U. A. M. IZTAPALAPA BIBLIOTECA

PROCESOS DE CONTROL MARKOVIANO CON GANANCIA PROMEDIO

TESIS QUE PRESENTA

MARIA TERESA ROBLES ALCARAZ

PARA LA OBTENCION DEL GRADO DE

MAESTRO EN MATEMATICAS

ABRIL 1991

105958

Director: Onésimo Hernández Lerma

División de Ciencias Básicas e Ingeniería

Universidad Autónoma Metropolitana

Unidad Iztapalapa

U. A. M. IZTAPALAPA BIBLIOTECA

Agradezco al Departamento de Matemáticas de la Universidad de Sonora su apoyo en cada una de las etapas de mis estudios de maestría, y al Departamento de Matemáticas de la Universidad Autónoma Metropolitana, Unidad Iztapalapa, por todas las facilidades prestadas.

Mis agradecimientos, también son, para el Consejo Nacional de Ciencia y Tecnología (CONACYT) de quién recibí una beca complementaria (no. reg. 55563).

INDICE

105958

INTRODUCCION. 1

NOTACION. 3

I. PROCESOS DE CONTROL MARKOVIANO. 5

0. Introducción. 5

1. Modelo de Control Markoviano. 5

2. Políticas Admisibles. 6

3. Indices de Funcionamiento. 9

4. Condiciones de Continuidad y Compacidad. 11

5. Condiciones de Recurrencia. 16

II. ECUACION DE OPTIMALIDAD. 24

0. Introducción. 24

1. Ecuación de Optimalidad. 24

2. Condiciones de sm-contracción. 26

3. Iteración de Valores. 30

4. Límite de Problema Descontados. 34

5. Comentarios. 42

III. POLITICAS OPTIMAS. 43

0. Introducción. 43

1. Maximizador de la Ecuación de Optimalidad. 44

2. Límite de Problemas en Horizonte Finito. 45

3. Condición de Doeblin. 48

CONCLUSIONES Y COMENTARIOS. 69

APENDICE. 71

REFERENCIAS. 79

INTRODUCCION

En el presente trabajo se estudia una subclase de los procesos de control markovianos en tiempo discreto. En general, estos procesos $\{x_t\}$ son modelos matemáticos de sistemas dinámicos que son observados en el tiempo, para los cuales después de cada observación x_t del estado del sistema, un control a_t es aplicado. Y basado en el estado y el control una ganancia (o costo) $r(x_t, a_t)$ es recibida .

El problema básico, consiste en determinar (o encontrar) una política (sucesión de controles) $\delta=\{a_t\}$ que optimize el comportamiento del sistema, de acuerdo a una función objetivo en un período determinado u horizonte del sistema.

Nos interesaremos por la subclase de estos procesos tales que:

(a) el espacio de estados del sistema es un <u>espacio de Borel</u>;

(b) la función objetivo es conocida como <u>ganancia promedio</u>:

$$J(\delta,x) := \limsup \frac{1}{n} E_x^{\delta} \sum_{t=0}^{n-1} r(x_t, a_t) ,$$

donde E_x^{δ} representa al operador esperanza, el estado inicial del sistema es $x_0=x$ y la política δ es usada. La función ganancia por etapa $r(x_t, a_t)$ la suponemos acotada.

Cuando el horizonte del sistema es infinito se usa, generalmente, el enfoque de la Ecuación de Optimalidad para estudiar el Problema de Control Optimo. Con este enfoque, se asegura la existencia de una política óptima si se tiene una solución de la Ecuación de Optimalidad y condiciones adecuadas de continuidad y compacidad.

Para estudiar el Problema de Control Optimo con ganancia Promedio es necesario imponer condiciones de 'recurrencia'. En particular, bajo este tipo de condiciones es posible asegurar la existencia de una solución acotada de la Ecuación de Optimalidad.

En el Capítulo I se introducen los procesos de control markoviano, se plantea el Problema de Control Optimo y se introducen las condiciones de recurrencia bajo las cuales se estudiará el Problema de Control Optimo con Ganancia Promedio.

En el Capítulo II se introduce la Ecuación de Optimalidad. El objetivo de este capítulo es estudiar la existencia de soluciones de la Ecuación de Optimalidad considerando los siguientes enfoques:

a) como punto fijo del Operador de Optimalidad;

b) límite de problemas con horizonte finito;

c) límite de problemas descontados.

Lo anterior se hace bajo las condiciones de recurrencia introducidas en el Capítulo I.

En el Capítulo III se asegura la existencia de una política óptima, usando los enfoques a) y b) señalados en el párrafo anterior. Finalmente, se estudia el Problema de Control Optimo bajo la Condición de Doeblin y la restricción de compacidad sobre los espacios de estados y acciones.

El análisis, en este trabajo, está basado en resultados de funciones multivaluadas, funciones semi-continuas superiormente, operadores de contracción, selectores medibles y topologías en espacios de medidas de probabilidad. Algunos de estos resultados son citados, sin demostración, en el Apéndice.

NOTACION

Para un espacio topológico X, se denota por $B(X)$ la σ-álgebra de Borel, esto es, la mínima σ-álgebra que contiene la familia de subconjuntos abiertos de X. Así, 'medibilidad' de conjuntos y funciones significa medibilidad de Borel.

Un espacio de Borel es un subconjunto de Borel de un espacio métrico, separable y completo.

Dado un espacio de Borel X, denotamos por

$B(X)$ el espacio de las funciones definidas en X, con valores en los reales, acotadas y medibles. Este espacio es de Banach, con la norma del supremo

$$\|v\| := \sup_{x\in X} |v(x)|, \quad v\in B(X).$$

$C(X)$ el subespacio de $B(X)$, de las funciones continuas.

Sea $v: X \longrightarrow \bar{\mathbb{R}}$ medible se dice que es <u>semi-continua superiormente</u> si para cada $x\in X$ y cada sucesión $(x_n)\subset X$ que converge a x, se cumple

$$\lim_n \sup v(x_n) \leq v(x).$$

$C_s(X)$ denota al espacio de las funciones acotadas semi-continuas superiormente.

$\mathbb{P}(X)$ el espacio de las medidas de probabilidad, dotado con la topología de la convergencia débil.

$\|\mu\|$	norma de la variación total para medidas con signo μ en X.
$I_B(x)$	la función indicadora del conjunto B definida por $I_B(x):=1$ si $x\in B$, $I_B(x):=0$ si $x\notin B$.
$q(\cdot/\cdot)$	probabilidad de transición (o probabilidad condicional) en X dado Y es una función tal que $q(\cdot/y)$ es una medida de probabilidad en X para cada $y\in Y$, y $q(B/\cdot)$ es una función medible en Y, para cada $B\in\mathcal{B}(X)$.
Δ	Conjunto de las políticas admisibles.
Π	Conjunto de las políticas estacionarias aleatorizadas
$\mathbb{F}$	Conjunto de las políticas estacionarias.
$\mathbb{H}_t$	Conjunto de las historias admisibles hasta el tiempo t.
P_ν^δ	medida de probabilidad inducida por la política δ y la distribución inicial ν.
E_ν^δ	operador esperanza con respecto a la medida P_ν^δ.

I. PROCESOS DE CONTROL MARKOVIANOS

0. Introducción

1. Modelo de control Markoviano

2. Políticas admisibles

3. Indices de Funcionamiento

4. Condiciones de continuidad y compacidad

5. Condiciones de recurrencia

0. INTRODUCCION.

El objetivo de este capítulo es introducir los procesos de control markovianos y plantear el problema de control óptimo (PCO), para lo que se requieren los siguientes elementos: un modelo de control, un conjunto de políticas admisibles y un índice de funcionamiento (Secciones 1,2 y 3). Además, se introducen condiciones de continuidad, compacidad y recurrencia sin las cuales, en general, no es posible resolver el PCO con Ganancia Promedio (Secciones 4 y 5), lo cual es el interés general del presente trabajo.

1. MODELO DE CONTROL MARKOVIANO.

Definición 1.1. Un modelo de control markoviano (MCM) consiste en los objetos (X, A, q, r), donde:

– X representa el espacio de estados y suponemos es un espacio de Borel, es decir, un subconjunto de Borel de un espacio métrico,

separable y completo;

– A el espacio de acciones es un espacio de Borel. Sea D una función que a cada estado $x\in X$ le asocia un subconjunto $D(x)\subseteq A$; los elementos de $D(x)$ representan las acciones o controles admisibles para el estado $x\in X$.

Sea $\mathbb{K}$ el conjunto de parejas (x,a) estado-acción definido por

$$\mathbb{K} := \{(x,a):\ x\in X, a\in D(x)\},$$

al que le pedimos sea un subconjunto medible del espacio producto XA.

– $q(B/x,a)$, $B\in\mathcal{B}(X)$ y $(x,a)\in\mathbb{K}$, representa la ley de tansición del sistema, es una probabilidad de transición sobre X dado $\mathbb{K}$;

– $r:\ \mathbb{K} \longrightarrow \mathbb{R}$, la función de ganancia en un paso, es una función medible y acotada.

Podemos interpretar al MCM (X,A,q,r), como un modelo matemático de un sistema estocástico controlado que se observa en los tiempos $t=0,1,2,\ldots$. El estado y la acción en el tiempo t serán representados por x_t y a_t, respectivamente. Ahora, si $x_t=x\in X$ y $a_t=a\in D(x)$, se recibe una ganancia $r(x,a)$ y el sistema 'se mueve' a un nuevo estado $x_{t+1}=x'$ de acuerdo a la ley de transición $q(\cdot/x,a)$, es decir,

$$q(B/x,a)=\text{Prob}\left[\, x_{t+1}\in B \mid x_t=x,\ a_t=a \,\right].$$

2. POLITICAS ADMISIBLES.

Para elegir las acciones de control para el sistema en cada tiempo t, se debe seguir alguna regla (o política) la cual puede o no depender de la historia del proceso. Para definir estos conceptos

en forma precisa consideremos los conjuntos

$$\mathbb{H}_0 := \mathbf{X}$$

$$\mathbb{H}_t := \{(x_0, a_0, \ldots, x_{t-1}, a_{t-1}, x_t) : (x_s, a_s) \in \mathbb{K}, 0 \leq s \leq t-1,\ x_t \in \mathbf{X}\}\ ,\ t \geq 1.$$

En cada tiempo t, $\mathbb{H}_t$ es el conjunto de historias admisibles hasta el tiempo t (t-historias), conteniendo en consecuencia, la información sobre la evolución del sistema hasta dicho tiempo; denotaremos genéricamente por h_t a los elementos del conjunto $\mathbb{H}_t$. Observe que

$$\mathbb{H}_t = \mathbb{K}^t \mathbf{X}.$$

Definición 2.1. Una política aleatorizada es una sucesión de probabilidades de transición $\delta = (\delta_t)$ sobre **A** dado $\mathbb{H}_t$ tales que

$$\delta_t(D(x_t)/h_t) = 1 \text{ para toda } h_t \in \mathbb{H}_t \text{ y } t \geq 0.$$

Denotaremos por Δ a la familia de políticas aleatorizadas.

Definición 2.2. Una política $(\delta_t) \in \Delta$ se dice que es estacionaria aleatorizada si existe una probabilidad de transición $\delta(\cdot/\cdot)$ sobre **A** dado **X** tal que

$$\delta_t(\cdot/h_t) = \delta(\cdot/x_t) \text{ para toda } h_t \in \mathbb{H}_t \text{ y } t \geq 0.$$

A esta familia la denotamos por Π.

Sea $\mathbb{F} := \{ f : \mathbf{X} \longrightarrow \mathbf{A} \mid \text{medible y tal que } f(x) \in D(x), \text{ para todo } x \in \mathbf{X}\}$; nos referiremos a esta familia de funciones como selectores medibles.

Definición 2.3. Una política de Markov es una sucesión (f_t) tal que $f_t \in \mathbb{F}$ para todo $t \geq 0$.

Definición 2.4. Una política estacionaria (f_t) es una política de Markov tal que $f_t = f \in \mathbb{F}$ para todo $t \geq 0$.

Al conjunto de las políticas estacionarias lo identificaremos con $\mathbb{F}$.

Notemos que cualquier política de Markov se puede representar como una política aleatorizada en la siguiente forma

$$\delta_t(B/h_t)=I_B(f_t(x_t)) \text{ para todo } B\in\mathcal{B}(A), h_t\in H_t \text{ y } t\geq 0,$$

donde I_B es la función indicadora del conjunto B.

En consecuencia, se satisface la siguiente relación de contenciones

$$\mathbb{F} \subset \Pi \subset \Delta.$$

Procesos Controlados de Markov

Sea (Ω,F) el espacio medible, donde $\Omega:=(XA)^{\infty}$ y F la σ-álgebra producto correspondiente. Los elementos $\omega\in\Omega$ son sucesiones infinitas de la forma

$$\omega=(x_0,a_0,x_1,a_1,\ldots,x_t,a_t,\ldots)$$

donde $x_t\in X$ y $a_t\in A$.

Definimos las variables de estado y control x_t y a_t como las proyecciones (en el tiempo t) de Ω a los conjuntos X y A, respectivamente; es decir,

$$x_t(\omega)=x_t \quad y \quad a_t(\omega)=a_t \quad \text{para todo } t\geq 0.$$

Por el teorema de C. Ionescu Tulcea (teorema 2.7.2 en Ash 1972), para cada $\delta\in\Delta$ y cada distribución inicial ν en X, existe una medida de probabilidad P_ν^δ dada por

$$P_\nu^\delta(dx_0,da_0,dx_1,\ldots)=\nu(dx_0)\delta_0(da_0/x_0)q(dx_1/x_0,a_0)\delta_1(da_1/x_0,a_0,x_1)\ldots$$

que satisface las siguientes propiedades:

(2.5) $\quad P_\nu^\delta(x_0\in B)=\nu(B), \quad B\in\mathcal{B}(X);$

(2.6) $P_\nu^\delta(a_t \in C / h_t) = \delta_t(C / h_t)$ para todo $C \in B(A)$ y $h_t \in H_t$;

(2.7) $P_\nu^\delta(x_{t+1} \in C / h_t, a_t) = q(C / x_t, a_t)$ para todo $C \in B(X)$ y $h_t \in H_t$.

Al proceso estocástico $(\Omega, F, P_\nu^\delta, \{x_t\})$ lo llamaremos proceso de Markov controlado (PMC).

Al operador esperanza con respecto a la medida de probabilidad P_ν^δ lo denotaremos por E_ν^δ; si la distribución inicial ν está concentrada en un punto $x \in X$, P_ν^δ y E_ν^δ serán denotadas por P_x^δ y E_x^δ, respectivamente.

3. INDICES DE FUNCIONAMIENTO.

Para determinar las políticas que son óptimas necesitamos un criterio de optimalidad, es decir, una función que mida la evolución del sistema la cual llamaremos índice de funcionamiento (o función objetivo).

Dada una política δ y una distribución ν en X, definimos el Indice de Ganancia Promedio $J(\delta, \nu)$ como

(3.1) $$J(\delta, \nu) := \limsup_n \frac{1}{n+1} E_\nu^\delta \sum_{k=0}^{n} r(x_k, a_k),$$

y al Indice de Ganancia Descontada $V_\beta(\delta, \nu)$ por

(3.2) $$V_\beta(\delta, \nu) := E_\nu^\delta \sum_{k=0}^{\infty} \beta^k r(x_k, a_k),$$

donde $\beta \in (0,1)$, representa un factor de descuento.

Observación 3.3.

a) Cuando la distribución inicial ν esté concentrada en un punto $x \in X$, esto es, $\nu(B) = I_B(x)$ denotaremos a (3.1) y (3.2) por

$J(\delta,x)$ y $V_\beta(\delta,x)$, respectivamente;

b) Puesto que r, la función de ganancia en un paso, es acotada por alguna constante M el lim sup en (3.1) es finito . Además,

$$|V_\beta(\delta,\nu)| \leq M(1-\beta)^{-1}, \tag{3.4}$$

para todo $\delta\in\Delta$ y $\nu\in P(X)$.

Ya que tenemos un modelo de control, un conjunto de políticas admisibles y un índice de funcionamiento, (3.1) o (3.2), pasamos a definir el problema de control óptimo.

Problema de Control Optimo

El problema de control óptimo (PCO) bajo el índice de Ganancia Promedio (GP) consiste en encontrar una política $\delta^*\in\Delta$ que maximize el índice de funcionamiento (3.1), esto es,

$$J(\delta^*,x) = \sup_{\delta\in\Delta} J(\delta,x) \text{ para todo } x\in X. \tag{3.5}$$

Una política que satisface (3.5), si existe, se dice que es óptima bajo el índice de GP, y a la función

$$J^*(x) := \sup_{\delta\in\Delta} J(\delta,x) \quad \text{para } x\in X, \tag{3.6}$$

se le conoce como función de valores óptimos en GP.

Analogamente, el problema de control óptimo para el Indice de Ganancia Descontada (GD) consiste en encontrar una política ${}^*\delta\in\Delta$ tal que

$$V_\beta({}^*\delta,x) = \sup_{\delta\in\Delta} V_\beta(\delta,x) \text{ para todo } x\in X. \tag{3.7}$$

Si ${}^*\delta\in\Delta$ satisface (3.7), entonces se dice que es óptima con el criterio β-descontado, y a la función

$$V_\beta(x) := \sup_{\delta\in\Delta} V_\beta(\delta,x), \quad x\in X, \tag{3.8}$$

se le llama función de valores óptimos con el indice GD, con factor de descuento β.

