Discrete-Time Stochastic Dynamic Programming

I. Notation and basic assumptions

We consider a problem defined in terms of

- t : a time index, with integer values
- C : a kx1 vector, called the control vector
- S : an nx1 vector, called the state vector
- $\Gamma(\cdot)$: a mapping from state vector values to subsets of R^k, defining constraints on the choice of C
- $\mathcal{I}_t$: the information set at t, consisting of $\{C(s), S(s), \varepsilon(s), \text{all } s \le t\}$
- ε_t : a px1 random vector of disturbances at time t.

The objective is to maximize

$$E\left[\sum_{t=0}^{\infty} \beta^t U(C_t, S_t)\right] \tag{1}$$

by choice of $\{C_t, S_{t+1}, t = 0,...,\infty\}$. We assume that the infinite sum inside the brackets in (1) is well-defined for each choice of C's that satisfies the constraints below and that the expectation of the sum is well-defined for each such choice of C's. The choice of C's is constrained in four ways:

A) S_0 is given, not subject to choice;

B) for each t=1,...,∞, S_t is determined from past history and current ε_t according to

$$S_t = f(C_{t-1}, S_{t-1}, \varepsilon_t) \ . \tag{2}$$

C) for each t, C_t is constrained to lie in the set $\Gamma(S_t)$;

D) for each t, C_t is allowed to depend only on information in $\mathcal{I}_t$, and only in such a way that C_t and $U(C_t, S_t)$ are well-defined random variables.

Note that (2) means that, though we describe the problem as that of choosing both C and S to maximize the objective subject to (A)-(D), effectively we choose only C, since at each t, once C_t is chosen, S_{t+1} is determined by (2). Note also that (D) means that the formal mathematical problem here is not choosing a sequence of numbers $\{C_t\}$, but choosing a sequence of functions

$\left\{C_t^*(\cdot)\right\}$ such that at each date t, $C_t^*(\omega_t) = C_t^*\left(\left\{C_{s-1}, S_s, \varepsilon_s\right\}, \text{all } s \le t\right)$ maps our position ω_t in the information set $\mathcal{I}_t$ into our best choice for C_t.[1]

To complete the specification we need assumptions on the random disturbances ε. The standard dynamic programming framework requires that

E) for each t, ε_{t+1} is independent of C_t and of all the random variables in $\mathcal{I}_t$[2] and that

F) the random variables $\left\{\varepsilon_t, t = 1, ..., \infty\right\}$ are mutually independent and identically distributed (i.i.d.).

This means that no choice made before time t can influence the realization of the random variable ε_t. Note that this does not mean that ε_t and C_s are independent for $s \ge t$. Since ε's dated earlier than s are in $\mathcal{I}_s$, they may influence our choice of C_s; and this will create dependence in the joint probability distribution of C_s and ε_t for $s \ge t$.

Now observe that the range of probability distributions we can generate for values of C_t and S_{t+1} for $t \ge 0$ through our choice of C^* functions depends on $\mathcal{I}_0$ only via S_0. Since we are allowed to make the choice of C_t depend on anything in $\mathcal{I}_t$ that we like, we can create dependencies between the actual future values of S and C and, say, ε_{-3} if we like. But since this dependence can take any form we like, and since the data in $\mathcal{I}_0$ are all fixed and known to us at the time 0 when we choose C_0, the range of distributions for future C and S that we can achieve does not depend on $\mathcal{I}_0$, except through the fact that S_0 enters the version of (2) for $t = 1$. We denote by $\mathcal{S}$ the set of all possible values of S, and we mean by calling $\mathcal{S}$ "all possible values of S" that not only is every value of S_0 with which we might be confronted in $\mathcal{S}$, but also for every S_0 in $\mathcal{S}$ and every way of choosing C's that satisfies (A)-(D), S_t lies in $\mathcal{S}$ for all t with probability one. Thus for every S_0 in $\mathcal{S}$ there will be a unique, possibly infinite, least upper bound for the attainable values of the objective function. [Note that, though the future C's and S's are unknown and random at time 0, the objective function includes an expectation operator, so its value is a number, not a random variable.] We denote by $V(\cdot)$ the function mapping S's in $\mathcal{S}$ into the least upper bound of achievable values of the objective function. If the problem is well defined, $V(S)$ exists for each S in $\mathcal{S}$, though it is important in practice to check that the infinite sum in (1) indeed converges for all feasible choices of actions. V is called the **value function**.

---

[1] The technically sophisticated reader may note that (D) directly rules out the kind of non-measurability that concerns Stokey and Lucas in their chapter 9.1.

[2] Since C_t is required by (D) above to depend only on random variables in $\mathcal{I}_t$, (E) could omit the "of C_t and" phrase.

