Chapter 7

Dynamic Programming and Filtering.

7.1 Optimal Control.

Optimal control or dynamic programming is a useful and important concept in the theory of Markov Processes. We have a state space X and a family π_α of transition probability functions indexed by a parameter $\alpha \in \mathcal{A}$. The parameter α is called the control parameter and can be chosen at will from the set $\mathcal{A}$. The choice is allowed to vary over time i.e. α_j can be the parameter of choice for the transition from x_j at time j to x_{j+1} at time $j+1$. The choice can also depend on the information available up to that point, i.e. α_j can be an $\mathcal{F}_j$ measurable function. Then the conditional probability $P\{x_{j+1} \in A|\mathcal{F}_j\}$ is given by $\pi_{\alpha_j(x_0,\cdots,x_j)}(x_j\,,A)$. Of course in order for things to make sense we need to assume some measurability conditions. We have a payoff function $f(x_N)$ and the object is to maximize $E\{f(x_N)\}$ by a suitable choice of the functions $\{\alpha_j(x_0,\cdots,x_j) : 0 \leq j \leq N-1\}$. The idea (Bellman's) of dynamic programming is to define recursively (by backward induction) for $0 \leq j \leq N-1$, the sequence of functions

$$V_j(x) = \sup_\alpha \int V_{j+1}(y)\pi_\alpha(x\,,dy) \tag{1}$$

with

$$V_N(x) = f(x)$$

as well as the sequence $\{\alpha_j^*(x) : 0 \le j \le N-1\}$ of functions that provide the supremum in (1).

$$V_j(x) = \int V_{j+1}(y)\pi_{\alpha_j^*(x)}(x\,,dy) = \sup_\alpha \int V_{j+1}(y)\pi_\alpha(x\,,dy)$$

We then have

Theorem 7.1. *If the Markov chain starts from x at time 0, then $V_0(x)$ is the best expected value of the reward. The 'optimal' control is Markovian and is provided by $\{\alpha_j^*(x_j)\}$.*

Proof. It is clear that if we pick the control as α_j^* then we have an inhomogeneous Markov chain with transition probability

$$\pi_{j,\,j+1}(x\,,dy) = \pi_{\alpha_j(x)}(x\,,dy)$$

and if we denote by P_x^*, the process corresponding to it that starts from the point x at time 0, we can establish by induction that

$$E^{P_x^*}\{f(x_N)|\mathcal{F}_{N-j}\} = V_{N-j}(x_{N-j})$$

for $1 \le j \le N$. Taking $j = N$, we obtain

$$E^{P_x^*}\{f(x_N)\} = V_0(x).$$

To show that $V_0(x)$ is optimal, for any admissible (not necessarily Markovian) choice of controls, if P is the measure on $\mathcal{F}_N$ corresponding to a starting point x,

$$E^P\{V_{j+1}(x_{j+1})|\mathcal{F}_j\} \le V_j(x_j)$$

and now it follows that

$$E^P\{f(x_N)\} \le V_0(x).$$

□

Exercise 7.1. The problem could be modified by making the reward function equal to

$$E^P\{\sum_{j=0}^N f_j(\alpha_{j-1}\,, x_j)\}$$

and thereby incorporate the cost of control into the reward function. Work out the recursion formula for the optimal reward in this case.

7.2 Optimal Stopping.

A special class of optimization problems are called optimal stopping problems. We have a Markov chain with transition probability $\pi(x, dy)$ and time runs from 0 to N. We have the option to stop at any time based on the history up to that time. If we stop at time k in the state x, the reward is $f(k\,,x)$, The problem then is to maximize $E_x\{f(\tau\,,x_\tau)\}$ over all stopping times $0 \le \tau \le N$. If $V(k\,,x)$ is the optimal reward if the game starts from x at time k, the best we can do starting from x at time $k-1$ is is to earn a reward of

$$V(k-1\,,x) = \max[f(k-1\,,x), \int V(k\,,y)\pi(x,dy)]$$

Starting with $V(N\,,x) = f(N\,,x)$, by backwards induction, we can get $V(j\,,x)$ for $0 \le j \le N$. The optimal stopping rule is given by

$$\bar{\tau} = \{\inf k : V(k\,,x_k) = f(k\,,x_k)\}$$

Theorem 7.2. *For any stopping time τ with $0 \le \tau \le N$,*

$$E_x\{f(\tau\,,x_\tau)\} \le V(0\,,x)$$

and

$$E_x\{f(\bar{\tau}\,,x_{\bar{\tau}})\} = V(0\,,x)$$

Proof. Because

$$V(k\,,x) \ge \int V(k+1\,,y)\pi(x\,,dy)$$

we conclude that $V(k\,,x_k)$ is a supermartingale and and an application of Doob's stopping theorem proves the first claim. On the other hand if $V(k\,,x) > f(k\,,x)$, we have

$$V(k\,,x) = \int V(k+1\,,y)\pi(x\,,dy)$$

and this means $V(\bar{\tau} \wedge k\,,x_{\bar{\tau}\wedge k})$ is a martingale and this establishes the second claim. □

Example 7.1. (The Secretary Problem.) An interesting example is the following game.