En el presente trabajo nos interesamos por el PCO con el Indice de Ganancia Promedio. Este criterio tiene la particularidad de no depender del comportamiento del sistema en las primeras etapas; es decir, para t arbitrario y fijo

$$\lim_{n} \frac{1}{n+1} E_x^{\delta} \sum_{k=0}^{t} r(x_k, a_k) = 0 \text{ para todo } x \in X \text{ y } \delta \in \Delta.$$

por lo que no distingue entre dos políticas que generan la misma ganancia a partir de cierta etapa en adelante, aunque difieran en las primeras.

Esta característica resulta ventajosa cuando se consideran problemas de control adaptable, esto es, problemas en los que el modelo de control $(X, A, q(\theta), r(\theta))$ depende del parámetro 'desconocido' θ, el cual debe ser 'estimado' al mismo tiempo que se 'controla' el sistema. Comúnmente, las propiedades de los estimadores se da en términos de algún tipo de convergencia, y se espera que en el límite las estimaciones sean 'suficientemente buenas', aunque en las primeras etapas no sean muy precisas.

4. CONDICIONES DE CONTINUIDAD Y COMPACIDAD.

Como es usual para estudiar los problemas de control óptimo, se imponen condiciones de continuidad y compacidad al modelo de control Markoviano (X, A, q, r) para asegurar la existencia de selectores medibles.

Para el PCO con el índice de Ganancia Descontada la existencia de una política óptima es asegurada si se satisface el siguiente conjunto de condiciones.

Condición 4.1. Para cada $x \in X$

a) $D(x)$ es compacto;

b) $r(x,\cdot)$ es semi-continua superiormente sobre $D(x)$;

c) $\int h(y)q(dy/x,a)$ es una función semi-continua superiormente en $a \in D(x)$ para toda $h \in B(X)$.

Bajo la Condición 4.1, V_β es la única función que satisface la ecuación

$$V_\beta(x) = \sup_{a \in D(x)} \{ r(x,a) + \beta \int V_\beta(y) q(dy/x,a) \}, \quad x \in X. \tag{4.2}$$

Esta ecuación se conoce como Ecuación de Optimalidad para el Indice de Ganancia β-Descontada (EOGD). También se tiene que una política $f_\beta \in F$ es β-óptima si y sólo si maximiza el lado derecho de la EOGD, esto es,

$$V_\beta(x) = r(x,f_\beta(x)) + \beta \int V_\beta(y) q(dy/x,f_\beta(x)) \}, \quad x \in X. \tag{4.3}$$

Para la demostración de estos resultados ver Hernández-Lerma (1989), pag. 19.

Para estudiar el PCO con el criterio GP, además de condiciones de continuidad y compacidad (en algunos caso 4.1) se necesitan algunas condiciones de recurrencia (Sección 5). En general, las políticas óptimas no siempre existen, aún cuando se satisfaga la Condición 4.1, como lo ilustra el siguiente ejemplo.

Ejemplo 4.4. Sea $X=\{1,\bar{1},2,\bar{2},3,\bar{3},\ldots\}$ y $A=\{1,2\}$ los espacios de estados y acciones. Supóngase que las probabilidades de transición están dadas por

$$q(i+1/i,1)=q(\bar{i}/i,2)=1$$

$$q(\bar{i}/\bar{i},1)=q(\bar{i}/\bar{i},2)=1$$

para todo $i \geq 0$.

La función de ganancia en un paso solo depende del estado

$$r(i,a)=0$$

$$r(\bar{i},a)=1-1/i$$

para todo $a \in A$ e $i \geq 0$.

Si el sistema inicia en $x_0=1$, entonces para cada política $\delta \in \Delta$ se cumple que $J(\delta,1)<1$.

Caso 1: Con probabilidad uno, la política δ siempre elige la acción 1, $J(\delta,1)=0$;

Caso 2: Con probabilidad positiva, δ elige la acción 2 en algún tiempo. En este caso para algún n, existe una probabilidad p_n de que δ tome la acción 2 cuando el sistema se encuentra en el estado n. Sea $\bar{n}$ el mínimo n con esta propiedad; entonces $J(\delta,1) \leq (1-1/n)$.

Bajo la política estacionaria

$$f(i)=\begin{cases} 1 & \text{si } i \neq n \\ 2 & \text{si } i=n \end{cases}$$

se satisface $J(f,1)=1-1/n$, por lo tanto el supremo de todas las políticas es igual a 1, en consecuencia la política óptima no existe.

■

En el ejemplo anterior, aunque la política óptima no existe, si existen políticas estacionarias que controlan al sistema 'tan bién' como se desee. Si esta fuera siempre la situación podríamos pensar que es posible restringir la atención a la familia de las políticas estacionarias. El siguiente ejemplo muestra que las políticas estacionarias no son suficientes.

Ejemplo 4.5. Tomemos $X=\{1,2,3,\dots\}$ y $A=\{1,2\}$; la probabilidad de transición y la función de ganancia en un paso son dados de la siguiente manera:

$$q(i+1/i,1)=q(i/i,2)=1$$

$$r(i,1)=0,\quad r(i,2)=1-1/i$$

Si el sistema se encuentra en el estado i, se le permite pasar al estado i+1 sin recibir ganancia alguna, ó permanecer en el estado i y recibir una ganancia $1-1/i$.

Tomemos $x_0=1$, $f\in F$ y consideremos los siguientes casos.

Caso 1: $f(i)=1$, para todo $i\in X$; entonces $J(f,1)=0$.

Caso 2: $f(i)=\begin{cases}1 & \text{si } i\neq n\\ 2 & \text{si } i=n\end{cases}$; entonces

$$J(f,i)=1-1/n<1\ .$$

Ahora, consideremos una política no estacionaria δ, tal que, elige la acción de control 2, i veces consecutivas y después elige la acción 1, cuando el proceso entra por primera vez al estado i.

Si $x_0=1$, entonces las ganancias son

$$0,0,1/2,1/2,2/3,2/3,2/3,0,3/4,3/4,3/4,3/4,0,4/5,4/5,\dots$$

entonces el promedio es 1,

$$J(\delta,1)=\lim_n \sup \frac{2}{n(n+1)+4n+4}\,\frac{n(n+1)}{2}=1,$$

y por lo tanto la política δ es estrictamente mejor que cualquier política estacionaria.

∎

Puesto que en general es más sencillo trabajar con la familia de las políticas estacionarias que con el conjunto de todas las políticas, es importante tener condiciones que aseguren la existencia de una política óptima estacionaria. En particular, esto se logra bajo las siguientes hipótesis.

Hipótesis 4.6.

a) Existe una función $V^*: X \longrightarrow \mathbb{R}$ y una constante j^* tales que

$$V^*(x)+j^* = \sup_{a\in D(x)} \{ r(x,a) + \int V^*(y)q(dy/x,a) \}, \quad x\in X;$$

b) Existe una política estacionaria $f^*\in F$ que maximiza el lado derecho de la ecuación anterior, es decir,

$$V^*(x)+j^* = r(x,f^*(x)) + \int V^*(y)q(dy/x,f^*(x)) \}, \quad x\in X;$$

c) Para cada política $\delta\in\Delta$ y $x\in X$, se cumple

$$\lim_n \frac{1}{n} E_x^{\delta} V^*(x_n)=0.$$

Observación 4.7. Consideremos que se verifica la Hipótesis 4.6; luego, para cada $h_t\in H_t$ (t-historia admisible), $\delta\in\Delta$ y $x\in X$ se sigue de (a) y la propiedad de Markov (2.7)

$$E_x^{\delta}\left[V^*(x_{n+1}) \mid h_t, a_t\right] = \int V^*(y)q(dy/x_t,a_t)$$

$$= r(x_t,a_t)+ \int_- V^*(y)q(dy/x_t,a_t)-r(x_t,a_t)$$

$$\leq j^* + V^*(x_t)-r(x_t,a_t)$$

En consecuencia,

$$\sum_{t=0}^{n-1} \left\{V^*(x_{t+1})-E_x^{\delta}\left[V^*(x_{t+1}) \mid h_t,a_t\right]\right\} \geq \sum_{t=0}^{n-1} r(x_t,a_t)+ V^*(x_n)-V^*(x_0)-nj^*$$

y tomando esperanza (ya que la esperanza del lado izquierdo es cero)

$$nj^* \geq E_x^{\delta} \sum_{t=0}^{n-1} r(x_t,a_t)+ E_x^{\delta}\left[V^*(x_n)-V^*(x_0)\right].$$

Si dividimos entre n y tomamos límite en la desigualdad anterior, se sigue de (c) que

$$j^*\geq J(\delta,x).$$

Ahora si tomamos $\delta=f^*\in F$ que satisface (b), se cumple la

igualdad en el desarrollo anterior, y por lo tanto,

$$j^* = J(f^*, x) \ (\geq J(\delta, x) \text{ para cada } \delta \in \Delta \ ), \ x \in X.$$

Es decir, f^* es óptima.

Entonces, en este enfoque, la idea es proporcionar condiones para que se satisfaga la Hipótesis 4.6.

5. CONDICIONES DE RECURRENCIA.

En esta sección se introducen varias condiciones de recurrencia que serán usadas en los siguientes capítulos; en particular, nos permitirán junto con la Condición 4.1, verificar la Hipótesis 4.6.

Antes de introducir dichas condiciones se hace la siguiente observación referente a la notación usada.

Observación 5.1.

a) Cuando se usa una política estacionaria $f \in F$ (definición 2.4) el proceso controlado $\{x_t\}$ es una cadena de Markov con función de ganancia en un paso y probabilidad de transición $r(x, f(x))$ y $q(\cdot / x, f(x))$, respectivamente; las que denotaremos en forma breve por $r(x, f)$ y $q(\cdot / x, f)$.

b) En este caso, definimos

$$q^n(B/x, f) := P_x^f(x_n \in B) \ , \ B \in \mathcal{B}(X)$$

a la probabilidad de que en la n-ésima etapa el sistema se encuentre en B dado que inició en el estado x, la cual puede ser expresada recursivamente como

$$q^n(B/x, f) := \int q^{n-1}(B/y, f) q(dy/x, f) \text{ para } n = 1, 2, \ldots,$$

donde $q^0(B/x, a) := I_B(x)$.

c) Si δ es una política aleatorizada estacionaria (definición 2.2), definimos

$$q(\cdot/x,\delta) := \int q(\cdot/x,a)\delta(da/x)$$

$$q^n(\cdot/x,\delta) := \int q^{n-1}(\cdot/y,\delta)q(dy/x,\delta) \text{ para } n=1,2,\ldots.$$

Las condiciones que a continuación se introducen, 5.2 a 5.5, se les conoce como condiciones de recurrencia.

Condición 5.2. Existen $x^* \in X$ y $\alpha_0>0$ tales que

$$q(x^*/x,f) \geq \alpha_0, \text{ para todo } x\in X \text{ y } f\in F.$$

Condición 5.3. Existe $\alpha<1$ tal que

$$\| q(\cdot/x,f)-q(\cdot/x',f') \| \leq 2\alpha$$

para cada $x,x'\in X$ y $f,f'\in F$, donde $\|\cdot\|$ denota la norma en variación para la medida $\lambda(\cdot) = q(\cdot/x,f)-q(\cdot/x',f')$; ver A.2.2.

Condición 5.4. La familia $\{h_\beta\}$ es uniformemente acotada, donde

$$h_\beta(\cdot) := V_\beta(\cdot)-V_\beta(x^*), \quad 0<\beta<1,$$

con $x^*\in X$ fijo y V_β es la función definida en (3.8).

Condición 5.5 (Condición de Doeblin). Existe una medida finita γ sobre X, un natural n y $\varepsilon>0$ tales que para todo $B\in B(X)$

$$q^n(B/x,f) \leq 1-\varepsilon \qquad \text{si} \quad \gamma(B)\leq\varepsilon$$

para cada $x\in X$ y $f\in F$.

En los siguientes lemas se exhiben algunas de las relaciones entre las condiciones de recurrencia expuestas párrafos arriba.

Lema 5.6.

a) La Condición 5.2 implica la Condición 5.3;

b) bajo la condición 4.1, la Condición 5.3 implica la Condición 5.4;

c) la Condición 5.2 implica la Condición 5.5.

Para la demostración de esta lema se hará uso del siguiente resultado [Ueno, 1957]:

$$\| q^t(\cdot/x,f) - q^t(\cdot/y,f) \| \leq \frac{1}{2^{t-1}} \sup_{x,y} \| q(\cdot/x,f) - q(\cdot/y,f) \|^t \tag{5.7}$$

para cada $f \in F$ y $x, y \in X$.

Demostración.

a) La Condición 5.2 implica la condición 5.3.

Sea $\lambda(\cdot) = q(\cdot/x,f) - q(\cdot/x',f')$. Por el Teorema de Descomposición de Jordan-Hahn (ver A.2.3), existen conjuntos medibles ajenos X^+ y X^- tales que $X = X^+ \cup X^-$, y además

$$\| \lambda \| = q(X^+/x,f) - q(X^+/x',f) - q(X^-/x,f) + q(X^-/x',f)$$

$$\leq q(X^+/x,f) - \mu(X^+) - \mu(X^-) + q(X^-/x',f) ,$$

donde $\mu(\cdot) := \alpha_0 I_{(\cdot)}(x^*)$.

Por 5.2 se tiene $q(\cdot/x,f) \geq \mu(\cdot)$, entonces

$$\| \lambda \| \leq 2 - \mu(X) = 2\alpha ,$$

donde $\alpha = (2 - \mu(X))/2$; claramente $\alpha < 1$ y por lo tanto se verifica la Condición 5.3.

b) La Condición 5.3 implica la Condición 5.4.

De (5.7) y la condición 5.3 se sigue

(5.8) $$\| q^t(\cdot/x,f) - q^t(\cdot/x^*,f) \| \leq 2\alpha^t \quad \text{para } t=0,1,2,\ldots .$$

Por la Condición 4.1, existe $f_\beta \in F$ una política óptima con el criterio GD con factor de descuento β (ver (4.3)) ; entonces la función de valores óptimos en (3.8) es

$$V_\beta(x) = \sum_{t=0}^{\infty} \beta^t \int r(y,f_\beta) q^t(dy/x,f_\beta) ,$$

de donde se deduce

$$\begin{aligned} |h_\beta(x)| &= |V_\beta(x) - V_\beta(x^*)| \\ &\leq \sum_{t=0}^{\infty} \beta^t \left| \int r(y,f_\beta) \left\{ q^t(dy/x,f_\beta) - q^t(dy/x^*,f_\beta) \right\} \right| \\ &\leq \|r\| \cdot \sum_{t=0}^{\infty} \beta^t \, \|q^t(dy/x,f_\beta) - q^t(dy/x^*,f_\beta)\| \\ &\leq 2 \|r\| \cdot \sum_{t=0}^{\infty} \alpha^t \leq 2\|r\| / (1-\alpha) , \text{ por (5.7) y } \beta<1. \end{aligned}$$

De esto último se sigue que la familia $\{h_\beta\}$ es uniformemente acotada.

c) La Condición 5.2 implica la Condición 5.5.

Tomemos $\gamma(\cdot) = \dfrac{\mu(\cdot)}{\mu(X)} = \dfrac{\mu(\cdot)}{\alpha_0}$, donde $\mu(\cdot) = \alpha_0 I_{(\cdot)}(x^*)$; luego $\gamma(B) = \dfrac{\mu(B)}{\alpha_0} = I_B(x^*)$. Ahora, si $\gamma(B) < \alpha_0$, entonces para $\varepsilon=\alpha_0$ se cumple 5.5. En efecto, $\gamma(B)=I_B(x^*) < \alpha_0 \leq 1$, luego $\gamma(B)=0$ por definición de la función indicadora y por lo tanto $x^* \in B^c$; en consecuencia

$$q(B^c/x,f) \geq q(x^*/x,f) \geq \alpha_0,$$

por lo tanto $q(B/x,f) \leq 1-\alpha_0$, de donde se sigue la Condición 5.5.

■

Lema 5.9. La Condición 5.3 implica cada una de las siguientes condiciones:

a) para cada $f\in F$ existe una medida de probabilidad P_f sobre X tal que para todo $x\in X$

$$\|q^t(\cdot/x,f)-P_f(\cdot)\| \longrightarrow 0 \quad \text{si } t\longrightarrow\infty ; \tag{5.10}$$

b) para cada $f\in F$, la medida P_f es invariante con respecto a $q(\cdot/x,f)$, es decir,

$$P_f(B)=\int q(B/x,f)P_f(dx) \ , \ B\in\mathcal{B}(X); \tag{5.11}$$

c) para cada $f\in F$, la función $J(f,x)$ es constante y

$$J(f,x)=\int r(y,f)P_f(dy)=: j(f) \quad \text{para todo } x\in X.$$

Demostración.

a) Primero demostraremos que la sucesión $\left\{q^t(\cdot/x,f)\right\}$ es una sucesión de Cauchy en la norma de la variación.

$$\begin{aligned}
\|q^t(\cdot/x,f)-q^{t+s}(\cdot/x,f)\| &= 2\sup_B |q^t(B/x,f)-q^{t+s}(B/x,f)| \ , \ \text{A.2.6,}\\
&= 2\sup_B \left| \int \left(q^t(B/x,f)-q^t(B/y,f) \right) q^s(dy/x,f) \right|\\
&\le 2\sup_B \sup_y |q^t(B/x,f)-q^t(B/y,f)| \ , \ \text{A.2.5,}\\
&= \sup_y \|q^t(B/x,f)-q^t(B/y,f)\| \ , \ \text{A.2.6,}\\
&\le 2^{1-t} \sup \|q(\ /x,f)-q(\ /y,f)\|^t \ , \ \text{por (5.7)}\\
&\le 2\,\alpha^t \ , \ \text{por 5.3.}
\end{aligned}$$

Por lo tanto $\left\{q^t(\cdot/x,f)\right\}$ es una sucesión de Cauchy. Por otra parte, puesto que el espacio de las medidas con signo (finitas) es un espacio de Banach con la norma de la variación, $q^t(\cdot/x,f)$

converge a una medida P_f sobre X, que no depende de x; por lo tanto se cumple (5.10).

b) para $B \in \mathcal{B}(X)$

$$|P_f(B) - \int q(B/x,f)P_f(dx)|$$

$$\leq |P_f(B) - q^n(B/y,f)| + |\int q(B/x,f)q^{n-1}(dx/y,f) - \int q(B/x,f)P_f(dx)|$$

$$\leq |P_f(B) - q^n(B/y,f)| + |\int q(B/x,f)\left[q^{n-1}(dx/y,f) - P_f(dx)\right]|$$

$$\leq |P_f(B) - q^n(B/y,f)| + \|q^{n-1}(\ /y,f) - P_f(\ )\|;$$

entonces de (5.10) se sigue el resultado deseado.

c) para cada $f \in F$,

$$|E_x^f\, r(x_t,a_t) - \int r(y,f)P_f(dy)| = |\int r(y,f)q^t(dy/x,f) - \int r(y,f)P_f(dy)|$$

$$= |\int r(y,f)\left[q^t(dy/x,f) - P_f(dy)\right]|$$

$$= \|r\|\ \|q^t(\ /x,f) - P_f(\ )\|\ .$$

Entonces por (5.10) se sigue

$$\lim_t E_x^f\, r(x_t,a_t) = \int r(y,f)P_f(dy) =: j(f).$$

De esto último y la definición (3.1) se sigue la conclusión del lema.

