UNIVERSIDAD DE SONORA

División de Ciencias Exactas y Naturales
Departamento de Matemáticas

"INTRODUCCIÓN A LA TEORÍA DE CONTROL ESTOCÁSTICO"

TESIS

Que para obtener el Título de

LICENCIADO EN MATEMÁTICAS

Presenta:

AROLDO PÉREZ PÉREZ

Hermosillo, Sonora. septiembre de 1996

Con amor a

mi esposa

mis padres

mis hermanos

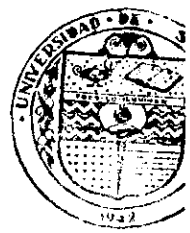
al resto de la familia y amigos.

En especial a mi padre.

Agradezco al director de tesis M.C. Oscar Vega Amaya, al comité revisor M.C. Fernando Luque Vazques, M.C. Jesus Adolfo Minjarez Soza y M.C. Maria Teresa Robles Alcaraz por sus valiosas sugerencias a este trabajo.

Agradezco de manera especial al sistema de Investigación del Mar de Cortés, por el apoyo recibido para la realización de este trabajo bajo el proyecto **SIMAC/94/CT-005.**

INDICE



INTRODUCCION.

La teoría de control trata con sistemas que evolucionan en el tiempo y cuyo comportamiento puede ser influenciado para alcanzar ciertas metas. En la década pasada se dió un notable resurgimiento en el estudio de las aplicaciones y teoría de los problemas de control estocástico, cuya raíz proviene de los años 50's. Los modelos de problemas de control estocástico tienen ganado reconocimiento en varias áreas tales como ecología, economía, ingeniería, etc.

Los problemas de control estocástico(programación dinámica estocástica), son modelos para tomar decisiones secuenciales. Estos modelos consisten de *estados, acciones, costos(o ganancia),* y *probabilidades de transición,* los cuales se relacionan de la manera siguiente: en un estado dado se elige una acción, generándose un costo(o ganancia) y el siguiente estado se determina por medio de la probabilidad de transición. Las acciones se eligen de acuerdo a una política o estrategia, así que el objetivo primordial de los Problemas de Control Óptimo es el de encontrar *políticas óptimas* en algún sentido.

El objetivo principal de este trabajo es analizar el artículo AVERAGE COST OPTIMAL STATIONARY POLICIES IN INFINITE STATE MARKOV PROCESSES WIHT UNBOUNDED COSTS de Linn I. Sennott, y dar una introducción a los Problemas de Control Óptimo (PCO) en tiempo discreto con espacio de estados numerable. Para ello, en el Capítulo 1 introducimos y definimos los elementos y terminología necesaria. En el Capítulo 2 estudiamos el PCO en horizonte finito y damos dos ejemplos: un modelo de juegos y un modelo de producción/inventario. En el Capítulo 3 estudiamos el PCO con horizonte infinito y costos descontados, lo cual es dado de la manera siguiente: en la sección 3.1 damos una breve introducción . En la sección 3.2 introducimos una ecuación funcional llamada Ecuación de Optimalidad y mostramos que existe una única solución a esta ecuación, a la cual llamaremos *función de valor óptimo,* y de la cual podemos encontrar una política óptima; además damos un método para calcular tal solución. En la sección 3.3 damos condiciones bajo las cuales considerando costos no acotados existe una solución a la Ecuación de Optimalidad y por lo tanto una política óptima; discutimos también el método de aproximaciones sucesivas para la función de valor óptimo. Finalmente, en la sección 3.4 damos un ejemplo de un modelo de reemplazo de máquinas. En el Capítulo 4 estudiamos el PCO con horizonte infinito en costo promedio, el cual es dado de la manera siguiente: en la sección 4.1 damos una breve introducción. En la sección 4.2, mostramos

mediante 2 contraejemplos que a diferencia de los capítulos anteriores aquí no es suficiente restringirse a las políticas estacionarias y que no necesariamente existen las políticas óptimas. En la sección 4.3 estudiamos condiciones bajo las cuales, considerando costos acotados, existe una política óptima y discutimos brevemente un enfoque de reducción al caso descontado. En la sección 4.4 estudiamos condiciones bajo las cuales, considerando costos no acotados, existe una política estacionaria óptima. Por último, en la sección 4.5 damos un ejemplo de un modelo de colas.

1. PROBLEMA DE CONTROL OPTIMO

1.1. INTRODUCCION

La teoría de control trata con sistemas dinámicos, es decir, sistemas que evolucionan en el tiempo y cuyo comportamiento puede ser influenciado o regulado para alcanzar ciertas metas; así uno de los temas centrales de la teoría de control es el *Problema de Control Óptimo*(PCO), el cual consiste en controlar un sistema de manera que su comportamiento sea óptimo en algún sentido especificado.

El estudio de los problemas de control puede clasificarse en problemas en tiempo continuo o discreto, dependiendo de si la evolución del sistema es observado en un intervalo de tiempo o solamente en un conjunto discreto de puntos en el tiempo; y como problemas estocásticos o deterministas, si los modelos que describen la dinámica del sistema incorporan o no componentes aleatorias.

Nosotros estamos interesados en el estudio de los sistemas estocásticos *markovianos* en tiempo discreto, cuya teoría básica se ubica en el contexto de la *Programación Dinámica*..

Para el planteamiento del PCO en forma precisa, tanto para el caso contínuo, como para el discreto, determinista o estocástico, se requieren los siguientes elementos:

(a) un modelo de control;
(b) un conjunto de estrategias o políticas de control admisibles; y
(c) un índice de funcionamiento o función objetivo.

El objetivo de este capítulo es introducir cada uno de estos elementos para los sistemas estocásticos markovianos en tiempo discreto. Para ello presentaremos primeramente un ejemplo, el cual nos permita ilustrar el tipo de problemas que nos interesan en este trabajo.

1.2. EJEMPLO: UN SISTEMA DE PRODUCCION/INVENTARIO

Consideremos un sistema de producción/inventario con capacidad $C < \infty$, para el cual:

a) x_t representa el nivel de inventario de cierto artículo al iniciar el período t $(= 0, 1, 2, \cdots)$.

b) a_t representa la cantidad de artículos ordenados a la unidad de producción para abastecer la unidad de inventario, al iniciar el periodo t $(= 0, 1, 2, \cdots)$, la cual suponemos es suministrada en forma inmediata.
c) w_t representa la demanda en el periodo t $(= 0, 1, 2, \cdots)$ y suponemos que la colección $\{w_t\}$ son v.a. *i.i.d.* con valores en los enteros no-negativos y función de probabilidad q.

Nos referiremos a las variables x_t y a_t, $t \geq 0$, como las variables de *estado* y *control*, respectivamente, del sistema de producción/inventario; mientras que a los conjuntos donde estas variables toman valores las denotaremos por $\mathbf{X}$ y $\mathbf{A}$, y les llamaremos *espacio de estados* y *espacio de controles*, respectivamente.

Observe que $\mathbf{X} = \mathbf{A} = \{0, 1, 2, \cdots, C\}$; además, si $x_t = x$, entonces sólo podemos ordenar a la unidad de producción una cantidad $a_t = a \in \mathbf{A}(x) := \{0, 1, 2, \cdots, C - x\}$, puesto que el sistema tiene capacidad C. Es decir, cada elemento $x \in \mathbf{X}$ tiene asociado un subconjunto no-vacio $\mathbf{A}(x)$ el cual especifica los *controles admisibles* para el sistema cuando este se encuentra en el estado x.

Si denotamos por y_t a la cantidad de producto vendida en el periodo t, la 'dinámica' de las variables de estado es modelada por el sistema de ecuaciones

$$x_{t+1} = x_t + a_t - y_t, \quad t = 1, 2, \cdots, \quad \text{y} \quad x_0 = x. \tag{1.1}$$

Por otro lado, y_t puede expresarse en términos de las variables x_t, a_t y w_t en la siguiente forma:

$$y_t = \min\left[x_t + a_t, w_t\right], \quad t = 1, 2, \cdots. \tag{1.2}$$

En vista de lo anterior, (1.1) se transforma en

$$x_{t+1} = \left(x_t + a_t - w_t\right)^+, \quad t = 1, 2, \cdots, \quad \text{y} \quad x_0 = x. \tag{1.3}$$

Ahora supongamos que la evolución del sistema de producción/inventario se ha observado hasta el tiempo t, de manera que se conocen los valores tomados por las variables $x_0, a_0, x_1, \cdots, x_t, a_t$ y supongamos también que $x_t = x$ y $a_t = a$. Lo que nos interesa encontrar es

$$\Pr\left[x_{t+1} = y \mid x_0, a_0, x_1, \cdots, x_t = x, \, a_t = a\right], \tag{1.4}$$

para cada $x, y \in \mathbf{X}$ y $a \in \mathbf{A}(x)$, en términos de la función de probabilidad $q(\cdot)$ de las variables aleatorias $w_0, w_1, \cdots$.

De (1.3) y la independencia de las variables $w_0, w_1, \cdots$, tenemos que

$$
\begin{aligned}
P: \quad &= \Pr\left[x_{t+1} = y \mid x_0, a_0, x_1, \cdots, x_t = x,\, a_t = a\right] \\
&= \Pr\left[(x_t + a_t - w_t)^+ = y \mid x_0, a_0, x_1, \cdots, x_t = x,\, a_t = a\right] \\
&= \Pr\left[(x + a - w_t)^+ = y\right] \\
&= \sum_{w \in B} q(w),
\end{aligned} \tag{1.5}
$$

donde $B = \left\{w \geq 0 : (x + a - w)^+ = y\right\}$.

Un hecho importante que se desprende de los cálculos anteriores es que la probabilidad en (1.4) sólo *depende* del último estado observado $x_t = x$ y del control $a_t = a$, sin importar la "historia" previa del proceso y el valor de t. Es decir,

$$
\Pr\left[x_{t+1} = y | x_0, a_0, x_1, \cdots, x_t = x,\, a_t = a\right] = \Pr\left[x_{t+1} = y | x_t = x,\, a_t = a\right], \tag{1.6}
$$

$\forall\, x, y \in \mathbf{X}$, $a \in \mathbf{A}(x)$, $t \geq 0$.

Por lo tanto, (1.3) puede expresarse equivalentemente por medio de la función

$$
p_{x,y}(a) := \Pr\left[x_{t+1} = y \mid x_t = x,\, a_t = a\right], \tag{1.7}
$$

para todo $x, y \in \mathbf{X}$, $a \in \mathbf{A}(x)$, $t \geq 0$, la cual representa la probabilidad de que el sistema (de producción inventario) se mueva en un 'paso' hacia el estado y dado que el sistema se encuentra en el estado x y se eligió el control $a \in \mathbf{A}(x)$. Es por esta razón que a la función en (1.7) se le llama *probabilidad de transición*(en un paso) del sistema.

Por otra parte, en cada periodo $t \geq 0$ se recibe una ganancia

$$
r_t := qy_t - ca_t - h\left[x_t + a_t\right], \quad t = 0, 1, 2, \cdots \tag{1.8}
$$

donde q, c, y h son constantes positivas y representan el precio de venta, el costo de producción y el costo de almacenamiento unitario, respectivamente.

De (1.2), tenemos que

$$
r_t = q \min\left[x_t + a_t, w_t\right] - ca_t - h\left[x_t + a_t\right], \quad t = 0, 1, 2, \cdots \tag{1.9}
$$

Supongamos que nuestro objetivo consiste en encontrar una *política de producción* que maximice la ganancia esperada en N etapas. Puesto que los niveles

de inventario son variables aleatorias, una política de producción queda bien especificada si nos dice que control debe elegirse cualquiera que sea el estado del inventario, para cada una de las etapas $t = 0, 1, 2, \cdots, N-1$. De esta manera, una política de producción puede representarse como una sucesión $\delta = \{f_t\}_{t=0}^{N-1}$ de funciones f_t :$\mathbf{X} \rightarrow \mathbf{A}$ tales que

$$f_t(x) \in \mathbf{A}(x), \quad \forall\, x \in \mathbf{X},\, t = 0, 1, \cdots, N-1 \tag{1.10}$$

La *ganancia esperada en N-etapas* bajo la política de producción (o control del sistema de inventario) $\delta = \{f_t\}_{t=0}^{N-1}$ dado que el estado inicial del sistema es $x_0 = x$, está dada por

$$J(\delta, x) := E_x \sum_{t=0}^{N-1} \{q \min[x_t + f_t(x_t), w_t] - c f_t(x_t) - h[x_t + f_t(x_t)]\}, \tag{1.11}$$

donde la esperanza es con respecto a la distribución conjunta de las variables $w_0, \cdots, w_{N-1}$.

De lo anterior nuestro problema puede plantearse de la forma siguiente: encontrar una política δ^* tal que

$$J(\delta^*, x) = \sup_{\delta} J(\delta, x), \quad \forall\, x \in \mathbf{X} \tag{1.12}$$

donde el supremo se toma sobre todas las políticas *admisibles*, esto es, que satisfacen (1.10).

1.3. MODELO DE CONTROL

Definición 1.3.1. Un modelo de control markoviano (MCM) en tiempo discreto consta de los objetos $(\mathbf{X}, \mathbf{A}, \{\mathbf{A}(x) \subset \mathbf{A},\, x \in \mathbf{X}\}, p, c)$ donde:

a) $\mathbf{X}$, el *espacio de estados*, es un conjunto numerable;

b) $\mathbf{A}$, el *espacio de controles*, es un conjunto numerable;

c) $\mathbf{A}(x)$ especifica el conjunto de *controles admisibles* para el estado $x \in \mathbf{X}$. Denotaremos por $\mathbf{K} := \{(x, a) : x \in \mathbf{X},\, a \in \mathbf{A}(x)\}$.

d) p es la ley de transición entre los estados, es decir,

i) $p_{xy}(a) \geq 0 \quad \forall\, x,\, y \in \mathbf{X}\,, \forall\, a \in \mathbf{A}(x)$ y

ii) $\sum_y p_{xy}(a) = 1$.

e) $c : \mathbf{K} \rightarrow \mathbf{R}$, es la función de costo.

Un MCM puede considerarse como la representación de un sistema estocástico, para el cual denotaremos por x_t al estado del sistema en el tiempo t y por a_t al control aplicado al sistema en dicho tiempo. La evolución del sistema ocurre en la forma siguiente: si $x_t = x \in \mathbf{X}$ y se elige un control $a_t = a \in \mathbf{A}(x)$, entonces se genera un costo $c(x,a)$ y el sistema "transita" a un nuevo estado $x_{t+1} = y$ con probabilidad $p_{xy}(a) = \Pr[x_{t+1} = y \mid x_t = x \text{ y } a_t = a]$. Una vez ocurrida la transición al estado $x_{t+1} = y$, se elige un nuevo control $a_{t+1} = a' \in \mathbf{A}(y)$ con un costo $c(y,a')$, y así sucesivamente.

1.4. POLITICAS DE CONTROL

Para definir las políticas de control admisibles introducimos los siguientes conjuntos:

$$H_0 \ : \ = \mathbf{X} \tag{1.13}$$

$$H_t \ : \ = \mathbf{K} \times H_{t-1}, \ t = 1, 2, \cdots. \tag{1.14}$$

Notemos que un elemento $h_t \in H_t$ (una $t-$historia) es un vector de la forma

$$h_t = (x_0, a_0, x_1, a_1, \cdots, a_{t-1}, x_t), \tag{1.15}$$

y describe la historia del proceso hasta el tiempo t. Introducimos también el conjunto

$$\mathbf{F} := \{f : \mathbf{X} \to \mathbf{A} \mid f(x) \in \mathbf{A}(x) \quad \forall\, x \in \mathbf{X}\}, \tag{1.16}$$

y a cada $f \in \mathbf{F}$ le llamaremos *selector*.

Definición 1.4.1.

a) Una *política de control (admisible)* es una sucesión $\delta = \{f_t\}$ de funciones $f_t : H_t \to \mathbf{A}$ tales que $f_t(h_t) \in \mathbf{A}(x_t) \quad \forall\, h_t \in H_t$, $t = 0, 1, 2, \cdots$.

b) Una *política markoviana* es una sucesión $\delta = \{f_t\}$ de selectores.

c) Una política markoviana es *estacionaria* si existe $f \in \mathbf{F}$ tal que $f_t = f \quad \forall\, t \geq 0$.

Denotaremos a las políticas estacionarias simplemente por f y Δ representará el conjunto de políticas de control admisibles.

