MATHEMATICAL ISSUES IN DYNAMIC PROGRAMMING

by

Dimitri P. Bertsekas and Steven E. Shreve †

1. INTRODUCTION

A general theory of dynamic programming must deal with the formidable mathematical questions that arise from the presence of uncountable probability spaces. These questions are explored at some length by means of a simple example in Section 2. With this example as motivation, the mathematical preliminaries necessary for the construction of a general finite horizon model are developed in Section 3. In Section 4, the results of Section 3 are applied to set up the model and to indicate how a valid dynamic programming algorithm can be defined.

The purpose of the paper is to provide some orientation for the development of a comprehensive and mathematically rigorous theory of dynamic programming, as given in the authors' book "Stochastic Optimal Control: The Discrete-Time Case," Academic Press, 1978 (republished by Athena Scientific, 1996). This book contains a detailed analysis of finite and infinite horizon problems, and provides references to earlier research.

2. A TWO-STAGE EXAMPLE

Suppose that we are given a point (state) x_0 on the real line $\Re$ and a system function $f(x_0, u_0, w_0)$, where u_0 and w_0 are also real numbers. Knowing x_0, we must choose a control $u_0 \in \Re$, a random

† Adapted from the expository paper "Dynamic Programming in Borel Spaces" by D. P. Bertsekas and S. E. Shreve, which appeared in the edited volume "Dynamic Programming and its Applications," by M. Puterman (Ed.), Academic Press, 1978.

disturbance w_0 is generated according to the probability measure p on the Borel sets of $\Re$, and the new state of the system $x_1 \in \Re$ is given by $x_1 = f(x_0, u_0, w_0)$. Knowing x_1, we must choose a control $u_1 \in \Re$ and then incur cost $g(x_1, u_1)$, where g is a real-valued function which is bounded below. Thus a cost is incurred only at termination. A policy $\pi = (\mu_0, \mu_1)$ is a pair of functions from state to control, i.e., if the policy π is employed and x_0 is the initial state, then we choose u_0 to be $\mu_0(x_0)$. If x_1 is the subsequent state, we choose u_1 to be $\mu_1(x_1)$. The *expected cost corresponding to a policy* $\pi = (\mu_0, \mu_1)$ when x_0 is the initial state is

$$J_\pi(x_0) = \int g\left(f[x_0, \mu_0(x_0), w_0], \mu_1(f[x_0, \mu_0(x_0), w_0])\right) p(dw_0). \tag{1}$$

We must insure that the integral in (1) is defined. A sufficient condition for this is that f, g and μ_1 are Borel measurable. However, our aim in this example is to point out the type of measure theoretic framework that must be adopted in order for specific results to hold. We thus leave unspecified at present the measurability restrictions on f, g, and μ_1 but always assume that μ_1 will be chosen from an appropriately measurable class of policies for which the cost in (1) is well defined. The *optimal cost* is

$$J^*(x_0) = \inf_\pi J_\pi(x_0), \tag{2}$$

where the infimum is over all policies $\pi = (\mu_0, \mu_1)$ such that μ_1 is measurable from $\Re$ to $\Re$ with respect to σ-algebras to be specified later. Given $\epsilon > 0$, a policy π is *ϵ-optimal* if

$$J_\pi(x_0) \leq J^*(x_0) + \epsilon \qquad x_0 \in \Re. \tag{3}$$

A policy π is *optimal* if (3) holds with $\epsilon = 0$.

The backward recursion of dynamic programming takes the following form:

$$J_1(x_1) = \inf_{u_1 \in \Re} g(x_1, u_1) \qquad x_1 \in \Re, \tag{4}$$

$$J_2(x_0) = \inf_{u_0 \in \Re} \int J_1[f(x_0, u_0, w_0)] p(dw_0) \qquad x_0 \in \Re. \tag{5}$$

The motivation for this algorithm is that under certain conditions $J_2(x_0) = J^*(x_0)$ for every $x_0 \in \Re$. An informal justification goes this way:

$$\begin{aligned}
J^*(x_0) &= \inf_\pi J_\pi(x_0) \\
&= \inf_{\mu_0} \inf_{\mu_1} \int g(f[x_0, \mu_0(x_0), w_0], \mu_1(f[x_0, \mu_0(x_0), w_0])) p(dw_0) && (6) \\
&= \inf_{\mu_0} \int \inf_{\mu_1} g(f[x_0, \mu_0(x_0), w_0], \mu_1(f[x_0, \mu_0(x_0), w_0])) p(dw_0) && (6a) \\
&= \inf_{\mu_0} \int J_1(f[x_0, \mu_0(x_0), w_0]) p(dw_0) && (6b) \\
&= J_2(x_0).
\end{aligned}$$

In order to make this rigorous, the interchange of infimization and integration in (6a) must be justified. As a part of this, it must be shown that the integral in (6a) is defined, that is, J_1 is sufficiently regular to allow the integral in (6b) to be defined.