■

Observación 5.12.

a) Un proceso controlado de Markov se dice <u>ergódico</u> si para cada $f \in F$, existe una medida de probabilidad invariante con respecto a $q(\cdot\ /x,f)$, esto es, si se satisface (5.11); se dice <u>fuertemente ergódico</u> si para cada $f \in F$, la medida de probabilidad invariante satisface (5.10).

Si el proceso es ergódico, la medida de probabilidad invariante satisface

$$P_f(B) = \lim_n \frac{1}{n} \sum_{t=0}^{n-1} q^t(B/x,f) \quad \text{para } B\in\mathcal{B}(X),\ f\in F \text{ y } x\in X.$$

Es decir, está dada como límite del 'promedio de las probabilidades de transición'. Y por la definición (3.1), $J(f,x)$ es el límite del 'promedio de las ganancias esperadas'; una relación entre estos promedios es dada por el Lema 5.9(c), es decir,

$$J(f,x) = \int r(y,f)P_f(dy),$$

lo que ilustra como los problemas en ganancia promedio están relacionados con las propiedades ergódicas de los procesos controlados.

b) En la Sección 4, Observación 4.7, se exhibió que bajo la Hipótesis 4.6 existe una solución al PCO con Ganancia Promedio. Entonces, bajo este enfoque el problema consiste en imponer condiciones para que se satisfaga dicha hipótesis.

Notese que si la función en 4.6(a) es acotada, entonces se satisface 4.6(c).

El objetivo en el Capítulo II es imponer condiciones (de recurrencia) para que existan j^* y V^*, acotada, tal que:

$$V^*(x)+j^* = \sup_{a\in D(x)} \{ r(x,a) + \int V^*(y)q(dy/x,a) \}, \quad x\in X;$$

es decir, para que se cumplan las condiciones a) y c) en la Hipótesis 4.6.

La existencia del maximizador f^* en Hipótesis 4.6(c), es asegurada por medio de teoremas de 'selección medible', bajo condiciones adecuadas de continuidad y compacidad (e.g., Condición 4.1), para esto ver el Capítulo III, Secciones 1 y 2.

c) Existen otras condiciones de recurrencia que implican a la Hipótesis 4.6(a), las que están relacionadas con las condiciones 5.2-5.5; ver [7],[8].

d) Gran parte de este material está apoyado en [7], los ejemplos de la Sección 4 se tomaron de [13].

II. ECUACION DE OPTIMALIDAD

0. Introducción

1. Ecuación de optimalidad

2. Condiciones de sm-contracción

3. Iteración de valores

4. Límite de problemas descontados

5. Comentarios.

0. INTRODUCCION.

Existen diferentes enfoques para resolver el problema de control óptimo bajo el Indice de Ganancia Promedio. Uno de ellos, quizás el más usado, es por medio de la Ecuación de Optimalidad (Definición 1.1), el cual asegura la existencia de una política óptima suponiendo que existe una solución de dicha ecuación (ver Observación I.4.7 y Capítulo III, Secciones 1 y 2).

Entonces el problema se reduce, bajo este enfoque, a imponer algún tipo de recurrencia que garantize una solución de la Ecuación de Optimalidad, lo cual es el objetivo de este capítulo.

1. ECUACION DE OPTIMALIDAD.

Sean $j^* \in \mathbb{R}$, $V^* \in B(X)$ tales que

$$(1.1) \qquad j^* + V^*(x) = \sup_{a \in D(x)} \left\{ r(x,a) + \int V^*(y)\, q(dy/x,a) \right\},$$

donde $B(X)$ es el espacio de las funciones definidas sobre X, medibles y acotadas.

A (1.1) se le conoce como Ecuación de Optimalidad Ganancia Promedio (EOGP), en este trabajo nos referiremos a ella como Ecuación de Optimalidad; y se dice que la pareja $\langle j^*, V^* \rangle$ es una solución de dicha ecuación.

Sea T el operador en B(X) definido por

$$(1.2) \qquad Tv(x) := \sup_{a \in D(x)} \left\{ r(x,a) + \int V^*(y)\, q(dy/x,a) \right\}, \quad x \in X.$$

Observación 1.3. Bajo la Condición I.4.1 se cumple que $Tv \in B(X)$ para toda $v \in B(X)$, es decir, T es un operador del espacio B(X) sobre si mismo. Al operador T se le conoce como Operador de Optimalidad (o de Programación Dinámica), ya que podemos escribir a (1.1) como

$$j^* + V^* = TV^*$$

Hipótesis 1.4. Existe un subespacio M(X) de B(X) tal que $Tv \in M(X)$ para toda $v \in M(X)$.

Es claro, por la Observación 1.3, que se puede tomar M(X)=B(X) si se verifica la Condición I.4.1.

Bajo la Hipótesis 1.4, para cada $v \in M(X)$ definimos a la función v^n por el siguiente procedimiento iterativo

$$(1.5) \qquad v^n := Tv^{n-1} \text{ para } n \geq 1 ,$$

donde $v^0 = v$.

Definición 1.6. Sea $u \in B(X)$; la semi-norma del rango de u se define por

$$sm(u) := \sup_x u(x) - \inf_x u(x).$$

Note que $sm(u)=0$ si y sólo si u es una constante.

Un operador $S:B(X) \longrightarrow B(X)$ es una sm-contracción o una contracción en la semi-norma del rango si para algún $\alpha \in [0,1)$ se tiene

$$sm(Su-Sv) \leq \alpha\, sm(u-v)$$

para toda $u,v \in B(X)$.

Definición 1.7. Una función $u \in B(X)$ se dice que es un sm-punto fijo de S, si

$$sm(Su-u)=0.$$

En la siguiente sección se dan condiciones para que el operador T en (1.2) sea de sm-contracción y posea un sm-punto fijo.

2. CONDICIONES DE SM-CONTRACCION.

En esta sección se asegura la existencia de una solución acotada $\{j^*,V^*\}$ de la Ecuación de Optimalidad como un sm-punto fijo del operador T.

Teorema 2.1. Bajo las condiciones I.4.1 y I.5.3 existe una solución acotada $\{j^*,V^*\}$ de la Ecuación de Optimalidad en Ganancia Promedio.

Demostración. Sean $x,x' \in X$, $f,f' \in F$ y λ la medida con signo definida por

$$\lambda(\cdot) := q(\cdot/x,f)-q(\cdot/x',f').$$

Por la Condición I.5.3 se sigue

$$\|\lambda\| = \|q(\cdot/x,f)-q(\cdot/x',f')\| \leq 2\alpha.$$

Por el Teorema de Descomposición de Jordan-Hahn (ver A.2.3(c)), existen conjuntos medibles ajenos X^+ y X^- cuya unión es X, tales que

$$\|\lambda\| = \lambda(X^+)-\lambda(X^-).$$

Además,

$$\lambda(X) = \lambda(X^+) + \lambda(X^-)$$

$$= q(X^+/x,f) - q(X^+/x',f') + q(X^-/x,f) - q(X^-/x',f')$$

$$= q(X/x,f) - q(X/x',f') = 0.$$

Entonces,

$$\|\lambda\| = 2\lambda(X^+) \leq 2\alpha ,$$

de donde $\lambda(X^+) \leq \alpha$.

Ahora, consideremos la integral

$$\int V(y)\lambda(dy) = \int_{X^+} V(y)\lambda(dy) + \int_{X^-} V(y)\lambda(dy)$$

$$\leq \int_{X^+} \sup_y V(y)\lambda(dy) + \int_{X^-} \inf_y V(y)\lambda(dy)$$

$$\leq \lambda(X^+)\, sm(V) + \inf_y V(y)\, \lambda(X)$$

$$\leq \alpha\, sm(V).$$

Por lo tanto,

(2.2) $$\int_X V(y)\lambda(dy) \leq \alpha\, sm(V).$$

Sea T el operador definido en (1.2) y u_1, u_2 funciones arbitrarias en $B(X)$; por la Proposición A.3.3 , existen selectores medibles g_1, g_2 tales que

$$Tu_i(x) = r(x,g_i(x)) + \int u_i(y)\, q(dy/x,g_i)$$

para $x \in X$, $i=1,2$.

Para $i \neq j$, se sigue

$$Tu_i(x) \geq r(x,g_j(x)) + \int u_i(y)\, q(dy/x,g_j) \quad \text{para } x \in X;$$

entonces,

$$(Tu_1-Tu_2)(x)-(Tu_1-Tu_2)(x') \leq$$

$$\leq \int u_1(y)\left[q(dy/x,g_1)-q(dy/x',g_2)\right]-\int u_2(y)\left[q(dy/x,g_1)-q(dy/x',g_2)\right]$$

$$\leq \int [u_1(y)-u_2(y)]\lambda(dy)$$

$$\leq \alpha\, sm(u_1-u_2)$$

donde λ es la medida con signo en (2.2), tomando $f=g_1$ y $f'=g_2$.

De la desigualdad anterior,

$$(Tu_1-Tu_2)(x)-(Tu_1-Tu_2)(x') \leq \alpha\, sm(u_1-u_2).$$

Como x y x' son arbitrarios, tomando supremo sobre x e ínfimo sobre x' en la desigualdad anterior, se sigue

$$sm(Tu_1-Tu_2) \leq \alpha\, sm(u_1-u_2) ,$$

para cada u_1 y u_2 en $B(X)$.

Por lo tanto, T es un operador de sm-contracción y, en consecuencia, tiene un sm-punto fijo; es decir, existe una función $V^*\in B(X)$ tal que $sm(TV^*-V^*)=0$ de manera que $TV^*-V^*=j^*$, donde j^* es una constante. Para ser más explícitos, denotemos por

$$B_0 := \{\, v\in B(X) : v(x^*)=0 \,\} ,$$

donde $x^*\in X$ es fijo.

Para $v\in B_0$, se define el operador T_0 de la siguiente manera

(2.3) $$T_0v(x) := Tv(x)-Tv(x^*).$$

Claramente T_0 es un operador de B_0 en él mismo y además es de sm-contracción, esto es,

$$sm(T_0u-T_0v)=sm(Tu-Tv)\leq \alpha\, sm(u-v) ,$$

para cualquier $u,v\in B_0$.

Observación 2.4. Si $v \in B_0$, $\inf_x v(x) \leq 0 \leq \sup_x v(x)$ en consecuencia la norma del supremo $\|v\|$ y la semi-norma del rango $sm(v)$ está relacionada de la siguiente manera

$$\|v\| \leq sm(v) \leq 2\|v\|$$

(la segunda desigualdad se sigue de $|v| = v^+ + v^-$).

Entonces la semi-norma del rango es una norma en B_0 equivalente a la norma del supremo, y por lo tanto $(B_0, sm(\cdot))$ es un espacio métrico completo. De aquí, se sigue por la propiedad de contracción de T_0 y el Teorema de Punto Fijo de Banach (ver Ross 1970), que existe una única función $V^* \in B_0$ tal que $T_0 V^* = V^*$. Por la definición de T_0 se sigue

$$T_0 V^* = TV^* - j^* = V^*,$$

con la constante $j^* := TV^*(x^*)$, luego

$$TV^* = V^* + j^*$$

y por la Observación 1.3 se sigue la conclusión del teorema. ∎

Observación 2.5. En la demostración del Teorema de Punto Fijo de Banach se verifican los siguientes pasos:

a) $T_0^n v$ es una sucesión de Cauchy para toda $v \in B_0$;

b) T_0 es un operador continuo;

(a) y b) se siguen de la propiedad de contracción)

c) por la completez del espacio $(B_0, sm(\ ))$ y a) existe $V^* \in B_0$ tal que $V^* = \lim_n T_0^n v$ (uniformemente).

Por b) y c) se sigue $T_0 V^* = V^*$.

Observemos que (c) nos provee de un procedimiento iterativo para aproximar uniformmente a V^*.

3 ITERACION DE VALORES.

En esta sección se usará el método de aproximaciones sucesivas o iteración de valores para aproximar uniformemente a $\langle j^*, V^* \rangle$, solución de la Ecuación de Optimalidad.

Observemos de la sección anterior que:

a) por la definición del operador T_o (ver 2.3) y la observación 2.5(c)

$$V^* = \lim_n \{ T^n v - Tv(x^*) \} , \quad v \in B_o.$$

b) por el teorema 2.1

$$TV^* = V^* + j^*,$$

donde T es el operador definido en (1.2). Iterando T se obtiene

$$T^n V^* = T^{(n-1)} V^* + j^*,$$

luego

$$T^n V^* - T^{(n-1)} V^* = j^*.$$

En consecuencia, lo anterior sugiere 'buscar' $\langle j^*, V^* \rangle$ como límites de la forma,

$$V^*(x) = \lim \{ V^n(x) - V^n(x^*) \}$$

(3.1)

$$j^* = \lim \{ V^n(x^*) - V^{n-1}(x^*) \},$$

donde $x^* \in X$, arbitrario pero fijo y $V^n = TV^{n-1}$, ver (1.5).

El objetivo de esta sección es asegurar la existencia de una solución de la ecuación de Optimalidad como límites de la forma (3.1); ver Teorema 3.9.

Por otra parte, es posible obtener propiedades adicionales (de continuidad) para la función V^*, modificando las condiciones de continuidad y compacidad sobre el MCM (Condición I.4.1) por la siguiente condición, bajo la cual se desarrolla esta sección.

105958

Condición 3.2.

a) $D(x)$ es un conjunto compacto para cada $x \in X$ y la función D es semi-continua superiormente, es decir, para cada conjunto abierto $G \subset A$,. el conjunto $\{x \in X: D(x) \subset G\}$ es un conjunto abierto en X;

b) r es semi-continua superiormente;

c) $(x,a) \longrightarrow \int u(y)q(dy/x,a)$ es una función semi-continua superiormente para toda $u \in C_s(X)$, donde $C_s(X)$ es el espacio de las funciones acotadas y semi-continuas superiormente en X.

Observación 3.3. Bajo la Condición 3.2, el subespacio $C_s(X) \subset B(X)$ satisface la Hipótesis 1.4 [ver A.2.11(c)], esto es el operador T en (1.2) está definido de $C_s(X)$ en si mismo. Entonces, es posible estudiar las funciones iteradas $V^n \in C_s(X)$ como en (1.5):

$$V^n = T^n(V^0) \ (= TV^{n-1}), \quad V^0 \in C_s(X).$$

Antes de citar el teorema principal de esta sección (Teorema 3.9), presentaremos de nuevo la Condición I.5.3 para mejor referencia, ya que tanto el Teorema 3.9 como las siguientes dos proposiciones se verifican bajo la misma.

Condición I.5.3. Existe $\alpha < 1$ tal que

$$\| q(\ /x,f) - q(\ /x',f') \| \leq 2\alpha$$

para cada $x, x' \in X$ y $f, f' \in F$, donde $\| \ \|$ denota la norma en variación para medidas finitas con signo.

Proposición 3.4. Bajo las condiciones 3.2 y I.5.3

Para cada $u_1, u_2 \in C_s(X)$ y $n \in \mathbb{N}$.

$$sm(T^n u_1 - T^n u_2) \leq \alpha^n \, sm(u_1 - u_2), \quad \alpha \in [0,1).$$

Demostración.

De la misma forma que en la demostración del Teorema 2.1 se sigue que

$$sm(T^n u_1 - T^n u_2) \leq \alpha \, sm(T^{n-1} u_1 - T^{n-1} u_2) \leq \ldots \leq \alpha^n \, sm(u_1 - u_2)$$

$u_1, u_2 \in C_s(X)$.

■

Sea $x^* \in X$ fijo, definimos

(3.5) $$V_n(x) = V^n(x) - V^n(x^*) \quad \text{para } x \in X \text{ y } n \in \mathbb{N} \cup \{0\},$$

donde V^n es como en la Observación 3.4.

Proposición 3.6. Bajo las condiciones I.5.3 y 3.2, existe $V^* = \lim_n V_n$, tal que $V^* \in C_s(X)$ y

(3.7) $$\|V^* - V_n\| \leq \alpha^n C,$$

donde $C = sm(V^1 - V^0) \sum_{k=0}^{\infty} \alpha^n$

Demostración.

