# 1 Stochastic Dynamic Programming

Formally, a stochastic dynamic program has the same components as a deterministic one; the only modification is to the state transition equation. When events in the future are uncertain, the state does not evolve deterministically; instead, states and actions today lead to a distribution over possible states in the future. We'll assume that $P(B, x, a)$ is the induced probability that tomorrow's state will lie in the set $B$ given that today's state is $x$ and today's action is $a$.

Let's take a look at the Bellman equation for a stochastic dynamic program:

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \int_X v(x') P(dx', x, a) \right\}.$$

We have to relate this expression to

$$V(x_0) = \max_{\{x_t, a_t\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t r(x_t, a_t),$$

the value function defined over the sequential problem. It turns out to be a bit more work than it was before because we need to be clear how to construct expectations. There are three important concepts buried in here that we need – conditional expectations, expectations over sequences of events that unfold over time, and the law of iterated expectations. To understand these concepts formally we need to take a detour into measure theory.

## 1.1 Measure Theory Detour

**Definition 1** *Let $S$ be a space. A family $\mathcal{F}$ of subsets of $S$ is called a **$\sigma$-algebra** if (i) $A \in \mathcal{F}$ implies $S \backslash A \in \mathcal{F}$; (ii) if $A_n \in \mathcal{F}$ $\forall n$ then $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$; and (iii) $S \in \mathcal{F}$. The pair $(S, \mathcal{F})$ is called a **measurable space**.*

Note that because the union of sets is equal to the intersection of the complements of those sets, $\sigma$-algebras are also closed under intersections. Also, $\emptyset \in \mathcal{F}$ because $\emptyset = S \backslash S$ and $S \in \mathcal{F}$. The largest $\sigma$-algebra on $S$ consists of all subsets of $S$ and the smallest consists of $S$ and the empty set. The $\sigma$-algebra generated by open sets (meaning that it is obtained by taking open sets and then adding whatever is needed to make the collection closed under unions and complements) is called the **Borel field** and is typically denoted $\mathcal{B}(S)$ (or $\mathcal{B}^n$ if $S = \mathcal{R}^n$); this field contains all open and closed sets as well as arbitrary unions of open sets and arbitrary intersections of closed sets (but not arbitrary unions of closed sets or arbitrary intersections of open sets, only countable ones). The restriction of $\mathcal{B}^n$ to a subset of $\mathcal{R}^n$ is also a $\sigma$-algebra. One vocabulary word you may come across: the probability space $(S, \mathcal{B}(S))$ where $S$ is a complete metric space is called a **Polish space**.

**Definition 2** *A sequence of $\sigma$-algebras $\{\mathcal{F}_t\}_t$ with $\mathcal{F}_t \subset \mathcal{F}_{t+1}$ is called a **filtration**.*

Filtrations are the natural method for expressing the flow of information. To see this, let's imagine a world that lasts two periods. Before period 1 nature selects a sequence of two outcomes, good or bad, in each period. Thus, the space of outcomes is

$$S = \{(G,G) \cup (G,B) \cup (B,G) \cup (B,B)\}.$$

at time 0 the agent knows only that nature chose; therefore, the appropriate $\sigma$-algebra is

$$\mathcal{F}_0 = \{\emptyset, S\}.$$

At time 1 the first part is revealed, meaning that the agent now has the $\sigma$-algebra

$$\mathcal{F}_1 = \{\{(G,G) \cup (G,B)\}, \{(B,G) \cup (B,B)\}, \emptyset, S\}.$$

Finally, at time 2 everything is revealed, yielding

$$\mathcal{F}_2 = \{(G,G), (G,B), (B,G), (B,B), \emptyset, S\}$$

unioned with all the appropriate unions and complements (which take up too much space to write out). Note that $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2$, so that this sequence is a filtration.

**Definition 3** *A **probability measure** on $(S, \mathcal{F})$ is a mapping $\mu : \mathcal{F} \rightarrow [0,1]$ such that (i) $\mu(\emptyset) = 0$; (ii) $\mu\left(\cup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n)$ if $A_n \cap A_m = \emptyset\ \forall n \neq m$ and $A_n \in \mathcal{F}$; and (iii) $\mu(S) = 1$.*

Thus, $\mu(A)$ is interpreted as the probability of event $A$. Measures are monotone: $\mu(A) \leq \mu(B)$ if $A \subseteq B$. Also, if $\{A_n\} \subseteq \mathcal{F}$ and $A_1 \subseteq A_2 \subseteq \cdots$ and $A = \cup_n A_n$, then $\mu(A_n) \nearrow \mu(A)$. Also, if $\{A_n\} \subseteq \mathcal{F}$ and $A_1 \supseteq A_2 \supseteq \cdots$ and $A = \cap_n A_n$, then $\mu(A_n) \searrow \mu(A)$. And finally, for any collection of sets $\{A_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ we have

$$\mu(\cup_n A_n) \leq \sum_n \mu(A_n)$$

where the $A_n$ may not be disjoint. You will prove these facts for homework.

**Definition 4** *A triple $(S, \mathcal{F}, \mu)$ is called a **probability space**.*

The key thing we need are concepts of integration and random variables. In particular, we need to know how to decide whether a function can be integrated with respect to a particular $\sigma$-algebra.

