, the reward function here is

$$\begin{aligned} y_0 &= 0 \\ y_n(x_1, \cdots, x_n) &= \rho_n(x_1, \cdots, x_n) + nc \\ y_\infty(x_1, \cdots, x_n, \ldots) &= \infty. \end{aligned}$$

Example 3: "(Finite-Horizon) one-armed bandit problem" (Bradt, Johnson & Karlin 1956)
There are M patients and two treatments available for the cure of a disease. Treatment $T1$ has an unknown probability p of cure, though its prior distribution $G(p)$ is known. Treatment $T2$ has known probability p_0 of cure. The goal is to maximize the number of patients cured.

At first, this does not appear to be an optimal stopping problem. However, we have
Fact: For very general reward functions (and, in particular, for maximizing # of patients cured), if you ever switch to $T2$, you should continue to use $T2$ for the rest of the time.

Using this fact, we can reformulate the problem as an optimal stopping problem. Define $X_j = 1$ if patient j is cured under T1. Let N be the time at which one switches to T2. Then this is an optimal stopping problem with reward functions:

$$y_n(x_1, \cdots, x_n) = \sum_{j=1}^{n} x_j + (M - n)p, \quad n \in \{0, \ldots, M\}.$$

Example 4: Secretary Problem There are n applicants for a position. It is assumed that

- applicants can be ranked linearly from best to worst;
- applicants are interviewed sequentially;
- all $n!$ permutations of ranks, in the order of arrivals, are equally likely;
- relative ranks of arrived interviewees are known, but not their asbolute ranks.

The problem is to maximize the probability of picking the highest rank candidate.
To put this in an optimal stopping framework, let $X_1, X_2, \cdots$ be the relative ranks of the candidates. Let a relatively best applicant be called a *candidate*, i.e., the jth applicant is a candidate if and only if $X_j = 1$. If a candidate at stage j is accepted, we win if and only if the probability is the best among all applicants. This has probability j/n due to the random ordering of arrivals. Thus the reward functions are given by $y_j(x_1, \cdots, x_j) = \begin{cases} \frac{0}{n} & \text{if applicant } j \text{ is a candidate} \\ 0 & \text{otherwise} \end{cases}$.

The basic problem can actually be solved rather simply, without using the optimal stopping framework mentioned above. We now show that the optimal stopping rule has the following "threshold" structure:

Claim: The optimal stopping rule is of the form N_r for some r, where N_r is the rule that rejects the first $r - 1$ applicants and then accepts the next relatively best applicant.

To see why this is the case, let w_j be the probability of win using an optimal rule among rules that pass up the first j applicants. Then, clearly, $w_j \geq w_{j+1}$ and it is optimal to stop with a candidate at stage j if $j/n \geq w_j$. In that case, it is also optimal to stop at k for all $k \geq j$ since

$$\frac{k}{n} \geq \frac{j}{n} \geq w_j \geq w_k.$$

To find the optimal value of the **threshold** r, we explicitly calculate the probability P_r of win using N_r:

$$\begin{aligned} P_r &= \sum_{k=r}^{n} \mathbb{P}(k\text{th applicant is best and selected}) \\ &= \sum_{k=r}^{n} \mathbb{P}(k\text{th is best})\mathbb{P}(k\text{th is selected}|k\text{th is best}) \\ &= \sum_{k=r}^{n} \frac{1}{n}\mathbb{P}(\text{best of the first } k-1 \text{ appears at stage } r) \\ &= \sum_{k=r}^{n} \frac{1}{n} \cdot \frac{r-1}{k-1} = \frac{r-1}{n} \sum_{k=r}^{n} \frac{1}{k-1} \text{ where } \frac{r-1}{r-1} = 1 \text{ if } r = 1. \end{aligned}$$

Noting that $P_{r+1} \leq P_r$ if and only if

$$\frac{r}{n} \sum_{k=r+1}^{n} \frac{1}{k+1} \leq \frac{r-1}{n} \sum_{k=r}^{n} \frac{1}{k-1} \Leftrightarrow \sum_{k=r+1}^{n} \frac{1}{k-1} \leq 1,$$

we conclude that the optimal rule is N_{r^*}, where

$$r^* \doteq \min \left\{ r \geq 1 : \sum_{k=r+1}^{n} \frac{1}{k-1} \leq 1 \right\}.$$

Note that for large n, $\sum_{k=r+1}^{n} \frac{1}{(k-1)} \sim \log(n/r)$, and so $\log(n/r^*) \approx 1$, which implies $r^*/n \approx e^{-1}$.

4.2. **Discrete-time Framework: Analysis.** We now describe certain conditions, which we show are necessary and sufficient for the existence of an optimal stopping rule:

A1. $\mathbb{E}\left[\sup_n Y_n\right] < \infty$.
A2. $\limsup_{n\to\infty} Y_n \leq Y_\infty$ a.s.

• **Necessity of Conditions**

We first provide counterexamples to show the above two conditions are necessary for an optimal rule to exist.

1. $X_1, X_2, \cdots$ be i.i.d. Bernoulli trials with prob $\frac{1}{2}$ of success, and define

$$Y_n = (2^n - 1) \prod_{i=1}^{n} X_1.$$

Then $Y \to 0$ a.s., and so A2 is satisfied. Therefore,

$$\sup_n Y_n = \begin{cases} 2^k - 1 & \text{with prob. } \frac{1}{2^{k+1}} \\ 0 & \text{with prob. } \frac{1}{2} \end{cases}$$