Tomemos $k \in \mathbb{N}$, haciendo uso repetido del Teorema 3.4 se sigue

$$sm(V_{n+k} - V_n) = sm(T^n V^k - T^n V^k(x^*) - TV^0 - T^n V^0(x^*))$$

$$= sm(T^n V^k - T^n V^0) \leq \alpha^n \, sm(V^k - V^0)$$

y

$$sm(V^k - V^0) \leq \sum_{n=0}^{\infty} sm(T^n V^1 - T^n V^0) \leq sm(V^1 - V^0) \sum_{k=0}^{\infty} \alpha^n =: C.$$

De (3.5) se sigue

$$\inf_x (V_{n+k} - V_n) \leq 0 \leq \sup_x (V_{n+k} - V_n)$$

lo cual implica

(3.8) $$\|V_{n+k}-V_n\| \leq sm\,(V_{n+k}-V_n) \leq \alpha^n\, C\ ,$$

donde $\|\cdot\|$ es la norma del supremo. En consecuencia existe $V^*=\lim_n V_n$; puesto que $\{V_n\}\subset C_s(X)$ y la convergencia es uniforme, entonces $V^*\in C_s(X)$. Haciendo $k \longrightarrow \infty$ en (3.8) se obtiene (3.7).

■

Teorema 3.9. Bajo las condiciones I.5.3 y 3.2 existe una solución $\{j^*,V^*\}$ de la Ecuación de Optimalidad, donde $V^*\in C_s(X)$, como límites de la forma (3.1).

Demostración.

Sea $g_n:=V^n(x^*)-V^{n-1}(x^*)$ para $n\in\mathbb{N}$, donde $x^*\in X$ fijo como en (3.5).

De la Proposición 3.4 tenemos

$$sm(V^n-V^{n-1}) \leq \alpha^{n-1}\, sm\,(V^1-V^0)\ ,$$

entonces

$$\lim_n sm(V^n-V^{n-1})=0$$

y por la definición de semi-norma (definición 1.6) se deduce que existe una constante, que denotaremos j^*, tal que

$$j^*=\lim_n [V^n-V^{n-1}]=\lim_n g_n.$$

Para cada $x\in X$ y $n\in\mathbb{N}$, definimos las funciones

(3.10) $$LV_n(x,a):=r(x,a)+\int V_n(y)q(y/x,a) \qquad a\in D(x).$$

Bajo las condiciones del teorema, $\{LV_n(x,\cdot)\}$ es una sucesión de funciones semi-continuas superiormente que convergen uniformemente a $LV^*(x,\cdot)$. Es decir,

$$|LV_n(x,a)-LV^*(x,a)| = \left|\int [V_n(y)-V^*(y)]q(dy/x,a)\right| \leq \|V_n-V^*\| \leq \alpha^n C,$$

donde $V^*=\lim_n V_n$; la última desigualdad se sigue de (3.7).

Además, las sucesiones $\langle V_n\rangle$ y $\langle g_n\rangle$ satisfacen la siguiente relación

$$g_{n+1}+V_{n+1}=TV^n \quad \text{para } n\in\mathbb{N}.$$

Tomando límite cuando $n \longrightarrow \infty$ en la igualdad anterior se sigue

$$\lim_n \{g_{n+1}+V_{n+1}\} = \lim_n TV^n = T(\lim_n V^n),$$

la última desigualdad se sigue de la continuidad del operador T (lo cual es una consecuencia de la propiedad de contracción, ver Proposición 3.4), o bién por la Proposición A.1.4(c).

En consecuencia,

$$j^*+V^*=TV^*;$$

luego, $\{j^*,V^*\}$ es una solución acotada de la Ecuación de Optimalidad, donde $V^*\in C_s(X)$.

■

4. LIMITES DE PROBLEMAS DESCONTADOS.

Se obtiene una solución $\{j^*,V^*\}$ de la Ecuación de Optimalidad Ganancia Promedio como 'límites de problemas descontados', cuando el factor de descuento se aproxima a uno.

Para ser más explícitos, recordemos que $V_\beta(\delta,x)$, $V_\beta(x)$ representan el índice de funcionamiento y la función de valores óptimos, respectivamente, en los problemas de ganancia descontada (con factor de descuento β), las cuales están definidas por

$$V_\beta(\delta,x) := E_x^\delta \sum_{k=0}^{\infty} \beta^k r(x_k,a_k)$$

y

$$V_\beta(x) := \sup_{\delta\in\Delta} V_\beta(\delta,x) \quad , \ \beta\in(0,1).$$

Obsevación 4.1. Recordemos que bajo la Condición I.4.1,

a) $V_\beta(x)$ satisface la Ecuación de Optimalidad en Ganancia β-Descontada

$$V_\beta(x) = \sup_{a\in D(x)} \{r(x,a)+\beta\int V_\beta(y)q(dy/x,a)\} \quad \text{para todo } x\in X;$$

b) una política $f_\beta\in F$ es β-óptima si y sólo si $f_\beta(x)\in D(x)$ maximiza el lado derecho de la ecuación anterior para todo $x\in X$, esto es,

$$V_\beta(x) = r(x,f_\beta(x)) + \beta\int V_\beta(y)q(dy/x,f_\beta(x)) \quad \text{para todo } x\in X;$$

ver I.4.2 y I.4.3, respectivamente.

Ahora si consideramos $h_\beta(x)=V_\beta(x)-V_\beta(x^*)$ y $(1-\beta)V_\beta(x^*)$, con $x^*\in X$ fijo; entonces por el inciso a) tenemos

$$(4.2) \qquad (1-\beta)V_\beta(x^*)+h_\beta(x) = \sup_{a\in D(x)} \{r(x,a) + \beta\int h_\beta(y)q(dy/x,a)\}.$$

Si existe una sucesión de factores de descuento $(\beta_n)\uparrow 1$ tal que se verifican las relaciones

$$(4.3) \qquad j^* = \lim_n (1-\beta_n)V_{\beta_n}(x^*) \ , \quad \bar{V}^*(x) = \lim_n h_{\beta_n}(x),$$

$$(4.4) \qquad \lim_n \sup_{a\in D(x)} \{r(x,a)+\beta_n\int h_{\beta_n}(y)q(dy/x,a)\}$$

$$= \sup_{a\in D(x)} \{r(x,a) + \int V^*(y)q(dy/x,a)\}$$

entonces, tomando límite en (4.2) a través de la sucesión $(\beta_n)\uparrow 1$, se cumple

y

$$V_\beta(x) := \sup_{\delta\in\Delta} V_\beta(\delta,x) \ , \ \beta\in(0,1).$$

Obsevación 4.1. Recordemos que bajo la Condición I.4.1,

a) $V_\beta(x)$ satisface la Ecuación de Optimalidad en Ganancia β-Descontada

$$V_\beta(x)= \sup_{a\in D(x)} \{r(x,a)+\beta\int V_\beta(y)q(dy/x,a)\} \quad \text{para todo } x\in X;$$

b) una política $f_\beta\in F$ es β-óptima si y sólo si $f_\beta(x)\in D(x)$ maximiza el lado derecho de la ecuación anterior para todo $x\in X$, esto es,

$$V_\beta(x)= r(x,f_\beta(x)) +\beta\int V_\beta(y)q(dy/x,f_\beta(x)) \quad \text{para todo } x\in X;$$

ver I.4.2 y I.4.3, respectivamente.

Ahora si consideramos $h_\beta(x)=V_\beta(x)-V_\beta(x^*)$ y $(1-\beta)V_\beta(x^*)$, con $x^*\in X$ fijo; entonces por el inciso a) tenemos

$$(4.2) \qquad (1-\beta)V_\beta(x^*)+h_\beta(x)= \sup_{a\in D(x)} \{r(x;a) + \beta \int h_\beta(y)q(dy/x,a)\}.$$

Si existe una sucesión de factores de descuento $(\beta_n)\uparrow 1$ tal que se verifican las relaciones

$$(4.3) \qquad j^*= \lim_n (1-\beta_n)V_{\beta_n}(x^*) \ , \quad V^*(x)= \lim h_{\beta_n}(x),$$

$$(4.4) \qquad \lim_n \sup_{a\in D(x)} \{r(x,a)+\beta_n\int h_{\beta_n}(y)q(dy/x,a)\}$$

$$= \sup_{a\in D(x)} \{r(x,a) + \int V^*(y)q(dy/x,a)\}$$

entonces, tomando límite en (4.2) a través de la sucesión $(\beta_n)\uparrow 1$, se cumple

$$j^*+V^*(x)= \sup_{a\in D(x)} \{r(x,a) + \int V^*(y)q(dy/x,a)\} ;$$

es decir, (j^*,V^*) es solución de la Ecuación Optimalidad Ganancia Promedio.

Así, el objetivo de esta sección es imponer condiciones adecuadas sobre el modelo de control Markoviano, de manera que se verifiquen (4.3) y (4.4), y entoncs obtener una solución de EOGP; ver Teorema 4.9.

Las condiciones 'adecuadas' a las que se hace alusión son I.4.1 y I.5.4, las que presentamos nuevamente para mayor facilidad en la referencia.

Condición I.4.1. Para cada $x\in X$,

a) $D(x)$ es un subconjunto compacto;

b) $r(x,\cdot)$ es semi-continua superiormente sobre $D(x)$;

c) $\int h(y)q(dy/x,a)$ es una función semi-continua superiormente en $a\in D(x)$ para toda $h\in B(X)$.

Condición I.5.4. La familia de funciones $\{h_\beta : \beta\in(0,1)\}$ es uniformmente acotada, donde

$$h_\beta(\cdot) := V_\beta(\cdot)-V_\beta(x^*) \qquad \text{con } x^*\in X \text{ fijo.}$$

También se hace uso de la siguiente condición.

Condición 4.5.

a) para cada $(x,a)\in K$, la distribución $q(\ /x,a)$ tiene función de densidad $p(x,a,\cdot)$ con respecto a una medida λ en X, es decir, para todo $B\in B(X)$ se tiene

$$q(B/x,a)= \int p(x,a,x')\lambda(dx');$$

$p(x,a,\cdot) \leq \bar{p}(x,\cdot)$, $\bar{p}(x,\cdot)$ λ-integrable

b) para cada $(x,x') \in XX$, $a \longrightarrow p(x,a,x')$ es continua en $D(x)$.

Antes de seguir adelante, es conveniente presentar algunas consecuencias de las condiciones anteriores.

Observación 4.6.

a) Bajo la Condición 4.5 se cumple lo siguiente

$$\lim_{a \to a_0} \int |p(x,a,x') - p(x,a_0,x')| \lambda(dx') =$$

$$= \int \lim_{a \to a_0} |p(x,a,x') - p(x,a_0,x')| \lambda(dx') = 0$$

la primera desigualdad se cumple por el Teorema de Convergencia Dominada, de donde claramente se cumple el inciso (c) de I.4.1.

b) Sea $\{h_\beta : \beta \in (0,1)\}$ como en la Condición I.5.4. Entonces la familia de funciones

$$a \longrightarrow \int p(x,a,x') h_\beta(x') \lambda(dx') \quad , \quad 0 < \beta < 1$$

es equicontinua.

En efecto, dado $\varepsilon > 0$ existe δ tal que:

$$\left| \int p(x,a,x') h_\beta(x') \lambda(dx') - \int p(x,a_0,x') h_\beta(x') \lambda(dx') \right|$$

$$\leq \int |p(x,a,x') - p(x,a_0,x')| \cdot h_\beta(x') | \lambda(dx') \leq \varepsilon$$

para

$$\int |p(x,a,x') - p(x,a_0,x')| \lambda(dx') \leq \varepsilon / k \quad \text{(por el inciso a))},$$

si $d(a,a_0) < \delta$, donde d es la métrica en A y k es la cota uniforme para la familia $\{h_\beta : \beta \in (0,1)\}$.

c) Bajo la condición I.4.1 , sea $f_\beta \in \mathbb{F}$ una política β-óptima; entonces satisface

$$V_\beta(x) = r(x,f_\beta) + \beta \int V_\beta(y) q(dy/x,f_\beta) \quad \text{para todo } x \in X;$$

ver I.4.2 u Observación 4.1(b).

Consideremos de nuevo a las familias $(1-\beta_n)V_{\beta_n}(x^*)$ y $h_{\beta_n}(\cdot)$; luego se sigue de un cálculo directo

$$h_{\beta_n}(x) + (1-\beta_n)V_{\beta_n}(x^*) = r(x,f_{\beta_n}) + \beta_n \int h_{\beta_n}(y) q(dy/x,f_{\beta_n});$$

y para $n \neq m$

$$h_{\beta_m}(x) + (1-\beta_m)V_{\beta_m}(x^*) \geq \{r(x,f_{\beta_n}) + \beta_m \int h_{\beta_m}(y) q(dy/x,f_{\beta_n})\};$$

y por la Condición 4.5(a) las últimas expresiones se pueden reescribir de la forma

(4.7) $$h_{\beta_n}(x) + (1-\beta_n)V_{\beta_n}(x^*) = r(x,f_{\beta_n}) + \beta_n \int h_{\beta_n}(y) p(x,f_{\beta_n},y)\lambda(dy)$$

(4.8) $$h_{\beta_m}(x) + (1-\beta_m)V_{\beta_m}(x^*) \geq r(x,f_{\beta_n}) + \beta_m \int h_{\beta_m}(y) p(x,f_{\beta_n},y)\lambda(dy)$$

las cuales se usarán en la demostración del siguiente teorema.

Teorema 4.9. Bajo las Condiciones I.4.1, I.5.4 y 4.5 existe una solución $\langle j^*, V^* \rangle$ de la EOGP que satisface (4.3), esto es,

$$j^* = \lim_{\beta_n \uparrow 1} (1-\beta_n)V_{\beta_n}(x^*)$$

y

$$V^*(x) = \lim_{\beta_n \uparrow 1} h_{\beta_n}(x).$$

Demostración.

Consideremos la sucesión $(1-\beta_n)V_{\beta_n}(x^*)$, $\beta_n \uparrow 1$, la cual es uniformemente acotada por M (ver I.3.4); entonces existe una

subsucesión convergente la cual denotaremos como la sucesión original, esto es,

$$j^* := \lim_n (1-\beta_n) V_{\beta_n}(x^*). \tag{4.10}$$

Ahora consideremos la familia $\{h_\beta : \beta\in(0,1)\}$ introducida en la Condición I.5.4. Entonces $h_\beta \in L^\infty(\lambda)$, y además, $L^\infty(\lambda)$ es dual del espacio de Banach $L^1(\lambda)$ (ver Royden, 1968, p.249) el cual es separable, y por lo tanto, la bola unitaria de $L^\infty(\lambda)$ es secuencialmente compacta en la topología débil de $\sigma(L^\infty(\lambda), L^1(\lambda))$. Esto implica que existe $u\in L^\infty(\lambda)$ tal que

$$u(x) = \lim_{\beta_n \uparrow 1} h_{\beta_n}(x) \quad \text{en la topología } \sigma(L^\infty(\lambda), L^1(\lambda)).$$

Es decir, para cada $(x,a)\in K$, se sigue

$$\lim_n \int h_{\beta_n}(y)p(x,a,y)\lambda(dy) = \int p(x,a,y)u(y)\lambda(dy). \tag{4.11}$$

Ahora para cada $x\in X$, la familia de funciones

$$a \longrightarrow \int h_{\beta_n}(y)p(x,a,y)\lambda(dy)$$

es equicontinua (ver 4.6(b)), y por el Teorema de Ascoli (ver Royden,1968, p. 179), la función

$$a \longrightarrow \int p(x,a,y)u(y)\lambda(dy)$$

es continua, y la convergencia en (4.11) es uniforme en el conjunto compacto $D(x)$.

Así, para $\varepsilon>0$ existe n_0 tal que

$(1-\beta_{n_0}) \le \varepsilon/M$, para $n\ge n_0$ y cada $a\in D(x)$;

$$\left|\int h_{\beta_n}(y)p(x,a,y)\lambda(dy) - \int p(x,a,y)u(y)\lambda(dy)\right| \le \varepsilon \quad \text{por (4.11) y}$$

$$\left|j^* - \lim_{\beta_n \uparrow 1} (1-\beta_n)V_{\beta_n}(x^*)\right| \le \varepsilon \quad \text{por (4.10)}.$$

Ahora se mostrará que la sucesión $\{h_{\beta_n}(x)\}$ es Cauchy. Para esto consideremos la diferencia entre las expresiones en (4.7) y (4.8) para $n \neq m$ mayores que n_0,

$$h_{\beta_n}(x) - h_{\beta_m}(x) + (1-\beta_n)V_{\beta_n}(x^*) - (1-\beta_m)V_{\beta_m}(x^*) \leq$$

$$\beta_n \int h_{\beta_n}(y)p(x, f_{\beta_n}, y)\lambda(dy) - \beta_m \int h_{\beta_m}(y)p(x, f_{\beta_m}, y)\lambda(dy)$$

entonces

$$h_{\beta_n}(x) - h_{\beta_m}(x) \leq (1-\beta_m)V_{\beta_m}(x^*) - (1-\beta_n)V_{\beta_n}(x^*) +$$

$$\beta_n \int [h_{\beta_n}(y) - u(y)]p(x, f_{\beta_n}, y)\lambda(dy) - \beta_m \int [h_{\beta_m}(y) - u(y)]p(x, f_{\beta_m}, y)\lambda(dy) -$$

$$(\beta_n - \beta_m) \int u(y)p(x, f_{\beta_n}, y)\lambda(dy) \leq 4\varepsilon.$$

En consecuencia $\{h_{\beta_n}(x)\}$ es una sucesión de Cauchy para $x \in X$, luego existe $V^* \in L^\infty(\lambda)$ tal que

$$V^*(x) = \lim_n h_{\beta_n}(x).$$

Por lo tanto

$$\int V^*(y)p(x,a,y)\lambda(dy) = \lim_n \int h_{\beta_n}(y)p(x,a,y)\lambda(dy) = \int u(y)p(x,a,y)\lambda(dy),$$

donde la primera igualdad es por el Teorema de Convergencia Dominada, y la segunda igualdad por la convergencia en la topología débil $\sigma(L^\infty(\lambda), L^1(\lambda))$.