**Definition 5** *Let $(S, \mathcal{F})$ and $(T, \mathcal{G})$ be measurable spaces. A function $f : S \rightarrow T$ is **measurable** if and only if $\forall A \in \mathcal{G}$ we have $f^{-1}(A) \in \mathcal{F}$, where*

$$f^{-1}(A) = \{s \in S : f(s) \in A\}.$$

In particular, if we have $\mathcal{F} = \mathcal{B}^n$ then $f$ is measurable if and only if the sets

$$\{s \in S : f(s) \leq r\}$$

are contained in $\mathcal{F}$ for each $r$. Any continuous function is measurable, since those sets are all closed and thus must be in $\mathcal{B}^n$; measurability is "not quite" continuity however: functions which are measurable on a set can be shown to be within $\epsilon$ of a step function and within $\epsilon$ of a continuous function, but there is still a "gap." This notion leads directly to the idea of a random variable.

**Definition 6** *If $(S, \mathcal{F}, \mu)$ is a probability space and $f : S \to \mathcal{R}^n$ is measurable, then we call $f$ a **random variable**.*

It is straightforward to apply the definition of measurability to prove the following facts:

**Proposition 7** *Suppose $f_n : S \to \mathcal{R}$ is measurable $\forall n$ and that $f_n(x) \to f(x)$ $\forall x$. Then $f$ is measurable.*

**Proposition 8** *If $f : S \to \mathcal{R}$ is measurable and $g : \mathcal{R} \to \mathcal{R}$ then $g \circ f$ is measurable where $(g \circ f)(s) = g(f(s))$.*

**Proposition 9** *If $f : S \to \mathcal{R}$ and $g : S \to \mathcal{R}$ are measurable then $f + g$ and $f - g$ are measurable.*

**Proposition 10** *If $f : S \to \mathcal{R}$ is measurable and $\alpha$ is a constant, then $\alpha f$ is measurable.*

**Proposition 11** *If $f : S \to \mathcal{R}$ and $g : S \to \mathcal{R}$ are measurable, then $fg$ is measurable where $(fg)(s) = f(s)g(s)$. $f/g$ is measurable if $g(s) \neq 0$ $\forall s$, where $(f/g)(s) = f(s)/g(s)$.*

To define integration, we need to define some types of functions.

**Definition 12** *A function $\mathbf{1}_A : S \to \mathcal{R}$ is called the **indicator function for** $A$ (sometimes called the **characteristic function of** $A$) if it takes the form*

$$\mathbf{1}_A(s) = \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{if } s \notin A \end{cases}.$$

*A function $f : S \to \mathcal{R}$ is called a **simple function** if it takes the form*

$$f(s) = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}(s)$$

*for some real numbers $\{\alpha_i\}$ and disjoint sets $A_i$ such that $\cup_i A_i = S$.*

Indicator functions are nonzero only over the set $A$, and simple functions are step functions created through linear combinations of indicator functions over disjoint sets. It is straightforward to see that indicator functions are measurable if and only if $A \in \mathcal{F}$; that is, the $\sigma$-algebra must allow you to "identify" whether the set $A$ "happened" or not. To see this, consider $r \geq 1$; then

$$\{s \in S : \mathbf{1}_A(s) \leq r\} = S \in \mathcal{F}$$

because indicator functions always take values less than or equal to 1. For $r < 0$ we have

$$\{s \in S : \mathbf{1}_A(s) \leq r\} = \emptyset \in \mathcal{F}.$$

For any $r \in (0, 1)$ we have

$$\{s \in S : \mathbf{1}_A(s) \leq r\} = S \backslash A,$$

so that $\mathbf{1}_A$ is measurable if and only if $A \in \mathcal{F}$ (because it needs to have $S \backslash A \in \mathcal{F}$). Simple functions are measurable if and only if every $A_i \in \mathcal{F}$; that is, we need to be able to identify whether each $A_i$ happened.

We then define the **Lebesgue integral** of a simple function $f$ (which is nonnegative) as

$$\int f(s)\, \mu(ds) = \sum_{i=1}^{n} \alpha_i \mu(A_i);$$

the notation $\mu(ds)$ is meant to make clear that we are summing the height of the simple function times the size of the interval $ds$. If $f$ is both positive and negative we simply define the positive and negative parts separately and then put them together:

$$f^{+}(s) = \begin{cases} f(s) & \text{if } f(s) \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$f^{-}(s) = \begin{cases} -f(s) & \text{if } f(s) \leq 0 \\ 0 & \text{otherwise} \end{cases}$$

so that

$$\int f(s)\, \mu(ds) = \int f^{+}(s)\, \mu(ds) + \int f^{-}(s)\, \mu(ds).$$

This function is not measurable if both terms are infinite.

For 'nonsimple' functions we define their integral as the limit of the integrals of simple functions that 'fit under' $f$. That is,

$$\int f(s)\, \mu(ds) = \sup_{g} \left\{ \int g(s)\, \mu(ds) : g \text{ is a simple function and } 0 \leq g(s) \leq f(s) \;\; \forall s. \right\}$$

Again, to integrate a function that has both negative and positive parts we simply do each piece separately and "reassemble" them.

**Definition 13** *A function $f$ is **integrable** if $\int f(s)\, \mu(ds) < \infty$.*

It is straightforward to see that

$$\int_A f(s)\,\mu(ds) = \int f(s)\,\mathbf{1}_A(s)\,\mu(ds),$$

which is the notion of integrating over a limited domain; obviously, if $f$ is integrable over $S$ it is integrable over any subset of $S$. One helpful property that we will use later is the fact that

$$\left|\int f(s)\,\mu(ds)\right| \leq \int |f(s)|\,\mu(ds);$$

this is just an application of the triangle inequality. Other facts are equally easy to establish, once we note the property of "almost everywhere."