II. The principle of optimality: necessity and sufficiency

Theorem 1: **(The Principle of Optimality)** Suppose that V is the value function for the problem of maximizing (1) subject to (A)-(F). Then for each S in $\mathcal{S}$, $E\left[V\left(f(C,S,\varepsilon)\right)\right]$ exists for all C in $\Gamma(S)$ (with ±infinity allowable values), and

$$V(S) = \underset{C \text{ in } \Gamma(S)}{\text{l.u.b.}} \left\{U(C,S) + \beta E\left[V\left(f(C,S,\varepsilon)\right)\right]\right\} \tag{3}$$

Remark: In (3) we have omitted dates on C, S, and ε, but we mean here to take the expectation with respect to the distribution of ε, which is the same for all t, and to treat S as non-random. If (3) is true in this form for every S in $\mathcal{S}$, then of course it will also be true at every t with C, S and the E operator given t subscripts and ε given a t+1 subscript. For (3) to make sense, $E\left[V\left(C,S,\varepsilon\right)\right]$ must be defined, though possibly infinite, as the theorem asserts.

Proof: Note that the objective (1) can be written as

$$U(C_0,S_0) + \beta E_0\left[E_1 \sum_{t=0}^{\infty} \beta^s U(C_{t+1},S_{t+1})\right]. \tag{4}$$

The term in (4) in brackets, together with the preceding E operator, is exactly the same in form as (1), except with all the time subscripts advanced by 1. Since the constraints are all of the same form at all dates, the least upper bound of this term for a given value of S_1 is $V(S_1)$. In a well-defined problem, (4), being the value of the objective function, must itself be well-defined for every S_0 in $\mathcal{S}$ and every feasible choice of actions. But one particular feasible choice of actions is to choose C_0 in $\Gamma(S_0)$ arbitrarily, then to choose C_t for dates t=1 and later so that the second additive term in (4), for every possible value of $S_1 = f(C_0,S_0,\varepsilon_1)$, is at least $V(S_1) - \delta$ when $V(S_1)$ is finite and at least $1/\delta$ when $V(S_1)$ is infinite, where δ is an arbitrarily small positive number. With this particular way of choosing C's, we will have, therefore, when $V(S_1)$ is finite with probability one,

$$\left\{U(C_0,S_0) + \beta V(S_1)\right\} - \left\{U(C_0,S_0) + \beta E_1\left[\sum_{t=0}^{\infty} \beta^s U(C_{t+1},S_{t+1})\right]\right\} \in [0,\delta]. \tag{5}$$

Since the second term in brackets is a random variable whose expectation we know exists, as it is the value of the objective function for a feasible choice of actions, (5) bounds its first term in brackets above and below by random variables whose expectations exist and are arbitrarily close to each other. Thus the expectation of the first term exists as well. When $V(S_1)$ is infinite with non-zero probability, our choice of C's gives an arbitrarily large value of the objective function, implying that the first term, being bounded below by random variables with arbitrarily large expectation, itself has a well-defined infinite expectation.

Now suppose (3) were not true. Then either there would be, for some S in $\mathcal{S}$, a choice $\tilde{C}(S)$ of C making the right-hand-side of (3) exceed $V(S)$, or there would be some S in $\mathcal{S}$ such that the right-hand side of (3) is bounded away from $V(S)$ from below. Suppose that (3) fails through

the right-hand side being larger than $V(\tilde{S})$, where $\tilde{S}$ is the particular value of S at which (3) fails. Now consider this way of choosing C's when $S_0 = \tilde{S}$: at time 0, choose $C_0 = \tilde{C}(\tilde{S})$; at all later dates, choose C's according to a scheme that makes the term in brackets in (4) very close to $V(S_1)$. By doing so, we can make (4) as close as we like to the right-hand side of (3). But then we will have succeeded in choosing C's in such a way that the objective function value exceeds $V(S_0)$, a contradiction.

If (3) fails the other way, so that for some $\tilde{S}$ the right-hand side of (3) is bounded away from $V(\tilde{S})$ from below, a parallel argument, again using (4) shows that there is no way to choose C's to bring the objective function value arbitrarily close to $V(\tilde{S})$ when $S_0 = \tilde{S}$, again a contradiction with the definition of V, which completes the proof.

There is an additional necessary condition on the value function that characterizes its long run rate of growth.

Theorem 2: Suppose that V is the value function for the problem of maximizing (1) subject to (A)-(F). Then for every $\delta > 0$, it is possible to choose a policy function $C^*_\delta(\cdot)$ such that, for each S in $\mathcal{S}$ with $V(S)$ finite, the value of the objective function attained using $C^*_\delta(\cdot)$ is at least $V(S) - \delta$, and the sequence of $S_t, t = 1, \ldots, \infty$ generated by setting $S_0 = S$ and

$$S_t = f(C^*_\delta(S_{t-1}), S_{t-1}, \varepsilon_t),\ t = 1, \ldots, \infty \tag{6}$$

satisfies

$$\beta^t E_0 V(S_t) \xrightarrow[t\to\infty]{} 0\,. \tag{7}$$

Remark: In the (usual) special case where there is a policy function $C^*(\cdot)$ that actually generates an objective function value equal to $V(S_0)$ (rather than just arbitrarily close to it), (7) must hold for the S sequence generated by $C^*(\cdot)$ from every initial S for which $V(S)$ is finite.

