Introduction to Mathematical Finance

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 500

Subsection 1

The Stochastic Dynamic Programming Problem

The Setup

- In the general **stochastic dynamic programming problem**, we suppose that a system is observed at the beginning of each period and its state is determined.
- Let $\mathcal{S}$ denote the set of all possible states.
- After observing the state of the system, an action must be chosen.
- If the state is x and action a is chosen, then:
  (a) A reward $r(x, a)$ is earned;
  (b) The next state, call it $Y(x, a)$, is a random variable whose distribution depends only on x and a.

Maximal Expected Return

- Suppose our objective is to maximize the expected sum of rewards that can be earned over N time periods.
- Let $V_n(x)$ denote the maximal expected sum of rewards that can be earned in the next n time periods given that the current state is x.
- If we initially choose action a, then:
  - A reward $r(x, a)$ is immediately earned;
  - The next state will be $Y(x, a)$.
- If $Y(x, a) = y$, then at that point:
  - There will be an additional $n - 1$ time periods to go;
  - So the maximal expected additional return would be $V_{n-1}(y)$.

Maximal Expected Return (Cont'd)

- Summarizing, assuming that:
  - The current state is x;
  - We initially choose action a,

  the maximal expected return that could be earned over the next n time periods is
  $$r(x,a)+E[V_{n-1}(Y(x,a))].$$
- Hence, the overall maximal expected return $V_n(x)$ satisfies
  $$V_n(x)=\max_a\{r(x,a)+E[V_{n-1}(Y(x,a))]\}.$$
- Starting with $V_0(x)=0$ the preceding equation can be used to recursively solve:
  - For the function $V_1(x)$;
  - For the function $V_2(x)$;

    $\vdots$
  - For the function $V_N(x)$.

Optimal Value Function

- The optimal policy, when there are n additional time periods to go with the current state being x, chooses the action (or one of the actions) that maximizes the right side of the preceding.
- We let $a_n(x)$ be the action maximizing $r(x, a) + E[V_{n-1}(Y(x, a))]$.
- This is written as

$$a_n(x) = \mathrm{argmax}_a\{r(x, a) + E[V_{n-1}(Y(x, a))]\}, \quad n = 1, \ldots, N.$$

- Then an optimal policy chooses, for all n and x, $a_n(x)$, when:
  - The state is x;
  - There are n time periods remaining.
- The function $V_n(x)$ is called the **optimal value function**.
- The equation for $V_n(x)$ is called the **optimality equation**.

Discrete Form

- Suppose $\mathcal{S}$ is a subset of the set of all integers.
- Let $P_{i,a}(j)$ denote the probability that the next state is j, when:
  - The current state is i;
  - Action a is chosen.
- In this case, the optimality equation can be written

$$V_n(i) = \max_a \left\{ r(i,a) + \sum_j P_{i,a}(j) V_{n-1}(j) \right\}.$$

Continuous Form

- Suppose, on the other hand, that $\mathcal{S}$ is a continuous set.
- Let $f_{x,a}(y)$ be the probability density of the next state given that:
  - The current state is x;
  - Action a is chosen.
- In this case, the optimality equation can be written

$$V_n(x) = \max_a \left\{ r(x,a) + \int f_{x,a}(y) V_{n-1}(y) dy \right\}.$$

Discounting

- In certain problems future costs may be discounted.
- Specifically, a cost incurred k time periods in the future may be discounted by the factor β^k.
- In such cases the optimality equation becomes

$$V_n(x) = \max_a \{r(x,a) + \beta E[V_{n-1}(Y(x,a))]\}.$$

- For instance, if we wanted to maximize the present value of the sum of rewards, then we would let $\beta = \frac{1}{1+r}$, where r is the interest rate per period.
- The quantity β is called the **discount factor**.
- It is usually assumed to satisfy $0 \leq \beta \leq 1$.