**Definition 14** *A property $P$ of $f$ is said to hold* ***almost everywhere (a.e)*** *if the set*

$$P^c = \{s \in S : f(s) \text{ does not have property } P\}$$

*has measure $\mu(P^c) = 0$.*

We then can easily establish the following.

**Proposition 15** *Let $f$,$g$ be integrable. Then $f+g$ is integrable, $\alpha f$ is integrable $\forall \alpha$, and*

$$\begin{aligned} \int (f+g)(s)\,\mu(ds) &= \int f(s)\,\mu(ds) + \int g(s)\,\mu(ds) \\ \int (\alpha f)(s)\,\mu(ds) &= \alpha \int f(s)\,\mu(ds). \end{aligned}$$

**Proposition 16** *If $f = 0$ almost everywhere, then $\int f(s)\,\mu(ds) = 0$.*

**Proposition 17** *If $f \leq g$ almost everywhere, then $\int f(s)\,\mu(ds) \leq \int g(s)\,\mu(ds)$.*

**Proposition 18** *If $f = g$ almost everywhere, then $\int f(s)\,\mu(ds) = \int g(s)\,\mu(ds)$.*

**Proposition 19** *If $f \geq 0$ and $\mu(\{s \in S : f(s) > 0\}) > 0$ then $\int f(s)\,\mu(ds) > 0$.*

Now we will define a space that will be relevant for our application to dynamic programming.

**Definition 20** *Let $(S, \mathcal{F}, \mu)$ be a probability space. The space $L^1(S, \mathcal{F}, \mu)$ is the space of all functions that are integrable with respect to $\mu$.*

We will now prove that $L^1$ is a Banach space. The proof requires two theorems.

**Theorem 21** ***(Monotone Convergence Theorem)*** *Suppose $E$ is a measurable set. Let $\{f_n\}$ be a sequence of measurable functions such that*

$$0 \leq f_1(x) \leq f_2(x) \leq \cdots$$

*for $x \in E$. Let $f$ be defined by $f_n(x) \to f(x)$ as $n \to \infty$. Then*

$$\int_E f_n(x)\,\mu(dx) \to \int_E f(x)\,\mu(dx).$$

***Proof.*** *As $n \to \infty$,*

$$\int_E f_n(x)\,\mu(dx) \to \alpha$$

*for some $\alpha$; since $\int f_n \leq \int f$, we have*

$$\alpha \leq \int_E f(x)\,\mu(dx).$$

*Choose $c$ such that $0 < c < 1$ and let $s$ be a simple measurable function such that $0 \leq s \leq f$. Put*

$$E_n = \{x : f_n(x) \geq cs(x)\}.$$

*By monotonicity, $E_1 \subset E_2 \subset \cdots$ and by the convergence of $f_n$ to $f$ we have*

$$E = \cup_{n=1}^{\infty} E_n.$$

*For every $n$,*

$$\int_E f_n(x)\,\mu(dx) \geq \int_{E_n} f_n(x)\,\mu(dx) \geq c \int_{E_n} s(x)\,\mu(dx).$$

*Let $n \to \infty$; since the integral is countably additive set function we can obtain*

$$\alpha \geq c \int_E s(x)\,\mu(dx).$$

*Letting $c \to 1$ we have*

$$\alpha \geq \int_E s(x)\,\mu(dx).$$

*Therefore,*

$$\alpha \geq \int_E f(x)\,\mu(dx).$$

■

**Theorem 22** ***(Fatou's Lemma)*** *Suppose $E$ is measurable. If $\{f_n\}$ is a sequence of nonnegative measurable functions and*

$$f(x) = \lim \inf_{n \to \infty} \{f_n(x)\}$$

*then*

$$\int_E f(x)\,\mu(dx) \leq \lim\inf_{n\to\infty} \int_E f_n(x)\,\mu(dx).$$

***Proof.*** *Set* $g_n(x) = \inf_{i\geq n}\{f_i(x)\}$. $g_n$ *is measurable on* $E$ *and*

$$\begin{aligned} 0 &\leq g_1(x) \leq g_2(x) \leq \cdots \\ g_n(x) &\leq f_n(x) \\ g_n(x) &\to f(x). \end{aligned}$$

*Therefore, by the Monotone Convergence Theorem*

$$\int_E g_n(x)\,\mu(dx) \to \int_E f(x)\,\mu(dx).$$

*The conclusion is now immediate.* ■

**Theorem 23** ***(Riesz-Fischer Theorem)*** $L^1$ *is a Banach space.*

***Proof.*** *We need to show that if* $\{f_n\}$ *is a Cauchy sequence in* $L^1$*, then there exists* $f \in L^1$ *such that* $\{f_n\} \to f$ *in* $L^1$. *Since* $f_n$ *is Cauchy, there exists a sequence* $\{n_k\}$, $k = 1, 2, \ldots$, *such that*

$$\left|\left|f_{n_k} - f_{n_{k+1}}\right|\right| < \frac{1}{2^k}.$$

*Choose* $g \in L^1$. *By the Schwarz inequality*

$$\int_X \left|g\left(f_{n_k} - f_{n_{k+1}}\right)\right| \mu(dx) \leq \frac{||g||}{2^k}.$$

*Hence,*

$$\sum_{k=1}^{\infty} \int_X \left|g\left(f_{n_k} - f_{n_{k+1}}\right)\right| \mu(dx) \leq ||g||.$$