En términos de las probabilidades de transición q, se cumple

$$\int V^*(y)q(dy/x,a) = \lim_n \int h_{\beta_n}(y)q(dy/x,a).$$

Considerese (4.2), con factor de descuento β_n, $n > n_0$. Esto es,

$$(1-\beta_n)V_{\beta_n}(x^*) + h_{\beta_n}(x) = \sup_{a \in D(x)} \{r(x,a) + \beta_n \int h_{\beta_n}(y)q(dy/x,a)\}.$$

Ahora tomando límite cuando $n \longrightarrow \infty$ ($\beta_n \uparrow 1$), en la igualdad anterior

$$j^*+V^*(x) = \lim_n \sup_{a\in D(x)} \{r(x,a) + \beta_n \int h_{\beta_n}(y)q(dy/x,a)\}$$

$$\geq \sup_{a\in D(x)} \lim_n \{r(x,a) + \beta_n \int h_{\beta_n}(y)q(dy/x,a)\}$$

$$= \sup_{a\in D(x)} \{r(x,a) + \int V^*(y)q(dy/x,a)\}$$

la desigualdad se sigue de Hinderer, Lema 3.4, p.18.

Para demostrar la desigualdad contraria, se escribe 4.7 en términos de la probabilidad de transición q, y se toma el límite cuando $n \longrightarrow \infty$

$$j^*+V^*(x) = \lim_n \{r(x,f_{\beta_n}) + \beta_n \int h_{\beta_n}(y)q(dy/x,f_{\beta_n})\}.$$

Ahora para $n\geq n_0$ y $\varepsilon>0$ se cumple:

$$r(x,f_{\beta_n})+\beta_n\int h_{\beta_n}(y)q(dy/x,f_{\beta_n})\leq r(x,f_{\beta_n})+\beta_n\int V^*(y)q(dy/x,f_{\beta_n})+\varepsilon$$

entonces

$$j^*+V^*(x)\leq \lim_n \left[r(x,f_{\beta_n})+\beta_n\int V^*(y)q(dy/x,f_{\beta_n})\right] + \varepsilon$$

$$\leq \lim_n \sup_{a\in D(x)} \{r(x,a) + \beta_n \int V^*(y)q(dy/x,a)\} + \varepsilon$$

$$\leq \sup_{a\in D(x)} \{r(x,a) + \int V^*(y)q(dy/x,a)\} + \varepsilon$$

como ε es arbitrario se sigue

$$j^*+V^*(x) \leq \sup_{a\in D(x)} \{r(x,a) + \int V^*(y)q(dy/x,a)\}\ ,\ x\in X$$

y de las dos desigualdades se sigue

$$j^*+V^*(x) = \sup_{a\in D(x)} \{r(x,a) + \int V^*(y)q(dy/x,a)\}\ ,\ x\in X.$$

5. COMENTARIOS.

5.1. En la Sección 2, el paso clave es la propiedad de contracción del operador de programación dinámica T, obtenida bajo condiciones de ergodicidad impuestas a la ley de transición **q**.

5.2. En la Sección 3, se aproxima uniformemente a la solución de la Ecuación de Optimalidad, por medio de las funciones $V^n(x^*)-V^{n-1}(x^*)$ y $V^n(\cdot)-V^n(x^*)$, introducidas por White(1963). Para otras aproximaciones, basada también en iteración de valores ver Hernández-Lerma, O. (1989, pag. 62).

Además, bajo este enfoque es posible demostrar la existencia de una política óptima $f^*\in F$, donde $f^*(x)$ es un punto de acumulación de alguna sucesión $\{f_n(x)\}$ tal que $f_n(x)$ satisface la ecuación de progración dinámica en horizonte finito, ver Capítulo III.2.

5.3. Existen otros conjuntos de condiciones para asegurar una solución de la Ecuación de Optimalidad, por ejemplo, ver Rieder (1979) Sección 5. Rieder supone la Condición I.5.4 y que el conjunto de acciones admisibles para cada estado $x\in X$, es todo el conjunto $A=D(x)$. Supone, además; 3.3(c) y que $|q(\ /x,a)-q(\ /x',a)| \longrightarrow 0$ cuando $x \longrightarrow x'$ uniformemente en $a\in D(x)$, lo cual se satisface bajo la Condición 4.1. Rieder usa el enfoque de la Sección 4, es decir, como límite de problemas descontados.

El material de este capítulo está apoyado en [5],[6],[7],[15].

III.- POLITICAS OPTIMAS

0. Introducción.

1. Maximizador de la Ecuación de Optimalidad.

2. Límite de problemas con horizonte finito.

3. Hipótesis de Doeblin.

0. INTRODUCCION.

El objetivo de este capítulo es demostrar, bajo ciertas condiciones, la existencia de políticas óptimas para el PCO con Ganacia Promedio. En el Capítulo II, Secciones 2,3 y 4, se demostró la existencia de una solución acotada de la Ecuación de Optimalidad considerando diferentes puntos de vista.

Ahora, usando el enfoque de la Ecuación de Optimalidad, Observación I.4.7, y las Secciones II.2 y II.3 se resolverá el Problema de Control Optimo.

En la primera sección se demuestra la existencia de una política estacionaria óptima como maximizador de la Ecuación de Optimalidad en Ganancia Promedio (EOGP). En la Sección 2, se demuestra que el maximizador de la Ecuación de Optimalidad (política estacionaria óptima) es el 'límite de políticas óptimas' de problemas de programación dinámica con horizonte finito. Finalmente, en la Sección 3 bajo la Condición de Doeblin y la restricción de que los espacios de estados y acciones sean compactos, se muestra la existencia de una política estacionaria óptima.

1. MAXIMIZADOR DE LA ECUACION DE OPTIMALIDAD.

Uno de los enfoques para resolver el PCO es el señalado en la Observación I.4.7 (enfoque de la Ecuación de Optimalidad), según el cual, se tiene una solución al PCO, si se satisface la Hipótesis I.4.6 que reproducimos a continuación:

Hipótesis I.4.6.

a) Existe $V^*: X \longrightarrow \mathbb{R}$ y una constante j^* tales que

$$j^* + V^*(x) = \sup_{a \in D(x)} \left\{ r(x,a) + \int V^*(y) q(dy/x,a) \right\}, \quad x \in X;$$

b) existe $f^* \in F$ que maximiza el lado derecho de la ecuación anterior, es decir,

$$j^* + V^*(x) = r(x,f^*(x)) + \int V^*(y) q(dy/x,f^*(x)) \quad x \in X;$$

c) para cada política $\delta \in \Delta$ y $x \in X$, se cumple

$$\lim_n \frac{1}{n} E_x^{\delta} V^*(x_n) = 0.$$

Por otra parte, en el Capítulo II se proporcionan condiciones bajo las cuales se satisface 4.6(a) y (c); mientras que 4.6(b) se sigue, bajo condiciones de continuidad y compacidad, de resultados de selección medible.

En el siguiente teorema se resume lo anterior.

Teorema 1.1. Bajo las Condiciones I.4.1 y I.5.3 se satisface la Hipótesis I.4.6 y, por lo tanto, existe una política estacionaria óptima.

Demostración.

Por el teorema II.2.1 existe (j^*, V^*), $V^* \in B(X)$, tal que se satisfacen I.4.6(a) y por ser V^* acotada se sigue también I.4.6(c). Del Teorema de Selección Medible A.3.3, existe $f^* \in F$ que maximiza el

lado derecho de la Ecuación de Optimalidad, por lo que se cumple I.4.6(b). Entonces, por la Observación I.4.7 se tiene que f^* es una política estacionaria óptima y que $J(f^*,x)=j^*$ para todo $x\in X$. ∎

2. LIMITE DE PROBLEMAS CON HORIZONTE FINITO.

En esta sección, usando los resultados de II.3 (Iteración de Valores), se demuestra la existencia de una política $f^*\in F$ óptima.

De nuevo se hará uso del enfoque de la Ecuación de Optimalidad (Observación I.4.7), por lo que el objetivo es verificar la Hipóteis I.4.6. La difererencia con la sección anterior, radica en que el procedimiento de iteración de valores (además de verificar la Hipótesis I.4.6) nos proporciona aproximaciones a la función de valores óptimos J^* y la política óptima $f^*\in F$ se obtiene como 'límite' de una sucesión de selectores $\{f_n(x)\}$, tales que para cada n, f_n maximiza a Tv^{n-1}.

Antes de seguir adelante, recordaremos algunas de las funciones introducidas en II.3.

$$V^n(x)=TV^{n-1}(x)=\sup_{a\in D(x)}\left\{ r(x,a)+\int V^{n-1}(y)q(dy/x,a)\right\},$$

$$LV^{n-1}(x,a)= r(x,a)+\int V^{n-1}(y)q(dy/x,a)\ ,\ y$$

$$V_n(x)=V^n(x)-V^n(x^*),$$

donde $x\in X$, $a\in D(x)$ y $x^*\in X$ es fijo.

Observe que

$$V^n(x)=\sup_{a\in D(x)} LV^{n-1}(x,a),\ y$$

$$TV_{n-1}(x)=\sup_{a\in D(x)} LV_{n-1}(x,a).$$

Para cada $x \in X$, consideremos los conjuntos

$$\Gamma_n(x) := \left\{ a \in D(x) : LV^{n-1}(x,a) = V^n(x) \right\},$$

$$L_\infty\left[\Gamma_n(x)\right] := \left\{ \begin{array}{l} a \in A(x) : a \text{ es un punto de acumulación de alguna sucesión} \\ \langle a_n \rangle \text{ con } a_n \in \Gamma_n(x), \text{ para cada } n \in N. \end{array} \right\}.$$

Notemos también que $\Gamma_n(x) := \left\{ a \in D(x) : LV_{n-1}(x,a) = TV_{n-1}(x) \right\}$.

Teorema 2.1. Bajo I.5.3 y II.3.2, existe $f^* \in F$ política esatacionaria óptima.

Demostración.

Por el Teorema II.3.9, existen j^* y V^* tales que

$$V^* = \lim_n V_n \ , \ V^* \in C_s(X);$$

$$j^* = \lim_n \{V^n(x^*) - V^{n-1}(x^*)\};$$

y además, son una solución a la Ecuación de Optimalidad, es decir,

$$j^* + V^* = TV^*.$$

En consecuencia, se verifica la Hipóteisis I.4.6(a) y (b).

Como es usual, para verificar I.4.6(b), se hará uso de teoremas de selección medible.

Sea

$$\Gamma(x) := \left\{ a \in D(x) : LV^*(x,a) = TV^*(x) \right\},$$

luego, por la Proposición A.1.4(b) se sigue que

$$L_\infty\left[\Gamma_n(x)\right] \subseteq \Gamma(x).$$

Ahora, por el Teorema de Selección A.3.3, existe $f_n \in F$ tal que

$$r(x,f_n)+\int V_{n-1}(y)q(dy/x,f_n)=\sup_{a\in D(x)}\left\{r(x,a)+\int V_{n-1}(y)q(dy/x,a)\right\}$$

$$=TV_{n-1}(x).$$

En consecuencia, $f_n(x)\in\Gamma_n(x)$, $x\in X$ y $n\in N$. Luego, por la Proposición A.3.4, existe $f^*\in F$ tal que para cada $x\in X$

$$f^*(x)\in L\infty\left[\Gamma_n(x)\right]\subseteq\Gamma(x).$$

En consecuencia, f^* es un maximizador de la Ecuación de Optimalidad y se verifica I.4.6(b);entonces por la Observación I.4.7 se sigue la conclusión del teorema. Además, la función de valores óptimos

$$J^*(x)=J(f^*,x)=\lim_n\{V^n(x^*)-V^{n-1}(x^*)\}=j^*.$$

■

Observación 2.2.

a) Aunque la conclusión en el Teorema 1.1 y el Teorema 2.1 es la misma, en la demostración del Teorema 2.1 se usó iteración de valores para verificar la Hipóteisis I.4.6, lo que nos permite aproximar la función de valores óptimos. Otras aproximaciones también basadas en iteración de valores son proporcionadas en Hernández-Lerma (1989, p. 62).

b) Usando el enfoque de 'problemas descontados', ver II.4, también es posible resolver el PCO. Bajo este enfoque Rieder (1979) demuestra:

i) la existencia de una política óptima estacionaria $f^*=\lim_n f_n$, donde f_n satisface

$$r(x,f_n)+\beta_n\int V_{\beta_n}(y)q(dy/x,f_n)=T_{\beta_n}V_{\beta_n}(x),$$

para $n\geq 0$ y $x\in X$, con T y V_β como en I.(1.2) y I.(3.8), repectivamente;

ii) la función de valores óptimos es dada por

$$J^*(x)=\lim_n (1-\beta_n)\,V_{\beta_n}(x^*) \text{ para todo } x\in X,$$

donde $x^*\in X$ es fijo y $\beta_n\uparrow 1$.

3. HIPOTESIS DE DOEBLIN.

El objetivo de esta sección es demostrar la existencia de una política estacionaria óptima en presencia de la Condición de Doeblin, la restricción de compacidad en los espacios de estados y acciones, y como es usual, con condiciones de continuidad adecuadas. Bajo estas restricciones se demuestra la existencia de un subconjunto C del espacio de estados y una política estacionaria bajo la cual alcanza el supremo de la ganancia promedio cuando el estado inicial del sistema es un elemento del conjunto C.

La presentación del material se hará con el siguiente esquema.

3(a). Se dan definiciones y resultados básicos (sin demostración) que serán usados en el resto del trabajo;

3(b). Para cada política $\delta\in\Delta$ y cada distribución inicial $\nu\in P(X)$, se demuestra la existencia de una política $\pi\in\Pi$ (política estacionaria aleatorizada) y $\lambda\in P(X)$ tales que

$$J(\pi,\lambda)\geq J(\delta,\nu).$$

Esto indica que es posible restringir la atención al conjunto de las políticas estacionarias aleatorizadas.

3(c). Se exhibe la existencia de un conjunto medible C y una política estacionaria $\bar{f}\in F$ tal que

$$J(\bar{f},x)= \sup_{\nu\in\mathbb{P}(X)} \sup_{\delta\in\Delta} J(\delta,\nu) \text{ para cada } x\in C.$$

por lo que es posible restringirse al conjunto de las políticas estacionarias, si el estado inicial del sistema es un elemento de C.

3(d). Finalmente, se demuestra la existencia de una política estacionaria óptima f^*.

3.(a). Definiciones y resultados básicos.

Definición 3.1. Para cada política estacionaria aleatorizada $\pi\in\Pi$ (definición I.2.3), consideremos el proceso de Markov controlado (PCM) introducido en I.2.

a) Un conjunto $E\in\mathcal{B}(X)$ es invariante con respecto a π si

$$q(E/x,\pi)=1 \quad , \text{ para cada } x\in E.$$

b) Un conjunto $S\in\mathcal{B}(X)$ es nulo con respecto a π si

$$\lim_n q^n(S/x,\pi)=0, \text{ para cada } x\in X,$$

donde

$$q(\cdot/x,\pi)=\int q(\cdot/x,a)\pi(da),$$

y

$$q^n(\cdot/x,\pi)=\int q^{n-1}(\cdot/y,\pi)q(dy/x,\pi).$$

En el resto del trabajo supondremos que se cumplen las siguientes condiciones.

Condición 3.2. (Doeblin). Existen una medida finita γ sobre X, con $\gamma(X)>0$, un entero ℓ y $\varepsilon>0$ tales que: para cada $B\in\mathcal{B}(X)$

$$q^{\ell}(B/x,\pi)\le 1-\varepsilon \text{ si } \gamma(B)\le\varepsilon \text{ para } x\in X, \pi\in\Pi.$$

Notese que esta condición implica a la Condición I.5.5, puesto que $\mathbb{F}\subset\Pi$.

Condición 3.3.

a) Los conjuntos X y $\mathbb{K}=\{(x,a): x\in X, a\in D(x)\}$ son compactos.

b) $r\geq 0$ semi-continua superiormente.

c) Si $(x_n,a_n)\longrightarrow (x,a)$, entonces $q(\cdot/x_n,a_n)\longrightarrow q(\cdot/x,a)$ débilmente; ver A.2.10.

De la teoría general de Procesos de Markov, en esta sección se hará uso de los siguientes teoremas.

Teorema 3.4.

Suponga que se satisface la Condición 3.2. Entonces para cada $\pi\in\Pi$, existen subconjuntos medibles $S, C_1, C_2, \ldots, C_k$ tales que:

a) $X = S \cup (\bigcup_{i=1}^{k} C_i)$;

b) S es nulo con respecto a π;

c) $\gamma(C_i)>\varepsilon$ y C_i es invariante con respecto a π, $1\leq i\leq k$.

A los conjuntos $C_1, C_2, \ldots, C_k$ se les llama <u>conjuntos ergódicos</u>; ver demostración en Doob (1953, V.5).