Sea $N < \infty$. Si $\delta = \{f_0, f_1, ..., f_{N-1}\}$ es una política de control admisible y $x_0 = x$ es el estado inicial, entonces la función

$$p_x^\delta(h_N) = \delta_x(x_0)\, p_{x_0,x_1}(\widehat{a}_0)\, p_{x_1,x_2}(\widehat{a}_1) ... p_{x_{N-1},x_N}(\widehat{a}_{N-1}) \tag{1.17}$$

donde $\widehat{a}_t = f_t(x_0,\ a_0,\ x_1, ..., x_{t-1},\ a_{t-1},\ x_t)$, $\quad t = 0,\ 1, ..., N-1$,
es una función de probabilidad definida en H_N. Si $N = \infty$, entonces el Teorema de Ionescu Tulcea(Ver Bertsekas-Shreve, pag. 140, 1978) nos garantiza la existencia de una única función de probabilidad p_x^δ definida en H_∞ tal que

$$\begin{aligned} p_x^\delta(H_\infty) &= 1 \\ p_x^\delta(x_{t+1} = y \mid h_t,\ a_t) &= p_{x_t,y}(\widehat{a}_t) \end{aligned} \tag{1.18}$$

Denotaremos al valor esperado con respecto a p_x^δ por E_x^δ.

En adelante utilizaremos la siguiente notación: para cada $x,\ y \in \mathbf{X}$ y $\ f \in \mathbf{F}$,

$$\begin{aligned} c(x,f) &:= c(x, f(x)) \\ p_{xy}(f) &:= p_{xy}(f(x)). \end{aligned}$$

Notemos que una política markoviana es una política de control (admisible) que sólo depende del estado actual del proceso, esto es, si $f \in \mathbf{F}$, entonces $f(h_t) = f(x_t) \quad \forall\ h_t \in H_t, t = 0, 1, ...$ Por lo tanto, cuando se aplica una política markoviana $\delta = \{f_t\}$, el proceso de estados $\{x_t\}$ es una cadena de Markov en el sentido usual y las probabilidades de transición en un paso son dadas por $p_{x,\cdot}(f_t)$; además si $\delta = f$ es una política estacionaria, entonces el proceso de estados $\{x_t\}$ es una cadena de Markov homogénea con probabilidad de transición $p_{x,\cdot}(f)$. Debido a estas características se dice que $\{x_t\}$ es un proceso de Markov controlado.

1.5. INDICES DE FUNCIONAMIENTO

Para plantear el PCO sólo nos resta introducir el índice de funcionamiento o función objetivo, por medio de la cual se evaluará el "rendimiento" de las políticas de control. Si denotamos por $J(\delta, x)$ tal función, entonces el problema de control óptimo consiste en hallar una política δ^* tal que

$$J(\delta^*, x) = \inf_\delta J(\delta, x) \quad \forall\ x \in \mathbf{X}. \tag{1.19}$$

A una política δ^* que cumpla (1.19) le llamaremos *política óptima*, y a

$$J(x) := \inf_{\delta} J(\delta, x), \quad x \in \mathbf{X} \tag{1.20}$$

le llamaremos *función de valor óptimo.*

Por ejemplo, si deseamos operar el sistema hasta un cierto tiempo fijo $N \geq 0$, podriamos considerar el *costo total esperado*

$$J_N(\delta, x) := E_x^{\delta}\left[\sum_{t=0}^{N-1} c(x_t, a_t) + c_T(x_N)\right] \tag{1.21}$$

como índice de funcionamiento, donde $c_T(x)$ es una función definida en $\mathbf{X}$ y puede interpretarse como un 'costo terminal'. En los capítulos 3 y 4 consideraremos problemas en horizonte infinito ($N = +\infty$) en cuyo caso tomaremos $c_T(\cdot) \equiv 0$.

2. PCO EN HORIZONTE FINITO

2.1. INTRODUCCION

En este capítulo consideramos el modelo de control markoviano

$$\text{MCM} = (\mathbf{X},\ \mathbf{A},\ \{\mathbf{A}(x) \subset \mathbf{A},\ x \in \mathbf{X}\}, p, c),$$

introducido en la definición 1.3.1 con un horizonte de planeación finito N. Luego, el Problema de Control Óptimo consiste en hallar una política $\delta^* \in \Delta$ tal que

$$J_N(\delta^*, x) = \inf_{\delta} J_N(\delta, x) =: J_N(x) \tag{2.1}$$

donde

$$J_N(\delta, x) := E_x^{\delta}\left[\sum_{t=0}^{N-1} c(x_t, a_t) + c_T(x_N)\right] \tag{2.2}$$

y c_T es el costo terminal.

La teoría básica para el estudio de los procesos markovianos de control en tiempo discreto se ubica en el contexto de la Programación Dinámica. En forma general, podemos decir que el enfoque de la Programación Dinámica para el PCO con horizonte finito N consiste en descomponer el problema de optimización en N etapas, en N problemas de optimización de una etapa. En otras palabras, el problema se "reduce" a encontrar una sucesión $\{f_t;\ t = 0, 1, \cdots, N-1\}$ de selectores, con la característica de que cada f_t haga mínimo el costo de la etapa correspondiente en adelante. Esta es la idea subyacente en el llamado *Algoritmo de la Programación Dinámica.*

Definimos así,

$$v_N(x) \ :\ = c_T(x), \tag{2.3}$$

$$v_t(x) \ :\ = \inf_{a \in \mathbf{A}(x)} \left[c(x,a) + \sum_y p_{xy}(a)\, v_{t+1}(y)\right],\ t = N-1, \cdots, 0. \tag{2.4}$$

Intuitivamente

$$v_{N-1}(x) = \inf_{a \in \mathbf{A}(x)} \left[c(x,a) + \sum_y p_{xy}(a)\, c_T(y) \right]$$

representa el costo mínimo en la última etapa y entonces esperaríamos que $v_t(x)$ represente el costo mínimo esperado en las últimas $N-t$ etapas. Veremos que esto es efectivamente así en el siguiente teorema.

2.2. ALGORITMO DE LA PROGRAMACION DINAMICA

Teorema.2.2.1. Si para cada $t = N-1, N-2, ..., 0$ existe un selector f_t^* tal que

$$v_t(x) = c(x,\ f_t^*) + \sum_y p_{xy}(f_t^*)\, v_{t+1}(y) \quad \forall\ x \in \mathbf{X},$$

entonces,
i) $v_0(x) = J_N(x)$ para todo $x \in \mathbf{X}$.
ii) $\delta^* = \left\{ f_0^*,\ f_1^*, \cdots, f_{N-1}^* \right\}$ es óptima, es decir, $J_N(\delta^*, x) = J_N(x) = v_0(x)$ para todo $x \in \mathbf{X}$.

Demostración. Sea $\delta = \{f_t\}_{t=0}^{N-1}$ una política admisible y definamos

$$C_t \ :\ = E_x^{\delta}\left[c_T(x_N) + \sum_{n=t}^{N-1} c(x_n, a_n) \mid h_t \right],\ t = 0, 1, \cdots, N-1 \qquad (2.5)$$

$$C_N \ :\ = E_x^{\delta}\left[c_T(x_N) \mid h_N \right]. \qquad (2.6)$$

Notemos que C_t representa el costo total del tiempo t al tiempo N, dada la historia h_t.

Probaremos que:
a) $C_t \geq v_t(x_t)$ para todo t y
b) Si $\delta = \delta^*$, entonces $C_t = v_t(x_t)$.

Puesto que a) implica que $C_0 \geq v_0(x_0)$ y así por T. 1.4.(iv) del Apéndice tendríamos

$$J_N(\delta, x) \geq v_0(x) \quad \forall\ x \in \mathbf{X}, \qquad (2.7)$$

y usando (b)

$$J_N(\delta^*, x) = v_0(x) \quad \forall\, x \in \mathbf{X}. \tag{2.8}$$

De (2.7) y (2.8) y puesto que δ es arbitrario se sigue que

$$v_0(x) = J_N(x) = J_N(\delta^*, x) \text{ para todo } x \in \mathbf{X},$$

que es lo que deseamos probar.

Haremos la prueba de (a) y (b) por inducción: por definición $C_N = c_T(x_N)$. Supongamos ahora que

$$C_{t+1} \geq v_{t+1}(x_{t+1}) \tag{2.9}$$

y probemos que se cumple

$$C_t \geq v_t(x_t).$$

Por T.1.4 (iii) del Apéndice tenemos que

$$C_t = E_x^{\delta}[c(x_t, a_t) \mid h_t] + E_x^{\delta}\left[\sum_{n=t+1}^{N-1} c(x_n, a_n) + c_T(x_N) \mid h_t\right]$$

y usando la fórmula (2.5) y T.2.6. del Apéndice tenemos

$$C_t = E_x^{\delta}[c(x_t, a_t) \mid h_t] + E_x^{\delta}[C_{t+1} \mid h_t].$$

Así, por la hipótesis de inducción:

$$C_t \geq c(x_t, f_t) + \sum_y v_{t+1}(y)\, p_{x_t y}(f_t) \tag{2.10}$$

y como δ es arbitraria

$$C_t \geq v_t(x_t), \tag{2.11}$$

lo cual prueba (a).

Por otra parte, si se tiene la igualdad en (2.9) y $\delta = \delta^*$, entonces se cumple la igualdad en (2.10) lo cual implica que también se cumple la igualdad en (2.11). Esto prueba (b). ■

Si nuestro objetivo es maximizar ganancias más bien que minimizar costos, podemos usar este algoritmo intercambiando inf por sup en todas partes, puesto que $\sup B = -\inf(-B) \quad \forall\, B \subset \mathbf{R}$. En adelante emplearemos la función $c(x, a)$ indistintamente para costos o ganancias.

2.3. EJEMPLOS

2.3.1. UN MODELO DE JUEGOS

En cada ronda de un juego, una persona puede apostar cualquier cantidad de su capital disponible y puede ganar o perder ese monto con probabilidad p y $q = 1-p$, respectivamente. El jugador va a hacer N apuestas y su objetivo es maximizar la esperanza del logaritmo de su fortuna final. ¿Cuál es la estrategia de apuesta para lograr este fin?.

Solución:

Para plantear este problema en el contexto del Algoritmo de la Programación Dinámica, denotemos por x_t al capital disponible del jugador en el tiempo t y por a_t la fracción del capital que el jugador a decidido apostar en dicho tiempo, esto es, $a_t = a$ para algún $a \in [0,1]$. Observemos que $\mathbf{X} = [0, +\infty)$ y $\mathbf{A} = \mathbf{A}(x) = [0,1] \quad \forall\, x \in \mathbf{X}$. Por otra parte las probabilidades de transición estan dadas por

$$p_{xy}(a) = \begin{cases} p, \text{ si } y = x + ax \\ q, \text{ si } y = x - ax. \end{cases}$$

Notemos que aunque en este problema el espacio de estados y el espacio de controles no son numerables, únicamente podemos pasar de un estado x al estado $x+ax$ o al $x-ax$; por lo cual, podemos emplear el Algoritmo de la Programación Dinámica(T.2.2.1) para la solución de este problema. Tomando

$$\begin{aligned} c(x,a) &:= 0 \\ c_T(x) &:= \ln(x), \end{aligned}$$

el problema consiste en encontrar una política de control δ^* (una estrategia de apuesta) tal que

$$\begin{aligned} J_N(\delta^*, x) &= \sup_{\delta} J_N(\delta, x) \\ &= \sup_{\delta} E_x^{\delta} \ln(x_N). \end{aligned}$$

Así, por el Algoritmo de la Programación Dinámica obtenemos

$$\begin{aligned} v_{N-t} &= \max_{0\leq a\leq 1}\left[c(x,ax)+\sum_{y}p_{xy}(a)\,v_{N-t+1}(y)\right] \\ &= \max_{0\leq a\leq 1}\left[pv_{N-t+1}(x+ax)+qv_{N-t+1}(x-ax)\right],\ t=1,\cdots,N, \quad (2.12) \end{aligned}$$

con costo terminal $v_N(x)=\ln x$.

Podemos ver fácilmente que si $p\leq \frac{1}{2}$, entonces $v_0(x)=\ln x$ y la estrategia óptima es apostar siempre 0. En efecto, usando

$$v_N(x)=\ln x,$$

tenemos por la ecuación (2.12) que

$$\begin{aligned} v_{N-1}(x) &= \max_{a}\left[p\ln(x+ax)+q\ln(x-ax)\right] \\ &= \max_{a}\left[p\ln(1+a)+q\ln(1-a)\right]+\ln x \quad (2.13) \end{aligned}$$

y como $1+a\leq \dfrac{1}{1-a}$ para todo $0\leq a<1$,

$$\ln(1+a)\leq -\ln(1-a),\ 0\leq a<1$$

así,

$$p\ln(1+a)+q\ln(1-a)\leq 0 \text{ para todo } 0\leq a\leq 1.$$

Por lo cual el máximo se alcanza en $a=0$, de donde obtenemos

$$v_{N-1}(x)=\ln x$$

y usando (2.12) con $t=2$ obtenemos de igual forma que

$$v_{N-2}(x)=\ln x,$$

siguiendo este mismo argumento podemos ver fácilmente que

$$J_N(x)=v_0(x)=\ln x \text{ con } a=0.$$

Supongamos ahora que $p > \frac{1}{2}$. De (2.13) encontramos por simple cálculo que el máximo es alcanzado en $a = p - q$ y así,

$$v_{N-1}(x) = C + \ln x$$

donde $C = \ln 2 + p \ln p + q \ln q$

Usando nuevamente (2.12) con $t = 2$ obtenemos

$$v_{N-2}(x) = \max_{\alpha} [p \ln (x + ax) + q \ln (x - ax)] + C$$

por lo tanto en comparación con (2.13), vemos que la decisión óptima está nuevamente dada por la fracción $p - q$ de su capital y

$$v_{N-2}(x) = 2C + \ln x.$$

Es fácil ver que

$$v_0(x) = nC + \ln x$$

y la acción óptima es apostar siempre la fracción $p - q$ de su capital disponible.

2.3.2. UN PROBLEMA DE INVENTARIO

Consideremos un sistema de producción/inventario como el introducido en el capítulo I, pero con espacio de estados $\mathbf{X} = \mathbf{Z}$ y espacio de acciones $\mathbf{A} = \mathbf{N}$. Además, supongamos que $w_0, w_1, ...$ son v.a. *i.i.d*, acotadas y con valores en los enteros no- negativos, función de probabilidad q y esperanza finita. Así,

$$\mathbf{A}(x) = \mathbf{N} \text{ para todo } x \in \mathbf{X}$$

y la dinámica esta dada por

$$\begin{aligned} x_0 &= x \\ x_{t+1} &= x_t + a_t - w_t, \quad t = 0, 1, 2, ... \end{aligned}$$

de donde

$$p_{xz}(a) = q(x + a - z).$$

Nuestro propósito es encontrar una política que minimice el costo esperado en cierto periodo N, el cual está determinado por las siguientes componentes:
a) ca_t, donde c es el costo por unidad de producción.
b) $h\max(0,\ x_t + a_t - w_t)$, donde h es el costo de almacenamiento por unidad y
c) $p\max(0,\ w_t - x_t - a_t)$, donde p es el costo por demanda no satisfecha en el momento solicitado. Suponemos que c, h, y p son constantes positivas con $p > c$.