We have a lottery with N tickets. Each ticket has a number on it. The numbers $a_1 \cdots , a_N$ are distinct, but the player has no idea of what they are. The player draws a ticket at random and looks at the number. He can either keep the ticket or reject it. If he rejects it, he can draw another ticket from the remaining ones and again decides if he wants to keep it. The information available to him is the numbers on the tickets he has so far drawn and discarded as well as the number on the last ticket that he has drawn and is holding. If he decides to keep the ticket at any stage, then the game ends and that is his ticket. Of course if he continues on till the end, rejecting all of them, he is forced to keep the last one. The player wins only if the ticket he keeps is the one that has the largest number written on it. He can not go back and claim a ticket that he has already rejected and he can not pick a new one unless he rejects the one he is holding. Assuming that the draws are random at each stage, how can the player maximize the probability of winning? How small is this probability?

It is clear that the strategy to pick the first or the last or any fixed draw has the probability of $\frac{1}{N}$ to win. It is not apriori clear that the probability p_N of winning under the optimal strategy remains bounded away from 0 for large N. It seems unlikely that any strategy can pick **the winner** with significant probability far large values of N. Nevertheless the following simple strategy shows that

$$\liminf_{N\to\infty} p_N \ge \frac{1}{4}.$$

Let half the draws go by, no matter what, and then pick the first one which is the highest among the tickets drawn up to the time of the draw. If the second best has already been drawn and the best is still to come, this strategy will succeed. This has probability nearly $\frac{1}{4}$. In fact the strategy works if the k best tickets have not been seen during the first half, $(k+1)$-th has been and among the k best the highest shows up first in the second half. The probability for this is about $\frac{1}{k2^{k+1}}$, and as these are disjoint events

$$\liminf_{n\to\infty} p_N \ge \sum_{k\ge 1} \frac{1}{k2^{k+1}} = \frac{1}{2}\log 2$$

If we decide to look at the first Nx tickets rather than $\frac{N}{2}$, the lower bound becomes $x \log \frac{1}{x}$ and an optimization over x leads to $x = \frac{1}{e}$ and the resulting lower bound

$$\liminf_{n\to\infty} p_N \geq \frac{1}{e}.$$

We will now use the method optimal stopping to decide on the best strategy for every N and show that the procedure we described is about the best. Since the only thing that matters is the ordering of the numbers, the numbers themselves have no meaning. Consider a Markov chain with two states 0 and 1. The player is in state 1 if he is holding the largest ticket so far. Otherwise he is in state 0. If he is in state 1 and stops at stage k, i.e. when k tickets have been drawn, the probability of his winning is easily calculated to be $\frac{k}{N}$. If he is in state 0, he has to go on and the probability of landing on 1 at the next step is calculated to be $\frac{1}{k+1}$. If he is at 1 and decides to play on the probability is still $\frac{1}{k+1}$ for landing on 1 at the next stage. The problem reduces to optimal stopping for a sequence $X_1, X_2, \cdots, X_N$ of independent random variables with $P\{X_i = 1\} = 1i+1$, $P\{X_i = 0\} = \frac{i}{i+1}$ and a reward function of $f(i\,,1) = \frac{i}{N}\,; f(i\,,0) = 0$. Let us define recursively the optimal probabilities

$$V(i\,,0) = \frac{1}{I+1}V(i+1,1) + \frac{i}{i+1}V(i+1\,,0)$$

and

$$V(i\,,1) = \max[\frac{i}{N}\,, \frac{1}{I+1}V(i+1,1) + \frac{i}{i+1}V(i+1\,,0)] = \max[\frac{i}{N}\,, V(i\,,0)]$$

It is clear what the optimal strategy is. We should draw always if we are in state 0, i.e. we are sure to lose if we stop. If we are holding a ticket that is the largest so far, we should stop provided

$$\frac{i}{N} > V(i\,,0)$$

and go on if

$$\frac{i}{N} < V(i\,,0).$$

Either startegy is acceptable in case of equality. Since $V(i{+}1\,,1) \geq V(i{+}1\,,0)$ for all i, it follows that $V(i\,,0) \geq V(i+1,0)$. There is therefore a critical $k(=k_N)$ such that $\frac{i}{N} \geq V(i\,,0)$ if $i \geq k$ and $\frac{i}{N} \leq V(i\,,0)$ if $i \leq k$. The best strategy is to wait till k tickets have been drawn, discarding every ticket, and then pick the first one that is the best so far. The last question is the determination of $k = k_N$. For $i \geq k$,

$$V(i\,,0) = \frac{1}{i+1}\frac{i+1}{N} + \frac{i}{i+1}V(i+1\,,0) = \frac{1}{N} + \frac{i}{i+1}V(i+1\,,0)$$

or

$$\frac{V(i\,,0)}{i} - \frac{V(i+1\,,0)}{i+1} = \frac{1}{N}\cdot\frac{1}{i}$$

telling us

$$V(i\,,0) = \frac{i}{N}\sum_{j=i}^{N-1}\frac{1}{j}$$

so that

$$k_N = \inf\left\{i : \frac{1}{N}\sum_{j=i}^{N-1}\frac{1}{j} < \frac{1}{N}\right\}$$