Optimal Return from a Call Option

- Assume the following discrete time model for the price movement of a security.
- Whatever the price history so far, the price of the security during the following period is its current price multiplied by a random variable Y.
- Assume an interest rate of $r > 0$ per period.
- Let $\beta = \frac{1}{1+r}$.
- We want to determine the appropriate value of an American call option having:
  - Exercise value K;
  - Expiration time at the end of n additional periods.

Comments

- We are not assuming that Y has only two possible values.
- So there will not be a unique risk-neutral probability law.
- Consequently, arbitrage considerations will not enable us to determine the value of the option.
- We make the additional assumption that the security cannot be sold short for the market price.
- So there will no longer be an arbitrage argument against early exercising.
- To determine the appropriate value of the option under these conditions, we will suppose that the successive Y's are independent with a common specified distribution.
- We aim to determine the maximal expected present-value return that can be obtained from the option.

Available Options and Returns

- The current state of the system will be the current price.
- Define the optimal value function $V_j(x)$ to equal the maximal expected present-value return from the option given that:
  - It has not yet been exercised;
  - A total of j periods remain before the option expires;
  - The current price of the security is x.
- Suppose the preceding describes the current situation.
  - If the option is exercised, then a return $x - K$ is earned and the problem ends;
  - If the option is not exercised, then the maximal expected present-value return will be $E[\beta V_{j-1}(xY)]$.

Optimal Policy

- The overall best is the maximum of the best one can obtain under the different possible actions.
- So the optimality equation is

$$V_j(x) = \max\{x - K, \beta E[V_{j-1}(xY)]\}.$$

- Moreover, the boundary condition is

$$V_0(x) = (x - K)^+ = \max\{x - K, 0\}.$$

- Consider the policy that, when the current price is x and j periods remain before the option expires:
  - Exercises if $V_j(x) = x - K$;
  - Does not exercise if $V_j(x) > x - K$.
- This is an optimal policy.
- So the optimal policy exercises in state x when j periods remain if and only if $V_j(x) = x - K$.

Structure of the Optimal Policy

- We determine the structure of the optimal policy.
- We show that:
  - If $E[Y] \geq 1 + r$, then the call option should never be exercised early;
  - If $E[Y] < 1 + r$, then there is a nondecreasing sequence x_j, $j \geq 0$, such that the policy:

    Exercise when j periods remain, if the current price is at least x_j.

    is an optimal policy.

First Case

Lemma

If $E[Y] \geq 1 + r$, then the policy that only exercises when no additional time remains and the price is greater than K is an optimal policy.

- It follows from the optimality equation that $V_j(x) \geq x - K$.

  We also have $\beta E[Y] \geq \beta(1 + r) = 1$.

  So, for $j \geq 1$,

  $$\beta E[V_{j-1}(xY)] \geq \beta E[xY - K] \geq x - \beta K > x - K.$$

  Thus, it is never optimal to exercise early.

An Auxiliary Lemma

Lemma

If $E[Y] < 1 + r$, then $V_j(x) - x$ is a decreasing function of x.

- The proof is by induction on j.

  For $j = 0$, $V_0(x) - x = \max\{-K, -x\}$. So the result holds.

  Assume that $V_{j-1}(x) - x$ is decreasing in x.

  Then, by the optimality equation,

$$\begin{aligned} V_j(x) - x &= \max\{-K, \beta E[V_{j-1}(xY)] - x\} \\ &= \max\{-K, \beta(E[V_{j-1}(xY)] - xE[Y]) + \beta x E[Y] - x\} \\ &= \max\{-K, \beta E[V_{j-1}(xY) - xY] + x(\beta E[Y] - 1)\}. \end{aligned}$$

  By the induction hypothesis, for all Y, $(V_{j-1}(xY) - xY) \searrow x$.

  Therefore $E[V_{j-1}(xY) - xY]$ is also decreasing in x.

  As $\beta E[Y] < 1$, $x(\beta E[Y] - 1)$ is decreasing in x.

  So $\beta E[V_{j-1}(xY) - xY] + x(\beta E[Y] - 1)$ is decreasing in x.