*Interchanging the order of integration and summation, we have*

$$|g(x)| \sum_{k=1}^{\infty} \left|f_{n_k}(x) - f_{n_{k+1}}(x)\right| < \infty$$

*almost everywhere on* $X$. *Therefore,*

$$\sum_{k=1}^{\infty} \left|f_{n_k}(x) - f_{n_{k+1}}(x)\right| < \infty,$$

*for otherwise we could take* $g(x) \neq 0$ *on a subset of* $E$ *of positive measure and contradict the previous inequality. Since the* $k$*th partial sum of*

$$\sum_{k=1}^{\infty} \left(f_{n_{k+1}}(x) - f_{n_k}(x)\right)$$

*is*

$$f_{n_{k+1}}(x) - f_{n_k}(x)$$

*and the series converges almost everywhere on $X$, we have*

$$f(x) = \lim_{k\to\infty} f_{n_k}(x)$$

*defines $f(x)$ for almost all $x \in X$; it does not matter how we define $f(x)$ on those other points.*

*To show that $f$ has the desired properties, let $\epsilon > 0$ be given and choose $N$ such that $n \geq N$ and $m \geq N$ implies $||f_n - f_m|| \leq \epsilon$ (which is possible because $f_n$ is Cauchy). If $n_k > N$, Fatou's lemma shows that*

$$||f_n - f_{n_k}|| \leq \liminf_{i\to\infty} ||f_{n_i} - f_{n_k}|| \leq \epsilon.$$

*Thus, $f - f_{n_k} \in L^1$ and since $f = (f - f_m) + f_{n_k}$, we have $f \in L^1$. Since $\epsilon$ is arbitrary,*

$$\lim_{k\to\infty} ||f - f_{n_k}|| = 0.$$

*Finally, the inequality*

$$||f - f_n|| \leq ||f - f_{n_k}|| + ||f_{n_k} - f_n||$$

*shows that $f_n$ converges to $f$ in $L^1$, for if we take $n$ and $n_k$ large enough each of the terms on the RHS can be made small.* ■

This point will be useful, because it will enable us to apply the CMT to a stochastic dynamic programming problem. The final step is to define an expectation, which is of course just an integral.

**Definition 24** *Let $(S, \mathcal{F}, \mu)$ be a probability space. If $f : S \to \mathcal{R}$ is a random variable then the **expectation of** $f$ **with respect to** $\mu$ is defined to be*

$$E_\mu[f] = \int f(s)\,\mu(ds).$$

When the measure is not ambiguous, it will be dropped from the expression leaving just $E[f]$. What we are interested in is the conditional expectation of a random variable, which is a bit more challenging to define since we must worry about the possibility that our conditioning event has probability 0 (say, if it is continuous). What we will do is define a conditional expectation as a random variable over a $\sigma$-algebra that is smaller than $\mathcal{F}$ so that $\mathcal{G} \subset \mathcal{F}$; that is, any set that is measurable with respect to $\mathcal{G}$ is also measurable with respect to $\mathcal{F}$.

**Definition 25** *A **measurable partition of** $S$ is a collection $\mathcal{G}$ of subsets of $S$ such that (i) $G \cap G' = \emptyset\ \forall G \in \mathcal{G}$; (ii) $\cup_{G\in\mathcal{G}} G = S$; and (iii) $G \in \mathcal{F}$ if $G \in \mathcal{G}$.*

If $g$ is a measurable function with respect to $\mathcal{G}$, then $g$ must take on constant values over sets in $\mathcal{G}$. This idea makes sense as a theory of information – if $\mathcal{G}$ is to represent the "things you know" then any function which you can take expectations of must not provide you any information about stuff you don't; that is, it cannot take on different values inside sets $G$ because then you would know more than you do. We can now define a conditional expectation.

**Definition 26** *The* ***conditional expectation of*** $f$ ***relative to*** $\mathcal{G}$ *is a function* $E_\mu [f; \mathcal{G}] (s)$ *that is measurable with respect to* $\mathcal{G}$ *and that satisfies the requirement*

$$\int_G E_\mu [f; \mathcal{G}] (s) \, \mu (ds) = \int_G f (s) \, \mu (ds)$$

*for any* $G \in \mathcal{G}$.

It can be shown that $E_\mu [f; \mathcal{G}]$ defined in this way is unique up to a set of measure zero. Since $E_\mu [f; \mathcal{G}]$ must take on a constant value over $G$, we have

$$\begin{aligned} \int_G E_\mu [f; \mathcal{G}] (s) \, \mu (ds) &= E_\mu [f; \mathcal{G}] \int_G \mu (ds) \\ &= E_\mu [f; \mathcal{G}] \, \mu (G) \, . \end{aligned}$$

If $\mu (G) > 0$, we have the familiar expression

$$E_\mu [f; \mathcal{G}] = \frac{\int_G f (s) \, \mu (ds)}{\mu (G)};$$

if $\mu (G) = 0$, we still can employ this concept although we cannot solve for the expression analytically. Conditional probabilities are then obtained by letting $f$ be an indicator function:

$$E_\mu [\mathbf{1}_A; \mathcal{G}] = \frac{\int_G \mathbf{1}_A (s) \, \mu (ds)}{\mu (G)}$$

has the interpretation as the probability of event $A$ conditional on $G$, provided again that $\mu (G) > 0$.