Entonces el costo por etapa es

$$\widehat{c}(x_t, a_t, w_t) = ca_t + h\max(0, x_t + a_t - w_t) + p\max(0,\ w_t - x_t - a_t), \quad t = 0, 1, ...$$

luego, la función de costo queda dada por

$$\begin{aligned} c(x, a\ ) &= E\left[\widehat{c}(x_t,\ a_t,\ w_t\ ) \mid x_t = x, a_t = a\right] \\ &= ca + hE\left[\max(0,\ x + a - w_t)\right] + pE\left[\max(0,\ w_t - x - a)\right] \end{aligned}$$

Ahora, si se aplica la política de control $\delta = (f_0, \cdots, f_{N-1})$, dado que el sistema inició en el estado $x \in \mathbf{X}$, el costo esperado en N etapas es

$$J_N(\delta, x) = E_x^{\delta}\left[\sum_{t=0}^{N-1} c(x_t,\ a_t)\right]$$

y el problema de control óptimo consiste en encontrar δ^* tal que

$$\begin{aligned} J_N(x) &= J_N(\delta^*, x) \\ &= \inf_{\delta} J_N(\delta, x) \text{ para todo } x \in \mathbf{X}. \end{aligned}$$

Definiendo

$$L(y) := hE\left[\max(0,\ y - w_t)\right] + pE\left[\max(0,\ w_t - y)\right] \tag{2.14}$$

observemos que

$$c(x, a) = ca + L(x + a).$$

Así, por el Teorema de la Programación Dinámica obtenemos

$$\begin{aligned} v_N(x) &= 0 \text{ para todo } x \in \mathbf{X}, \\ v_t(x) &= \inf_{a \geq 0} \left[ca + L(x+a) + \sum_z v_{t+1}(z)\, q(x+a-z) \right]. \end{aligned}$$

Haciendo el cambio de variable $y = x + a$ obtenemos

$$v_t(x) = \inf_{y \geq x} \left[cy + L(y) + \sum_{z \leq y} v_{t+1}(z)\, q(y-z) \right] - cx. \tag{2.15}$$

Luego,

$$v_{N-1}(x) = \inf_{y \geq x} [cy + L(y)] - cx.$$

Como las funciones $\max(0,\ y - w_t)$ y $\max(0,\ w_t - y)$ son convexas en y para cualquier w fijo, se sigue que $L(y)$ es convexa (Ver T.3.2 y T.3.3 del Apéndice), por lo tanto la función $G_{N-1}(y) := cy + L(y)$ es convexa.

Probaremos ahora que $G_{N-1}(y) \to +\infty$ cuando $|y| \to +\infty$. Para hacer esto, consideraremos a la función G_{N-1} como una función definida sobre los números reales. De la suposición de que las w_t son acotadas, tenemos que existe una constante $M > 0$ tal que $w \leq M \quad \forall\ w$. Luego, de (2.14) y el T.1.4.(iii) del Apéndice se sigue que la derivada de $G_{N-1}(y)$ esta dada por

$$G'_{N-1}(y) = \begin{cases} c - p, \text{ si } y < 0 \\ c + h, \text{ si } y > M \end{cases}$$

y como $p > c$ observamos que $\lim_{|y| \to +\infty} G_{N-1}(y) = +\infty$. Por lo tanto del T.3.4. del Apéndice, tenemos que $G_{N-1}(y)$ tiene un mínimo en $\mathbf{R}$; por lo tanto, también tiene un mínimo cuando su dominio se restringe a los enteros, al cual denotamos por S_{N-1}.

Si $x < S_{N-1}$, entonces $y^* = S_{N-1}$ alcanza el mínimo en $\inf_{y \geq x} [cy + L(y)]$ lo cual implica que

$$a^* = S_{N-1} - x \tag{2.16}$$

y si $x \geq S_{N-1}$, entonces $y^* = x$ alcanza el mínimo en $\inf_{y \geq x} [cy + L(y)]$ lo cual implica que

$$a^* = 0. \tag{2.17}$$

De (2.16) y (2.17) tenemos:

$$a^* = f^*_{N-1}(x) = \begin{cases} S_{N-1} - x, \text{ si } x < S_{N-1} \\ 0, \text{ si } x \geq S_{N-1} \end{cases}$$

Luego,

$$v_{N-1}(x) = \begin{cases} c(S_{N-1} - x) + L(S_{N-1}), \text{ si } x < S_{N-1} \\ L(x), \text{ si } x \geq S_{N-1} \end{cases}$$

Probaremos ahora que $\lim_{|y| \to +\infty} v_{N-1}(y) = +\infty$:

Denotemos por L'_+ a la derivada por la derecha de la función L. Como

$$v'_{N-1}(y) = \begin{cases} -c, \text{ si } y < S_{N-1} \\ L'(y), \text{ si } y \geq S_{N-1} \end{cases}$$

y puesto que S_{N-1} es el mínimo de $G_{N-1}(y)$ tenemos que

$$G'_{N-1+}(y) = c + L'_+(y) > 0 \quad \text{para } y \geq S_{N-1}.$$

Por lo tanto $L'_+(y) > -c$, así $\lim_{|y| \to +\infty} v_{N-1}(y) = +\infty$.

Además como $L(y)$ es convexa, v_{N-1} lo es. Por lo tanto, podemos de (2.15) determinar v_t para $t = N-2, ..., 0$ usando argumentos similares al anterior; como sigue:

Supongamos que v_{t+1} es convexa y que $\lim_{|y| \to +\infty} v_{t+1}(y) = +\infty$. Entonces $G_t(y) := cy + L(y) + \sum_{z \leq y} v_{t+1}(z)\, q(y - z)$ es también convexa.

Además como por hipótesis $\lim_{|y| \to +\infty} v_{t+1}(y) = +\infty$ y las w_t toman un número finito de valores por ser acotadas, se sigue que $\lim_{|y| \to +\infty} G_t(y) = +\infty$. Luego $G_t(y)$ tiene un mínimo en un punto S_t, así

$$f^*_t(x) = \begin{cases} S_t - x, \text{ si } x < S_t \\ 0, \text{ si } x \geq S_t \end{cases} \tag{2.18}$$

Luego,

$$v_t(x) = \begin{cases} L(x) + \sum_{z \leq y} v_{t+1}(z)\, q(x-z), \text{ si } x \geq S_t \\ c(S_t - x) + L(S_t) + \sum_{z \leq S_t} v_{t+1}(z)\, q(S_t - z), \text{ si } x < S_t \end{cases}$$

Así, por T.2.2.1

$$v_0(x) = J_N(\delta^*, x) = J_N(x)$$

con $\delta^* = \left(f_0^*, \cdots, f_{N-1}^*\right)$ determinada por (2.18).

3. EL PCO CON HORIZONTE DE PLANEACION INFINITO Y COSTOS DESCONTADOS.

3.1. INTRODUCCION.

En este capítulo estudiamos el PCO cuando $N = +\infty$ e intoducimos un factor de descuento $0 < \alpha < 1$ en los costos. El uso de un factor de descuento está económicamente motivado por el hecho de que un costo en el futuro puede tener menos valor que un costo en el presente. Así,

$$J_\alpha(\delta, x) := E_x^\delta \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right], \quad \delta \in \Delta, \ x \in \mathbf{X}, \tag{3.1}$$

cuando la expresión de la parte derecha esté bien definida. En este caso a una política δ^* tal que

$$J_\alpha(\delta^*, x) = \inf_\delta J_\alpha(\delta, x), \quad x \in \mathbf{X}, \tag{3.2}$$

le llamaremos *política óptima $\alpha-$descontada, y a*

$$J_\alpha(x) := \inf_\delta J_\alpha(\delta, x), \quad x \in \mathbf{X}, \tag{3.3}$$

le llamaremos función de valor óptimo ($\alpha-$descontada).

En la sección 3.2.1 probamos que la función de valor óptimo $J_\alpha(x)$ satisface una ecuación funcional(ecuación de optimalidad) y mostramos que existe una política óptima $\alpha-$descontada estacionaria, a la cual llamaremos *política estacionaria $\alpha-$óptima* o simplemente *política $\alpha-$óptima*. En la sección 3.2.2 discutimos un método para calcular $J_\alpha(x)$ conocido como Método de *A*proximaciones Sucesivas y en la sección 3.3 consideramos lo anterior(con una suposición adicional) para el caso en que los costos son positivos no-acotados, esto es, $0 \leq c(x, a) \leq \infty$ para todo $x \in \mathbf{X}$ y $a \in \mathbf{A}$. En la sección 3.4 consideramos un ejemplo para ilustrar el Método de Aproximaciones Sucesivas.

3.2. CRITERIO EN COSTO DESCONTADO CON COSTOS ACOTADOS.

Supondremos en esta sección que se cumplen las siguientes hipótesis:

Condición 3.2.1.
(a) $\mathbf{A}(x)$ es finito para cada $x \in \mathbf{X}$.
(b) $|c(x,a)| < M$, para todo $x \in \mathbf{X}$ y $a \in \mathbf{A}(x)$, para alguna constante positiva M.

Notemos que bajo la C.3.2.1(b), el costo total esperado $\alpha-$descontado

$$J_\alpha(\delta, x) := E_x^\delta \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right], \quad \delta \in \Delta, \ x \in \mathbf{X},$$

está bien definido, ya que en este caso

$$|J_\alpha(\delta, x)| < \frac{M}{1-\alpha}.$$

3.2.1. LA ECUACION DE OPTIMALIDAD Y POLITICA $\alpha-$OPTIMA

Probaremos aquí, que bajo la condición 3.2.1 existe una política $\alpha-$óptima. Para ello mostraremos que

$$J_\alpha(x) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a) J_\alpha(y) \right] \tag{3.4}$$

y que f es una política $\alpha-$óptima si y sólo si, f minimiza el lado derecho de (3.4). A (3.4) se le conoce como Ecuación de Optimalidad.

Para este propósito necesitamos introducir los siguientes conceptos y resultados:

Sea

$$B(\mathbf{X}) = \{u : \mathbf{X} \to \mathbf{R} \mid u \text{ es acotada}\}.$$

Definamos la norma de $u \in B(\mathbf{X})$ por

$$\|u\| = \sup_{x \in \mathbf{X}} |u(x)|,$$

y consideremos la métrica d dada por

$$d(u,v) = \|v - u\|, \quad u, v \in B(\mathbf{X}).$$

Observemos que con esta métrica , $B(\mathbf{X})$ es un espacio de Banach.

Definamos los operadores T_f y T de $B(\mathbf{X})$ en $B(\mathbf{X})$ por

$$T_f u(x) := c(x,f) + \alpha \sum_y p_{xy}(f)\, u(y) \tag{3.5}$$

y

$$Tu(x) := \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a)\, u(y) \right], \tag{3.6}$$

respectivamente.

Proposición 3.2.2.
(a) T_f y T son operadores monótonos, es decir, si $u \leq v$, entonces $T_f u \leq T_f v$ y $Tu \leq Tv$
(b) Si $u \in B(\mathbf{X})$ y k es una constante, entonces
(i) $T(u+k)(x) = Tu(x) + \alpha k$ y
(ii) $T_f(u+k)(x) = T_f u(x) + \alpha k$
(c) T_f y T son operadores de contracción.

Demostración. Las demostraciones de (a), (b) y la primera parte de (c) son directas, por lo cual únicamente probaremos que T es de contracción.

Como

$$\begin{aligned} v(x) &\leq u(x) + \|u - v\| \quad \text{y} \\ u(x) &\leq v(x) + \|u - v\| \end{aligned}$$

tenemos por (a) y (b) que

$$\begin{aligned} Tu(x) &\leq Tv(x) + \alpha \|u - v\| \quad \text{y} \\ Tv(x) &\leq Tu(x) + \alpha \|u - v\|. \end{aligned}$$

Luego,

$$\begin{aligned} Tu(x) - Tv(x) &\leq \alpha \|u - v\| \quad \text{y} \\ Tv(x) - Tu(x) &\leq \alpha \|u - v\| \end{aligned}$$

de donde

$$|Tu(x) - Tv(x)| \leq \alpha \|u - v\| \text{ para todo } x \in \mathbf{X}.$$

Por lo tanto

$$\|Tu - Tv\| \leq \alpha \|u - v\|,$$

lo cual prueba lo deseado. ■

Proposición 3.2.3.
(a) T tiene un único punto fijo en $B(\mathbf{X})$
(b) Para cada selector f, el operador T_f tiene un único punto fijo en $B(\mathbf{X})$.

Demostración. Se sigue de la proposición anterior y el Teorema de Punto Fijo para operadores de contracción(T.6.4 Apéndice) ■

Lema 3.2.4.
(a) El punto fijo de T_f es $J_\alpha(f, \cdot)$, es decir,

$$J_\alpha(f, x) = T_f J_\alpha(f, x) \text{ para todo } x \in \mathbf{X}.$$

(b) Una política $\delta = \{f_t\}_{t=0}$ es α-óptima si, y solo si, $J_\alpha(\delta, x)$ es punto fijo de T.

Demostración (a):
Como

$$J_\alpha(f, x) = E_x^f \sum_{t=0}^{\infty} \alpha^t c(x_t,\ a_t),$$

usando el T.1.4(iii) y el T.2.3 del Apéndice se obtiene que

$$\begin{aligned} J_\alpha(f,x) &= c(x,f) + \alpha E_x^f J_\alpha(f,x_1) \\ &= c(x,f) + \alpha \sum_y p_{xy}(f)\, J_\alpha(f,y) \\ &= T_f J_\alpha(f,x). \end{aligned}$$

Por lo tanto $J_\alpha(f,\cdot)$ es el punto fijo de T_f.

(b) Supongamos que $u(x) = J_\alpha(\delta, x)$ es punto fijo de T.

Tenemos entonces que

$$u(x) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a)\, u(y) \right].$$

Sea $\delta' = \left\{ f_t' \right\}_{t=0}$ una política arbitraria y consideremos

$$E_x^{\delta'} \left[\alpha^{t+1} u(x_{t+1}) \mid h_t, a_t \right] = \alpha^{t+1} \sum_y p_{x_t y}\left(f_t'\right) u(y)$$

donde h_t es la historia del proceso hasta el tiempo t.

Entonces

$$\begin{aligned} E_x^{\delta'} \left[\alpha^{t+1} u(x_{t+1}) \mid h_t, a_t \right] &= \alpha^t \left[c\left(x_t,\ f_t'\right) + \alpha \sum_y p_{x_t y}\left(f_t'\right) u(y) \right] - \alpha^t c\left(x_t, f_t'\right) \\ &\geq \alpha^t u(x_t) - \alpha^t c\left(x_t, f_t'\right). \end{aligned}$$

Así,

$$\alpha^t u(x_t) - E_x^{\delta'} \left[\alpha^{t+1} u(x_{t+1}) \mid h_t, a_t \right] \leq \alpha^t c\left(x_t, f_t'\right)$$

y por T.1.4(iii) y T.2.3 del Apéndice

$$\alpha^t E_x^{\delta'} u(x_t) - \alpha^{t+1} E_x^{\delta'} u(x_{t+1}) \leq \alpha^t E_x^{\delta'} c(x_t, a_t)$$

sumando desde $t = 0$ hasta n obtenemos

$$E_x^{\delta'} u(x_0) - \alpha^{n+1} E_x^{\delta'} u(x_{n+1}) \leq E_x^{\delta'} \sum_{t=0}^{n} \alpha^t c(x_t, a_t) \tag{3.7}$$

como $u(x)$ es acotada y $\alpha \in (0,\ 1)$ obtenemos tomando límite en (3.7) cuando $n \to \infty$ que

$$u(x) \leq J_\alpha\left(\delta', x\right),$$

es decir,

$$J_\alpha(\delta, x) \leq J_\alpha\left(\delta', x\right)$$

y puesto que δ' fue arbitraria

$$J_\alpha(\delta, x) = J_\alpha(x).$$

Por lo tanto δ es α-óptima.

Supongamos ahora que δ es α-óptima, es decir, $u(x) = J_\alpha(\delta, x) = J_\alpha(x)$.
Probaremos:
(i) $u \geq Tu$
(ii) $u \leq Tu$

Para demostrar (i) consideremos $u(x) = E_x^{\delta} \sum_{t=0}^{\infty} \alpha^t c(x_t, a_t)$.

Así,

$$u(x) = c(x, f_0) + \alpha E_x^{\delta} J_\alpha\left(\delta',\ x_1\right)$$

donde $\delta' = \{f_1, \cdots,\ f_n, \cdots\}$.

Y por lo tanto

$$u(x) \geq c(x, f_0) + \alpha E_x^{\delta} u(x_1)$$

de donde se obtiene que

$$u(x) \geq \min_{a \in \mathbf{A}(x)} \left[c(x, a) + \alpha \sum_{y} p_{xy}(a)\, u(y) \right],$$

lo cual prueba (i).

Para demostrar (ii), sea g un selector arbitrario y definamos $\delta' = \{g,\ \delta\}$. Así,

$$u(x) \leq J_\alpha\left(\delta', x\right)$$

de donde

$$u(x) \leq c(x,g) + \alpha E_x^{\delta'} J_\alpha(\delta, x_1)$$

y como $u(x_1) = J_\alpha(\delta, x_1)$ se sigue que

$$u(x) \leq c(x,g) + \alpha \sum_y p_{xy}(g)\, u(y)$$

y ya que g fue arbitrario

$$u(x) \leq \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a)\, u(y) \right],$$

lo cual prueba (ii). ■

Definición 3.2.5. Una función $u \in B(\mathbf{X})$ se dice ser una *solución* de la *Ecuación de Optimalidad* si $u = Tu$, es decir,

$$u(x) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a)\, u(y) \right], \quad x \in \mathbf{X}. \tag{3.8}$$

Probaremos ahora el siguiente Teorema.