Second Case

Proposition

If $E[Y] < 1 + r$, then there is a increasing sequence x_j, $j \geq 0$, such that the policy that exercises when j periods remain, whenever the current price is at least x_j, is an optimal policy.

- Let $x_j = \min\{x : V_j(x) = x - K\}$ be the minimal price at which it is optimal to exercise when j periods remain.

  By the preceding lemma, for $x' > x_j$,
  $$V_j(x') - x' \leq V_j(x_j) - x_j = -K.$$
  But the optimality equation yields that $V_j(x') \geq x' - K$.

  So we see that
  $$V_j(x') = x' - K.$$
  This shows that it is optimal to exercise when j stages remain and the current price is x' if and only if $x' \geq x_j$.

Second Case (Cont'd)

- We show, next, that x_j increases in j.

  We use that $V_j(x)$ is increasing in j.

  This follows from the fact that having additional time before the option expires cannot reduce the maximal expected return.

  Using $V_j(x) \nearrow j$, yields

  $$V_{j-1}(x_j) \leq V_j(x_j) = x_j - K.$$

  By the optimality equation, $V_{j-1}(x_j) \geq x_j - K$.

  So the preceding equation shows that $V_{j-1}(x_j) = x_j - K$.

  But x_{j-1} is the smallest value of x for which $V_{j-1}(x) = x - K$.

  So the preceding yields that $x_{j-1} \leq x_j$ and completes the proof.

Example

- An urn initially has:
  - n red balls;
  - m blue balls.
- At each stage the player may randomly choose a ball from the urn.
  - If the ball is red, then 1 is earned;
  - If it is blue, then 1 is lost.
- The chosen ball is discarded.
- At any time the player can decide to stop playing.
- We maximize the player's total expected net return.
- We analyze this as a dynamic programming problem with the state equal to the current composition of the urn.

Example (Optimality Equation)

- We let $V(r,b)$ denote the maximum expected additional return given that there are currently:
  - r red balls in the urn;
  - b blue balls in the urn.
- The expected immediate reward if a ball is chosen in state (r,b) is
$$\frac{r}{r+b}-\frac{b}{r+b}=\frac{r-b}{r+b}.$$
- The best one can do after the initial draw is:
  - $V(r-1,b)$ if a red ball is chosen;
  - $V(r,b-1)$ if a blue ball is chosen.
- So the optimality equation is
$$V(r,b)=\max\left\{0,\frac{r-b}{r+b}+\frac{r}{r+b}V(r-1,b)+\frac{b}{r+b}V(r,b-1)\right\}.$$
- We start with $V(r,0)=r$ and $V(0,b)=0$.
- Then use the optimality equation to obtain $V(n,m)$.

Example

- Suppose we can make up to n bets in sequence.
- Each bet consists of choosing a stake amount s, which can be any nonnegative value less than or equal to the current fortune.
- The result of the bet is that the amount sY is returned, where Y is a nonnegative random variable with a known distribution.
- We wish to maximize the expected value of the logarithm of the final fortune after n bets have taken place.
- The state is the current fortune.
- Let $V_n(x)$ be the maximal expected logarithm of the final fortune if:
    - The current fortune is x;
    - n bets remain.
- Let the decision be the fraction α of the current wealth to stake.

Example (Optimality Equation)

- After betting the amount αx:
  - The fortune is $\alpha x Y + x - \alpha x = x(\alpha Y + 1 - \alpha)$;
  - $n - 1$ bets remain.
- So the optimality equation becomes

$$V_n(x) = \max_{0 \leq \alpha \leq 1} E[V_{n-1}(x(\alpha Y + 1 - \alpha))].$$

Example (Fist Step)

- We assumed $V_0(x) = \log(x)$.
- So we get

$$\begin{aligned} V_1(x) &= \max_{0 \leq \alpha \leq 1} E[\log(x(\alpha Y + 1 - \alpha))] \\ &= \log(x) + \max_{0 \leq \alpha \leq 1} E[\log(\alpha Y + 1 - \alpha)] \\ &= \log(x) + C, \end{aligned}$$

  where $C = \max_{0 \leq \alpha \leq 1} E[\log(\alpha Y + 1 - \alpha)]$.
- Denote again the value of α that maximizes $E[\log(\alpha Y + 1 - \alpha)]$ by

$$\alpha^* = \text{argmax}_\alpha E[\log(\alpha Y + 1 - \alpha)].$$

- Then the optimal policy when only one bet can be made is to bet $\alpha^* x$ if your current wealth is x.