The law of iterated expectations can now be defined. Suppose we have two $\sigma$-algebras contained in $\mathcal{F}$ such that $\mathcal{G}_1 \subset \mathcal{G}_2$. Define the random variables $E_\mu [f; \mathcal{G}_1] (s)$ and $E_\mu [f; \mathcal{G}_2] (s)$, which are the conditional expectations with respect to each $\sigma$-algebra. By definition we therefore have

$$\int_G E_\mu [f; \mathcal{G}_i] (s) \, \mu (ds) = \int_G f (s) \, \mu (ds)$$

over any set $G \in \mathcal{G}_i$. Let's now consider the conditional expectation of $E_\mu [f; \mathcal{G}_2] (s)$ with respect to $\mathcal{G}_1$, which must by definition be

$$\int_G E_\mu [E_\mu [f; \mathcal{G}_2] ; \mathcal{G}_1] (s) \, \mu (ds) = \int_G E [f; \mathcal{G}_2] (s) \, \mu (ds) \, .$$

Since $G \in \mathcal{G}_1$ implies $G \in \mathcal{G}_2$ the RHS is

$$\int_G E\left[f; \mathcal{G}_2\right](s)\, \mu(ds) = \int_G f(s)\, \mu(ds).$$

For any $G \in \mathcal{G}_1$ we must also have

$$\int_G E\left[f; \mathcal{G}_1\right](s)\, \mu(ds) = \int_G f(s)\, \mu(ds).$$

Therefore, for any $G \in \mathcal{G}_1$

$$\int_G E_\mu\left[E_\mu\left[f; \mathcal{G}_2\right]; \mathcal{G}_1\right](s)\, \mu(ds) = \int_G E\left[f; \mathcal{G}_1\right](s)\, \mu(ds).$$

In the sequential problem this law enables us to get rid of expectations of future events with respect to information we gain only in the future; that is, taking the expectation with respect to information today is not inconsistent with the fact that the agent knows that some things will be learned and used to form expectations in the future.

We can now be very precise about the term stochastic transition.

**Definition 27** *A mapping $P : \mathcal{F} \times X \times A \to [0,1]$ is a* ***stochastic transition function*** *on $(X, \mathcal{F})$ if, $\forall x \in X$ and $a \in A$, $P(\cdot, x, a)$ is a probability measure on $\mathcal{F}$ and for each $B \in \mathcal{F}$ $P(B, \cdot, \cdot)$ is a measurable function on $X \times A$.*

An important class of stochastic processes are called Markov processes with stationary transitions. 'Markov' means that the transition probabilities only depend on a finite history of past realizations and 'stationary' means that calender time is irrelevant. These notions are important for dynamic programming, as they imply that $t$ is again irrelevant and that the state vector is finite-dimensional. Some familiar stochastic processes are Markovian with stationary transitions. For example, let $\Phi(x)$ denote the cdf of a standard normal. Then the transition function

$$P(\{x' : x' \leq r\}, x) = 1 - \Phi(r - \rho x)$$

describes an AR(1) process

$$x_{t+1} = \rho x_t + \epsilon_{t+1}$$

with $\epsilon_{t+1} \sim \text{iid}\, N(0,1)$. Another familiar and useful stochastic process is a Markov chain, which arises when $X$ is a finite set. The transition function is usually written as $P = [p_{ij}]_{i,j=1,\ldots,N}$, with the interpretation that

$$p_{ij} = \Pr\{x_{t+1} = x_i : x_t = x_j\};$$

to be consistent with rules of probability requires $p_{ij} \geq 0$ and $\sum_i p_{ij} = 1$.

Stochastic transition functions can be used to define probability measures over sequences, which need to do to define the sequential problem. That is, we can define

$$\mu_T\left(B;x_0\right)=\int_{A_1}\int_{A_2}\cdots\int_{A_T}P\left(dx_T,x_{T-1}\right)P\left(dx_{T-1},x_{T-2}\right)\cdots P\left(dx_1,x_0\right)$$

as the probability of $B$ given the initial state $x_0$, so

$$\mu_T\left(\cdot;x_0\right)=\Pr\left\{x_1\in A_1,x_2\in A_2,\cdots,x_T\in A_T|x_0\right\}.$$

It is not hard to show that $\mu_T$ is in fact a probability measure on $\mathcal{F}^T=\mathcal{F}\otimes\mathcal{F}\otimes\cdots\otimes\mathcal{F}$. As $T\to\infty$ we can define $\mu_\infty$ which is technically not a measure on $\mathcal{F}^\infty$ but something a bit weaker; we use the $\sigma$-algebra $\mathcal{H}$ which is the smallest $\sigma$-algebra that contains all finite unions of sets of the form

$$B=A_1\times A_2\times A_3\times\cdots\times A_T\times X\times X\cdots$$

where $A_t\in\mathcal{F}$ for each $t$ and all but finitely many of the sets equal $X$ so that

$$\mu_\infty\left(B;x_0\right)=\int_{A_1}\int_{A_2}\cdots\int_{A_T}P\left(dx_T,x_{T-1}\right)P\left(dx_{T-1},x_{T-2}\right)\cdots P\left(dx_1,x_0\right)$$

where $T$ is the largest date beyond which $A_t=X\ \forall t>T$. We can uniquely extend this measure to all events (that is, those that do not require $A_t=X$ $\forall t\geq T$) in $\mathcal{H}$ using one of the Hahn-Caratheodory extension theorems.