Teorema 3.2.6.

(a) J_α es la única solución acotada de la Ecuación de Optimalidad.

(b) f es α-óptima si, y sólo si, f minimiza el lado derecho de la Ecuación de Optimalidad, es decir,

$$J_\alpha(x) = c(x,f) + \alpha \sum_y p_{xy}(f)\, J_\alpha(y).$$

Demostración (a):

Como T es de contracción y $B(\mathbf{X})$ es completo, existe una única función $u \in B(\mathbf{X})$ tal que

$$u(x) = Tu(x) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a)\, u(y) \right].$$

Sea g el selector que alcanza el mínimo, entonces $u(x) = T_g u(x)$ y por el Lema 3.2.4(a)

$u(x) = J_\alpha(g,x)$, lo cual implica que g es α-óptima y así, que $u(x) = J_\alpha(x)$.

(b) Supongamos que f es α-óptima. Entonces por el Lema 3.2.4(b)

$$TJ_\alpha(f,x) = J_\alpha(f,x) = J_\alpha(x)$$

y así,

$$\min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a)\, J_\alpha(y) \right] = \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a)\, J_\alpha(f,y) \right]$$

y como por el Lema 3.2.6(a) $J_\alpha(f,x) = T_f J_\alpha(f,x)$,

$$c(x,f) + \alpha \sum_y p_{xy}(f)\, J_\alpha(y) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a)\, J_\alpha(y) \right].$$

Por lo tanto f minimiza el lado derecho de la Ecuación de Optimalidad.

Supongamos ahora que f minimiza el lado derecho de la Ecuación de Optimalidad.

Entonces,

$$J_\alpha(x) = T_f J_\alpha(x)$$

y por el Lema 3.2.4(a)

$$J_\alpha(x) = J_\alpha(f,x),$$

lo cual implica que f es α-óptima. ■

3.2.2. METODO DE APROXIMACIONES SUCESIVAS

Del Teorema 3.2.6 observamos que conociendo la función de valor óptimo J_α podriamos conocer la política α-óptima, esto es, la política estacionaria que cuando el proceso esta en estado x, elige un control que minimiza

$$c(x,a) + \alpha \sum_y p_{xy}(a) J_\alpha(y).$$

Discutimos enseguida un método para obtener J_α como un límite de la función de valor óptimo en $n-$etapas.

El método es como sigue:

Sea $u \in B(\mathbf{X})$, y definamos

$$J_{\alpha,0}(x) := u(x)$$

y para $t \geq 1$,

$$J_{\alpha,t}(x) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a) J_{\alpha,t-1}(y) \right].$$

Notemos que $J_{\alpha,t}$ es el costo mínimo esperado α-descontado en $t-$etapas con costo terminal $J_{\alpha,0}(y)$ si $x_t = y$.

En la siguiente proposición mostramos que $J_{\alpha,t}$ converge uniformemente a J_α cuando $t \to \infty$.

Proposición 3.2.7. Para cualquier función acotada $u \in B(\mathbf{X})$, $J_{\alpha,t}(x) \to J_\alpha(x)$ uniformemente cuando $t \to \infty$.

Demostración. Por el Teorema de punto fijo(T.6.4 Apéndice) y T.3.2.6(a). ■

3.3. CRITERIO EN COSTO DESCONTADO CON COSTOS NO ACOTADOS.

Suponemos aquí la condición 3.2.1(a) de la sección anterior, así como las siguientes hipótesis:

Condición 3.3.1.

(a) $c(x,a) \geq 0$ para todo $x \in \mathbf{X}$ y $a \in \mathbf{A}(x)$.

(b) $J_\alpha(x)$ es finito para cualquier estado $x \in \mathbf{X}$ y factor de descuento α.

Nótese que bajo la C.3.3.1(a), el costo total esperado $\alpha-$descontado

$$J_\alpha(\delta, x) = E_x^\delta \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right]$$

está bien definido (posiblemente como función en los reales extendidos) para todo $0 < \alpha < 1$ y $\delta \in \Delta$.

Verificamos enseguida que si $J_\alpha(x)$ es finito para cualquier $x \in \mathbf{X}$ y $0 < \alpha < 1$, entonces $J_\alpha(x)$ satisface la Ecuación de Optimalidad y que la política estacionaria determinada por ella, es una política óptima $\alpha-$descontada; veremos también que $J_\alpha(x)$ es la menor solución no-negativa de la Ecuación de Optimalidad.

3.3.1. LA ECUACION DE OPTIMALIDAD Y POLITICA $\alpha-$OPTIMA.

Teorema 3.3.2. Si se satisfacen las condiciones 3.2.1(a) y 3.3.1, entonces $J_\alpha(x)$ satisface la Ecuación de Optimalidad, es decir,

$$J_\alpha(x) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a) J_\alpha(y) \right], \quad x \in X. \tag{3.9}$$

Demostración. Sea $\delta = \{f_t\}$ una política arbitraria.
Se sigue del T.1.4(iii) y T.2.3 del Apéndice que

$$\begin{aligned} J_\alpha(\delta, x) &= c(x,\ f_0) + E_x^\delta E_x^\delta \left[\sum_{t=1}^{\infty} \alpha^t c(x_t,\ a_t) \mid x_0, a_0, x_1 \right] \\ &= c(x,\ f_0) + \alpha E_x^\delta E_{x_1}^{\delta'} \left[\sum_{t=1}^{\infty} \alpha^{t-1} c(x_t,\ a_t) \right] \end{aligned}$$

donde $\delta' = (f_1, \cdots, f_n, \cdots)$.
Luego,

$$J_\alpha(\delta, x) = c(x,\ f_0) + \alpha E_x^\delta J\left(\delta',\ x_1\right)$$

y como $J_\alpha(x) \leq J_\alpha\left(\delta', x\right)$ para todo $x \in \mathbf{X}$ tenemos que

$$J_\alpha(x_1) \leq J_\alpha\left(\delta', x_1\right),$$

de donde se sigue por el T.1.4(iv) del Apéndice que

$$\begin{aligned} J_\alpha(\delta, x) &\geq c(x,\ f_0) + \alpha E_x^\delta J_\alpha(x_1) \\ &= c(x, f_0) + \alpha \sum_y p_{xy}(f_0)\, J_\alpha(y) \end{aligned}$$

así,

$$J_\alpha(\delta, x) \geq \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a)\, J_\alpha(y) \right] \text{ para toda } \delta \in \Delta.$$

Por lo tanto,

$$J_\alpha(x) \geq \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a)\, J_\alpha(y) \right] \tag{3.10}$$

Recíprocamente, sea a_0 tal que

$$c(x, a_0) + \alpha \sum_y p_{xy}(a_0)\, J_\alpha(y) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a)\, J_\alpha(y) \right]. \tag{3.11}$$

Sea $\varepsilon > 0$ dado, y sea δ_0 la política que elige a_0 en el tiempo 0 y si el siguiente estado es y, consideremos δ_y de manera que

$$J_\alpha(\delta_y, y) \leq J_\alpha(y) + \varepsilon$$

y definamos $\delta = \delta_0$ en $t = 0$ y $\delta = \delta_y$ de $t = 1$ en adelante si $x_1 = y$.

Así,

$$\begin{aligned} J_\alpha(\delta, x) &= c(x, a_0) + \alpha \sum_y p_{xy}(a_0)\, J_\alpha(\delta_y,\ y) \qquad (3.12) \\ &\leq c(x, a_0) + \alpha \sum_y p_{xy}(a_0)\, J_\alpha(y) + \alpha\varepsilon. \end{aligned}$$

Como $J_\alpha(x) \leq J_\alpha(\delta, x)$, tenemos de (3.12) que

$$J_\alpha(x) \leq c(x, a_0) + \alpha \sum_y p_{xy}(a_0) J_\alpha(y) + \alpha\varepsilon$$

por lo tanto, de (3.11) se sigue que

$$J_\alpha(x) \leq \min_{a \in \mathbf{A}(x)} \left[c(x, a) + \alpha \sum_y p_{xy}(a) J_\alpha(y) \right] + \alpha\varepsilon,$$

haciendo $\varepsilon \to 0$ obtenemos

$$J_\alpha(x) \leq \min_{a \in \mathbf{A}(x)} \left[c(x, a) + \alpha \sum_y p_{xy}(a) J_\alpha(y) \right] \tag{3.13}$$

y (3.9) se sigue de (3.10) y (3.13).

Teorema 3.3.3. Supongamos que se cumplen las condiciones 3.2.1(a) y 3.3.1, esto es, se satisface la Ecuación de Optimalidad. Si f_α es un selector que minimiza la parte derecha de (3.9) , entonces f_α es una política estacionaria $\alpha-$óptima, es decir, $J_\alpha(f_\alpha, x) = J_\alpha(x)$ para todo $x \in \mathbf{X}$.

Demostración. Sea f_α un selector que minimiza la parte derecha de (3.9). Entonces

$$J_\alpha(x) = c(x, f_\alpha) + \alpha \sum_y p_{xy}(f_\alpha) J_\alpha(y) \text{ para todo } x \in \mathbf{X},$$

luego

$$J_\alpha(x) = c(x, f_\alpha) + \alpha \sum_y p_{xy}(f_\alpha) c(y, f_\alpha) + \alpha^2 \sum_y \sum_z p_{xy}(f_\alpha) p_{yz}(f_\alpha) J_\alpha(z)$$

lo cual se puede escribir como

$$J_\alpha(x) = E_x^{f_\alpha} \sum_{t=0}^{1} \alpha^t c(x_t, a_t) + \alpha^2 E_x^{f_\alpha} J_\alpha(x_2)$$

y siguiendo este mismo procedimiento, podemos ver que

$$J_\alpha(x) = E_x^{f_\alpha} \sum_{t=0}^{n-1} \alpha^t c(x_t, a_t) + \alpha^n E_x^{f_\alpha} J_\alpha(x_n). \tag{3.14}$$

Como $J_\alpha(x) \geq 0$ (todos los costos son no-negativos) vemos que $E_x^{f_\alpha} J_\alpha(x_n) \geq 0$, así

$$E_x^{f_\alpha}\left[\sum_{t=0}^{n-1}\alpha^t c\left(x_t,a_t\right)\right]\leq J_\alpha\left(x\right),$$

haciendo $n\to\infty$,

$$J_\alpha\left(f_\alpha,x\right)\leq J_\alpha\left(x\right)$$

lo cual prueba que f_α es α-óptima. ■

Veremos ahora que $J_\alpha\left(x\right)$ es la menor solución no-negativa de (3.9).

Proposición 3.3.4. Bajo las condiciones del Teorema anterior, $J_\alpha\left(x\right)$ es la mínima solución no-negativa de (3.9)

Demostración. Sea $u\left(x\right)$ una función no-negativa tal que

$$u\left(x\right)=\min_{a\in\mathbf{A}(x)}\left[c\left(x,a\right)+\alpha\sum_y p_{xy}\left(a\right)u(y)\right]. \tag{3.15}$$

Si g es la política estacionaria determinada por (3.15), entonces

$$\begin{aligned} c\left(x,g\right)+\alpha\sum_y p_{xy}\left(g\right)u(y) &= \min_{a\in\mathbf{A}(x)}\left[c\left(x,a\right)+\alpha\sum_y p_{xy}\left(a\right)u(y)\right] \\ &= u\left(x\right). \end{aligned}$$

Por los mismos argumentos dados para obtener (3.14), se deduce

$$u\left(x\right)=E_x^g\left[\sum_{t=0}^{n-1}\alpha^t c\left(x_t,a_t\right)\right]+\alpha^n E_x^g u\left(x_n\right),$$

lo cual implica, puesto que $u\geq 0$, que

$$E_x^g\left[\sum_{t=0}^{n-1}\alpha^t c\left(x_t,a_t\right)\right]\leq u\left(x\right),$$

haciendo $n\to\infty$,

$$J_\alpha\left(g,x\right)\leq u\left(x\right).$$

Como $J_\alpha\left(x\right)\leq J_\alpha\left(g,x\right)$, el Teorema queda probado. ■

3.3.2. METODO DE APROXIMACIONES SUCESIVAS.

Sea

$$J_{\alpha,0}(x) = 0$$

y para $t \geq 1$,

$$J_{\alpha,t}(x) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a) J_{\alpha,t-1}(y) \right], \quad x \in \mathbf{X}.$$

Observemos que como los costos son no-negativos, $J_{\alpha,t}(x) \leq J_{\alpha,t+1}(x)$.
Definamos $J_\infty(x) = \lim_{t\to\infty} J_{\alpha,t}(x)$. Entonces, como

$$J_{\alpha,t}(x) \leq J_\alpha(x) \text{ para todo } t,$$

vemos que

$$J_\infty(x) \leq J_\alpha(x).$$

Teorema 3.3.5. Bajo las condiciones 3.2.1(a) y 3.3.1, $J_\infty(x) = J_\alpha(x), \quad x \in \mathbf{X}$.

Demostración. Ya mostramos que $J_\infty \leq J_\alpha$.
Para probar la otra desigualdad demostraremos que J_∞ satisface la Ecuación de Optimalidad y así, por la proposición 3.3.4, $J_\alpha \leq J_\infty$.
Como cada $\mathbf{A}(x)$, $x \in \mathbf{X}$ es finito se sigue que

$$\begin{aligned} J_\infty(x) &= \lim_{t\to\infty} J_{\alpha,t}(x) \\ &= \lim_{t\to\infty} \min_{a\in\mathbf{A}(x)} \left[c(x,a) + \alpha \sum_y p_{xy}(a) J_{\alpha,t-1}(x) \right] \\ &= \min_{a\in\mathbf{A}(x)} \left\{ \lim_{t\to\infty} \left[c(x,a) + \alpha \sum_y p_{xy}(a) J_{\alpha,t-1}(y) \right] \right\}, \end{aligned}$$

y usando el Teorema de Convergencia Monótona (puesto que $\{J_{\alpha,t}(x)\}_{t=0}^{\infty}$ es una sucesión creciente) obtenemos,

$$\begin{aligned} J_{\infty}(x) &= \min_{a\in \mathbf{A}(x)}\left[c(x,a)+\alpha\sum_{y}p_{xy}(a)\lim_{t\to\infty}J_{\alpha,t-1}(y)\right] \\ &= \min_{a\in \mathbf{A}(x)}\left[c(x,a)+\alpha\sum_{y}p_{xy}(a)J_{\infty}(y)\right] \end{aligned}$$

Por lo tanto J_{∞} satisface la Ecuación de Optimalidad, lo cual prueba lo deseado. ■

3.4. EJEMPLO: UN MODELO DE REEMPLAZO DE MAQUINAS.

Dependiendo del estado de cierta máquina, debemos tomar la decisión de cambiarla o no por una nueva. Si tomamos la desición de reemplazarla, obtenemos un costo R y el estado del siguiente período de tiempo es 0, el estado de una máquina nueva. Si el estado actual es $x \in \mathbf{X} = \{0,\ 1,...\}$ y hemos decidido no cambiarla, entonces la probabilidad de que el siguiente estado sea y es p_{xy}. Además, ya que en el inicio de cada periodo de tiempo la máquina se encuentra en un cierto estado x, en cada periodo de tiempo existe un costo operativo $c(x)$.

Si $J_{\alpha}(x)$ es el mínimo costo total esperado α-descontado, dado que el estado inicial es x. Entonces J_{α} satisface la Ecuación de Optimalidad

$$J_{\alpha}(x) = c(x) + \min\left[R+\ \alpha J_{\alpha}(0)\,,\ \alpha\sum_{y}p_{xy}J_{\alpha}(y)\right].$$

Parece razonable esperar que J_{α} sea una función creciente de x. Veremos que esto es efectivamente así, bajo las siguientes condiciones:
(i) $c(x)$ es creciente en x.
(ii) Para cada k, $\sum_{y=k}^{\infty}p_{xy}$ crece en x.

Proposición 3.4.1. Bajo las condiciones (i) y (ii), $J_{\alpha}(x)$ es creciente.