Given $\epsilon > 0$, suppose we can find a function $\bar{\mu}_1 : \Re \to \Re$, which is measurable with respect to appropriate σ-algebras and is such that

$$g[x_1, \bar{\mu}_1(x_1)] \leq J_1(x_1) + \frac{\epsilon}{2} \qquad x_1 \in \Re. \tag{7}$$

Let $\bar{\mu}_0 : \Re \to \Re$ be such that

$$\int J_1(f[x_0, \bar{\mu}_0(x_0), w_0])p(dw_0) \leq J_2(x_0) + \frac{\epsilon}{2} \qquad x_0 \in \Re. \tag{8}$$

Formally, we have for $\bar{\pi} = (\bar{\mu}_0, \bar{\mu}_1)$ and $x_0 \in \Re$,

$$\begin{aligned} J_{\bar{\pi}}(x_0) &= \int g\left(f[x_0, \bar{\mu}_0(x_0), w_0], \bar{\mu}_1(f[x_0, \bar{\mu}_0(x_0), w_0])\right) p(dw_0) \\ &\leq \int J_1(f[x_0, \bar{\mu}_0(x_0), w_0])p(dw_0) + \frac{\epsilon}{2} \\ &\leq J_2(x_0) + \epsilon \\ &\leq J^*(x_0) + \epsilon, \end{aligned} \tag{9}$$

so $\bar{\pi}$ is ϵ-optimal. Furthermore, if an appropriately measurable $\bar{\mu}_1$ satisfying (7) exists, then

$$\begin{aligned} \int J_1\left(f[x_0, \mu_0(x_0), w_0]\right) p(dw_0) &= \int \inf_{\mu_1} g\left(f[x_0, \mu_0(x_0), w_0], \mu_1(f[x_0, \mu_0(x_0), w_0])\right) p(dw_0) \\ &\leq \inf_{\mu_1} \int g\left(f[x_0, \mu_0(x_0), w_0], \mu_1(f[x_0, \mu_0(x_0, w_0])\right) p(dw_0) \\ &\leq \int g\left(f[x_0, \mu_0(x_0), w_0]\bar{\mu}_1(f[x_0, \mu_0(x_0), w_0])\right) p(dw_0) \\ &\leq \int J_1\left(f[x_0, \mu_0(x_0), w_0]\right) p(dw_0) + \frac{\epsilon}{2}, \end{aligned} \tag{10}$$

so the interchange of integration and infimization in (6a) is valid provided that the integral in (6b), or equivalently, the integral in (5) is defined.

We observe that if the probability measure $p(dw_0)$ has *countable support*, i.e., is concentrated on a countable number of points, then the integrals in (1), (5), (6), and (8)-(10) reduce to (possibly infinite) summations. Thus, all the integrals are defined without the imposition of measurability restrictions on μ_1, f and g, and $\bar{\mu}_1$ and $\bar{\mu}_0$ satisfying (7), (8) exist since g is bounded below.

If $p(dw_0)$ does not have countable support, two approaches have been used. The first is to *expand the notion of integration*, and the second is to place *appropriate measurability restrictions on f, g and μ_1*. Expanding the notion of integration can be done by interpreting the integrals in (1), (5), (6) and (8)-(10) as outer integrals. Since the outer integral can be defined for any

function, measurable or not, there is no need to require that f, g, μ_0 and μ_1 be measurable in any sense, and the arguments advanced above go through just as in the countable disturbance case. We do not discuss this approach further except to mention that the Bertsekas and Shreve book shows that the basic results for finite and infinite horizon problems of perfect state information carry through within an outer integration framework. However, there are inherent limitations in this approach centering around the pathologies of outer integration. For example, the value of the cost function corresponding to a policy may depend on the definition of outer integral, i.e., two different (but natural) definitions of outer integration may result in different cost functions. Difficulties also occur in the treatment of imperfect information problems using sufficient statistics. The other approach was initiated in more general form by Blackwell in 1965. We discuss it at length in the subsequent sections.