For a Markov chain $\mu_T$ is straightforward to see – it is just the transition matrix raised to the $T$ power, so that

$$\Pr\left\{x_T=x_j|x_0=x_i\right\}=p_{ij}^T,$$

the $(i,j)$th element of $P^T$ (not the $(i,j)$th element of $P$ raised to the $T$ power). Under certain circumstances $P^\infty$ exists and is unique and therefore can be used to formulate expectations over infinite sequences; this matrix defines the invariant distribution of the Markov chain (the unconditional distribution of states).

For $P^\infty$ to exist we need to rule out the existence of cyclically-moving subsets of the Markov chain – that is, subsets through which the chain moves deterministically. As an example, consider

$$P=\begin{bmatrix}0&1\\1&0\end{bmatrix}.$$

This Markov chain does not possess a stationary distribution because $P^n=P$ for odd $n$ and

$$P^n=\begin{bmatrix}1&0\\0&1\end{bmatrix}$$

for even $n$.

For $P^{\infty}$ to be unique we need to rule out the presence of absorbing states (a state $i$ such that $p_{ii} = 1$). Consider the matrix

$$P = \begin{bmatrix} 0.5 & 0.25 & 0.25 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Once this process leaves state 1 it moves to a state which it never leaves; as a result, the unconditional distribution is not unique, since it depends on which state gets chosen. The required condition for uniqueness is often stated as the requirement that there exists $T$ such that $p_{ij}^{T} > 0$ for every $(i, j)$, so that it is possible to move from state $i$ today to state $j$ today in a finite $T$ number of moves. This requirement is actually a bit stronger since it also rules out transient states (a state $i$ such that $p_{ij} = 0 \; \forall j \neq i$), which are states that are never revisited once left. Each absorbing state is associated with an invariant distribution that assigns it probability 1, so uniqueness is lost above.

For the AR(1) $\mu_{\infty}$ exists if and only if $|\rho| < 1$, so that the unconditional distribution has well-defined moments. If that condition is not satisfied, in some cases it may be possible to alter the state space to generate a modified problem that does not have this problem, for example by defining the state as a ratio of two other states.

Consider the functions

$$F_T(x_1, x_2, ..., x_T) = \sum_{t=1}^{T} \beta^{t-1} f(x_t)$$

and

$$F(x_1, x_2, ...) = \sum_{t=1}^{\infty} \beta^{t-1} f(x_t).$$

Each $F_T$ is measurable with respect to $\mathcal{H}$ and $\{F_T\}$ converges pointwise to $F$, so $F$ is measurable with respect to $\mathcal{H}$. We can therefore write

$$\begin{aligned} E[F; x_0] = & \int_X f(x_1) P(dx_1, x_0) + \\ & \beta \int_X \left[ \int_X f(x_2) P(dx_2, x_1) \right] P(dx_1, x_0) + \\ & \beta^2 \int_X \left[ \int_X \left[ \int_X f(x_3) P(dx_3, x_2) \right] P(dx_2, x_1) \right] P(dx_1, x_0) + \\ & \vdots \end{aligned}$$

so that

$$E[F; x_0] = \int_X [f(x_1) + \beta E[F; x_1]] P(dx_1, x_0).$$

That is exactly what we want to have – the lifetime expected value is the sum of the first-period return and the lifetime expected value from tomorrow onward, discounted back to today.

## 1.2 Back to Dynamic Programming Again

Consider the functional equation

$$v(x) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \beta \int_X v(x') P(dx', x, a) \right\}$$

where $x'$ is a set of events tomorrow. We want to relate this functional equation to the sequential problem

$$V(x_0) = \max_{\{x_t, a_t\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t r(x_t, a_t)$$

subject to the restrictions embedded in the state transition and feasible control sets. The sequences which solve the sequential problem should be interpreted as conditional on outcomes – that is, they unfold over event trees. This expression is often written as

$$V(x_0) = \max_{\{x_t, a_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \pi\left(s^t | s_0\right) r(x_t, a_t)$$

where $s^t$ is the history of events up through time $t$ ($s^t = \{s_0, s_1, ..., s_t\}$) and $\pi(s^t | s_0)$ is the probability of that history conditional on $s_0$.

We want to prove that $v(x)$ exists and is unique. Define

$$(Tw)(x) = \max_{a \in \Gamma(x)} \left\{ r(x, a) + \beta \int_X w(x') P(dx', x, a) \right\}.$$

We make the same continuity assumptions as before. To apply the Blackwell-Boyd conditions we need a $\phi(x) > 0$ function; this function must additionally now satisfy an integrability condition.

**Theorem 28** *Assume **(1)** the one-period return function $r$ is continuous on $X \times A$ and the feasible actions correspondence $\Gamma$ is continuous and compact-valued; **(2)** there is a function $\phi \in \Xi$ with $\phi(x) > 0$ such that (i) there exists an $M < \infty$ with*

$$\max_{a \in \Gamma(x)} \{|r(x, a)|\} \leq M\phi(x) \tag{1}$$

*$\forall x \in X$, (ii)*

$$\beta \sup_{x \in X} \left\{ \max_{a \in \Gamma(x)} \left\{ \int_X \frac{\phi(x')}{\phi(x)} P(dx', x, a) \right\} \right\} < 1, \tag{2}$$

*and (iii)*

$$\int_X \phi(x') P(dx', x, a) < \infty$$

*$\forall x \in X$ and $a \in \Gamma(x)$; and **(3)** for any continuous $w \in \Xi_\phi$*

$$\int_X w(x') P(dx', x, a)$$

*is continuous on $X \times A$. Under these conditions the Bellman operator is a strict contraction map.*

Condition (2)(iii) guarantees the integrability of any member of $\Xi_\phi$, since

$$||f||_\phi = \sup_{x \in X} \left\{ \frac{|f(x)|}{\phi(x)} \right\}$$

and therefore $f$ is bounded in absolute value by an integrable function $\phi$.