Demostración.
Sea

$$J_{\alpha,1}(x) = c(x)$$

y para $t > 1$

$$J_{\alpha,t}(x) = c(x) + \min\left[R + \alpha J_{\alpha,t-1}(0), \alpha \sum_y p_{xy} J_{\alpha,t-1}(y)\right].$$

De la condición (i), se sigue que $J_{\alpha,1}(x)$ es creciente en x y suponiendo que $J_{\alpha,t-1}(y)$ es creciente en y, de la condición 2, vemos que $\sum_y p_{xy} J_{\alpha,t-1}(y)$ crece en x, y así, $J_{\alpha,t}(x)$ crece en x. Por lo tanto, por inducción $J_{\alpha,t}(x)$ es creciente para toda t, y como $J_\alpha(x) = \lim_t J_{\alpha,t}(x)$, se sigue que $J_\alpha(x)$ es creciente. ■

Veremos ahora que:

Proposición 3.4.2. Bajo las condiciones (i) y (ii), existe un $\widehat{x}$, $\widehat{x} \leq \infty$, tal que la política α-óptima reemplaza cuando el estado es x si $x \geq \widehat{x}$ y no reemplaza si $x < \widehat{x}$.

Demostración. Se sigue de la Ecuación de Optimalidad que es óptimo reemplazar en x si

$$\alpha \sum_y p_{xy} J_\alpha(y) \geq R + \alpha J_\alpha(0).$$

Puesto que $J_\alpha(y)$ es creciente en y, vemos que

$$\sum_y p_{xy} J_\alpha(y) \text{ es creciente en } x.$$

Por lo tanto

$$\widehat{x} = \min\left[x : \alpha \sum_y p_{xy} J_\alpha(y) \geq R + \alpha J_\alpha(0)\right],$$

con $\widehat{x} = \infty$ si el conjunto anterior es vacio. ■

4. EL PCO CON HORIZONTE DE PLANEACION INFINITO EN COSTO PROMEDIO.

4.1. INTRODUCCION.

En este capítulo consideramos el PCO con índice de funcionamiento dado por

$$V(\delta, x) := \limsup_{n\to\infty} \frac{J_n(\delta, x)}{n} \tag{4.1}$$

donde

$$J_n(\delta, x) := E_x^{\delta}\left[\sum_{t=0}^{n-1} c(x_t, a_t)\right]$$

es el costo total esperado para las primeras $n-$etapas usando la política δ y el estado inicial $x_0 = x$.

Así,

$$\frac{J_n(\delta, x)}{n}$$

es el costo promedio esperado por unidad de tiempo y nos referiremos a (4.1) simplemente como el *costo promedio(*CP*)*. En este contexto a una política δ^* tal que

$$V(\delta^*, x) = \inf_{\delta} V(\delta, x) \text{ para todo } x \in \mathbf{X}$$

le llamaremos *política* CP-*óptima(*o simplemente óptima) y como en el capítulo anterior,

$$V(x) := \inf_{\delta} V(\delta, x)$$

es la *función de valor óptimo*.

En la sección 4.2 presentamos dos contraejemplos: el primero de ellos nos muestra que no necesariamente existe una política óptima y el segundo nos muestra que en el caso promedio, no es suficiente restringirse a las políticas estacionarias.

En la sección 4.3 estudiamos el Criterio en Costo Promedio con Costos Acotados y damos condiciones bajo las cuales existe una política estacionaria óptima, así como una condición bajo la cual el Problema en Costo Promedio puede reducirse al Criterio en Costo Descontado. En la sección 4.4 abordamos el Criterio en Costo Promedio con Costos no-Acotados, y en la sección 4.5 damos un ejemplo(un modelo de colas).

4.2. CONTRAEJEMPLOS

El siguiente ejemplo muestra que para el criterio en costo promedio no necesariamente existe una política óptima.

Contraejemplo 4.2.1. Consideremos un proceso con espacio de estados $\mathbf{X} = \mathbf{Z} - \{0\}$, conjunto de acciones $\mathbf{A} = \{1, 2\}$ y con probabilidades de transición

$p_{x,x+1}(1) = p_{x,-x}(2) = 1, \quad x \geq 1,$

$p_{-x,-x}(1) = p_{-x,-x}(2) = 1, \quad x \geq 1.$

Y costos dados pòr

$c(x, a) = 1, \quad x \geq 1, a \in \mathbf{A}(x)$

$c(-x, a) = \dfrac{1}{x}, \quad x \geq 1, a \in \mathbf{A}(x).$

Es decir, podemos decidir pasar del estado x al estado $x+1$ con un costo igual a 1 o bien, podemos decidir pasar al estado $-x$ con un costo igual a $\dfrac{1}{x}$ para cada periodo de tiempo posterior.

Obviamente, $V(\delta, 1) > 0$ para cualquier política δ. Sin embargo, podemos obtener un costo arbitrariamente cercano a cero escogiendo la acción 2 por primera vez en el periodo n (para n grande), así

$$\inf_{\delta} V(\delta, 1) = 0$$

por lo tanto, no existe $\delta^* \in \Delta$ tal que

$$V(\delta^*, 1) = \inf_{\delta} V(\delta, 1)$$

Esto muestra que para el caso promedio no necesariamente existen las políticas óptimas.

Contraejemplo 4.2.2.

Consideremos ahora un proceso con espacio de estados los enteros positivos y con unicamente 2 acciones admisibles, tal que las probabilidades de transición y costos son dados por

$p_{x,x+1}(1) = 1 = p_{x,x}(2)$

$c(x,1) = 1,$

$c(x,2) = \frac{1}{x}.$

Es decir, el sistema se mueve del estado x al estado $x+1$ con un costo igual a 1, o bien, permanece en x con un costo igual a $\frac{1}{x}$.

Si $x_0 = 1$ y f es una política estacionaria, existen 2 posibilidades:
(i) Que f siempre elija el control 1. En cuyo caso, $V(f,1) = 1$.
(ii) Que f elija el control 2 por vez primera en el estado n. En cuyo caso, el proceso va del estado 1 al estado n y como en este estado f elige la acción 2, el proceso nunca deja tal estado y siempre se obtiene un costo igual a $\frac{1}{n}$ en todos los pasos siguientes. Por lo tanto $V(f,1) = \frac{1}{n} > 0$.

Si ahora consideramos una política no-estacionaria δ, tal que en cada estado x, elija x veces consecutivas el control 2 y luego elija el control 1, se sigue(ya que el estado inicial es $x_0 = 1$) que los costos son dados por la sucesión:

$$1, 1, \frac{1}{2}, \frac{1}{2}, 1, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 1, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}, 1, \frac{1}{5}, \cdots.$$

Como el valor promedio de esta sucesión es igual a 0, $V(\delta,1) = 0$ y por lo tanto la política no-estacionaria δ es estrictamente mejor que cualquier política estacionaria. De esta manera, hemos probado que no es suficiente restringirse a las políticas estacionarias.

4.3. CRITERIO EN COSTO PROMEDIO CON COSTOS ACOTADOS.

Reasumimos aquí la condición 3.2.1:
(a) $\mathbf{A}(x)$ es finito para cada $x \in \mathbf{X}$.
(b) $|c(x,a)| < M, \quad \forall\, x \in \mathbf{X}$ y $a \in \mathbf{A}(x)$, para alguna constante positiva M, y presentamos un teórema que puede considerarse como la versión en Costo Promedio de la Ecuación de Optimalidad.

Teorema 4.3.1. Si existen una función acotada $H(x)$, $x \in \mathbf{X}$ y una constante g tal que

$$g + H(x) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \sum_{y=0}^{\infty} p_{xy}(a) H(y) \right], \quad x \in \mathbf{X}, \tag{4.2}$$

entonces cualquier selector f^* que minimize la parte derecha de (4.2) es tal que

$$g = V(f^*, x) = \inf_{\delta} V(\delta, x) \text{ para todo } x \in \mathbf{X}.$$

Demostración. Sea $h_t = (x_0,\ a_0, \cdots,\ x_t,\ a_t)$ la historia del proceso hasta el tiempo t.

Por T.1.4(iii) y T.2.3 del Apéndice, se sigue que, para cualquier política $\delta \in \Delta$,

$$E_x^{\delta}\left\{\sum_{t=0}^{n}\left[H(x_{t+1}) \ - E_x^{\delta}(H(x_{t+1}) \mid h_t)\right]\right\} = 0. \tag{4.3}$$

Pero,

$$\begin{aligned} E_x^{\delta}[H(x_{t+1}) \mid h_t] &= \sum_{y=0}^{\infty} H(y) p_{x_t y}(a_t) \\ &= c(x_t,\ a_t) + \sum_{y=0}^{\infty} H(y) p_{x_t y}(a_t) - c(x_t, a_t) \qquad (4.4) \\ &\geq \min_{a \in \mathbf{A}(x)}\left[c(x_t, a) \ + \ \sum_{y=0}^{\infty} H(y) p_{x_t y}(a)\right] - c(x_t, a_t) \\ &= g + H(x_t) - c(x_t, a_t), \end{aligned}$$

por lo tanto, sustituyendo esto en (4.3) obtenemos

$$0 \leq E_x^{\delta}\left\{\sum_{t=0}^{n}[H(x_{t+1}) - g - H(x_t) + c(x_t, a_t)]\right\}$$

o lo que es lo mismo

$$g \leq E_x^{\delta}\left[\frac{H(x_{n+1})}{n+1}\right] - E_x^{\delta}\left[\frac{H(x_0)}{n+1}\right] + E_x^{\delta}\left[\frac{\sum_{t=0}^{n} c(x_t, a_t)}{n+1}\right].$$

Haciendo $n \to \infty$ y usando el hecho de que H es acotada, obtenemos que

$$g \leq V(\delta, x) \text{ para todo } x \in \mathbf{X}.$$

Si suponemos en los cálculos anteriores que $\delta = f^*$, es la política que minimiza la parte derecha de (4.2), entonces obtenemos la igualdad en (4.4) y por lo tanto, también obtenemos la igualdad en los siguientes pasos, es decir

$$g = V(f^*, x) \text{ para todo } x \in \mathbf{X}.$$

De esta manera, el teorema queda demostrado. ■

Observemos que si se satisfacen las condiciones del teorema anterior, entonces existe una política estacionaria óptima, la cual puede ser caracterizada por la ecuación funcional (4.2).

En lo que resta de esta sección, estudiaremos condiciones suficientes para la existencia de las hipótesis del Teorema anterior.

Recordemos que bajo la condición 3.2.1,

$$J_\alpha(x) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) \ + \alpha \sum_y p_{xy}(a) J_\alpha(y) \right], \quad x \in \mathbf{X}. \tag{4.5}$$

Teorema 4.3.2. Si existe un $N < \infty$ tal que $|J_\alpha(x) \ - J_\alpha(0)| < N$ para todo $0 < \alpha < 1$ y todo $x \in \mathbf{X}$, entonces existen una función acotada $H(x)$ y una constante g tal que

$$g + H(x) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) \ + \sum_{y=0}^{\infty} p_{xy}(a) H(y) \right], \quad x \in \mathbf{X}.$$

Demostración. Sea $H_\alpha(x) = J_\alpha(x) - J_\alpha(0)$.

Por hipótesis $H_\alpha(x)$ es uniformemente acotada en x y α. Como cualquier sucesión acotada posee una subsucesión convergente (T. Bolzano-Weierstrass), existe una subsucesión $\alpha_{1,n} \to 1$ tal que $\lim_{n\to\infty} H_{\alpha_{1,n}}(1) = H(1)$ existe. Similarmente, como la sucesión $H_{\alpha_{1,n}}(2) : n \geq 1$ es acotada, existe una subsucesión $\{\alpha_{2,n}\}$ de $\{\alpha_{1,n}\}$ tal que $\lim_{n\to\infty} H_{\alpha_{2,n}}(2) = H(2)$ existe. Continuando el proceso, se obtiene una subsucesión $\{\alpha_{3,n}\}$ de $\{\alpha_{2,n}\}$ tal que $\lim_{n\to\infty} H_{\alpha_{3,n}}(3) = H(3)$ existe, y así se continua. Tomando ahora $\alpha_n = \alpha_{n,n}$, vemos que $H_{\alpha_n}(x) \to H(x)$ cuando $n \to \infty$ para cada $x \in \mathbf{X}$, ya que $\{\alpha_n\}$ es una subsucesión de cada sucesión $\alpha_{k,n} : k = 1, 2, \cdots$. Y como los costos son acotados, se sigue que $(1 - \alpha_n) J_{\alpha_n}(0)$ es acotado, por lo tanto existe una subsucesión $\{\alpha_{\overline{n}}\}$ de $\{\alpha_n\}$ para la cual $\lim_{n\to\infty} (1 - \alpha_{\overline{n}}) J_{\alpha_{\overline{n}}}(0) = g$ existe. Así, de (4.5) obtenemos que

$$(1-\alpha_{\overline{n}})\,J_{\alpha_{\overline{n}}}(0)\,+H_{\alpha_{\overline{n}}}(x)=\min_{a\in\mathbf{A}(x)}\left[c(x,a)\;+\;\alpha_{\overline{n}}\sum_{y=0}^{\infty}p_{xy}(a)\,H_{\alpha_{\overline{n}}}(y)\right].$$

De aquí, haciendo $n\to\infty$ y notando que como $H_{\alpha_{\overline{n}}}(y)$ es acotada,

$$\sum_{y=0}^{\infty}p_{xy}(a)\,H_{\alpha_{\overline{n}}}(y)\to\sum_{y=0}^{\infty}p_{xy}(a)\,H(y)\text{ cuando }n\to\infty\text{ (T.4.3 Apéndice)}$$

el resultado se sigue. ■

Observación: como $(1-\alpha)\,J_{\alpha}(0)$ es acotado, para cualquier sucesión $\{\alpha_n\}$ existe una subsucesión $\{\alpha_{\overline{n}}\}$ tal que $\lim_{n\to\infty}(1-\alpha_{\overline{n}})\,J_{\alpha_{\overline{n}}}(0)$ existe, y por la prueba del Teorema anterior se sigue que este límite debe ser g. Por lo tanto, $g=\lim_{\alpha\to 1}(1-\alpha)\,J_{\alpha}(0)$.

Para la demostración del siguiente Teorema necesitamos el siguiente resultado, conocido como desigualdad de Jensen y cuya prueba damos en el Apéndice.

Lema 4.3.3. Sean f una función convexa y Y una variable aleatoria. si $E[Y]$ y $E[f(Y)]$ existen, entonces

$$E[f(Y)]\geq f(E[Y]).$$

El siguiente Teorema nos da una condición suficiente para que $J_{\alpha}(x)-J_{\alpha}(0)$ sea uniformemente acotado, lo cual implica la conclusión del Teorema 4.3.2.

Sea $m_{x0}(f)$ el tiempo medio para ir del estado x al estado 0 usando la política f.

Teorema 4.3.4. Sea $\alpha\in(0,1)$ y f_{α} una política $\alpha-$óptima. Si existe una constante $N<\infty$ tal que $m_{x0}(f_{\alpha})<N$ para todo $0<\alpha<1$ y $x\in\mathbf{X}$, entonces $J_{\alpha}(x)-J_{\alpha}(0)$ es uniformemente acotado.

Demostración. Puesto que estamos considerando costos acotados, podemos sin pérdida de generalidad suponer que

$$0\leq c(x,a)\leq M\text{ para todo }x\in\mathbf{X}\text{ y }a\in\mathbf{A}(x).$$

Sea $T:=\min\{t:x_t=0\}$, entonces

$$\begin{aligned}J_{\alpha}(x)&=E_x^f\left[\sum_{t=0}^{T-1}\alpha^t c(x_t,\,a_t)\right]+E_x^f\left[\sum_{t=T}^{\infty}\alpha^t c(x_t,a_t)\right]\\&\leq ME_x^f(T)+J_{\alpha}(0)\,E_x^f\left(\alpha^T\right)\\&\leq MN+J_{\alpha}(0).\end{aligned}\tag{4.6}$$

Por otra parte, ya que

$$J_\alpha(x) \geq J_\alpha(0)\, E_x^f\left(\alpha^T\right),$$

obtenemos que

$$J_\alpha(0) \leq J_\alpha(x) + \left[1 - E_x^f\left(\alpha^T\right)\right] J_\alpha(0)$$

y así, como

$$J_\alpha(0) \leq \frac{M}{1-\alpha} \quad \text{y } E\left(\alpha^T\right) \geq \alpha^{E(T)} \geq \alpha^N$$

(desigualdad de Jensen) se sigue que

$$\begin{aligned} J_\alpha(0) &\leq J_\alpha(x) + \left(1-\alpha^N\right)\tfrac{M}{1-\alpha} \\ &\leq J_\alpha(x) + MN \end{aligned} \tag{4.7}$$

pues

$$\frac{1-\alpha^N}{1-\alpha} = 1 + \alpha + ... + \alpha^{N-1} \leq N.$$

De las desigualdades (4.6) y (4.7) obtenemos la desigualdad deseada. ■

ENFOQUE DE REDUCCION AL CASO DESCONTADO.