Example (Next Step)

- Now suppose the current fortune is x and two bets remain.
- Then the maximal expected logarithm of the final fortune is

$$\begin{aligned} V_2(x) &= \max_{0\leq\alpha\leq 1} E[V_1(x(\alpha Y+1-\alpha))] \\ &= \max_{0\leq\alpha\leq 1} E[\log(x(\alpha Y+1-\alpha))+C] \\ &= \log(x)+C+\max_{0\leq\alpha\leq 1} E[\log(\alpha Y+1-\alpha)] \\ &= \log(x)+2C. \end{aligned}$$

- Once again, it is optimal to stake the fraction α^* of the total wealth.
- Using mathematical induction, we can show:
  - For all n,
  $$V_n(x)=\log(x)+nC;$$
  - It is optimal, no matter how many bets remain, to always stake the fraction α^* of the total wealth.

Subsection 2

Infinite Time Models

Setup

- We look at stochastic dynamic programming problems in which the total expected reward earned over an infinite time horizon is to be maximized.
- The problem begins at time 0.
- X_n is the state at time n.
- A_n is the action chosen at time n.
- A policy π is a rule for choosing actions.
- E_π indicates that we are taking the expectation under the assumption that policy π is employed.
- We want to choose the policy π that maximizes

$$V_\pi(x) = E_\pi\left[\sum_{n=0}^{\infty} r(X_n, A_n) | X_0 = x\right].$$

- We will assume that the sum is well defined and finite.

Setup (Cont'd)

- Suppose the one stage rewards $r(x,a)$ are bounded, $|r(x,a)| < M$.
- Assume a discount factor β, with $0 \leq \beta < 1$.
- The expected total discounted cost of a policy π is $\leq \frac{M}{1-\beta}$.
- Now consider the optimal value function

$$V(x) = \max_{\pi} V_{\pi}(x).$$

- $V(x)$ satisfies the optimality equation

$$V(x) = \max_{a} \{r(x,a) + E[V(Y(x,a))]\}.$$

Example: An Optimal Asset Selling Problem

- Suppose we receive an offer each day for an asset we want to sell.
- When the offer is received, we must:
  - Pay a cost $c > 0$;
  - Decide whether to accept or to reject the offer.
- Suppose that successive offers are independent with probability mass function
$$p_j = P(\text{offer is } j), \quad j \geq 0.$$
- We want to determine the policy that maximizes the expected net return.
- The state is the current offer.
- Let $V(i)$ denote the maximal additional net return from here on, given that an offer of i has just been received.

Example (Optimality Equation)

- If the offer is accepted, then $-c+i$ is received and the problem ends.
- If the offer is rejected, then c is paid and we wait for the next offer.
- The next offer will equal j with probability p_j.
- If the next offer is j, then the maximal expected return from that point on would be $V(j)$.
- So the maximal expected net return if the offer of i is rejected is $-c+\sum_j p_j V(j)$.
- The maximum expected net return is the maximum of the maximum in the two cases.
- So the optimality equation is

$$V(i)=\max\left\{-c+i,-c+\sum_j p_j V(j)\right\}.$$

- Setting $v=\sum_j p_j V(j)$, we get $V(i)=-c+\max\{i,v\}$

Example (Solution)

- It follows from the preceding that the optimal policy is to accept offer i if and only if it is at least v.
- To determine v, note that

$$V(i) = \begin{cases} -c + v, & \text{if } i \leq v, \\ -c + i, & \text{if } i > v. \end{cases}$$