Approximately $\log N - \log k_N = 1$ or $k_N = \frac{N}{e}$.

7.3 Filtering.

The problem in filtering is that there is an underlying stochastic process that we cannot observe. There is a related stochastic process ‘driven’ by the first one that we can observe and we want to use our information to make conclusions about the state of the unobserved process. A simple but extreme example is when the unobserved process does not move and remains at the same value. Then it becomes a parameter. The driven process may be a sequence of i.i.d random variables with densities $f(\theta, x)$ where θ is the unobserved, unchanging underlying parameter. We have a sample of n independent observations $X_1, \cdots, X_n$ from the common distribution $f(\theta, x)$ and our goal is then nothing other than parameter estimation. We shall take a Bayesian approach. We have a prior distribution $\mu(d\theta)$ on the space of parameters Θ and this can be modified to an ‘aposteriori’ distribution after the sample is observed. We have the joint distribution

$$\prod_{i=1}^{n} f(\theta, x_i)\,dx_i\,\mu(d\theta)$$

and we calculate the conditional distribution of

$$\mu_n(d\theta|x_1\cdots x_n)$$

given $x_1,\cdots,x_n$. This is our best informed guess about the nature of the unknown parameter. We can use this information as we see fit. If we have an additional observation x_{n+1} we need not recalculate everything, but we can simply update by viewing μ_n as the new prior and calculating the posterior after a single observation x_{n+1}.

We will just work out a single illustration of this known as the Kallman-Bucy filter. Suppose $\{x_n\}$ the unobserved process is a Gaussian Markov chain

$$x_{n+1}=\rho x_n+\sigma\xi_{n+1}$$

with $0<\rho<1$ and the noise term ξ_n are i.i.d normally distributed random variables with mean 0 and variance 1. The observed process y_n is given by

$$y_n=x_n+\eta_n$$

where the $\{\eta_j\}$ are again independent standard Gaussians that are independent of the $\{\xi_j\}$ as well. If we start with an initial distribution for x_0 say one that is Gaussian with mean m_0 and variance σ_0^2, we can compute the joint distribution of x_0, x_1 and y_1 and then the conditional of x_1 given y_1. This becomes the new distribution of the state x_1 based on the observation y_1. This allows us te calculate recursively at every stage.

Let us do this explicitly now. The distribution of x_1, y_1 is jointly normal with mean $(\rho m_0\,,\rho m_0)$ variances $(\rho^2\sigma_0^2+\sigma^2\,,\rho^2\sigma_0^2+\sigma^2+1)$ and covariance $(\rho^2\sigma_0^2+\sigma^2)$. The posterior distribution of x_1 is again Normal with mean

$$\begin{aligned}m_1&=\rho m_0+\frac{(\rho^2\sigma_0^2+\sigma^2)}{(\rho^2\sigma_0^2+\sigma^2+1)}(y_1-\rho m_0)\\&=\frac{1}{(\rho^2\sigma_0^2+\sigma^2+1)}m_0+\frac{(\rho^2\sigma_0^2+\sigma^2)}{(\rho^2\sigma_0^2+\sigma^2+1})y_1\end{aligned}$$

and variance

$$\begin{aligned}\sigma_1^2&=(\rho^2\sigma_0^2+\sigma^2)(1-\frac{(\rho^2\sigma_0^2+\sigma^2)}{(\rho^2\sigma_0^2+\sigma^2+1)}\\&=\frac{(\rho^2\sigma_0^2+\sigma^2)}{(\rho^2\sigma_0^2+\sigma^2+1)}\end{aligned}$$

After a long time while the recursion for m_n remains the same

$$m_n = \frac{1}{(\rho^2\sigma_0^2 + \sigma^2 + 1)} m_{n-1} + \frac{(\rho^2\sigma_0^2 + \sigma^2)}{(\rho^2\sigma_0^2 + \sigma^2 + 1})y_n$$

the variance σ_n^2 has an asymptotic value σ_∞^2 given by the solution of

$$\sigma_\infty^2 = \frac{(\rho^2\sigma_\infty^2 + \sigma^2)}{(\rho^2\sigma_\infty^2 + \sigma^2 + 1)} \ .$$