Proof: First observe that we can certainly find $C^*_\delta(\cdot)$. Since V satisfies (3) by Theorem 1, we can for each S that delivers finite $V(S)$ choose $C^*_\delta(S)$ to satisfy

$$U(C^*_\delta(S), S) + \beta E\left[V\left(f(C^*_\delta(S), S, \varepsilon\right)\right] \geq V(S) - (1-\beta)\delta\,. \tag{8}$$

If (8) holds for each S with finite $V(S)$, then we can apply (8) to the term in brackets on its own left-hand side to obtain

$$U(C^*_\delta(S_0), S_0) + \beta E_0\left[U(C^*_\delta(S_1), S_1) + \beta V\left(S_2\right)\right] \geq V(S_0) - (1-\beta)\delta - (1-\beta)\delta\beta\,, \tag{9}$$

where we are assuming that the S_t sequence is being generated from (6). Repeatedly applying (8) this way will give us the desired conclusion, that the realized value of the objective function using $C^*_\delta(\cdot)$ is at least $V(S_0) - \delta$.

This result suggests a method for solving these problems: keep guessing forms for the V function until we find one that satisfies (3) for all S in $\mathcal{S}$. Since in checking whether (3) is satisfied for all S, we will ordinarily be finding, for every S, the $C^*(S)$ that maximizes the right-hand side, we will have the policy rule $C^*(\cdot)$ immediately at hand when we have found the right V.

The problems with this strategy are, first, that it is hopelessly inefficient until we find some systematic way to locate V's that might satisfy (3), and, second, that so far we know only that the V that represents the maximum attainable objective function value -- the value function of the problem -- satisfies (3). We have not yet shown that there cannot be other functions V that also satisfy (3). It turns out that generally there are other V's, besides the actual value function, that satisfy (3), but that we can pick out the true value function by applying some additional side conditions.

Theorem 3: Suppose there is a function V^* that satisfies (3) for every S in $\mathcal{S}$ and that in addition

i) for every S in $\mathcal{S}$ there is a value $C^*(S)$ for C that attains the maximum on the right-hand side of (3) with $V = V^*$;

ii) for every S_0 in $\mathcal{S}$, if S_t, $t = 1,\ldots,\infty$, is generated from

$$S_{t+1} = f(C^*(t), S_t, \varepsilon_{t+1}) \tag{10}$$

then

$$\beta^t E_0\left[V^*(S_t)\right] \xrightarrow[t\to\infty]{} 0 \text{ ; and} \tag{11}$$

iii) for any $V^{**} \neq V^*$ that solves (3) for every S in $\mathcal{S}$, there is some $\delta>0$ such that for any associated C^{**}_δ satisfying

$$U(C^{**}_\delta(S), S) + \beta E\left[V^{**}\left(f(C^{**}_\delta(S), S, \varepsilon)\right)\right] \geq V^{**}(S) - \delta \tag{12}$$

for every S_0 in $\mathcal{S}$, if S_t, $t = 1,\ldots,\infty$, are generated from

$$S_t = f(C^{**}_\delta(S_{t-1}), S_{t-1}, \varepsilon_t),\ t = 1,\ldots,\infty \text{ , then} \tag{13}$$

$$\lim_{t\to\infty} \beta^t E_0 V^*(S_t) \geq 0\,. \tag{14}$$

Then V^* is the value function for the problem.

Remark: The theorem asserts that if the necessary conditions of Theorems 1 and 2 are met for V^*, then V^* is the value function unless there is some other solution to (3) that violates (14). It is straightforward, but somewhat tedious, to extend this theorem to the case where no C^* is available and we instead have to settle for a sequence of C^*_δ's that approach the upper bound, as we used in Theorem 2.