- Hence,

$$\begin{gathered} v = \sum_i p_i V(i) = -c + \sum_{i \leq v} v p_i + \sum_{i > v} i p_i \\ v \sum p_i - v \sum_{i \leq v} p_i = -c + \sum_{i > v} i p_i \\ v \sum_{i > v} p_i = -c + \sum_{i > v} i p_i \\ \sum_{i > v} (i - v) p_i = c \\ c = \sum_i (i - v)^+ p_i. \end{gathered}$$

Example (Optimal Policy)

- Let X be a random variable having the distribution of an offer.
- Then the preceding states that
$$c = E[(X - v)^+].$$
- That is, v is that value that makes $E[(X - v)^+]$ equal to c.
- In most cases, v will have to be numerically determined.
- The optimal policy is to accept the first offer that is at least v.
- Since $v = \sum_i p_i V(i)$, v is the maximum expected net return before the initial offer is received.

Example: A Machine Replacement Model

- Suppose that at the beginning of each period a machine is evaluated to be in some state i, $i = 0, \ldots, M$.
- After the evaluation, one must decide whether to pay the amount R and replace the machine or leave it alone.
  - If the machine is replaced, then a new machine, whose state is 0, will be in place at the beginning of the next period.
  - If a machine in state i is not replaced, then at the beginning of the next time period that machine will be in state j with probability $P_{i,j}$.
- Suppose that an operating cost $C(i)$ is incurred whenever the machine in use is evaluated as being in state i.
- Assume a discount factor $0 < \beta < 1$.
- The objective is to minimize the total expected discounted cost over an infinite time horizon.

Example (Optimality Equation)

- Let $V(i)$ be the minimal expected discounted cost when starting in i.
- If the machine is replaced:
  - We incur an immediate cost $C(i) + R$;
  - The minimal expected additional cost from then on is $\beta V(0)$.
- If the machine is not replaced:
  - Our immediate cost is $C(i)$;
  - The best we can do, if the next state is j, is $\beta V(j)$.

  So, if we continue in state i, the minimal expected total discounted cost is $C(i) + \beta \sum_j P_{i,j} V(j)$.
- The optimality equation is

$$V(i) = C(i) + \min \left\{ R + \beta V(0), \beta \sum_j P_{i,j} V(j) \right\}.$$

- Moreover, the policy that replaces a machine in state i if and only if $\beta \sum_j P_{i,j} V(j) \geq R + \beta V(0)$ is an optimal policy.

Example (Increasing Minimal Expected Discounted Cost)

- Suppose we want to determine conditions that imply that $V(i)$ is increasing in i.
- One condition we might want to assume is that the operating costs $C(i)$ are increasing in i.

  Assumption 1: $C(i+1) \geq C(i)$, $i \geq 0$.
- After some thought, we can see that Assumption 1 by itself would not imply that $V(i)$ increases in i.
  - Assume, e.g., that $C(10) < C(11)$.
  - Even though state 11 has a higher operating cost than state 10, it may be more likely to get us to a better state.
  - So it is possible that state 11 is preferable to state 10.

Example (Assumption 2)

- To rule this out, we assume that $N(i)$, the next state of a not replaced machine, currently in state i, is stochastically increasing in i.

  Assumption 2: $N_{i+1} \geq_{\text{st}} N_i$, $i \geq 0$.

- Recall that $N_{i+1} \geq_{\text{st}} N_i$ means

$$P(N_{i+1} \geq k) \geq P(N_i \geq k), \quad \text{for all } k.$$

- This can be written as

$$\sum_{j \geq k} P_{i+1,j} \geq \sum_{j \geq k} P_{i,j}, \quad \text{for all } k.$$

- By a previous proposition, Assumption 2 is equivalent to

  Assumption 2: $E[h(N_i)]$ increases in i whenever h is an increasing function.

Example (Theorem)

Theorem

Under Assumptions 1 and 2:

(a) $V(i)$ is increasing in i.