In conclusion, we point out that if the infima in (4) and (5) are attained for every x_1 and x_0, respectively, and appropriately measurable $\bar{\mu}_1 : \Re \to \Re$ and $\bar{\mu}_0 : \Re \to \Re$ can be found such that (7) and (8) are satisfied with $\epsilon = 0$, then $J_2(x_0) = J^*(x_0)$ for every $x_0 \in \Re$ and $\pi = (\bar{\mu}_0, \bar{\mu}_1)$ is optimal, provided only that the integral in (5) is defined. This can be seen by setting $\epsilon = 0$ in (9) and (10).

3. MEASURABLE SELECTION

The example of the preceding section shows that if measurability restrictions are placed on μ_1 and μ_0, then measurable selection becomes a crucial part of the analysis. We discuss this in the framework of Borel spaces. Given a topological space Y, we denote by $\mathcal{B}_Y$ the σ-algebra generated by the open subsets of Y and refer to the members of $\mathcal{B}_Y$ as the *Borel subsets* of Y. A topological space Y is a *Borel space* if it is homeomorphic to a Borel subset of a complete separable metric space. The concept of Borel space is quite broad, containing any "reasonable" subset of n-dimensional Euclidean space. Any Borel subset of a Borel space is again a Borel space, as is any homeomorphic image of a Borel space and any finite or countable Cartesian product of Borel spaces. However, even in the unit square, there exist Borel sets whose projections onto an axis are not Borel subsets of that axis. This leads us to the analytic sets. A subset A of a Borel space Y is said to be *analytic* if there exists a Borel space Z and a Borel subset B of $Y \times Z$ such that $A = \text{proj}_Y(B)$, where proj_Y is the projection mapping from $Y \times Z$ to Y. It is clear that every Borel subset of a Borel space Y is also an analytic subset of Y.

We list some of the properties of analytic sets that are relevant to our development. Let Y

and Z be Borel spaces.

(i) If $A \subset Y$ is analytic and $h : Y \to Z$ is Borel measurable, then $h(A)$ is analytic. In particular, if Y is a product of Borel spaces Y_1 and Y_2 and $A \subset Y_1 \times Y_2$ is analytic, then $\text{proj}_{Y_1}(A)$ is analytic.

(ii) If $A \subset Z$ is analytic and $h : Y \to Z$ is Borel measurable, then $h^{-1}(A)$ is analytic.

(iii) If $A_1, A_2, \ldots$ are analytic subsets of Y, then $\cup_{k=1}^{\infty} A_k$ and $\cap_{k=1}^{\infty} A_k$ are analytic. It is not always true, however, that the complement of an analytic set is analytic, so the collection of analytic subsets of Y need not constitute a σ-algebra.

Let Y be a Borel space and let $h : Y \to [-\infty, \infty]$ be a function. We say that h is *lower semianalytic* if $\{y \in Y | h(y) < c\}$ is analytic for every $c \in \Re$.

Theorem 1: Let Y and Z be Borel spaces, and let $h : Y \times Z \to [-\infty, \infty]$ be lower semianalytic. Then $h^* : Y \to [-\infty, \infty]$ defined by

$$h^*(y) = \inf_{z \in Z} h(y, z) \tag{11}$$

is lower semianalytic.

It turns out that if $h^* : Y \to [-\infty, \infty]$ is a given lower semianalytic function and Z is any uncountable Borel space, then a Borel measurable function $h : Y \times Z \to [-\infty, \infty]$ can be found for which (11) holds. A comparison of (4) with (11) shows how lower semianalytic functions can arise in dynamic programming. We give as lemmas two useful properties of these functions.

Lemma 1: Let Y be a Borel space and let $h, l : Y \to [-\infty, \infty]$ be lower semianalytic functions. Suppose that for every $y \in Y$, the sum $h(y) + l(y)$ is defined, i.e., is not of the form $\infty - \infty$. Then $h + l$ is lower semianalytic.

Lemma 2: Let Y and Z be Borel spaces, $h : Y \to Z$ Borel measurable, and $l : Z \to [-\infty, \infty]$ lower semianalytic. Then the composition $l \circ h$ is lower semianalytic.