Proof: Suppose there is some solution $V^{**} \neq V^*$ to (3). If for all S in $\mathcal{S}$, $V^{**}(S) \leq V^*(S)$, with strict inequality for some S, then V^{**} can't be the value function. This follows because, since it satisfies (ii), V^* does represent an attainable value of the objective function for each possible value of its argument, so a policy that delivers a lower V^{**} instead cannot be optimal. But then suppose that, for some $\bar{S}$, $V^{**}(\bar{S}) - V^*(\bar{S}) = \gamma > 0$. Consider the following policy: for $t = 0,...,T-1$, set $C_t = C_\delta^{**}(S_t)$ chosen as in (iii), then for $t \geq T$ set $C_t = C^*(S_t)$, . The value of the objective function under this policy starting from $S_0 = \bar{S}$ is

$$E_0\left[\sum_{t=0}^{T-1} U(C_\delta^{**}(S_t), S_t)\beta^t\right] + \beta^T E_0 V^*(S_T) \xrightarrow[T\to\infty]{} V_\delta^{**}(\bar{S}), \tag{15}$$

where by making δ small enough we can make $V_\delta^{**}(\bar{S})$ as close as we like to $V^{**}(\bar{S})$. The convergence in (15) follows from (iii) and the definition of C_δ^{**}. But notice that

$$V^*(\bar{S}) = U(C^*(\bar{S}), \bar{S}) + \beta E_0 V^*(S_1) \geq U(C_\delta^{**}(\bar{S}), \bar{S}) + \beta E_0 V^*(S_1) \tag{16}$$

(Note that in (16) we are implicitly treating the two S_1 values as generated by the C choice at time zero in each expression -- so the S_1's in the two expressions are not the same.) Applying the same argument again to $V^{**}(S_1)$ in (16), and so on recursively T times allows us to conclude

$$E_0\left[\sum_{t=0}^{T-1} U(C_\delta^{**}(S_t), S_t)\beta^t\right] + \beta^T E_0 V^*(S_T) < V^*(\bar{S}) \ . \tag{17}$$

But with δ arbitrarily small, (15) and (17) together imply $V^*(\bar{S}) > V^{**}(\bar{S})$, which contradicts our initial assumption. So $V^{**}(S) \leq V^*(S)$ for all S, which we have already noted means that V^{**} is not the value function. Since the argument applies to arbitrary $V^{**} \neq V^*$, V^* is the unique optimal solution and we have completed the proof.

Corollary: If U is bounded below, the necessary conditions of Theorems 1 and 2 are also sufficient.

Proof: Because the objective function is discounted, it is bounded below when U is bounded below. This makes (14) hold automatically.

III. Value Iteration

The arguments of the previous section point to a conceptually simple method for approximating the value function and hence C^*, the optimal policy rule. The method is called, for reasons that will be obvious, **value function iteration,** and proceeds as follows. Begin by guessing a form V_0 for the value function. Then for each iteration n, n=1,2,3,..., set, for each S in $\mathcal{S}$,

$$V_n(S) = \underset{C \text{ in } \Gamma}{\text{l.u.b.}}\left\{U(C,S) + \beta E\left[V_{n-1}(f(C,S,\varepsilon))\right]\right\} \ . \tag{18}$$

Continue until $V_n(S) = V_{n-1}(S)$, all S, to within some criterion for numerical accuracy. At that point a V satisfying the principal of optimality will have been arrived at, and the other necessary and sufficient conditions can be checked.

It is not necessarily true that value function iterations converge, however. When U is unbounded either above or below, it can easily happen that value function iteration convergence fails. Since many of the standard utility functions of macroeconomic models -- logarithmic $U(C) = \log(C)$ and CRR $U(C) = \frac{C^{1-\gamma}}{1-\gamma}$ for example -- fail to satisfy such a boundedness condition, we must generally be wary that value iteration might not converge. It is worth knowing, though, that when U is bounded (which we already know is a sufficient condition to guarantee that a solution to the optimality equation is the value function) value iteration necessarily converges. The argument goes as follows. First we note that

$$\left|V_n(S) - V_{n-1}(S)\right| \le \left| \underset{C \text{ in } \Gamma(S)}{\text{l.u.b.}} \left\{ U(C,S) - U(C,S) + \beta E\left[V_{n-1}\left(f(C,S,\varepsilon)\right) - V_{n-2}\left(f(C,S,\varepsilon)\right)\right]\right\}\right| . \quad (19)$$

This follows from (18) and the fact that

$$\left|\text{l.u.b.}\{a(\cdot)\} - \text{l.u.b.}\{b(\cdot)\}\right| \le \text{l.u.b.}\{|a(\cdot) - b(\cdot)|\} . \quad (20)$$

If we introduce as a norm on the space of V's

$$\|V\| = \sup_S |V(S)| , \quad (21)$$

(19) can be used to produce

$$\|V_n - V_{n-1}\| \le \beta \|V_{n-1} - V_{n-2}\| . \quad (22)$$

This implies that the sequence of value function iterates $\{V_n\}$ is a Cauchy sequence on the space of bounded functions, and hence that it converges to some bounded function. The reason the argument fails when U is unbounded is that then V is generally unbounded (not infinite -- it just gets arbitrarily large or small as we change its argument S) and therefore does not allow us to use (21). Even if we start with a bounded V_0, the unboundedness of U generally can make V_1 unbounded.