We now show that $T : \left(\Xi_\phi, ||\cdot||_\phi\right) \to \left(\Xi_\phi, ||\cdot||_\phi\right)$. If $w \in \Xi_\phi$ then $Tw$ is continuous by the maximum theorem, so $T : \left(\Xi_\phi, ||\cdot||_\phi\right) \to \left(\Xi, ||\cdot||_\phi\right)$. To see that $Tw$ has finite $\phi$-norm, note that

$$\begin{aligned}
|(Tw)(x)| &= \left| \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \int_X w(x') P(dx', x, a) \right\} \right| \\
&\leq \max_{a \in \Gamma(x)} \{|r(x,a)|\} + \beta \max_{a \in \Gamma(x)} \left\{ \int_X |w(x')| P(dx', x, a) \right\} \\
&= \max_{a \in \Gamma(x)} \{|r(x,a)|\} + \beta \max_{a \in \Gamma(x)} \left\{ \int_X \frac{|w(x')|}{\phi(x')} \phi(x') P(dx', x, a) \right\} \\
&\leq \max_{a \in \Gamma(x)} \{|r(x,a)|\} + \beta ||w||_\phi \max_{a \in \Gamma(x)} \left\{ \int_X \phi(x') P(dx', x, a) \right\}
\end{aligned}$$

which is finite in the $\phi$-norm by conditions (i) and (iii).

Now we check monotonicity. Let $v \geq w$. Then

$$\begin{aligned}
(Tv)(x) &= \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \int_X v(x') P(dx', x, a) \right\} \\
&\geq \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \int_X w(x') P(dx', x, a) \right\} \\
&= (Tw)(x).
\end{aligned}$$

And then discounting constant functions:

$$\begin{aligned}
(T(w + A\phi))(x) &= \max_{a \in \Gamma(x)} \left\{ r(x,a) + \beta \int_X [w(x') + A\phi(x')] P(dx', x, a) \right\} \\
&\leq (Tw)(x) + \beta A \max_{a \in \Gamma(x)} \left\{ \int_X \phi(x') P(dx', x, a) \right\} \\
&\leq (Tw)(x) + A\phi(x) \beta \max_{a \in \Gamma(x)} \left\{ \int_X \frac{\phi(x')}{\phi(x)} P(dx', x, a) \right\} \\
&\leq (Tw)(x) + \theta A \phi(x)
\end{aligned}$$

with $\theta < 1$ by condition (ii). Therefore, we know that $v$ is the unique function that satisfies the Bellman equation.

It is not as straightforward to prove that $v = V$; it turns out that it could be the case that $V$ cannot be integrated, meaning that it could not satisfy the Bellman equation (the example is due to Blackwell and can be found in SLP).

We will now show that our assumptions rule out this pathological case. To begin, let $\Pi$ denote the collection of all policies $\pi : X \to A$ such that

$$\pi(x) \in \Gamma(x),$$

the function

$$Q(B,x) = P(B,x,\pi(x))$$

is a stochastic transition on $(X,\mathcal{F},\mu)$, and $r(x,\pi(x))$ is a measurable function with respect to $\mathcal{F}$. This set is nonempty under the conditions that $r$ is continuous, $P$ is continuous (condition 3 guarantees this), and that $\Gamma$ admits a continuous selection (continuity of $\Gamma$ guarantees this).

A given $\pi \in \Pi$ defines a probability measure over realizations $\{x_t\}_{t=1}^{T}$ for any finite $T$ defined by

$$\begin{aligned} Q(B,x) &= P(B,x,\pi(x)) \\ &= \Pr\{x_{t+1} \in B | x_t = x\}; \end{aligned}$$

call this measure $\mu_T(\cdot; x_0, \pi)$ for each $T$. That is,

$$\mu_T\left(\times_{t=1}^{T} A_t; x_0, \pi\right) = \int_{A_1}\int_{A_2}\cdots\int_{A_T} Q(dx_T, x_{T-1})\cdots Q(dx_1, x_0).$$

For finite $T$ we therefore have

$$\begin{aligned} E_0 \sum_{t=0}^{T} \beta^t r(x_t, a_t) &= r(x_0, \pi(x_0)) + \\ & \beta \int_X r(x_1, \pi(x_1)) Q(dx_1, x_0) + \\ & \beta^2 \int_X \int_X r(x_2, \pi(x_2)) Q(dx_2, x_1) Q(dx_1, x_0) + \\ & \vdots \\ & \beta^T \int_X \int_X \cdots \int_X r(x_T, \pi(x_T)) Q(dx_T, x_{T-1}) \cdots Q(dx_1, x_0). \end{aligned}$$

We need to show that this sum converges to a finite number as $T \to \infty$.