Si en un proceso dado, además de la condición 3.2.1 se cumple la siguiente condición:

Condición 4.3.5. Existe un estado, el cual denotamos por 0 y un número $0 < \beta < 1$ tal que

$$p_{x0}(a) \geq \beta, \quad \forall\, x \in \mathbf{X},\, a \in \mathbf{A}.$$

Entonces podemos reducir el problema en Costo Promedio a uno en Costo Descontado, donde podemos emplear el Método de Aproximaciones Sucesivas para la función de valor óptimo $J_\alpha(x)$. Este enfoque de reducción es como sigue: dado un proceso que cumple la condición 3.2.1 y la condición 4.3.5, consideramos un nuevo proceso con el mismo espacio de estados y espacio de controles, así como los mismos costos; pero con probabilidades de transición dadas por

$$\overline{p}_{xy}(a) = \begin{cases} \dfrac{p_{xy}(a)}{1-\beta}, & y \neq 0 \\ \dfrac{p_{x0}(a)-\beta}{1-\beta}, & y = 0. \end{cases} \tag{4.8}$$

Tenemos entonces que

Teorema 4.3.6. La política $(1-\beta)$-óptima de este nuevo proceso es la política CP-óptima del proceso original.

Demostración. Sea

$$\alpha := 1-\beta$$

y denotemos por $\overline{J}_\alpha(x)$ a la función de valor óptimo para este nuevo proceso.

Haciendo

$$H_\alpha(x) = \overline{J}_\alpha(x) - \overline{J}_\alpha(0),$$

obtenemos de (4.5) que

$$\begin{aligned} \beta\overline{J}_\alpha(0) + H_\alpha(x) &= \min_{a\in\mathbf{A}(x)}\left[c(x,a) + \alpha\sum_y \overline{p}_{xy}(a)\,H_\alpha(y)\right] \\ &= \min_{a\in\mathbf{A}(x)}\left[c(x,a) + \sum_y p_{xy}(a)\,H_\alpha(y)\right], \end{aligned} \tag{4.9}$$

donde la última ecuación se sigue de (4.8) puesto que $H_\alpha(0) = 0$. Así, por el Teorema 4.3.1, $g = \beta\overline{J}_\alpha(0)$, y la política óptima en costo promedio es la que elige el control que minimiza la parte derecha de (4.9). Pero esta es precisamente la política $(1-\beta)$ –óptima del nuevo proceso. ■

4.4. CRITERIO EN COSTO PROMEDIO CON COSTOS NO ACOTADOS.

En esta sección asumimos las hipótesis dadas en C.3.2.1(a) y C.3.3.1:

(a) $\mathbf{A}(x)$ es finito para cada $x \in \mathbf{X}$.

(b) $c(x,a) \geq 0$ para todo $x \in \mathbf{X}$ y $a \in \mathbf{A}(x)$.

(c) $J_\alpha(x)$ es finito para cualquier estado $x \in \mathbf{X}$ y factor de descuento α.
Así, como las siguiente condición:

Condición 4.4.1.

(a) Existe una constante no-negativa N tal que

$$-N \leq H_\alpha(x), \quad \forall\, x \in \mathbf{X},\, \alpha \in (0,1),$$

donde $H_\alpha(x) := J_\alpha(x) - J_\alpha(0)$.

(b) Existe una función no-negativa $b : \mathbf{X} \to \mathbf{R}$ tal que

$$H_\alpha(x) \leq b(x), \quad \forall\, x \in \mathbf{X},\, \alpha \in (0,1).$$

(c) Para cada $x \in \mathbf{X}$, existe una acción $a(x)$ tal que $\sum_y p_{xy}(a(x))\, b(y) < \infty$.

Veremos aquí que bajo estas condiciones existe una política estacionaria óptima en costo promedio; pero la Ecuación de Optimalidad (4.2) puede no tenerse. Sin embargo, si además de las hipótesis anteriores se cumple lo siguiente:

Condición 4.4.2. $\sum_y p_{xy}(a)\, b(y) < \infty, \quad \forall\, x \in \mathbf{X},\, a \in \mathbf{A}(x)$.

Entonces se cumple la Ecuación de Optimalidad (4.2).

Lema 4.4.3. Si $\{\alpha_n\} \subset (0,1)$ es una sucesión de factores de descuento tal que $\alpha_n \uparrow 1$ y $\{f_{\alpha_n}\}$ es una sucesión de políticas estacionarias óptimas α_n−descontadas, entonces existe una sucesión $\{\beta_n\} \subset \{\alpha_n\}$ y una política estacionaria f tal que

$$f(x) = \lim_{n\to\infty} f_{\beta_n}(x) \quad \forall x \in \mathbf{X}. \tag{4.10}$$

Demostración. Como para cada $x \in \mathbf{X}$, $\mathbf{A}(x)$ es un conjunto compacto con respecto a la topologia discreta, puesto que es un conjunto finito (Ver C.3.2.1(a)). se sigue del Teorema de Tychonoff(Ver Ash 1972, p.383) que el conjunto

$$\prod_{x\in\mathbf{X}} \mathbf{A}(x) \tag{4.11}$$

es compacto. Por otro lado, toda política estacionaria puede considerarse como un punto del espacio (4.11), de manera que$\{f_{\alpha_n}\}$ es una sucesión en (4.11) y por lo tanto admite una subsucesión convergente; es decir, existe una sucesión $\{\beta_n\}$ y

una política estacionaria f que satisfacen la relación (4.10). ■

Lema 4.4.4. Si existen una constante g, una función $H(x)$, con $-N \leq H(x)$ para todo $x \in \mathbf{X}$ y una política estacionaria f tal que

$$g + H(x) \geq c(x, f) + \sum_y p_{xy}(f) H(y), \quad x \in \mathbf{X}, \tag{4.12}$$

entonces ,

$$V(f, x) \leq g \text{ para todo } x \in \mathbf{X}.$$

Si en (4.12) se tiene la desigualdad contraria y $H(x) \leq N$ para todo x, entonces

$$V(f, x) \geq g \text{ para todo } x \in \mathbf{X}.$$

Demostración. Sean $x_0 = x$, x_1, x_2, ... la sucesión de estados del proceso bajo la política estacionaria f. De (4.12) se sigue que

$$g + H(x_t) \geq c(x_t, f) + E_x^f(H(x_{t+1}) \mid x_t), \quad t \geq 0. \tag{4.13}$$

Mostraremos por inducción que

$$E_x^f(H(x_t)) \leq tg + H(x), \quad t \geq 0.$$

El resultado es obviamente cierto para $t = 0$. Supongamos que es válido para t.
De (4.13)

$$E_x^f(H(x_{t+1}) \mid x_t) \leq g + H(x_t),$$

por lo tanto, tomando E_x^f en ambos lados y usando la hipótesis de inducción,

$$\begin{aligned} E_x^f(H(x_{t+1})) &\leq g + E_x^f(H(x_t)) \\ &\leq g + tg + H(x) \\ &= (t+1)g + H(x). \end{aligned}$$

Tomando E_x^f en ambos lados de (4.13) obtenemos que

$$E_x^f(c(x_t,\ f(x_t))) \leq g + E_x^f(H(x_{t+1})),\ t \geq 0. \tag{4.14}$$

Sumando los términos de (4.14) desde $t = 0, \cdots, n-1$ y dividiendo por n obtenemos

$$\begin{aligned}\sum_{t=0}^{n-1}\frac{E_x^f\left(c\left(x_t,f\right)\right)}{n} &\leq g+\frac{H\left(x\right)}{n}-\frac{E_x^f\left(H\left(x_n\right)\right)}{n}\\ &\leq g+\frac{H\left(x\right)}{n}+\frac{N}{n}.\end{aligned}$$

Tomando el limsup en ambos lados se obtiene el resultado. La demostración de la segunda afirmación es similar. ■

Lema 4.4.5. Para cualquier política $\delta \in \Delta$,

$$\limsup_{\alpha\uparrow 1}\left(1-\alpha\right)J_\alpha\left(x\right)\leq V\left(\delta,x\right),\ x\geq 0.$$

Demostración. Para cada $x \in \mathbf{X}$ y $\delta \in \Delta$ se sigue por el Teorema Tauberiano(T.7.2 del Apéndice) tomando

$$c_t := E_x^\delta c\left(x_t,a_t\right) \quad \text{y} \quad s_n := \sum_{t=0}^{n-1} E_x^\delta c\left(x_t,a_t\right),$$

que

$$\begin{aligned}\limsup_{\alpha\uparrow 1}\left(1-\alpha\right)J_\alpha\left(x\right) &\leq \limsup_{\alpha\uparrow 1}\left(1-\alpha\right)J_\alpha\left(\delta,x\right)\\ &= \limsup_{\alpha\uparrow 1}\left(1-\alpha\right)E_x^\delta\left[\sum_{t=0}^{\infty}\alpha^t c\left(x_t,a_t\right)\right]\\ &= \limsup_{\alpha\uparrow 1}\left(1-\alpha\right)\sum_{t=0}^{\infty}\alpha^t E_x^\delta c\left(x_t,a_t\right)\\ &= \limsup_{\alpha\uparrow 1}\left(1-\alpha\right)\sum_{t=0}^{\infty}\alpha^t c_t\\ &\leq \limsup_{n\to\infty}\frac{s_n}{n}\\ &= V\left(\delta,x\right).\end{aligned}$$

Donde la primera desigualdad se debe a que $J_\alpha\left(x\right)\leq J_\alpha\left(\delta,x\right)$, $\delta\in\Delta$ y la segunda igualdad es por el Teorema de Convergencia Monótona (T.4.1 del Apéndice), pues

$c(x,a) \geq 0, \quad \forall\, x \in \mathbf{X},\, a \in \mathbf{A}.$ ■

Teorema 4.4.6. Supongamos que se satisfacen las hipótesis dadas en C.3.2.1(a), C.3.3.1 y C.4.4.1. Entonces
(a) existen un selector $f \in \mathbf{F}$, una constante g y una función $H(x)$ tal que

$$-N \leq H(x) \leq b(x) \quad \forall\, x \in \mathbf{X}$$

que satisfacen

$$\begin{aligned} g + H(x) &\geq c(x,f) + \sum_y p_{xy}(f)\, H(y) \qquad (4.15) \\ &\geq \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \sum_y p_{xy}(a)\, H(y) \right], \quad x \in \mathbf{X}. \end{aligned}$$

(b) El selector f es una política estacionaria óptima en costo promedio con un costo promedio g. Más aún, cualquier política que alcance el mínimo en (4.15) es óptima en costo promedio.
(c) Si además se cumple la condición 4.4.2, entonces g y $H(x)$ satisfacen la Ecuación de Optimalidad

$$g + H(x) = \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \sum_y p_{xy}(a)\, H(y) \right], \quad x \in \mathbf{X}. \qquad (4.16)$$

Demostración. (a). Sean $\{\alpha_n\}$, $\{\beta_n\}$ y f como en el Lema 4.4.3. Para cada factor de descuento α, $(H_\alpha(x))_x$ es un punto en el espacio producto

$$\prod_x [-N, b(x)]. \qquad (4.17)$$

Por el Teorema de Tychonoff, el producto de espacios compactos es compacto, por lo tanto, existe una sucesión $\{\delta_n\} \subset \{\beta_n\}$ y un punto $(H(x))_x$ en el espacio (4.17) tal que $\lim_{n\to\infty} H_{\delta_n}(x) = H(x)$ para cada x.

Usando el T.3.3.2 y T.3.3.3, tenemos que

$$\begin{aligned} 0 &\leq (1-\alpha)\, J_\alpha(x) \\ &= c(x, f_\alpha) + \alpha \sum_y p_{xy}(f_\alpha)\, H_\alpha(y) - H_\alpha(x) \end{aligned}$$

$$\begin{aligned} &\leq c(x,a) + \alpha \sum_y p_{xy}(a(x)) H_\alpha(y) - H_\alpha(x) \qquad (4.18) \\ &\leq c(x,a(x)) + \sum_y p_{xy}(a(x)) b(y) + N, \quad x \in \mathbf{X}. \end{aligned}$$

donde la última desigualdad se sigue de la condición 4.4.1.

La parte derecha de (4.18) es un número D_x. Por lo tanto, $((1-\alpha) J_\alpha(x))_x$ es un punto en el espacio compacto

$$\prod_x [0, D_x]. \qquad (4.19)$$

Por lo tanto, existe una sucesión $\{\varepsilon_n\} \subset \{\delta_n\}$, y un punto $(g(x))_x$ en el espacio (4.19) tal que $\lim_{n\to\infty} (1-\varepsilon_n) J_{\varepsilon_n}(x) = g(x)$ para cada x.
Entonces

$$\begin{aligned} 0 &\leq |g(x) - g(0)| \\ &\leq \lim_n (1-\varepsilon_n) |H_{\varepsilon_n}(x)| \\ &\leq \lim_n (1-\varepsilon_n) \max\{N, b(x)\} = 0. \end{aligned}$$

Por lo tanto, $g(x)$ es una constante g.

Fijemos un estado $x \in \mathbf{X}$. Como $\{\varepsilon_n\}$ es una subsucesión de $\{\beta_n\}$, tenemos $f_{\varepsilon_n}(x) = f(x)$ para n suficientemente grande y por lo tanto de (3.1)

$$(1-\varepsilon_n) J_{\varepsilon_n}(x) + H_{\varepsilon_n}(x) = c(x,f) + \varepsilon_n \sum_y p_{xy}(f) H_{\varepsilon_n}(y). \qquad (4.20)$$

Haciendo $n \to \infty$ en (4.20) obtenemos

$$g + H(x) = c(x,f) + \lim_{n\to\infty} \sum_y p_{xy}(f) H_{\varepsilon_n}(y) \qquad (4.21)$$

y por lo tanto, el límite de la parte derecha de (4.21) existe y así, aplicando el Lema de Fatou (L.4.2 del Apéndice) obtenemos

$$\begin{aligned} g + H(x) &\geq c(x,f) + \sum_y p_{xy}(f) H(y) \\ &\geq \min_{a \in \mathbf{A}(x)} \left[c(x,a) + \sum_y p_{xy}(a) H(y) \right], \ x \in \mathbf{X}. \qquad (4.22) \end{aligned}$$

Lo cual prueba (4.15).

(b). De los Lemas 4.4.4 y 4.4.5 tenemos

$$\begin{aligned} V(f,x) &\leq g = \lim_{n\to\infty}(1-\varepsilon_n)\,J_{\varepsilon_n}(x) \\ &\leq \limsup_{\alpha\uparrow 1}(1-\alpha)\,J_\alpha(x) \\ &\leq V(\delta,x) \text{ para cualquier política } \delta\in\Delta. \end{aligned}$$

Esto prueba que f es óptima en costo promedio. Haciendo $\delta = f$ mostramos que g es el costo promedio mínimo. Además, por el Lema 4.4.4, cualquier política que alcance el mínimo en (4.22) es óptima en costo promedio.

(c). Para probar (4.16), observemos de (4.22) y la condicion 4.4.2, que para cada acción $a\in \mathbf{A}(x)$

$$\begin{aligned} (1-\varepsilon_n)\,J_{\varepsilon_n}(x)+H_{\varepsilon_n}(x) &= c(x,f_{\varepsilon_n})+\varepsilon_n\sum_y p_{xy}(f_{\varepsilon_n})\,H_{\varepsilon_n}(y) \\ &\leq c(x,a)+\varepsilon_n\sum_y p_{xy}(a)\,H_{\varepsilon_n}(y) \end{aligned}$$

Haciendo $n\to\infty$ y usando el Teorema de Convergencia Dominada (T.4.3 del Apéndice) en cada $a\in\mathbf{A}(x)$ obtenemos (4.16). Esto completa la demostración del Teorema. ■

VERIFICACION DE LAS CONDICIONES.

Generalmente la condición 4.4.1(a) se verifica mostrando que $J_\alpha(x)$ es creciente en x. Algunas veces es posible verificar las condiciones 3.3.1(b) y 4.4.1(b)-(c)(o 4.4.2) directamente, si no, el siguiente desarrollo es útil. Antes probaremos un resultado general para Cadenas de Markov.