(b) For some $0 \leq i^* \leq \infty$, the policy that replaces when in state i if and only if $i \geq i^*$ is an optimal policy.

- Let $V_n(i)$ denote the minimal expected discounted costs over an n-period problem that starts with a machine in state i. Then

$$V_n(i) = C(i) + \min\left\{R + \beta V_{n-1}(0), \beta \sum_j P_{i,j} V_{n-1}(j)\right\}, \quad n \geq 1.$$

We argue by induction that $V_n(i)$ is increasing in i, for all n.

Example (Part (a))

- Suppose $n = 1$. We have $V_1(i) = C(i)$.

  By Assumption 1, the result holds when $n = 1$.

  Assume that $V_{n-1}(i)$ is increasing in i.

  By Assumption 2, $E[V_{n-1}(N_i)]$ increases in i.

  But we have:
  - $E[V_{n-1}(N_i)] = \sum_j P_{i,j} V_{n-1}(j)$;
  - $V_n(i) = C(i) + \min\left\{R + \beta V_{n-1}(0), \beta \sum_j P_{i,j} V_{n-1}(j)\right\}$.

  Hence, using Assumption 1, $V_n(i)$ increases in i.

  Now $V(i) = \lim_{n\to\infty} V_n(i)$.

  So $V(i)$ increases in i.

Example (Part (b))

- We prove (b) by using that the optimal policy is to replace the machine in state i if and only if

$$\beta \sum_j P_{i,j} V(j) \geq R + \beta V(0).$$

This can be written as

$$E[V(N_i)] \geq \frac{R + \beta V(0)}{\beta}.$$

But $E[V(N_i)]$ is, by Part (a) and Assumption 2, increasing in i. Let

$$i^* = \min\left\{ i : E[V(N_i)] \geq \frac{R + \beta V(0)}{\beta} \right\}.$$

Then $E[V(N_i)] \geq \frac{R+\beta V(0)}{\beta}$ if and only if $i \geq i^*$.

Subsection 3

Optimal Stopping Problems

Optimal Stopping Problems

- An **optimal stopping problem** is a two-action problem.
- When in state x, one can choose between:
    - Pay $c(x)$ and continue to the next state $Y(x)$, whose distribution depends only on x;
    - Stop and earn a final reward $r(x)$.
- Let $V(x)$ be the maximal expected net additional return given that the current state is x.
- The optimality equation is

$$V(x) = \max\{r(x), -c(x) + E[V(Y(x))]\}.$$

A Special Case

- Suppose the state space is the set of integers.
- Let $P_{i,j}$ be the probability of going from state i to state j, if one decides not to stop in state i.
- Then the optimality equation takes the form

$$V(i) = \max\left\{r(i), -c(i) + \sum_j P_{i,j} V(j)\right\}.$$

The Finite Time Version

- Let $V_n(i)$ denote the maximal expected net return given that:
  - The current state is i;
  - One is only allowed to go at most n additional time periods before stopping.
- Then, by the usual argument,

$$\begin{array}{rcl} V_0(i) & = & r(i); \\ V_n(i) & = & \max\{r(i), -c(i) + \sum_j P_{i,j} V_{n-1}(j)\}. \end{array}$$

- Having additional time periods before one must stop cannot hurt.
- So we get that:
  - $V_n(i)$ increases in n;
  - $V_n(i) \leq V(i)$.

Stability

Definition

If $\lim_{n\to\infty} V_n(i) = V(i)$, the stopping problem is said to be **stable**.