Ahora, se define una medida de probabilidad γ_j en los conjuntos ergódicos C_j, $j=1,2,\ldots,k$, de la siguiente manera: para cada $\pi\in\Pi$, $x\in C_j$

(3.5) $$\gamma_j(\cdot) = \lim_{n} P^n(\cdot/x,\pi),$$

$$\text{donde } P^n(\cdot/x,\pi) = \frac{1}{n}\sum_{t=0}^{n-1} q^t(\cdot/x,\pi).$$

Teorema 3.6. Si se satisface la Condición 3.2, entonces para cada conjunto C_j, $1\leq j\leq k$, existen subconjuntos $C_j^1, C_j^2, \ldots, C_j^d$ de C_j, tal que para cada $x\in C_j^1$ existe el límite

a) $\lim_{t} q^{td+m-1}(\cdot/x,\pi) := \gamma_j^m(\cdot)$; y satisface

b) $\gamma_j(\cdot) = \sum_{m=0}^{d} \gamma_j^m(\cdot)/d$;

Además, la convergencia es uniforme y exponencial. A los conjuntos $\{C_j^1, C_j^2, \ldots, C_j^a\}$ se le llama clase cíclica de C_j; ver demostración de este teorema en Doob (1953, V.5).

Sea ν una medida de probabilidad en X. Se dice que ν es invariante con respecto a $q(\cdot/x,\pi)$ si

$$\nu(\cdot)=\int q(\cdot/y,\pi)\nu(dy).$$

Observación 3.7. Sea ν una medida de probabilidad invariante con respecto a $q(\cdot/x,\pi)$. Entonces se cumple

a) $\nu(\cdot)=\int q^t(\cdot/y,\pi)\nu(dy)$ para $t=1,2,\ldots$, y por lo tanto

b) $\nu(\cdot)=\int P^n(\cdot/x,\pi)\nu(dx)$ para $n=1,2,\ldots$, donde P^n es la medida en 3.5.

c) Si S es un conjunto nulo con respecto a π, tomando límite cuando $t\longrightarrow\infty$ en a), se sigue por el teorema de la Convergencia Dominada;

$$\nu(S)=\int \lim_t q^t(S/y,\pi)\nu(dy)=0.$$

Para cada política $\delta\in\Delta$ y distribución inicial $\nu\in P(X)$ se definen:

(3.8) $$J(\nu):=\sup_{\delta\in\Delta} J(\delta,\nu),$$

y

(3.9) $$J:=\sup\{J(\nu):\ \nu\in P(X)\}.$$

3(b). Restricción al conjunto Π.

Ahora se demostrará que es posible restringirse al conjunto de las políticas estacionarias aleatorizadas, ver Proposición 3.17.

Denotaremos por M a una familia numerable de funciones $g\in C(X)$, tales que si $P_1, P_2\in P(X)$ y para cada $g\in M$ se tiene que

$$\int g dP_1 = \int g dP_2,$$

entonces $P_1 = P_2$. La existencia del conjunto M con tal propiedad es posible asegurarla si **X** es un espacio de Borel; ver A.2.8.

En la demostración de la Proposición 3.17 se hará uso del siguiente lema.

Lema 3.10. Para cada $(\delta, \nu) \in \Delta \times \mathcal{P}(X)$ existe $\mu \in \mathcal{P}(K)$ tal que:

(3.11) $$\int r(x,a)\mu(d(x,a)) \geq J(\delta,\nu),$$

(3.12) $$\int g(x)\mu(d(x,a)) = \int \mu(d(x,a)) \int g(x')q(dx'/x,a).$$

para cada $g \in M$.

Demostración.

Para cada $(\delta, \nu) \in \Delta \times \mathcal{P}(X)$, denotemos por

$$P = P_\nu^\delta \text{ y } E = E_\nu^\delta$$

Por el Teorema de Estabilidad de Loeve (1978), p.53, se cumple

(3.13) $$\lim_N \frac{1}{N}\left[\sum_{k=1}^{N-1} \left\{ g(x_k) - E[g(x_k)/B_{k-1}] \right\} \right] = 0,$$

P-casi seguramente para cada $g \in M$, donde $B_k = \sigma(x_0, a_0, x_1, \ldots, x_k, a_k)$.

Sea $J_N = \frac{1}{N} \sum_{k=1}^{N-1} r(x_k, a_k)$, $N \geq 1$, entonces

$$E\left[\limsup_N J_N \right] = E\left[\limsup_N \frac{1}{N} \sum_{k=1}^{N-1} r(x_k, a_k) \right]$$

$$\geq \limsup_N E\left[\frac{1}{N} \sum_{k=1}^{N-1} r(x_k, a_k) \right]$$

$$= J(\delta, \nu) ,$$

la desigualdad se sigue por el Lema de Fatou-Lebesgue, ver Loeve (1978, p. 126).

Sea $A = \left\{ \limsup_N J_N \geq J(\delta,\nu) \right\}$ y supóngase que $P(A) = 0$; luego,

$$E\left[\limsup_N J_N\right] = \int_{A^c} \limsup_N J_N \, dP < \int_{A^c} J(\delta,\nu) \, dP = J(\delta,\nu),$$

lo cual es una contradicción y, por lo tanto, $P(A)>0$; es decir,

$$P\left[\limsup_N J_N \geq J(\delta,\nu)\right] > 0.$$

Ahora construiremos una medida de probabilidad que satisfaga 3.11 y 3.12, usando una trayectoria $\omega\in\Omega$ fija.

Sea $\omega\in\Omega$ tal que satisface 3.13 y

$$\limsup_N J_N \geq J(\delta,\nu).$$

Para este $\omega\in\Omega$ fijo existe una sub-sucesión $\{N_j\}$ para la cual se cumple 3.13 y además

(3.14) $$\lim_j J_{N_j} \geq J(\delta,\nu).$$

Para cada $D\in\mathcal{B}(\mathbb{K})$, definimos la medida (aleatoria) de probabilidad μ_N por

$$\mu_N(D) := \frac{1}{N} \sum_{k=0}^{N-1} I_D(x_k, a_k) \quad , \text{ para } N\geq 1$$

donde I_D es la función indicadora del conjunto D.

Ahora consideremos la sucesión de medidas (aleatorias) de probabilidad $\{\mu_{N_j}\}$.

Por otra parte, puesto que $\mathbb{K}$ es compacto, en la topología de la convergencia débil $\mathcal{P}(\mathbb{K})$ también lo es (ver A.2.7), y por lo tanto existe $\mu\in\mathcal{P}(\mathbb{K})$ tal que μ_{N_j} converge a μ débilmente a través de una sub-sucesión de N_j (la cual denoteremos como N_j de nuevo) para $\omega\in\Omega$ fijo, de manera que se satisface 3.13 y 3.14.

Ahora verificaremos que μ satisface (3.11) y (3.12).

De la definición de μ_N, 3.13 y 3.14 se pueden reescribir como

(3.15) $$\lim_j \left[\int g(x)\mu_{N_j}(d(x,a)) - \int \mu_{N_{j-1}} d(x,a) \int g(x')q(dx'/x,a) \right] = 0$$

para cada $g \in M$, y

(3.16) $$\lim_j \int r(x,a)\mu_{N_j}(d(x,a)) \geq J(\delta,\nu).$$

Por la continuidad débil de **q** , ver 3.3(c), para cada $g \in M$, se tiene que

$$\int g(x')q(dx'/x,a)$$

es continua en $(x,a) \in \mathbf{XA}$, y por lo tanto cuando $j \longrightarrow \infty$ en 3.15 se tiene (3.12).

Se denotará por $C(\mathbf{XA})$ y $C_s(\mathbf{XA})$ el conjunto de funciones acotadas definidas sobre **XA** continuas y semi-continuas, respectivamente.

Como $r \in C_s(\mathbf{XA})$, entonces existe una sucesión $\{g_k\} \subset C_s(\mathbf{XA})$, no-creciente tal que $g_k \longrightarrow r$.

Para cada k y cada j,

$$\int g_k(x,a)\mu_{N_j}(d(x,a)) \geq \int r(x,a)\mu_{N_j}(d(x,a)).$$

Tomando límite cuando $j \longrightarrow \infty$, la convergencia débil de medidas implica para cada k,

$$\lim_j \int g_k(x,a)\mu_{N_j}(d(x,a)) = \int g_k(x,a)\mu(d(x,a)).$$

Por lo tanto,

$$\int g_k(x,a)\mu(d(x,a)) \geq \lim_j \int r(x,a)\mu_{N_j}(d(x,a)) \ ,$$

tomando límite cuando $k \longrightarrow \infty$, por ser g_1 acotada y por 3.16, se sigue que

$$\int r(x,a)\mu(d(x,a)) \geq J(\delta,\nu),$$

lo cual demuestra (3.11).

∎

Proposición 3.17. Para cada $(\delta,\nu)\in\Delta\Pi(X)$, existe $(\pi,\lambda)\in\Pi\Pi(X)$ tal que

$$J(\pi,\lambda)\geq J(\delta,\nu).$$

Demostración.

Para cada $(\delta,\nu)\in\Delta\Pi(X)$, sea $\mu\in\Pi(K)$ tal que se cumplen 3.11 y 3.12. Entonces μ se puede descomponer en una medida $\lambda\in\Pi(X)$ y $\pi\in\Pi$ tales que para cada $D_1\in B(X)$ y $D_2\in B(A)$,

$$\mu(D_1D_2) = \int_{D_1} \pi(D_2/x)\lambda(dx),$$

para la demostración de este resultado ver Bertsekas and Shreve (1978, p. 140).

Luego, para $g\in M$,

$$\int g(x)\mu(d(x,a)) = \int_X \int_A \pi(da/x)g(x)\lambda(dx) = \int_X g(x)\lambda(dx),$$

y por otra parte,

$$\int_{XA} \mu(d(x,a)) \int_X g(x')q(dx'/x,a) =$$

$$= \int_X \int_A \int_X g(x')q(dx'/x,a)\pi(da/x)\lambda(dx)$$

$$= \int_X \int_X g(x')q(dx'/x,\pi)\lambda(dx).$$

Por lo tanto, de 3.12 se sigue, para cada $g\in M$;

$$\int g(x)\lambda(dx) = \int_{XX} g(x')q(dx'/x,\pi)\lambda(dx).$$

En consecuencia,

$$\lambda(\cdot) = \int_X q(\cdot/x,a)\lambda(dx).$$

Por lo tanto, λ es una medida de probabilidad invariante para el $P\in M$ inducido por π.

En consecuencia,

$$J(\pi,\lambda) = \limsup_n \frac{1}{n} E_\lambda^\pi \left[\sum_{k=0}^{n-1} r(x_k,a_k) \right]$$

$$= \limsup_n \frac{1}{n} \sum_{k=0}^{n-1} \int r(x_k,a_k)\pi(da_k/x_k)q(dx_k/x_0,\pi)\lambda(dx)$$

$$= \limsup_n \frac{1}{n} \sum_{k=0}^{n-1} \int r(x_k,a_k)\pi(da_k/x_k)\lambda(dx_k)$$

$$= \int r(x,a)\mu(d(x,a)).$$

Luego, por (3.11),

$$J(\pi.\lambda)\geq J(\delta,\nu).$$

■

3(c). Política óptima para estados iniciales en C.

En esta sub-sección se demuestra la existencia de un conjunto C y de una política óptima, para todo estado inicial en C.

Teorema 3.18. Existe $C\in B(X)$, con $\gamma(C)\geq\varepsilon$ y $\bar{f}\in F$ tal que :

$$J(\bar{f},x)=J,\ J \text{ en (3.9)}$$

y

$$q(C/x,\bar{f})=1 \quad \text{para cada } x\in C.$$

(γ y ε como en la Condición 3.2)

Demostración.

Sea (ε_n) una sucesión de reales no negativos, decreciente y que converge al cero. De la definición de J en (3.9), existe una sucesión $\{(\delta_n,\nu_n)\}\subset\Delta P(X)$, para la cual se verifica la desigualdad

(3.19) $$J(\delta_n,\nu_n)\geq J-\varepsilon_n$$

Por el Lema 3.10, para cada (δ_n,ν_n) existe una medida $\mu^n\in P(K)$ tal que se verifican las relaciones (3.11) y (3.12) con respecto a μ^n, es decir,

$$\int r(x,a)\mu^n(d(x,a)) \geq J(\delta_n, \nu_n)$$

y

$$\int g(x)\mu^n(d(x,a)) = \int \mu^n(d(x,a)) \int g(x')q(dx'/x,a) \quad \text{para } g \in M$$

Por (3.19) y la desigualdad anterior tenemos

$$\int r(x,a)\mu^n(d(x,a)) \geq J-\varepsilon_n \tag{3.20}$$

Por la compacidad de $\mathbb{P}(K)$, existe $\bar{\mu} \in \mathbb{P}(K)$ tal que para alguna subsucesión de $\{\mu^n\}$, la cual denotaremos de nuevo por $\{\mu^n\}$, converge débilmente a $\bar{\mu}$. De esto último y al tomar límite en la igualdad anterior se concluye que $\bar{\mu}$ satisface la relación (3.11), esto es,

$$\int g(x)\bar{\mu}(d(x,a)) = \int \bar{\mu}(d(x,a)) \int g(x)q(dx'/x,a)$$

para cada $g \in M$.

Sea $\{g_k\} \subset C(XA)$ una sucesión no-creciente que converge a r; luego por (3.19)

$$\int g_k(x,a)\mu^n(d(x,a)) \geq \int r(x,a)\mu^n(d(x,a)) \geq J-\varepsilon_n.$$

Tomando límite cuando $n \longrightarrow \infty$, para cada k (fijo), en la desigualdad anterior, y luego tomando límite cuando $k \longrightarrow \infty$, se sigue de la convergencia débil y del Teorema de Convergencia Monótona, respectivamente, que

$$\int r(x,a)\bar{\mu}(d(x,a)) \geq J. \tag{3.21}$$

Además, como $\bar{\mu}$ satisface (3.12). Por lo tanto si se descompone $\bar{\mu}$ en la forma $\bar{\mu}(d(x,a)) = \pi(da/x)\lambda(dx)$, con $\lambda \in \mathbb{P}(X)$ y $\pi \in \Pi$, resulta que λ es invariante con respecto a $q(\cdot/x,\pi)$. Por (3.21) se sigue,

$$J(\pi,\lambda)=\int r(x,a)\bar{\mu}(d(x,a))\geq J\ ,$$

lo cual implica, por la definición de J, que

(3.22) $$J(\pi,\lambda)=J.$$

Ahora se considerará el proceso de Markov inducido por la política π.

Bajo la Condición (3.2), por el Teorema (3.4), existe una descomposición del espacio de estados **X** en un conjunto nulo S y un número finito de conjuntos ergódicos $C_1, C_2, \ldots, C_k$, $\mathbf{X}=\bigcup_{j=1}^{k} C_j \cup S$, con $\gamma(C_j)>\varepsilon$ para $1\leq j\leq k$.

Definimos $\gamma_j\in\mathbb{P}(\mathbf{X})$, $1\leq j\leq k$, y P^n como en (3.5).

Por ser λ invariante con respecto a $q(\cdot/x,\pi)$ y por la Observación (3.7)(b) se sigue, para $n\geq 1$:

$$\lambda(\cdot)=\int P^n(\cdot/x,\pi)\lambda(dx)$$

$$=\sum_{j=1}^{k}\int_{C_j} P^n(\cdot/x,\pi)\lambda(dx)+\int_S P^n(\cdot/x,\pi)\lambda(dx);$$

además, por (3.7)(c), $\lambda(S)=0$. Entonces tomando límites

$$\lambda(\cdot)=\sum_{j=1}^{k}\int_{C_j}\lim_n P^n(\ /x,\pi)\lambda(dx)$$

$$=\sum_{j=1}^{k}\int_{C_j}\gamma_j(\cdot)\lambda(dx),$$

donde la primera igualdad se sigue del Teorema de Convergencia Dominada. Es decir,

(3.23) $$\lambda(\cdot)=\sum_{j=1}^{k}\gamma_j(\cdot)\lambda(C_j).$$

Sea $r(x,\pi)=\int r(x,a)\pi(da/x)$. Por la definición de $J(\pi,x)$, para cada $x\in C_j$ se cumple;

$$J(\pi,x)=\limsup_n \frac{1}{n} E_x^{\pi}\left[\sum_{k=0}^{n-1} r(x_k,a_k)\right]$$

$$= \limsup_n \frac{1}{n} \sum_{k=0}^{n-1} \int r(y,\pi) q^k(dy/x,\pi) \quad , \quad x \in C_j$$

$$= \limsup_n \frac{1}{n} \sum_{k=0}^{n-1} \int_{C_j} r(y,\pi) q^k(dy/x,\pi) \quad , \text{ ya que } C_j \text{ es invar-}$$

$$= \limsup_n \int_{C_j} r(y,\pi) \frac{1}{n} \sum_{k=0}^{n-1} q^k(dy/x,\pi) \quad , \quad x \in C_j$$

$$= \int_{C_j} r(y,\pi) \gamma_j(dy).$$

Es decir, para cada $x \in C_j$

(3.24) $$J(\pi,x) = \int_{C_j} r(y,\pi) \gamma_j(dy).$$

Por otra parte:

$$J(\pi,\lambda) = \int r(x,\pi) \lambda(dx)$$

$$= \int r(x,\pi) \sum_{j=1}^{k} \gamma_j(dx) \lambda(C_j) \quad , \text{ por (3.23)},$$

$$= \sum_{j=1}^{k} \lambda(C_j) \int_{C_j} r(x,\pi) \gamma_j(dx)$$

$$= \sum_{j=1}^{k} \lambda(C_j) J(x_j,\pi) \quad , \text{ por (3.24), para cada } x_j \in C_j, \ 1 \le j \le k.$$

Además, $J(x,\pi) \le J$ para cada $x \in X$. Luego, por (3.22), para cada C_j con $\lambda(C_j) > 0$,

(3.25) $$J(x,\pi) = J, \text{ para cada } x \in C_j.$$

Ahora se muestra que $\pi \in \Pi$ puede ser reemplazada por una política estacionaria. Para esto primero veremos que para cada C_j con $\lambda(C_j) > 0$, existe $f_j : C_j \longrightarrow A$ medible, con $f_j(x) \in D(x)$ para cada $x \in C_j$, y una función medible v_j tal que :

(3.26) $$q(C_j/x,f_j) = 1 \quad , y$$

(3.27) $$v_j(x) + J \le r(x,f_j) + \int_{C_j} v_j(y) q(dy/x,f_j)$$

para cada $x \in C_j$.