If the function g in (4) is lower semianalytic, then J_1 defined by (4) is lower semianalytic (Theorem 1). If f in (5) is Borel measurable, then for fixed (x_0, u_0), the function $J_1[f(x_0, u_0, w_0)]$ is lower semianalytic in w_0 (Lemma 2). In the example of Section 2, there is no cost $g_0(x_0, u_0, w_0)$ incurred in the first stage of the system operation. When such a cost is incurred and g_0 is lower semianalytic, the integrand in (5) becomes $g_0(x_0, u_0, w_0) + J_1[f(x_0, u_0, w_0)]$, which is still lower semianalytic in w_0 for fixed (x_0, u_0) (Lemma 1).

In order to carry out the integration in (5), we must discuss the measurability of lower semianalytic functions. There are at least three natural σ-algebras in a Borel space Y. The

first is the Borel σ-algebra $\mathcal{B}_Y$ mentioned earlier. The second is the σ-algebra generated by the analytic subsets of Y, called the *analytic σ-algebra* and denoted by ${}_Y$. The third is the *universal σ-algebra* $\mathcal{U}_Y$, which is the intersection of all completions of $\mathcal{B}_Y$ with respect to all probability measures. Thus, $E \in \mathcal{U}_Y$ if and only if, given any probability measure p on $(Y, \mathcal{B}_Y)$, there is a Borel set B and a p-null set N such that $E = B \cup N$.

Theorem 2: Let Y be a Borel space. Then

$$\mathcal{B}_Y \subset \mathcal{A}_Y \subset \mathcal{U}_Y.$$

Corresponding to the three σ-algebras in Borel spaces, we have three classes of measurable functions. Suppose X, Y and Z are Borel spaces and $h : Y \to Z$ is given. The function h is said to be *Borel, analytically*, or *universally measurable* if for every $B \in \mathcal{B}_Z$, the set $h^{-1}(B)$ is Borel, analytically, or universally measurable respectively. It can be shown that if $U \subset Z$ is universally measurable and h is universally measurable, then $h^{-1}(U)$ is also universally measurable.

As a result, if $g : X \to Y$, $h : Y \to Z$ are Borel or universally measurable functions, then the composition $(g \circ h) : X \to Z$ is Borel or universally measurable, respectively. However, if g and h are analytically measurable, then $(g \circ h)$ *need not* be analytically measurable and this is a primary source of difficulty in working with analytically measurable policies in dynamic programming. If $h : Y \to [-c, \infty]$ is universally measurable, where $c \in \Re$, and p is a probability measure on $(Y, \mathcal{B}_Y)$, then p has a unique extension to a probability measure $\bar{p}$ and $\int h d\bar{p}$ is defined. We write simply p instead of $\bar{p}$ and $\int h dp$ in place of $\int h d\bar{p}$. In particular, if h is lower semianalytic, then $\int h dp$ can be defined in this manner. Thus defined, the integral of universally measurable functions operates linearly and obeys the classical convergence theorems. We understand the integration in (5) to be defined in this way.

We investigate now the existence of a measurable (in one of the three senses defined above) function $\bar{\mu}_1$ satisfying (7), where we assume that g is lower semianalytic and f is Borel measurable. This issue is resolved by the following selection theorem. Part (c) of the theorem addresses the question raised in the last paragraph of Section 2.

Theorem 3: Let Y and Z be Borel spaces and let $h : Y \times Z \to [-\infty, \infty]$ be a lower semianalytic function. Define $h^* : Y \to [-\infty, \infty]$ by (11). Let $I = \{y \in Y \mid \text{there exists a } z_y \in Z \text{ for which } h(y, z_y) = h^*(y)\}$, i.e., I is the set of points y for which the infimum in (11) is actually attained. Choose $\epsilon > 0$.

(a) For every probability measure p on $(Y, \mathcal{B}_Y)$, there exists a *Borel measurable* $\phi_p : Y \to Z$

such that

$$h[y, \phi_p(y)] \leq \begin{cases} h^*(y) + \epsilon & \text{if} \qquad h^*(y) > -\infty, \\ -1/\epsilon & \text{if} \qquad h^*(y) = -\infty, \end{cases} \tag{12}$$

for p-almost every $y \in Y$.

(b) There exists an *analytically measurable* $\phi : Y \to Z$ such that

$$h[y, \phi(y)] \leq \begin{cases} h^*(y) + \epsilon & \text{if} \qquad h^*(y) > -\infty, \\ -1/\epsilon & \text{if} \qquad h^*(y) = -\infty, \end{cases} \tag{13}$$

for every $y \in Y$.