- Most, though not all, stopping-rule problems that arise are stable.
- A sufficient condition for the stopping problem to be stable is the existence of constants $c > 0$ and $r < \infty$ such that

$$c(x) > c \quad \text{and} \quad r(x) < r, \text{ for all } x.$$

One-Stage Lookahead Policy

- **One-Stage Lookahead Policy**: Stop in state i if stopping would give a return that is at least as large as the expected return obtained by continuing for exactly one more period and then stopping.
- Suppose we are at state i.
  - Immediate stopping yields a final return $r(i)$;
  - Going exactly one more period and then stopping results in an expected additional return of $-c(i)+\sum_j P_{i,j}r(j)$.
- Let

$$B=\left\{i: r(i)\geq -c(i)+\sum_j P_{i,j}r(j)\right\}$$

  be the set of states for which immediate stopping is at least as good as continuing for one period and then stopping.
- The one-stage lookahead policy is the policy that:
  - Stops when the current state i is in B;
  - Continues when the current state i is not in B.

Optimality of One-Stage Lookahead

- Consider an optimal stopping problem.
- Assume that it is stable.
- Assume that the set of states B is closed.
- This means that, if the current state is in B, and one chooses to continue, then the next state will necessarily also be in B.
- We show that, for optimal stopping problems satisfying those two conditions, the one state lookahead policy is an optimal policy.

Theorem

Suppose the problem satisfies the following:

- It is stable;
- $P_{i,j} = 0$ for $i \in B$, $j \notin B$.

Then the one stage lookahead policy is an optimal policy.

Optimality of One-Stage Lookahead (Cont'd)

- Note first that it cannot be optimal to stop in state i when $i \notin B$.

  This is so because better than stopping is to continue exactly one additional stage and then stop.

  So we need to prove that it is optimal to stop in state i when $i \in B$.

  I.e., that $V(i) = r(i)$, $i \in B$.

  We prove this by showing, by induction, that for all n,

  $$V_n(i) = r(i), \quad i \in B.$$

  We have $V_0(i) = r(i)$. So the preceding is true when $n = 0$.

  Assume that $V_{n-1}(i) = r(i)$, for all $i \in B$.

Optimality of One-Stage Lookahead (Cont'd)

- Then, for $i \in B$,

$$\begin{aligned} V_n(i) &= \max\left\{r(i), -c(i) + \sum_j P_{i,j} V_{n-1}(j)\right\} \\ &= \max\left\{r(i), -c(i) + \sum_{j \in B} P_{i,j} V_{n-1}(j)\right\} \quad (B \text{ closed}) \\ &= \max\left\{r(i), -c(i) + \sum_{j \in B} P_{i,j} r(j)\right\} \quad (\text{induction}) \\ &= r(i). \quad (i \in B) \end{aligned}$$

Hence, $V_n(i) = r(i)$ for $i \in B$.

By stability, we obtain the result.

Example

- Consider a burglar each of whose attempted burglaries is successful with probability p.
  - If successful, the amount of loot earned is j with probability p_j, $j = 0, \ldots, m$.
  - If unsuccessful, the burglar is caught and loses everything he has accumulated to that time, and the problem ends.
- The burglar's problem is to decide whether to attempt another burglary or to stop and enjoy his accumulated loot.
- We find the optimal policy.

Example (Optimality Equation)

- The state is the total loot so far collected.
    - If the current total loot is i and the burglar decides to stop, then he receives a reward i and the problem ends.
    - If he decides to continue, then, if successful, the new state will be $i+j$ with probability p_j.
- Let $V(i)$ is the burglar's maximal expected reward, given that the current state is i.
- The optimality equation is

$$V(i) = \max\left\{i, p\sum_j p_j V(i+j)\right\}.$$

Example (Cont'd)

- Define
$$B = \left\{ i : i \geq p \sum_j p_j (i+j) \right\}.$$
- The one-stage lookahead policy calls for stopping in state i if $i \in B$.
- Let $\mu = \sum_j j p_j$ be the expected return from a successful burglary.
- Then
$$B = \{i : i \geq p(i+\mu)\} = \left\{ i : i \geq \frac{p\mu}{1-p} \right\}.$$
- The state cannot decrease (unless the burglar is caught and then no additional decisions are needed).
- So B is closed.
- It follows that the one-stage lookahead policy that stops when the total loot is at least $\frac{p\mu}{1-p}$ is an optimal policy.