Se usará la desigualdad anterior para demostrar que f_j es óptima cuando el estado inicial del sistema es un elemento de C_j.

Para $T>1$ y $x \in C_j$ sea:

$$v_j^T(x) = \sum_{k=0}^{[T/d]-1} E_x^{\pi} \{r(x_k, a_k) - J\},$$

donde $[z]$ es el mayor entero menor o igual a z, para z real.

Por (3.24), (3.25) y el Teorema 3.6, la función v_j^T se puede reescribir para cada $x \in C_j$

$$v_j^T(x) = \sum_{k=0}^{[T/d]d-1} \left[\int r(y,\pi) q^k(dy/x,\pi) - \int_{C_j} r(y,\pi) \gamma_j(dy) \right], \quad (3.24) \text{ y } (3.25)$$

$$= \sum_{k=0}^{[T/d]d-1} \left[\int r(y,\pi) \{ q^k(dy/x,\pi) - \gamma_j(dy) \} \right]$$

$$= \sum_{k=0}^{[T/d]d-1} \left[\int r(y,\pi) \{ q^k(dy/x,\pi) - \sum_{m=1}^{d} \gamma_j^m(dy)/d \} \right], \text{ por } 3.6 \text{ b)}$$

$$= \sum_{k=0}^{[T/d]d-1} \left[\sum_{m=1}^{d} \int r(y,\pi) \{ q^{kd+m-1}(dy/x,\pi) - \gamma_j^m(dy) \} \right].$$

Como la convergencia en el Teorema 3.6 es uniforme y exponencial, y además r es acotada, $\lim_T v_j^T$ existe y es finito.

Sea $v_j(x) := \lim_T v_j^T(x)$ para cada $x \in C_j$.

Por la definición de las funciones v_j^T y v_j se sigue

$$v_j^T(x) = r(x,\pi) - J + \sum_{k=1}^{[T/d]-1} E_x^{\pi} \{r(x_k, a_k) - J\},$$

de manera que

$$v_j^T(x) + J = \int r(x,a) \pi(da/x) + \int \pi(da/x) \left\{ \int \left[\sum_{k=0}^{[T/d]-1} E_y^{\pi} \{r(x_k, a_k) - J\} \right] q(dy/x,a) \right\}.$$

Tomando limite cuanto $T \longrightarrow \infty$

$$v_j(x) + J = \int r(x,a) \pi(da/x) + \int \pi(da/x) \left\{ \int v_j(y) q(dy/x,a) \right\}$$

en consecuencia

(3.28) $$v_j(x)+J=\int\pi(da/x)\left\{r(x,a)+\int v_j(y)q(dy/x,a)\right\}.$$

Además, C_j es invariante con respecto a $q(\ /x,\pi)$; es decir,

$$\int q(C_j/x,a)\pi(da/x)=1$$

para cada $x\in C_j$. Luego

(3.29) $$q(C_j/x,a)=1 \qquad \pi(\ /x)\text{-casi seguramente, } x\in C_j.$$

Sea $R:=\{(x,a)\in K: q(C_j/x,a)=1,\ v_j(x)+J\le r(x,a)+\int v_j(y)q(dy/x,a)\}$,

por (3.28) y (3.29) $\pi(\ R_x\ /x)>0$ para cada $x\in C_j$, donde $R_x:=\{a\in D(x):(x,a)\in R\}$. En efecto, si suponemos que $\pi(R_x/x)=0$, para algún $x\in C_j$, entonces

$$v_j(x)+J > r(x,a)+\int v_j(y)q(dy/x,a)$$

$\pi(\ /x)$- casi seguramente. Por lo tanto,

$$v_j(x)+J > \int\pi(da/x)\left\{r(x,a)+\int v_j(y)q(dy/x,a)\right\}$$

lo cual contradice (3.28).

Por el Teorema de Selección Medible A.3.5, existe $f_j: C_j \longrightarrow A$ tal que $f_j(x)\in R_x$ para cada $x\in C_j$; de aquí se sigue (3.26) y (3.27).

Sea $C=\bigcup_j C_j$ donde la unión es sobre $j\in\{i:\lambda(C_i)>0\}$. Para cada $f\in F$, se define $\bar{f}\in F$ por

$$\bar{f}(x)=\begin{cases} f_j(x) & \text{si } x\in C_j \text{ y } \lambda(C_j)>0 \\ f(x) & \text{en otro caso} \end{cases}$$

Por (3.26) y (3.27)

(3.30) $$q(C/x,\bar{f})=1 \qquad \text{para cada } x\in C$$

(3.31) $$v(x)+J \le r(x,\bar{f})+\int v(y)q(dy/x,\bar{f}) \quad , \quad x\in X,$$

donde $v(x)=v_j(x)$ si $x\in C_j$ con $\lambda(C_j)>0$.

Finalmente para cada $x\in C$, por (3.31), se sigue para $k\geq 0$

$$E_x^{\bar{f}}\{r(x_k,a_k)\} \geq J + E_x^{\bar{f}}\{ v(x_k)-v(x_{k+1}) \};$$

de aquí se sigue

$$J(\bar{f},x)= \limsup_n \frac{1}{n} E_x^{\bar{f}}\left[\sum_{k=0}^{n-1} r(x_k,a_k) \right]$$

$$\geq J + \limsup_n \{v(x)-E_x^{\bar{f}}\, v(x_n)\}/n \ ,$$

por lo que

$$J(\bar{f},x)=J \qquad \text{para cada } x\in C.$$

∎

3(d). Extensión de la política óptima.

Se demuestra la existencia de una política estacionaria óptima (Teorema 3.34) en presencia de las siguientes condiciones adicionales.

Condición 3.32. Para cada $x\in X$ y $D\in B(X)$ con $\gamma(D)>\varepsilon$, existe $\delta\in\Delta$ tal que :

$$P_x^{\delta}\left[\bigcup_{k=1}^{\infty} \{x_k\in D\} \right]=1,$$

donde γ es como en la Condición 3.2.

Condición 3.33. Para cada $D\in B(X)$ con $\gamma(D)>\varepsilon$, $q(D/x,a)$ es continua en $(x,a)\in XA$.

Además de estas condiciones, suponemos que también se verifican las Condiciones 3.2 y 3.3.

Sean C y $\bar{f} \in F$ como en el Teorema 3.18 y F=X-C.

Sea $\tau = \inf\{ k \geq 0 : x_k \in D \}$, $\inf \phi = \infty$.

Notemos que para cada $x \in F$, por la Condición 3.32 existe δ tal que

$$P_x^{\delta}\left[\bigcup_{k=0}^{\infty} \{x_k \in C\} \right] = 1 \ .$$

Se define la política $\delta^* = (\delta^*_0, \delta^*_1, \ldots)$ por

$$\delta_k^*(D/h_k) = \begin{cases} I_D(\bar{f}(x_k)) & \text{si } k \geq \tau \\ \delta(D/h_k) & \text{si } k < \tau \end{cases}$$

para cada $D \in \mathcal{B}(A)$ y $h_k \in H_k$.

Obsérvese que

$$P_x^{\delta^*}\left[\bigcup_{k=0}^{\infty} \{x_k \in C\} \right] = 1 .$$

Teorema 3.34. Bajo las Condiciones 3.32 y 3.33 existe una política estacionaria óptima.

Para la demostración del Teorema 3.34, se usaran los siguientes lemas.

Lema 3.35. $J(\delta^*, x) = J$.

Demostración.

Por el Teorema 3.18, $J(\bar{f}, x) = J$ para $x \in C$.

Sea $E = E_x^{\delta^*}$. Para $k \geq 0$ tal que $P_x^{\delta^*}[\tau = k] > 0$,

$$\limsup_n \frac{1}{n} E\left[\sum_{s=k}^{n-1} r(x_s, a_s) \mid \tau = k \right] = J(\bar{f}, x) = J \tag{3.36}$$

con $x = x_k \in C$.

Además,

$$J(\delta^*,x) = \limsup_n \frac{1}{n} E\left[\sum_{k=0}^{n-1} r(x_k,a_k)\right]$$

$$= \sum_{k=0}^{\infty} E\left[\limsup_n \frac{1}{n} E\left[\sum_{s=0}^{n-1} r(x_s,a_s) \mid \tau=k\right]\right]$$

$$\geq \sum_{k=0}^{\infty} E\left[\limsup_n \frac{1}{n} E\left[\sum_{s=(n-1)\wedge(k+1)}^{n-1} r(x_s,a_s) \mid \tau=k\right]\right],$$

donde $a\wedge b = \min\{a,b\}$. Como $r\geq 0$ y acotada, de (3.36) se sigue que:

$$J(\delta^*,x) \geq J\, P_x^{\delta^*}\left[\bigcup_{k=0}^{\infty}\{x_k\in C\}\right] = J.$$

∎

Ahora, para cada $u\in C_s(F)$, se define Ku para $x\in F$ por:

$$(3.37) \qquad Ku(x) := \sup_{a\in D(x)}\left\{Jq(C/x,a) + \int_F u(y)q(dy/x,a)\right\}$$

Lema 3.38. Sea $u\in C_s(F)$, entonces $Ku\in C_s(F)$.

Demostración.

Por el Teorema 3.18 $\gamma(C)\geq\varepsilon$, y luego por la Condición 3.33 $q(C/x,a)$ es continua en $(x,a)\in XA$.

Además, existen funciones $\phi_n\in C(F)$ que convergen decrecientemente a u, ya que $u\in C_s(F)$; por lo tanto,

$$\int_F u(y)q(dy/x_n,a_n) \leq \int_F \phi_k(y)q(dy/x_n,a_n)$$

para cada $(x_n,a_n)\in K$. Si $(x_n,a_n)\longrightarrow(x,a)$, tomando límite superior cuando $n\longrightarrow\infty$ en la desigualdad anterior se sigue

$$\limsup_n \int_F u(y)q(dy/x_n,a_n) \leq \limsup_n \int_F \phi_k(y)q(dy/x_n,a_n)$$

$$= \int_F \phi_k(y)q(dy/x,a)$$

la igualdad se sigue de la Condición 3.3 c). Ahora tomando límite cuando $k\longrightarrow\infty$ por el Teorema de Convergencia Monótona

$$\limsup_n \int_F u(y)q(dy/x_n,a_n) \leq \int_F u(y)q(dy/x,a).$$

Es decir, la función

$$(x,a) \longrightarrow \int_F u(y)q(dy/x,a)$$

es semi-continua superiormente y, por lo tanto, la función dentro el corchete en (3.37) es semi-continua superiormente. De aquí se sigue que $Ku \in C_s(F)$.

■

Lema 3.39. Sea v la función medible y acotada en F, definida como

(3.40) $$v = Ku .$$

Entonces $v(x) \geq J$ para cada $x \in F$.

Demostración.

Para cada $x \in F$ y δ^* como en el Lema 3.35 se denotará $E = E_x^{\delta^*}$ y $P = P_x^{\delta^*}$.

Sea $B_n(\delta^*,x) = J\ P\left[x_k \in C, \text{ para alguna } k \leq n \right]$

$$= J\ P\left[\bigcup_{k=0}^{n} \{x_k \in C\} \right]$$

$$= J\ P\left[\bigcup_{k=0}^{n} \{x_k \in C,\ x_0 \in F, \ldots, x_{k-1} \in F\} \right]$$

$$= J \sum_{k=0}^{n-1} P\left[x_0 \in F, \ldots, x_k \in F, x_{k+1} \in C \right]$$

$$= E\left[\sum_{k=0}^{n-1} J\ q(C/x_k,a_k) I_{\langle x_0 \in F, \ldots, x_k \in F \rangle} \right].$$

Por (3.40) y la definición (3.37), para cada $x \in F$

$$v(x) \geq J\ q(C/x,a) + \int_F u(y)q(dy/x,a),$$

luego

$$v(x) - E\left[u(y) I_{(y \in F)} \mid (x, a) \right] \geq J\, q(C \mid x, a),$$

sustituyendo esta desigualdad en B_n, se sigue para $x \in F$

$$B_n(\delta^*, x) \leq E\left\{ \sum_{k=0}^{n-1} \left\{ v(x_k) - E\left[v(x_{k+1}) I_{(x_{k+1} \in F)} \mid (x_k, a_k) \right] \right\} I_{(x_0 \in F, \ldots, x_k \in F)} \right\}$$

$$= v(x) - E\left[E\left\{ v(x_n) I_F \mid x_{n-1}, a_{n-1} \right\} I_{(x_0 \in F, \ldots, x_{n-1} \in F)} \right]$$

$$\leq v(x) - H\, P\left[x_k \in F,\ 0 \leq k \leq n \right]$$

para alguna constante H, ya que v es acotada. Tomando límite cuando $n \longrightarrow \infty$ se sigue que

$$\lim_n B_n(\delta^*, x) = J \cdot P\left[\bigcup_{k=0}^{\infty} \{x_k \in C\} \right] = J \leq v(x)$$

■

Demostración del Teorema 3.34.

Sea $\alpha < J$, se definen las siguientes funciones sobre F por:

$$v_{n+1} = K v_n \quad n \geq 1 \quad , v_0 \equiv \alpha, \text{ y K como en (3.37)}.$$

Luego por el Lema 3.38, $v_n \in C_s(F)$ para cada $n \geq 0$, y además,

$$v_n \leq v_{n+1} \leq J, \quad n \geq 0.$$

En consecuencia, el límite $v = \lim_n v_n$ está bién definido y

$$v_n \leq v \leq J.$$

Luego, como $v \geq v_n$, se sigue $Kv \geq Kv_n = v_{n-1}$, $n \geq 1$, y tomando límite cuando $n \longrightarrow \infty$ $Kv \geq v$.

Por otra parte, para cada $\eta > 0$, y por la definición de Kv,

(3.41) $$-\eta + Kv < J\, q(C \mid x, a) + \int_F v(y) q(dy \mid x, a)$$

para alguna $a \in D(x)$. Por el Teorema de Convergencia Monótona,

$$\lim_n \int v_n(y)q(dy/x,a) = \int v(y)q(dy/x,a)$$

y por (3.41) se sigue:

$$-\eta+Kv < J\ q(C/x,a) + \lim_n \int v_n(y)q(dy/x,a)$$

$$= \lim_n \{J\ q(C/x,a) + \int v_n(y)q(dy/x,a)\}$$

$$\leq \lim_n v_{n+1}(x) = v$$

como η es arbitrario, entonces $Kv \leq v$, luego $Kv = v$. Por el Lema 3.39,

$$v(x) \geq J\ ,\ x\in F.$$

Para cada $k=0,1,2,\ldots$, se cumple $\alpha\leq v_k\leq J$ y $\alpha<J$.

Se define $\bar{k}(x) := \min\{\ k:\ v_k(x)>\alpha\ \}$.

Sean los conjuntos $F_k=\{\ x\in F:\ \bar{k}(x)=k\ \}$ para $k\geq 1$ y $F_0=C$. Entonces $F=\bigcup_{k=1}^{\infty} F_k$ y $F_i\cap F_j=\phi$ si $i\neq j$.

Por el Teorema de Selección Medible A.3.3, para $k\geq 1$ existe $f_k: F_k\longrightarrow A$ (medible y $f_k(x)\in D(x)$) tal que :

(3.42) $$v_k(x) = J\ q(C/x,f_k) + \int v_{k-1}(y)q(dy/x,f_k),\ x\in F_k.$$

Ahora se define la política estacionaria $f^*\in F$ por:

$$f^*(x) := \begin{cases} \bar{f}(x) & \text{si } x\in C \\ f_k(x) & \text{si } x\in F_k,\ \text{para } k\geq 1 \end{cases}$$

Considérese el proceso controlado de Markov inducido por la política f^*.

De la definición de F_k se puede verificar que

$$q(\bigcup_{s=0}^{k-1} F_s\ /x,f^*)>0 \text{ para } x\in F_k,\ k\geq 1 \text{ y } F_0=C.$$

En efecto, si suponemos para $x\in F_k$ y $k\geq 1$, que

$$q(\bigcup_{s=0}^{k-1} F_s\ /x,f^*) = q(C/x,f^*)+q(\bigcup_{s=1}^{k-1} F_s\ /x,f^*)=0,$$

entonces para $x \in F_k$

$$v_k(x) = J\, q(C/x, f^*) + \int_F v_{k-1}(y)\, q(dy/x, f^*)$$

$$= \int_{s=k}^{\infty} \bigcup F_s \quad v_{k-1}(y)\, q(dy/x, f^*) = \alpha,$$

por definición de F_s; lo que implica que $x \notin F_k$, lo cual es una contradicción.

En consecuencia, F es un conjunto nulo con respecto a f^*, lo cual implica que

$$P_x^{f^*}\left[\bigcup_{k=0}^{\infty} \{x_k \in C\} \right] = 1,$$

para cada $x \in F$.

Por el Lema 3.35 $J(g,x) = J$, para cada $x \in X$.

■

Observación 3.43.

a) La Condición 3.33, en la demostración del Teorema 3.34, puede ser reemplazada por la siguiente condición:

Sea γ como en la condición 3.2. Para cada $D \in B(X)$, $\gamma(\partial D) = 0$ si $\gamma(D) > 0$, donde ∂D es la frontera del conjunto D.