Since we have assumed that $r$ is bounded by $\phi$, we have

$$\frac{|r(\widetilde{x}, \pi(\widetilde{x}))|}{\phi(\widetilde{x})} \leq M$$

for every $\widetilde{x} \in X$ and $\pi \in \Pi$. Take last term in the summation and consider the innermost integral:

$$\begin{aligned}\left|\int_X r\left(x_T,\pi\left(x_T\right)\right)Q\left(dx_T,x_{T-1}\right)\right| &\leq \int_X \left|r\left(x_T,\pi\left(x_T\right)\right)\right|Q\left(dx_T,x_{T-1}\right)\\ &\leq \int_X \frac{\left|r\left(x_T,\pi\left(x_T\right)\right)\right|}{\phi\left(x_T\right)}\phi\left(x_T\right)Q\left(dx_T,x_{T-1}\right)\\ &\leq M\int_X \phi\left(x_T\right)Q\left(dx_T,x_{T-1}\right)\\ &\leq M\phi\left(x_{T-1}\right)\sup_{x\in X}\left\{\int_X \frac{\phi\left(x'\right)}{\phi\left(x\right)}Q\left(dx',x\right)\right\}.\end{aligned}$$

By assumption $\beta\sup_{x\in X}\left\{\int_X \frac{\phi(x')}{\phi(x)}Q\left(dx',x\right)\right\}=\beta A<1$. Thus, the innermost integral is bounded by $MA\phi\left(x_{T-1}\right)$ in absolute value. If we repeat this procedure for every integral we get that the entire last term of the summation is bounded by $M\left(\beta A\right)^T\phi\left(x\right)$ in absolute value. If we then move to the next-to-last term we get it is bounded by $M\left(\beta A\right)^{T-1}\phi\left(x\right)$, and so on. Thus, the entire summation is bounded by

$$M\phi\left(x\right)\left(1+\beta A+\beta^2A^2+\cdots\beta^TA^T\right)=M\phi\left(x\right)\frac{1-\left(\beta A\right)^{T+1}}{1-\beta A}.$$

As $T\to\infty$ we therefore have that the summation is bounded by $M\phi\left(x\right)\frac{1}{1-\beta A}$ and the infinite expected value is well-defined, since $\beta A<1$.

Define

$$w_\pi\left(x\right)=E_\pi\left[\sum_{t=0}^{\infty}\beta^t r\left(x_t,\pi\left(x_t\right)\right)\middle| x_0=x\right],$$

the lifetime expected utility at $x$ obtained by using $\pi$. The value function is $V\left(x\right)=\sup_{\pi\in\Pi}\left\{w_\pi\left(x\right)\right\}$; since we know that $w_\pi\in\Xi_\phi$ then so is $V$. We want to know if $v=V$ and if $V=w_{\pi^*}$ for some optimal policy $\pi^*$.

For any $\pi\in\Pi$ it is true that there exists a unique $V_\pi\in\Xi_\phi$ that satisfies

$$V_\pi\left(x\right)=r\left(x,\pi\left(x\right)\right)+\beta\int_X V_\pi\left(x'\right)P\left(dx',x,\pi\left(x\right)\right)$$

by immediate application of the CMT; that is, $V_\pi$ is the unique fixed point of the strict contraction mapping $T_\pi$ defined as

$$\left(T_\pi u\right)\left(x\right)=r\left(x,\pi\left(x\right)\right)+\beta\int_X u\left(x'\right)P\left(dx',x,\pi\left(x\right)\right).$$

Now we need to show that $V_\pi=w_\pi$; that is, $w_\pi$ is the unique function in $\Xi_\phi$ that satisfies

$$w_\pi\left(x\right)=\left(T_\pi w_\pi\right)\left(x\right)=r\left(x,\pi\left(x\right)\right)+\beta\int_X w_\pi\left(x'\right)P\left(dx',x,\pi\left(x\right)\right).$$

That this is true is again an immediate implication of the CMT.

Take any function $u \in \Xi_\phi$ and any policy $\pi \in \Pi$. It is true that

$$||w_\pi - u||_\phi \leq \frac{||T_\pi u - u||_\phi}{1-\theta}$$

where $\theta < 1$; this property is an immediate corollary of the CMT. Also $\forall \epsilon > 0$ there exists a policy $\widehat{\pi} \in \Pi$ such that

$$||w_{\widehat{\pi}} - v||_\phi \leq \epsilon;$$

that is, there exists a policy that "nearly" achieves the function $v$, which should be easy to accept given the form of the fixed point operators that generate $v$ and $w_\pi$.

Now suppose $v > V$. Since there exists a policy $\widehat{\pi}$ such that $w_{\widehat{\pi}}$ is arbitrarily close to $v$ we can find one such that $w_{\widehat{\pi}} > V$. This contradicts the fact that $V$ is the value function and therefore the supremum over all such policies. So we know that $v \leq V$.

By the definitions of $T$ and $T_\pi$ we must have $T_\pi v \leq Tv = v$ (because $T$ performs a maximization and $T_\pi$ does not). $T_\pi$ is monotone so that $T_\pi^n v \leq v$ as well. Taking the limit yields $w_\pi \leq v$, which holds for arbitrary $\pi$. If $v(x) \geq w_\pi(x)$ for every $x$ and $\pi$, we must have

$$v(x) \geq V(x) = \sup_{\pi \in \Pi} \{w_\pi(x)\}.$$

Therefore, $v \geq V$. Combined with the previous result we get $v = V$.

To check the optimal policy condition $v = w_{\pi^*}$, note that

$$v(x) = r(x, \pi^*(x)) + \beta \int_X v(x') P(dx', x, \pi^*(x));$$

that is, $\pi^*$ attains the maximum on the RHS of the Bellman equation. Since $v = V$, we have

$$T_{\pi^*} V = T_{\pi^*} v = Tv = v = V,$$

so that

$$T_{\pi^*} V = V.$$

Thus, the value function is the fixed point of $T_{\pi^*}$, which is equal to $w_{\pi^*}$. Therefore, $V = w_{\pi^*}$.