Example

- Recall the Optimal Asset Selling Problem.
- We receive an offer each day for an asset we desire to sell.
- When the offer is received, we must:
    - Pay a cost $c > 0$;
    - Decide whether to accept or reject the offer.
- Successive offers are independent with probability mass function

$$p_j = P(\text{offer is } j), \quad j \geq 0.$$

- The problem is to determine the policy that maximizes the expected net return.

Example (One-Stage Lookahead Policy)

- Let $E[X]$ be the expected value of a new offer.
- Define
$$B = \{j : j \geq -c + E[X]\}.$$
- The one-stage lookahead policy of a previous example calls for accepting an offer j if $j \in B$.
- B is not a closed set of states (because successive offers need not be increasing).
- So the one-stage lookahead policy would not necessarily be an optimal policy.

The Recall Problem

- Suppose we allow the seller to be able to recall any past offer.
- So a rejected offer is not lost, but may be accepted at any future time.
- In this case, the state after a new offer is observed would be the maximum offer ever received.
- Suppose j is the current state.
- Suppose X is the offer in the final stage.
- The selling price, if we go exactly one more stage, is $j + (X - j)^+$.
- Hence, the set of stopping states of the one-stage lookahead policy is

$$B = \{j : j \geq j + E[(X - j)^+] - c\} = \{j : E[(X - j)^+] \leq c\}.$$

- We have
  - $E[(X - j)^+]$ is a decreasing function of j;
  - The state, being the maximum offer so far received, cannot decrease.
- So B is a closed set of states.
- Hence, the one-stage lookahead policy is optimal in this problem.

The Recall Problem (Cont'd)

- Let v be such that
$$E[(X-v)^+] = c.$$
- Then the one-stage lookahead policy in the recall problem is to accept the first offer that is at least v.
- But this policy can be employed even when no recall is allowed.
- So it must also be an optimal policy in the no recall problem.

  Suppose it were not an optimal policy for the no-recall problem.

  Then the maximum expected net return in the no-recall problem would be strictly larger than in the recall problem.

  This is clearly not possible.

Example

- Consider a tournament involving k players, in which player i, $i = 1, \ldots, k$, starts with an initial fortune of $n_i > 0$.
- In each period, two of the players are chosen to play a game.
- The game is equally likely to be won by either player.
- The winner of the game receives 1 from the loser.
- A player whose fortune drops to 0 is eliminated.
- The tournament continues until one player has the entire fortune of

$$\sum_{i=1}^{k} n_i.$$

- For fixed i and j, let $N_{i,j}$ be the number of games in which i plays j.
- We are interested in $E[N_{i,j}]$.

Example (Cont'd)

- We set up a stopping rule problem.
- After two players have been chosen for a game, they may:
  - Stop and receive a final reward equal to the product of the current fortunes of players i and j;
  - Continue, receiving a reward of:
    - 1 in that period, if the two contestants are i and j;
    - 0, if the contestants are not i and j.
- Suppose the current fortunes of i and j are n and m.
  - Stopping at this time will yield a final reward of nm.
  - If we continue for one additional period and then stop, we receive:
    - A total reward of nm, if i and j are not the competitors in the current round (0 during that period and, then, nm when we stop the following period);
    - The expected amount $1 + \frac{1}{2}(n+1)(m-1) + \frac{1}{2}(n-1)(m+1) = nm$, if i and j are the competitors.

Example (Cont'd)

- Hence, in all cases the return from immediately stopping is exactly the same as the expected return from going exactly one more period and then stopping.
- Thus, the one-stage lookahead policy always calls for stopping.
- So its set of stopping states is closed.
- It follows that it is an optimal policy.
- But continuing on for an additional period and then stopping yields the same expected return as immediately stopping.
- So always continuing is also optimal.
- Now observe that:
  - The total return from the policy that always continues is the number $N_{i,j}$ of times that i and j play each other;
  - The return from immediately stopping is $n_i n_j$.
- We conclude that $E[N_{i,j}] = n_i n_j$.
- This holds no matter how the contestants in each round are chosen.