(c) There exists a *universally measurable* $\phi : Y \to Y$ such that (13) holds for every $y \in Y$, and

$$h[y, \phi(y)] = h^*(y) \qquad y \in I.$$

There is one additional measure theoretic difficulty, which is encountered in the general model of Section 4 but not in the simple example of Section 2. It often is the case that the distribution of the disturbance at the k^{th} stage is parameterized by the k^{th} state and control, i.e., has the form $p(dw_k \mid x_k, u_k)$. Equation (5) would then become

$$J_2(x_0) = \inf_{u_0 \in R} \int J_1[f(x_0, u_0, w_0)] p(dw_0 \mid x_0, u_0).$$

The measurability of J_2 is of no consequence in the example of Section 2, since the dynamic programming algorithm terminates at this stage. If, however, more than two stages are involved,then J_2 would become part of an integrand in the next iteration, and we must check that it, like J_1, is lower semianalytic. In light of Theorem 1, it suffices to verify that $\int J_1[f(x_0, u_0, w_0)] p(dw_0 \mid x_0, u_0)$ is a lower semianalytic function of (x_0, u_0). The relevant definition and theorem follow.

Let Y and Z be Borel spaces. A *stochastic kernel* $q(dz \mid y)$ on Z given Y is a collection of probability measures on $(Z, \mathcal{B}_Z)$ parameterized by the elements of Y. If for each Borel set $B \in \mathcal{B}_Z$, the function $q(B \mid y)$ is Borel measurable in y, the stochastic kernel $q(dz \mid y)$ is said to be *Borel measurable*.

Theorem 4: Let Y and Z be Borel spaces, let $h : Y \times Z \to [-\infty, \infty]$ be a lower semianalytic function which is bounded above or bounded below, and let $q(dz \mid y)$ be a Borel measurable stochastic kernel on Z given Y. Then the function $\lambda : Y \to [-\infty, \infty]$ defined by

$$\lambda(y) = \int_Z h(y, z) q(da \mid y)$$

is lower semianalytic.

4. THE GENERAL FINITE HORIZON MODEL

The general model consists of a Borel state space S, a Borel control space C, a Borel disturbance space W, a Borel measurable system function $f : S \times C \times W \to S$, a Borel measurable stochastic disturbance kernel $p(dw \mid x, u)$ on W given $S \times C$, a lower semianalytic one stage cost function $g : S \times C \times W \to [-\infty, \infty]$, and a horizon N, which is a positive integer. We assume that either g is bounded below (P), or else g is bounded above (N). Thus, there are really two models, and the symbols (P) and/or (N) will precede a result to indicate that it is valid for the corresponding model. The boundedness assumption on g is made for convenience. Stronger results as well as infinite horizon and imperfect state information counterparts are possible. A *policy* is a sequence $\pi = (\mu_0, \ldots, \mu_{N-1})$ of measurable mappings, where $\mu_k : S \times C \times \ldots \times C \times S \to S$ determine the k^{th} control u_k as

$$u_k = \mu_k(x_0, u_0, \ldots, u_{k-1}, x_k), \qquad k = 0, \ldots N-1.$$

If for each k, μ_k has the form $u_k = \mu_k(x_0, x_k)$, the policy π is said to be *semi-Markov*. If μ_k has the form $u_k = \mu_k(x_k)$, π is said to be *Markov*. We say that a *policy is Borel, analytically, or universally measurable* if each component of the policy is Borel, analytically, or universally measurable, respectively. We consider only policies that are measurable in one of these senses. For convenience we admit only nonrandomized policies. Results relating to randomized policies are also available. The *cost corresponding to a policy* π at $x_0 \in S$ is

$$J_{N,\pi}(x_0) = E_{\pi,x_0}\left(\sum_{k=0}^{N-1} g(x_k, u_k, w_k)\right), \tag{15}$$

where the expectation is with respect to the probability measure determined by the disturbance distributions $p(dw_k \mid x_k, u_k), k = 0, \ldots, N-1$, the sytem equation

$$x_{k+1} = f(x_k, u_k, w_k), \qquad k = 0, \ldots, N-2,$$

and the policy

$$u_k = \mu_k(x_0, u_0, \ldots, u_{k-1}, x_k), \qquad k = 0, \ldots, N-1.$$

The *optimal N-stage cost at* x_0 is

$$J_N^*(x_0) = \inf_{\pi} J_{N,\pi}(x_0). \tag{16}$$

This optimal cost can be shown to be the same regardless of whether the infimum in (16) is over all universally measurable policies, only the Borel measurable Markov policies, or any collection