Proposición 4.4.8. Sea $\{x_n\}$ una Cadena de Markov ergódica e irreducible en $0,1,2,...$ con probabilidades de transición (p_{xy}) y distribución estacionaria (π_x). Supongamos que en el estado x se obtiene un costo no-negativo $c(x)$ y sea c_{x0} (respectivamente, m_{x0}) el costo esperado (respectivamente, tiempo) de pasar del estado x al estado 0. Los siguientes incisos son equivalentes:

(a) $\sum_x \pi_x c(x) < \infty$

(b) para cada x, $c_{x0} < \infty$
(c) existe un entero no-negativo N y una función no-negativa $r(x)$ tal que

$$\sum_y p_{xy} r(y) < \infty, \quad 0 \leq x \leq N$$

y

$$\sum_y p_{xy} r(y) - r(x) \leq -c(x), \quad x > N.$$

Demostración. Para probar la equivalencia de (a) y (b), supongamos que la cadena comienza en el estado 0 y sea $\rho_x(0)$ el número promedio de visitas de la cadena al estado x antes de retornar al estado 0., Sabemos por L.5.16 y T.5.18 que

$$\pi_x = \frac{\rho_x(0)}{m_{00}}.$$

Por lo tanto

$$\sum_x \pi_x c(x) = \sum_x \frac{\rho_{x0} c(x)}{m_{00}} = \frac{c_{00}}{m_{00}}. \tag{4.23}$$

Como la cadena es irreducible, $c_{x0} < \infty$ para todo x si y solo si $c_{00} < \infty$ y por lo tanto de (4.23) se sigue que (a) y (b) son equivalentes.

Supongamos (b). Sabemos que

$$c_{x0} = c(x) + \sum_{y>0} p_{xy} c_{y0}, \quad x \geq 0, \tag{4.24}$$

por lo tanto, escogiendo $r(0) = 0$, $r(x) = c_{x0} \quad x \geq 1$ y $N = 0$ se tiene (c).

Sólo nos resta demostrar que (c) implica (a).

Si $x_0 = x$, $x_1, \cdots, x_n = 0$, con $x_j > 0$ para $1 \leq j \leq n-1$, $K = \max_{0 \leq k \leq N} c(k)$ y $R = \max_{0 \leq k \leq N} \sum_y p_{ky} r(y)$ mostraremos que

$$c_{x0} \leq r(x) + (K + R)n.$$

Lo cual probará que

$$c_{x0} \leq r(x) + (K + R) m_{x0} < \infty$$

puesto que la cadena es ergódica.

Primero mostraremos que $E\left(r\left(x_m\right)\right) < \infty$ para $m \geq 0$.

Como

$$E\left(r\left(x_0\right)\right) = r(x) < \infty,$$

el resultado es cierto para $m = 0$.

Supongamos que vale para m. Si $0 \leq x_m \leq N$, entonces tomando valor esperado obtenemos

$$E\left(r\left(x_{m+1}\right)\right) < \infty.$$

Si $x_m > N$, entonces por (c),

$$E\left(r\left(x_{m+1}\right) \mid x_m\right) \leq r\left(x_m\right) - c\left(x_m\right) \leq r\left(x_m\right),$$

por lo tanto, tomando valor esperado en ambos lados obtenemos el resultado.

Fijemos un x_j, $0 \leq j < n$. Si $x_j > N$, entonces de (c) tenemos

$$c(x_j) \leq r(x_j) - E(r(x_{j+1}) \mid x_j).$$

Tomando valor esperado de ambos lados, obtenemos

$$E(c(x_j)) \leq E(r(x_j)) - E(r(x_{j+1})).$$

Si $x_j \leq N$, entonces $E(r(x_{j+1})/x_j) \leq R$,
por lo tanto,

$$E(r(x_{j+1})) \leq R \quad \text{y } R + E(r(x_j)) - E(r(x_{j+1})) \geq 0.$$

Luego, en cualquier caso,

$$E(c(x_j)) \leq K + R + E(r(x_j)) - E(r(x_{j+1})).$$

Sumando desde $j = 0$ a $n - 1$, encontramos que

$$c_{x0} \leq r(x) + (K + R)\, n,$$

lo cual, junto con los comentarios anteriores completan la demostración. ■

Proposición 4.4.9.
(a) Si un proceso de decisión Markoviano tiene una política estacionaria f que induce una Cadena de Markov ergódica e irreducible, el cual cumple alguno de los incisos de la proposición 4.4.8. Entonces, se cumplen las condiciones 3.3.1(b) y 4.4.1(b)-(c).
(b) Si para cada $x \in \mathbf{X}$ y $a \in \mathbf{A}$, existe una política $f_{x,a}$ que elige una acción a en el estado x y satisface las hipotesis de (a), entonces se cumple la condición 4.4.2.

Demostración. Sea (π_x) la distribución estacionaria bajo f.
Entonces,

$$g_f = \sum_x \pi_x c(x, f) < \infty \text{ por la proposición 4.4.8.}$$

Por el Lema 4.4.5, tenemos que

$$\limsup_{\alpha \uparrow 1} (1 - \alpha) J_\alpha(x) \leq g_f,$$

por lo tanto, $J_\alpha(x)$ es finito para α suficientemente cercano a 1. Notemos que si $\beta < \alpha$, entonces

$$J_\beta(x) \leq J_\alpha(x)$$

y por lo tanto, se cumple la condición 3.3.1(b).
Para demostrar que se cumple la condición 4.4.1(b)-(c),
sea $a(x) = f(x)$, $b(0) = 0$ y para $x \geq 1$, $b(x) = c_{x0}(f)$.
Modificando ligeramente la prueba del T.4.3.4, tenemos que

$$H_\alpha(x) \leq c_{x0}(f),$$

por lo tanto,

$$H_\alpha(x) \leq b(x) \text{ para todo } x. \text{ (Notemos que } H_\alpha(0) = 0 \text{).}$$

De (4.24) se sigue que

$$\sum_y p_{xy}(f) b(y) = \sum_{y>0} p_{xy}(f(x)) c_{y0}(f) \leq c_{x0}(f) < \infty,$$

por lo tanto, se cumple la condición 4.4.1(b)-(c).

Para demostrar (b), sea $b(0) = 0$ y para $x \geq 1$, $b(x) = \inf_f c_{x0}(f)$, donde el ínfimo se toma sobre todas las políticas estacionarias que inducen una Cadena de Markov ergódica e irreducible. Entonces para cualquier $x \in \mathbf{X}$ y $a \in \mathbf{A}$,

$$\begin{aligned} \sum_y p_{xy}(a)\, b(y) &= \sum_{y>0} p_{xy}(a)\, b(y) \\ &\leq \sum_{y>0} p_{xy}(a)\, c_{y0}(f_{x,a}) \\ &\leq c_{x0}(f_{x,a}) < \infty \end{aligned}$$

y por lo tanto, se cumple la condición 4.4.2. ■

4.5. EJEMPLO: UN MODELO DE COLAS

Consideremos un servicio de transmisión de paquetes. En este sistema, los paquetes llegan al servidor para ser eventualmente transmitidos, formandose una cola. Suponemos periodos de tiempo de manera que si hay paquetes en el inicio de un periodo, el servidor pueda transmitir un solo paquete durante ese periodo y si la cola es vacia en el inicio de un periodo, ningun paquete puede transmitirse durante ese periodo. Además, suponemos que el número de paquetes que llegan en cada periodo son independientes de la cola e independientes del número de paquetes generados en cualquier periodo. Sea p_x, $x \geq 0$, la probabilidad de que se generen x paquetes en cualquier periodo, y el estado del sistema, el número de paquetes en el inicio de un periodo. En el inicio de cada periodo, el transmisor elige o bien aceptar los nuevos paquetes (a) o rechazarlos (b). Las probabilidades de transición estan dadas por

$$p_{0y}(a) = p_y, \quad p_{0y}(b) = 1 \quad \text{si } y = 0 \quad \text{y } p_{0y}(b) = 0 \quad \text{si } y \geq 1,$$

y para $x \geq 1$

$$p_{x,x+y-1}(a) = p_y, \quad p_{x,x-1}(a) = p_0, \quad p_{x,x-1}(b) = 1 \quad \text{y } p_{x,x+y-1}(b) = 0,\ y \geq 1.$$

Luego, las Ecuaciones de Optimalidad α-descontadas son

$$J_\alpha(0) = \min\left[c(0,a) + \alpha\sum_y p_y J_\alpha(y),\, c(0,b) + \alpha J_\alpha(0)\right]$$

$$J_\alpha(x) = \min\left[c(x,a) + \alpha\sum_y p_y J_\alpha(x-1+y),\, c(x,b) + \alpha J_\alpha(x-1)\right], \quad x \geq 1.$$

Queremos determinar cuando existe una política óptima en costo promedio. Definamos

$$\begin{aligned} b_\alpha(x) &= \sum_{y=1}^{x} \alpha^{x-y} c(y,b), \\ b(x) &= \sum_{y=1}^{x} c(y,b), \quad x \geq 1 \quad \text{y} \\ b_\alpha(0) &= b(0) = 0 \end{aligned}$$

Si δ es una política que siempre rechaza los nuevos paquetes. Entonces

$$J_\alpha(\delta, x) = b_\alpha(x) + \frac{\alpha^x c(0,b)}{1-\alpha}, \quad x \geq 0.$$

Como estas cantidades son finitas, se cumple la condición 3.3.1(b).

Si suponemos que $c(x,a)$ y $c(x,b)$ son crecientes en x, entonces podemos probar por inducción en n que $J_{\alpha,n}(x)$ es creciente en x para cualquier n. Así, por el T.3.3.5, se sigue que $J_\alpha(x)$ es creciente en x y por lo tanto, se cumple la condición 4.4.1(a).

Para verificar la condición 4.4.1(b)-(c), observemos que para $x \geq 1$, $J_\alpha(x) - J_\alpha(x-1) \leq c(x,b)$, por lo tanto $H_\alpha(x) \leq b(x)$. Escogiendo $a(x) = b$, se cumple la condición 4.4.1(b)-(c). Para que se cumpla la suposición 4.4.2, se requiere que $\sum_y p_y b(x-1+y) < \infty$ para todo x.

APENDICE

1. Esperanza de Variables Aleatorias Discretas.

Definición 1.1. Sea X una variable aleatoria discreta con función de probabilidad f_X. Si al menos una de las condiciones (i) o (ii) se satisface:
(i) $\sum_{x_i>0} x_i f_X(x_i) < \infty$;
(ii) $\sum_{x_i<0} x_i f_X(x_i) > -\infty$.
Se define la esperanza (o valor esperado) de X por

$$EX := \sum_i x_i f_X(x_i). \tag{1.1}$$

Definición 1.2. Si tanto (i) como (ii) de la definición anterior se cumplen, se dice que X tiene esperanza finita.

Teorema 1.3. Sea X un vector aleatorio discreto $n-$dimensional con función de probabilidad f_X y sea Φ una función de valor real definida en $\mathbf{R}^n$. Si $Z = \Phi(X)$ es tal que su esperanza esta definida, entonces

$$EZ = \sum_x \Phi(x) f_X(x). \tag{1.2}$$

Demostración. Si $\{z_i\}$ y $\{x_i\}$ denotan los distintos valores de Z y X respectivamente. Entonces para cualquier z_i hay al menos un x_j tal que $z_i = \Phi(x_j)$. Denotemos por A_i la colección de tales x_j's, esto es, $A_i = \{x_j : \Phi(x_j) = z_i\}$. Entonces $\{x \in A_i\}$ y $\{Z = z_i\}$ denotan exactamente el mismo evento. Así,

$$\Pr(Z = z_i) = \Pr(x \in A_i) = \sum_{x \in A_i} f_X(x).$$

Consecuentemente,

$$\begin{aligned}\sum_i z_i f_Z(z_i) &= \sum_i z_i \Pr(Z = z_i) \\ &= \sum_i z_i \sum_{x \in A_i} f_X(x) \\ &= \sum_i \sum_{x \in A_i} z_i f_X(x)\end{aligned}$$

como $\Phi(x) = z_i$, para $x \in A_i$, entonces

$$\sum_i z_i f_Z(z_i) = \sum_i \sum_{x \in A_i} \Phi(x) f_X(x).$$

Por definición, los conjuntos A_i son disjuntos para valores distintos de i y su unión es el conjunto de todos los valores posibles de X. Por lo tanto

$$\sum_i z_i f_Z(z_i) = \sum_x \Phi(x) f_X(x). \blacksquare$$

Teorema 1.4. Sean X y Y variables aleatorias con esperanza finita.
(i) Si c es una constante y $\Pr(X = c) = 1$, entonces $EX = c$.
(ii) Si c es una constante, entonces cX tiene esperanza finita y $E(cX) = cEX$.
(iii) $X + Y$ tiene esperanza finita y $E(X + Y) = EX + EY$.
(iv) Si $\Pr(X \geq Y) = 1$, entonces $EX \geq EY$.
(v) $|EX| \leq E|X|$.

Demostración. Para probar (i) fijémonos que como $\Pr(X = c) = 1$, entonces X tiene densidad $f_X(x) = 0$ para $x \neq c$ y $f_X(c) = 1$. Así por (1.1)

$$EX = \sum_x x f_X(x) = c f_X(c) = c.$$

Para la demostración de (ii) definamos $\Phi(x) = cx$,
así

$$\sum_x |cx| f_X(x) = |c| \sum_x |x| f_X(x) < \infty,$$

por lo que cX tiene esperanza finita y por (1.2)

$$E(cX) = \sum_x (cx) f_X(x) = c \sum_x x f_X(x) = cEX.$$

Para la prueba de (iii). Sea $\Phi(x,y) = x + y$ y sea f la densidad conjunta de X y Y.
Entonces,

$$\begin{aligned}\sum_{x,y} |x+y| f(x,y) &\leq \sum_{x,y} |x| f(x,y) + \sum_{x,y} |y| f(x,y) \\ &= \sum_{x} |x| \sum_{y} f(x,y) + \sum_{y} |y| \sum_{x} f(x,y) \\ &= \sum_{x} |x| f_X(x) + \sum_{y} |y| f_Y(y) < \infty\end{aligned}$$

y por lo tanto $X + Y$ tiene esperanza finita.

Aplicando (1.2) vemos que

$$\begin{aligned}E(X+Y) &= \sum_{x,y} (x+y) f(x,y) \\ &= \sum_{x,y} x f(x,y) + \sum_{x,y} y f(x,y) \\ &= EX + EY.\end{aligned}$$

Para la demostración de la parte (iv) sea $Z = X - Y = X + (-Y)$.
Por (ii) y (iii) tenemos que

$$EX - EY = E(X-Y) = EZ = \sum_{z} z f_Z(z).$$

Y como $\Pr(Z \geq 0) = \Pr(X \geq Y) = 1$, entonces $z_i \geq 0$ para toda i .

Así,

$$\sum_{z} z f_Z(z) \geq 0,$$

por lo tanto

$$EX - EY \geq 0,$$

lo cual prueba la parte (iv).

Finalmente, (v) se sigue de (iv) y (ii) pues $-|x| \leq x \leq |x|$. Esto completa la prueba del teorema. ■

2. Esperanza Condicional de Variables Aleatorias Discretas.

Definición 2.1. Sea (X,Y) un vector aleatorio discreto. Para $x \in \mathbf{R}$ tal que $f_X(x) > 0$, sea $f_{Y|X}(y|x)$ la función de prob. condicional de Y dado $X = x$, y supongamos que EY está definida. Entonces la esperanza condicional de Y dado $X = x$ se define como

$$E(Y \mid X = x) := \sum_y y f_{Y|X}(y|x). \tag{2.1}$$

Definición 2.2. La esperanza condicional de Y dado X se define como

$$E(Y \mid X) := g(X) \tag{2.2}$$

donde $g(X) = E(Y \mid X = x)$.

Teorema 2.3. $E(Y \mid X)$ tiene la propiedad de la doble esperanza, esto es $E[E(Y \mid X)] = EY$.

Demostración. Por el Teorema 1.3 tenemos que

$$E[E(Y \mid X)] = \sum_{x \in R_X} g(x) f_X(x)$$

donde R_X es el conjunto de los posibles valores de X (rango de X) y R_Y es el rango de Y.