Este resultado puede consultarse en Kurano(1989).

b) La primera y segunda sección de este capítulo fueron tomadas de [6],[7] y [15]. La tercera sección se tomó basicamente de [10], con la modificación de que en lugar de considerar ganancias se consideran costos y, por lo tanto, el problema es minimizar la función objetivo.

CONCLUSIONES Y COMENTARIOS

El estudio de los Procesos Controlados de Markov con Ganancia Promedio se realiza, generalmente, imponiendo condiciones de recurrencia, continuidad y compacidad al Modelo de Control Markoviano.

En este tabajo, se usó el enfoque de la Ecuación de Optimalidad, según el cual el Problema de Control Optimo tiene solución si se verifica la existencia:

a) de una solución acotada $\{j^*, V^*\}$ de la Ecuación de Optimalidad (EO);

b) de un selector medible f^* que maximize el lado derecho de la misma ecuación.

Para verificar la existencia de la solución acotada $\{j^*, V^*\}$ se imponen condiciones de recurrencia, y para la existencia del selector f^* señaldo en b) se imponen condiciones de compacidad y continuidad. Además, la función de valores óptimos es igual a la constante $J^*(x)=j^*$ para todo $x \in X$.

En las secciones II.2, II.3 y II.4 se discutieron diferentes procedimientos y conjuntos de condiciones para verificar la existencia de la solución $\{j^*, V^*\}$ de la EO. Estos procedimientos nos proveen de varias formas de caracterizar o aproximar a la función de valores óptimos $J^*(x)=j^*$.

Otra condición de recurrencia comúnmente usada es la de Doeblin. Kurano (1989) bajo esta condición e imponiendo compacidad sobre los espacios de estados y acciones resuelve el PCO.

La introducción de la condición de Doeblin en el estudio de los Procesos de Control Markoviano con espacios de estados numerables fue hecha por Hordijk (1977). Otras condiciones de recurrencia relacionadas con la de Doeblin se exhiben en Hernández-Lerma, Montes-de Oca, Cavazos-Cadena (1991).

Otro enfoque para obtener políticas óptimas sin usar EO es imponer condiciones para que se cumpla lo siguiente:

a) la función $f: \longrightarrow J(f,x) = \int r(y,f)\, P_f(dy)$ es semi-continua superiormente en $\mathbb{F}$ (ver I.5.11 (c));

b) $\mathbb{F}$ es compacto;

c) $\sup_{f\in\mathbb{F}} J(f,x) \geq \sup_{\delta\in\Delta} J(\delta,x)$.

Hordijk (1973) usó este enfoque para espacios de estados numerables.

El caso en que la función ganancia $r(x,a)$ no necesariamente es acotada se puede ver en Hordijk (1977, §5), para el caso numerable; y en Hernández-Lerma (1990, §4.5), para espacios de estados en general.

APENDICE

1. Funciones semi-continuas superiormente
2. Medidas de probabilidad y probabilidades de transición
3. Funciones multivaluadas y teoremas de selección

1. FUNCIONES SEMI-CONTINUAS SUPERIORMENTE.

Sea X un espacio de Borel.

Definición 1.1. Una función medible $V: X \longrightarrow \bar{\mathbb{R}}$ se dice que es semi-continua superiormente si para cada $x \in X$ y cada sucesión $\{x_n\} \subset X$ que converge a x, se cumple

$$\limsup_n V(x_n) \leq V(x).$$

Recordemos que $C(X)$ y $C_s(X)$ denotan a las familias de funciones acotadas $V: X \longrightarrow \bar{\mathbb{R}}$, continuas y semi-continuas superiormente, respectivamente.

Proposición 1.2. $V \in C_s(X)$ si y solo si existe $\{V_n\} \in C(X)$ tal que $V_n \downarrow V$.

Demostración. Bertsekas and Shreve, 1978, § 7.5.

Proposición 1.3. Si X es compacto y $v \in C_s(X)$, entonces v alcanza el supremo; esto es, existe $x^* \in X$ tal que $v(x^*) = \sup_x v(x)$.

Demostración. Bertsekas and Shreve (978, §7.5)

Sea $\{B_n\}$ una sucesión de conjuntos de A, y denotemos por $L\infty\{B_n\}$ al conjunto formado por los puntos de acumulación de cualquier sucesión $\{b_n\}\subset A$, tal que $b_n\in B_n$ para cada n.

Sea A un espacio de Borel.

Proposición 1.4. Sea $\{\omega_n\}\subset C_s(A)$ y $\{\varepsilon_n\}$ una sucesión de reales no-negativos convergentes a cero, tales que

$$\omega_n(a)\leq \omega_m(a)+\varepsilon_m \qquad \text{para } m<n \text{ y } a\in A.$$

Entonces:

a) $\omega_\infty := \lim \omega_n$ y $\omega_\infty\in C_s(X)$;

b) $L\infty\{B_n\}\subset B\infty$, donde

$$B_n := \{a\in A:\ \omega_n(a)= \sup_{a'} \omega_n(a')\} \text{ para } n\in\mathbb{N} \text{ ó } n=\infty;$$

c) si $L\infty\{B_n\}\neq\phi$, entonces

$$\lim_n \sup_a \omega_n(a)= \sup_a \lim_n \omega_n(a).$$

Si $\{\omega_n\}$ es uniformemente convergente, se puede tomar $\varepsilon_n= \sup_{n\geq m} \|\omega_m-\omega_n\|$.

Demostración. Schäl (1975), Proposición 10.1.

2. MEDIDAS DE PROBABILIDAD Y PROBABILIDADES DE TRANSICION.

Sea $(X,\mathcal{B}(X))$ un espacio medible.

Una función

$$\mu : \mathcal{B}(X) \longrightarrow \bar{\mathbb{R}}$$

se dice que es una <u>medida con signo</u> en X si:

a) es σ-aditiva ;

b) $\mu(\phi)=0$.

Si $\mu(B)\geq 0$ para cada $B\in\mathcal{B}(X)$, solo se dice que μ es una medida.

Definición 2.1. Sea μ una medida con signo; definimos la parte positiva y negativa de μ por

$$\mu^+(B) := \sup \{ \mu(A) : A \in \mathcal{B}(X) \}, \text{ y}$$

$$\mu^-(B) := \inf \{ \mu(A) : A \in \mathcal{B}(X) \},$$

respectivamente. Estas medidas tienen las siguientes propiedades:

a) al menos una de ellas es finita;

b) $\mu := \mu^+ - \mu^-$.

A la igualdad en b) se le conoce como Descomposición de Jordan-Hahn.

Sea la norma definida por:

(2.2) $$\|\mu\| := \mu^+(X) - \mu^-(X) = \sup_B \mu(B) - \inf_B \mu(B), \quad B \in \mathcal{B}(X)$$

en el espacio de las medidas con signo finitas en X, la cual se conoce como norma de la variación total de μ. (ver Ash (1972)).

Teorema 2.3. (Teorema de Descomposición de Jordan-Hahn)

Si μ es una medida con signo en X, entonces existen conjuntos ajenos medibles X^+ y X^-, tales que

a) $X = X^+ \cup X^-$;

b) $\mu(X^+ \cap B) \geq 0$ y $\mu(X^- \cap B) \leq 0$ para cada $B \in \mathcal{B}(X)$;

c) $\|\mu\| := \mu^+(X^+) - \mu^-(X^-)$.

Demostración. Ash (1972) p. 60.

La integración con respecto a una medida con signo finita μ está definida para cada $v \in B(X)$ por:

(2.4) $$\int v d\mu := \int v d\mu^+ - \int v d\mu^-, \text{ y satisface}$$

105958

(2.5) $$\left|\int v d\mu\right| \leq \|v\| \, \|\mu\|$$ U. A. M. IZTAPALAPA BIBLIOTECA

Sean P y Q medidas de probabilidad en $(X, \mathcal{B}(X))$. Entonces P-Q es una medida con signo finita y su norma de variación total es dada por:

(2.6) $$\|P-Q\| = 2 \sup_{B} |P(B)-Q(B)| .$$

Convergencia Débil

Sea $\{P_n\}$ una sucesión de medidas de probabilidad en X, se dice que $\{P_n\}$ converge debilmente a P si

$$\int f dP_n \longrightarrow \int f dP$$

para toda $f \in C(X)$.

El espacio $\mathbb{P}(X)$ de las medidas de probabilidad en X, es un espacio topológico con la topología de la convergencias débil. En tal caso:

(2.7)(a) si X es un espacio de Borel, $\mathbb{P}(X)$ también lo es;

(2.7)(b) si X es compacto y metrizable, $\mathbb{P}(X)$ también lo es.

Ver Bertsekas and Shreve, (1978), p. 127.

Para X un espacio de Borel, $\{x_i\}$ un conjunto denso en X, sean $f_{i,j} \in C(X)$ definidas por

$$f_{i,j}(x) := 2(1 - j\, d(x, x_i)) \vee 0 \qquad i, j = 1, 2, \ldots ,$$

con d la métrica en X y $x \vee y = \max(x, y)$.

Sea $M := \{f_{i,j} : i, j = 1, 2, \ldots\}$

Proposición 2.8. Sean P_1, $P_2 \in \mathcal{P}(X)$. Si $\int f_{i,j} dP_1 = \int f_{i,j} dP_2$ para $i,j=1,2,\ldots,$ entonces $P_1 = P_2$.

Demostración. Ethier and Kurtz (1986), Bertsekas and Shreve (1978).

Probabilidades de transición

Definición 2.9. Sean X y Y espacios de Borel. Una probabilidad de transición $q(\ /\ )$ en X dado Y es una función tal que $q(\ /y)$ es una medida de probabilidad en X para cada $y \in Y$, y $q(B/\ )$ es una función medible en Y para cada $B \in \mathcal{B}(X)$.

Definición 2.10. Sean X y Y espacios de Borel y $q(\ /\ )$ una probabilidad de transición en X dado Y.

a) q es fuertemente continua si la función

$$y \longrightarrow \int v(x) q(dx/y)$$

es acotada y continua en y, para cada $v \in \mathcal{B}(X)$;

b) q es débilmente continua si la función en a) es acotada y continua para cada $v \in C(X)$;

c) q es semi-continua superiormente, si la función en a) es semi-continua superiormente para cada $v \in C_s(X)$.

Proposición 2.11. Sea $q(\ /\ )$ una probabilidad de transición en X dado Y y $f: XY \longrightarrow \mathbb{R}$ una función medible. Se define $g: Y \longrightarrow \mathbb{R}$ por

$$g(y) := \int f(x,y) q(dx/y) \qquad \text{(siempre que la integral exista)}$$

a) si $f \in B(XY)$ entonces $g \in B(Y)$;

b) si q es débilmente continua y $f \in C(XY)$, entonces $g \in C(Y)$;

c) si q es débilmente continua y $f \in C_s(XY)$, entonces $g \in C_s(Y)$.

Demostración. Bertsekas and Shreve (1978), p. 148.

3. FUNCIONES MULTIVALUADAS Y TEOREMAS DE SELECCION.

Sea $(X, B(X))$ un espacio medible y (A, ρ) un espacio métrico separable.

Para C, C' subconjuntos cerrados de A, se define

$$d(C,C') := \max \left\{ \sup_{c \in C} \rho(c,C'), \sup_{c' \in C'} \rho(c',C) \right\}$$

donde $\rho(x,B) := \inf \{ \rho(x,y) : y \in B \}$ $x \in X$ y $B \subset A$.

La función d satisface las propiedades de una métrica en la familia de los subconjuntos cerrados (no-vacios) de A, se le conoce como métrica de Hausdorff.

Sea $\mathbb{C}(A)$ la familia de todos los subconjuntos compactos (no-vacios) de A.

Proposición 3.1. Si (A, ρ) es separable, entonces el espacio métrico $(\mathbb{C}(A), d)$ también lo es.

Demostración. Schäl (1975).

Sean X y A espacios de Borel. A una correspondencia D que asocia a cada $x \in X$ un subconjunto no-vacío $D(x) \subseteq A$ se le llama <u>función multivaluada</u> de X en A. Al conjunto $K := \{(x,a) : x \in X, a \in D(x)\}$ se le conoce como la <u>gráfica de D</u>.

Para cada $B \subset A$ (no-vacío) se define $D^{-1}(B) := \{x \in X : D(x) \cap B \neq \phi\}$.

Sea D una función multivaluada de X en A. Se dice que es:

a) Borel-medible si $D^{-1}(B)$ es un subconjunto de Borel en X para

cada $B \subset A$ cerrado;

b) semi-continua superiormente si para cada subconjunto abierto $G \subset A$, el conjunto $\{x \in X: D(x) \subset G\}$ es abierto en X.

Un selector medible con respecto a D es una función medible $f: X \longrightarrow A$ tal que $f(x) \in D(x)$ para cada $x \in X$.

Proposición 3.2. Sea $D: X \longrightarrow \mathcal{C}(A)$ una función multivaluada. Las siguientes condiciones son equivalentes.

a) D es Borel-medible;

b) $D^{-}(B)$ es un conjunto de Borel en X para cada subconjunto abierto B de A;

c) la gráfica de D es un conjunto Borel-medible en XA;

d) D es una función Borel-medible de X a $\mathcal{C}(A)$ (con la topología generada por la métrica de Hausdorff).

Demostración. Himmelberg et al (1976)

Teoremas de Selección

Teorema 3.3. Sea $D: X \longrightarrow \mathcal{C}(A)$ una función multivaluada y $v(x,a) \in C_s(\mathbb{K})$. Entonces

a) existe un selector f con respecto a D tal que

$$v(x,f(x)) = \sup_{a \in D(x)} v(x,a) \quad \text{para cada } x \in X,$$

y la función

$$v^*(x) := \sup_{a \in D(x)} v(x,a) \quad \text{es medible.}$$

b) si D es semi-continua superiormente y $v \in C_s(XA)$, entonces $v^* \in C_s(X)$.

Demostración. Bertsekas and Shreve (1978); Schäl (1975).

Teorema 3.4. Sea $D: X \longrightarrow \mathcal{C}(A)$ una función multivaluada medible y $\langle f_n \rangle$ una sucesión de selectores medibles con respecto a D. Entonces existe $f: X \longrightarrow A$ medible tal que

$$f(x) \in L_{\infty}\{f_n(x)\} \subset D(x) \quad , \quad x \in X.$$

Demostración. Schäl (1975)

Teorema 3.5. Sean X, Y espacios de Borel y q una probabilidad de transición en X dado Y. Si para cualquier subconjunto $S \in \mathcal{B}(XY)$,

$$q(S_y / y) > 0 \text{ para cada } y \in Y,$$

donde $S_y := \{x \in X: (x,y) \in S\}$, entonces existe $g: X \longrightarrow Y$ medible, cuya gráfica es un subconjunto de S, esto es, $(x, g(x)) \in S$, $x \in X$.

Demostración. D. Blackwell and C. Ryll-Nardzewski (1963).

REFERENCIAS

1. Ash, R. (1972).

   Real Analysis and Probability.

   Academic Press, New York.

2. Bertsekas, D.P.(1987).

   Dynamic Programming: Deterministic and Stochastic Control.

   Prentice-Hall, Englewood Cliffs, New Jersey.

3. Bertsekas, D.P., Shreve, S.E.(1978).

   Stochastic Optimal Control: The Discrete Time Case.

   Academic Press, New York.

4. Doob, J.L.(1953).

   Stochastic Processes.

   Wiley, New York.

5. Georgin, J.P.(1978).

   Controle de chaines de Markov sur des espaces arbitraires.

   Ann. Inst. H. Poincaré 14, Sect. B, 255-277.

6. Hernández-Lerma, O. (1989).

   Adaptive Markov Control Processes.

   Springer-Verlag, New York.

7. Hernández-Lerma, O.(1990).

Lecture Notes on Discrete-Time Markov Control Processes.

Departamento de Matemáticas, CINVESTAV-IPN.

8. Hernández-Lerma, O, Montes de Oca, R., Cavazos-Cadena, R.(1991).

Recurrence conditions for Markov decisions processes with Borel space state: A survey.

Ann. Oper. Res.

9. Hinderer, K.(1970).

Foundations of Non-stationary Dynamic Programming with Discrete Time Parameter.

Lectures Notes Oper. Res. 33, Springer-Verlag, New York.

10. Kurano, M.(1989).

The existence of a minimum pair of state and policy for Markov decision processes under the hipothesis of Doeblin.

Siam J. Control and Optimization, 27, no.2, 296-307.

11. Loeve, M.(1978).

Probability Theory II, 4th edition.

Springer-Verlag, Berlin.

12. Neveu, J.(1965).

Mathematical Foundations of the Calculus of Probability.

Holden-Day, San Francisco.

13. Ross, S.M.(1970).

Applied Probability Models with Optimization Applications.

Holden-Day, San Francisco.

14. Royden, H.L.(1968).

Real Analysis, 2nd edition.

Mcmillan, New York.

15. Rieder, U.(1979).

On non-discounted dynamic programming with arbitrary state space.

University of Ulm, FRG.

16. Schäl, M.(1975).

Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal.

Z. Whars. verw. Geb. 32, 179-196.

17. Ueno, (1957).

Some limit therems for temporally discrete Markov processes.

J. Fac. Sciences, University Tokyo.

18. Blackwell, D., Ryll-Nardzewski (1963).

Non-existence of everywhere proper conditional distributions.

Ann. Math. Statist., 34, 223-225.

19. Hordijk, A. (1973).

A sufficient condition for the existence of an optimal policy with respect to the averge cost criterion in markovian decision processes.

Trans. 6th Prague Conf. on information theory, 263-274.

20. Hordijk, A. (1977).

Dynamic Programming and Markov Potential Theory.

Math. Centre Tracs 51, Mathematisch Centrum Amsterdam.