of policies lying between these two extremes. This follows from the fact that when x_0 is fixed and a universally measurable policy π is given, an "equivalent" Borel measurable policy $\bar{\pi}$ exists, i.e., a policy $\bar{\pi}$ for which

$$J_{N,\pi}(x_0) = J_{N,\bar{\pi}}(x_0).$$

It is only when properties which hold uniformly in the initial state x_0 are considered that the difference between universally measurable and Borel measurable policies manifests itself. Given $\epsilon > 0$ and $x_0 \in S$, we say a policy π is *ϵ-optimal at x_0* if

$$J_{N,\pi}(x_0) \leq \begin{cases} J_N^*(x_0) + \epsilon & \text{if } J_N^*(x_0) > -\infty, \\ -1/\epsilon & \text{if } J_N^*(x_0) = -\infty. \end{cases}$$

We say π is *optimal at x_0* if

$$J_{N,\pi}(x_0) = J_N^*(x_0).$$

If p is a probability measure on $(X, \mathcal{B}_S)$ and π is ϵ-optimal (optimal) at p-almost every x_0, we say π is *p-ϵ-optimal (p-optimal)*. If π is ϵ-optimal (optimal) at every x_0, we say π is *ϵ-optimal (optimal)*.

The optimal cost functions can be generated by the dynamic programming algorithm in both models (P) and (N).

Theorem 5: (P, N) For $K = 1, \ldots, N$, J_K^* is lower semianalytic, and

$$J_K^*(x) = \inf_{u \in C} \int_W \big(g(x,u,w) + J_{K-1}^*[f(x,u,w)]\big) p(dw \mid x, u), \tag{17}$$

where $J_0^*(x) = 0$ for every $x \in S$.

The existence results depend along the lines of Theorem 3 on the type of measurability of policies allowed.

Theorem 6: Let $\epsilon > 0$ and a probability measure p on $(S, \mathcal{B}_S)$ be given.

(P) There exists a Borel measurable p-ϵ-optimal Markov policy.

(N) There exists a Borel measurable p-ϵ-optimal semi-Markov policy.

Theorem 7: (P, N) Suppose for $K = 1, \ldots, N$, and for every $x \in S$, the infimum in (17) is attained, and let p be a probability measure on $(S, \mathcal{B}_S)$. Then there exists a Borel measurable p-optimal Markov policy.

If under (N) we have $J_N^*(x) > -\infty$ for all $x \in S$, the policy in Theorem 6 can actually be taken to be Markov. The dependence on p in Theorem 6 can be eliminated by admitting analytically measurable policies, but the semi-Markov property under (N) is lost.

Theorem 8: Let $\epsilon > 0$ be given.

(P) There exists an analytically measurable ϵ-optimal Markov policy.

(N) There exists an analytically measurable ϵ-optimal policy.

There is apparently no stronger version of Theorem 7 for analytically measurable policies. The fact that the composition of two analytically measurable functions need not be analytically measurable is the primary source of difficulty here. However, if universally measurable policies are allowed, both Theorems 7 and 8 can be strengthened.

Theorem 9: Let $\epsilon > 0$ be given.

(P) There exists a universally measurable ϵ-optimal Markov policy.

(N) There exists a universally measurable ϵ-optimal semi-Markov policy.

As in Theorem 6, if under (N) we have $J_N^*(x) > -\infty$ for all $x \in S$, the policy in Theorem 9 can be taken to be Markov.

Theorem 10: (P, N) Suppose for $K = 1, \ldots, N$, and for every $x \in S$, the infimum in (17) is attained. Then there exists a universally measurable optimal Markov policy $\pi = (\mu_0, \ldots, \mu_{N-1})$, such that for $K = 1, \ldots, N$, and every $x \in S$, $\mu_{N-k}(x)$ attains the infimum in (17).

A similar development is possible for infinite horizon dynamic programming models (see the Bertsekas and Shreve book). Again, the key point in this development is the use of universally measurable policies. This class of policies is sufficiently rich to ensure the existence of ϵ-optimal policies, and to allow the development of a general and comprehensive dynamic programming theory that is as powerful and easy to use as its deterministic counterpart.