Así por la definición 2.1

$$\begin{aligned} E[E(Y \mid X)] &= \sum_{x \in R_X} \left(\sum_{y \in R_Y} y f_{Y|X}(y|x) \right) f_X(x) \\ &= \sum_{y \in R_Y} y \sum_{x \in R_X} f_{Y|X}(y|x) f_X(x) \\ &= \sum_{y \in R_Y} y f_Y(y) \\ &= EY. \blacksquare \end{aligned}$$

Definición 2.4. Sea (X,Y,Z) un vector aleatorio discreto, si para $x, y \in \mathbf{R}$ $\Pr(X = x, Y = y) > 0$ y EZ está definida. Entonces la esperanza condicional de Z dado $X = x$ y $Y = y$ se define como

$$E(Z \mid X=x, Y=y) := \sum_z z \Pr(Z=z \mid X=x, Y=y).$$

Definición 2.5. La esperanza condicional de Z dado X y Y se define como

$$E(Z \mid X,Y) := g(X,Y)$$

donde $g(X,Y) = E(Z \mid X=x, Y=y)$.

Teorema 2.6. $E[E(Z \mid X,Y) \mid X] = E(Z \mid X)$.
Demostración. Por el Teorema 1.3 sabemos que

$$E[E(Z \mid X,Y) \mid X=x] = \sum_y g(x,y) \Pr(Y=y \mid X=x)$$

así, por la definición de g, tenemos que

$$\begin{aligned} E[E(Z \mid X,Y) \mid X=x] &= \sum_y \sum_z z \Pr(Z=z|X=x, Y=y) \Pr(Y=y|X=x) \\ &= \sum_z z \sum_y \Pr(Z=z|X=x, Y=y) \Pr(Y=y|X=x) \\ &\sum_z z \sum_y \frac{\Pr(Z=z,\, Y=y,\, X=x)}{\Pr(X=x)} \\ &= \sum_z z \Pr(Z=z \mid X=x) \\ &= E(Z \mid X=x)\ , \quad \forall\, x \in R_X, \end{aligned}$$

así

$$E[E(Z \mid X,Y) \mid X] = E(Z \mid X). \blacksquare$$

3. Convexidad.

Definición 3.1. Una función g:$\mathbf{R}\rightarrow\mathbf{R}$, se dice *convexa* si,
i) para todo $\alpha \in [0,1]$ se cumple

$$g(\alpha x + (1-\alpha) y) \leq \alpha g(x) + (1-\alpha) g(y), \quad \forall\, x,\ y \in \mathbf{R}. \tag{3.1}$$

o equivalentemente (Ver D. Stirzaker pag. 99)
ii) para cada a, existe $\lambda(a)$ tal que

$$g(x) \geq g(a) + \lambda(a)(x-a), \quad \forall\, x. \tag{3.2}$$

En el siguiente teorema usaremos la definición de función convexa dada en i), pero donde el dominio de la función es el conjunto de los enteros, el cual no es un conjunto convexo y entonces tenemos que adecuar la noción de convexidad a tales dominios. En este caso entenderemos que $g : \mathbf{Z} \rightarrow \mathbf{R}$ es convexa si, para cada $x,\ y \in \mathbf{Z}$ y $x \leq z \leq y,\ z \in \mathbf{Z}$ se cumple

$$g(z) \leq \alpha g(x) + (1-\alpha) g(y),$$

donde $z = \alpha x + (1-\alpha) y$ con $\alpha \in [0,1]$.

Teorema 3.2. Sean $g,\ h$:$\mathbf{Z}\rightarrow\mathbf{R}$ con g convexa. Si $H(z) = \sum_{z\in\mathbf{Z}} g(y)\, h(z-y)$ es convergente para cada z. Entonces H es convexa.
Demostración. Tomemos $u = z - y$, así

$$H(z) = \sum_{u} g(z-u)\, h(u).$$

Luego para $z_1,\ z_2 \in \mathbf{Z}$ y $\alpha \in [0,1]$ tal que $\alpha z_1 + (1-\alpha) z_2 \in \mathbf{Z}$, usando que $H(z)$ es convergente para cada z se obtiene

$$\begin{aligned}
H(\alpha z_1 + (1-\alpha) z_2) &= \sum_{u} g(\alpha z_1 + (1-\alpha) z_2 - u)\, h(u) \\
&= \sum_{u} g(\alpha(z_1 - u) + (1-\alpha)(z_2 - u))\, h(u) \\
&\leq \sum_{u} [\alpha g(z_1 - u) + (1-\alpha) g(z_2 - u)]\, h(u) \\
&= \alpha \sum_{u} g(z_1 - u)\, h(u) + (1-\alpha) \sum_{u} g(z_2 - u)\, h(u) \\
&= \alpha H(z_1) + (1-\alpha) H(z_2). \blacksquare
\end{aligned}$$

Teorema 3.3. Si $g,\ h$:$\mathbf{Z}\rightarrow\mathbf{R}$ son funciones convexas, entonces
(i) Para c constante positiva, cg es convexa.
(ii) $g+h$ es convexa.
(iii) $\max(g,h)$ es convexa.
Demostración. Directa de las definiciones. ■

Teorema 3.4. Si g :$\mathbf{Z}\rightarrow\mathbf{R}$ es convexa y $\lim_{|x|\rightarrow\infty} g(x)=\infty$, entonces g tiene un mínimo.

Demostración. Como $\lim_{|x|\rightarrow\infty} g(x)=\infty$ la convexidad es hacia arriba. Por lo tanto $g(x)$ tiene un mínimo. ■

Teorema 3.5
.(desigualdad de Jensen). Sea X una variable aleatoria con esperanza finita, y $g(x)$ una función convexa. Entonces

$$E(g(X))\geq g(E(X)). \tag{3.3}$$

Demostración. Tomando $a=E(X)$ en (3.2), tenemos

$$g(X)\geq g(E(X))+\lambda(X-E(X)).$$

Luego, tomando valor esperado en ambos lados obtenemos (3.3). ■

4. Teoremas de Convergencia.

Usualmente, estos teoremas son dados en la notación de integral de Lebesgue, pero como una suma infinita es simplemente un caso especial de una de tales integrales y nosotros requerimos estos teoremas solamente en el caso de sumas infinitas, enunciaremos y probaremos tales teoremas para sumas.

Teorema 4.1(Teorema de Convergencia Monótona). Sea $p=(p_1,\ p_2,...)$ una distribución de probabilidades bajo un conjunto numerable S denotado por $S=\{1,2,...\}$. Sea $\{h_n\}$ una sucesión de funciones de valor real extendida en S tal que $0\leq h_n(x)\leq h_{n+1}(x)\quad\forall\ x,\quad n=1,\ 2,...$ y sea $h:S\rightarrow[0,+\infty]$ la función límite $h(x)=\lim_{n\rightarrow\infty} h_n(x)$. Entonces

$$\lim_{n\to\infty}\sum_{x=1}^{\infty}p_x h_n(x)=\sum_{x=1}^{\infty}p_x\lim_{n\to\infty}h_n(x)=\sum_{x=1}^{\infty}p_x h(x).$$

Demostración. Tenemos

$$\sum_{x=1}^{\infty}p_x h_n(x)\leq\sum_{x=1}^{\infty}p_x h(x).$$

Tomando límite, obtenemos

$$\lim_{n\to\infty}\sum_{x=1}^{\infty}p_x h_n(x)\leq\sum_{x=1}^{\infty}p_x h(x).$$

Así, sólo resta probar la desigualdad inversa. Ahora, para cualquier entero $N\geq 1$ tenemos

$$\lim_{n\to\infty}\sum_{x=1}^{\infty}p_x h_n(x)\geq\lim_{n\to\infty}\sum_{x=1}^{N}p_x h_n(x)=\sum_{x=1}^{N}p_x h(x),$$

y tomando límite cuando $N\to+\infty$ la desigualdad inversa se sigue. ■

Teorema 4.2. Sea $p=(p_1,\ p_2,...)$ una distribución de probabilidades bajo un conjunto numerable S denotado por $S=\{1,2,...\}$. Sea $\{h_n\}$ una sucesión de funciones de valor real extendida en S tal que $0\leq h_n(x)\quad\forall\, x,\quad n=1,\ 2,....$Entonces

$$\sum_{x=1}^{\infty}p_x\operatorname*{liminf}_{n\to\infty}h_n(x)\leq\operatorname*{liminf}_{n\to\infty}\sum_{x=1}^{\infty}p_x h_n(x).$$

Demostración. Sea $H_m=\inf\{h_m,\ h_{m+1},...\}$. Entonces $H_m\leq h_n\quad\forall\, m\leq n$. Por lo tanto

$$\sum_{x=1}^{\infty}p_x H_m(x)\leq\sum_{x=1}^{\infty}p_x h_n(x),\quad m\leq n,$$

luego

$$\sum_{x=1}^{\infty}p_x H_m(x)\leq\operatorname*{liminf}_{n\to\infty}\sum_{x=1}^{\infty}p_x h_n(x).$$

Como la sucesión $H_m(x)$ es creciente y converge a $\operatorname*{liminf}_{n\to\infty}h_n(x)$, tenemos por el Teorema 4.1 que

$$\sum_{x=1}^{\infty} p_x \liminf_{n\to\infty} h_n(x) = \lim_{m\to\infty} \sum_{x=1}^{\infty} p_x H_m(x) \leq \liminf_{n\to\infty} \sum_{x=1}^{\infty} p_x h_n(x). \blacksquare$$

Teorema 4.3. Sea $p = (p_1,\ p_2, ...)$ una distribución de probabilidades bajo un conjunto numerable S denotado por $S = \{1, 2, ...\}$. Sea $\{h_n\}$ una sucesión de funciones de valor real extendida en S. Sea g una función también de valor real extendida en S tal que $\sum_{x=1}^{\infty} p_x g(x) < \infty$ y $|h_n(x)| \leq g(x) \quad \forall\ x, \quad n = 1,\ 2, ...$ y sea $h : S \to [0, +\infty]$ la función límite $h(x) = \lim_{n\to\infty} h_n(x)$. Entonces

$$\lim_{n\to\infty} \sum_{x=1}^{\infty} p_x h_n(x) = \sum_{x=1}^{\infty} p_x h(x).$$

Demostración. Como $g(x) + h_n(x) \geq 0 \quad \forall\ x, \quad n = 1,\ 2, ...$, tenemos usando el T.4.2 que

$$\begin{aligned}
\sum_{x=1}^{\infty} p_x g(x) + \sum_{x=1}^{\infty} p_x h(x) &= \sum_{x=1}^{\infty} p_x (g(x) + h(x)) \leq \liminf_{n\to\infty} \sum_{x=1}^{\infty} p_x (g(x) + h_n(x)) \\
&= \liminf_{n\to\infty} \left(\sum_{x=1}^{\infty} p_x g(x) + \sum_{x=1}^{\infty} p_x h_n(x) \right) \\
&= \sum_{x=1}^{\infty} p_x g(x) + \liminf_{n\to\infty} \sum_{x=1}^{\infty} p_x h_n(x).
\end{aligned}$$

Por lo tanto, tenemos

$$\sum_{x=1}^{\infty} p_x h(x) \leq \liminf_{n\to\infty} \sum_{x=1}^{\infty} p_x h_n(x). \tag{4.1}$$

Ahora, como $g(x) - h_n(x) \geq 0 \quad \forall\ x, \quad n = 1,\ 2, ...$, tenemos aplicando nuevamente el T.4.2 que

$$\begin{aligned}
\sum_{x=1}^{\infty} p_x g(x) - \sum_{x=1}^{\infty} p_x h(x) &= \sum_{x=1}^{\infty} p_x (g(x) - h(x)) \leq \liminf_{n\to\infty} \sum_{x=1}^{\infty} p_x (g(x) - h_n(x)) \\
&= \sum_{x=1}^{\infty} p_x g(x) - \limsup_{n\to\infty} \sum_{x=1}^{\infty} p_x h_n(x),
\end{aligned}$$

de lo cual se sigue que

$$\limsup_{n\to\infty} \sum_{x=1}^{\infty} p_x h_n(x) \leq \sum_{x=1}^{\infty} p_x h(x). \tag{4.2}$$

De (4.1) y (4.2) obtenemos la igualdad deseada. ■

5. Cadenas de Markov.

Definición 5.1. Sea X= $\{X_n,\ n \geq 0\}$ una sucesión de variables aleatorias con valores en un conjunto numerable S, llamado el *espacio de estados.*

a) Si, para toda $n \geq 0$ y cualesquiera valores $x, y, x_0, ..., x_{n-1}$, se cumple

$$\Pr[X_{n+1} = y \mid X_0 = x_0, ..., X_n = x] = \Pr[X_{n+1} = y \mid X_n = x]$$

entonces X se dice ser una *cadena de Markov,* o tener la propiedad de Markov.

b) Si, además se cumple

$$\Pr[X_{n+1} = y \mid X_n = x] = \Pr[X_1 = y \mid X_0 = x]$$

entonces X se dice ser una cadena de Markov *homogénea.*

A la función p_{xy}^n dada por

$$p_{xy}^n = \Pr[X_n = y \mid X_0 = x],\ x, y \in S$$

para $n \geq 1$ y por

$$p_{xy}^0 = \begin{cases} 1, \text{ si } x = y \\ 0, \text{ si } x \neq y \end{cases}$$

se le llama función de transición (en n-pasos) de la cadena. Denotamos p_{xy}^1 simplemente como p_{xy}.

Teorema 5.2(Ecuación de Chapman-Kolmogorov). Si X es una cadena de Markov con espacio de estados S, entonces para cualesquiera m y $n \in \mathbf{Z}^+$, tenemos

$$p_{xy}^{m+n} = \sum_{z\in S} p_{xz}^m p_{zy}^n,\ x, y \in S.$$

Definición 5.3. Una cadena de Markov X es *irreducible* si para cada x y $y \in S$ existe un $n < \infty$ tal que

$$p_{xy}^n > 0.$$

Definición 5.4. Sea X una cadena de Markov con espacio de estados S y sea $x \in S$ tal que $p_{xx}^n > 0$ para algún $n > 0$. Definimos el periodo d_x por

$$d_x := mcd\{n > 0 : p_{xx}^n > 0\}.$$

Definición 5.5.
a) $x \in S$ es periódico si $d_x > 1$, y
b)$x \in S$ es aperiódico si $d_x = 1$.

Definición 5.6. Para una cadena de Markov X con $X_0 = x$, definimos

$$T_{xy} = \min\{n \geq 0 : X_n = y \mid X_0 = x\},$$

y $m_{xy} = E(T_{xy})$ es el tiempo medio de pasar del estado x al estado y.

Definición 5.7. Sea X una cadena de Markov con espacio de estados S. Si denotamos $\rho_{xy:=} \Pr[T_{xy} < \infty]$, entonces:
a) Si, para cualquier $x \in S$,

$$\rho_{xx} = 1$$

la cadena se llama *recurrente.*
b) Si, para cualquier $x \in S$,

$$\rho_{xx} < 1$$

la cadena se llama *transitoria.*

Definición 5.8. Sea X una cadena recurrente.
a) Si, para cualquier $x \in S$,

$$m_{xx} = \infty$$

la cadena se llama *recurrente nula.*

b) Si, para cualquier $x \in S$,

$$m_{xx} < \infty$$

la cadena se llama *recurrente positiva.*

Definición 5.9. Una cadena de Markov se dice ser *ergódica* si es recurrente positiva y aperiódica.

Lema 5.10. Sea X una cadena de Markov. Si $N(y) = \sum_{n=1}^{\infty} 1_y(X_n)$, entonces

$$E_x(N(y)) = \sum_{n=1}^{\infty} p_{xy}^n.$$

Notemos que $N(y)$ es el número de visitas de la cadena al estado y.

Lema 5.11. Si $y \in S$ es transitorio, entonces

$$E_x(N(y)) = \frac{\rho_{xy}}{1-\rho_{yy}} < +\infty, \ x \in S.$$

Teorema 5.12. Si $y \in S$ es un estado transitorio, entonces

$$\lim_{n\to\infty} p_{xy}^n = 0, \ x \in S.$$

Definición 5.13. Sea X una cadena de Markov con espacio de estados S y funcion de transición p. Entonces $\Pi = \{\pi_x, x \in S\}$ es una distribución estacionaria de la cadena si

a) $\pi_x \geq 0 \quad \forall\, x \in S, \quad \sum_{x\in S} \pi_x = 1.$

b) $\sum_{x\in S} \pi_x p_{xy} = \pi_y, \quad y \in S.$

Teorema 5.14. Si $\Pi = \{\pi_x, x \in S\}$ es una distribución estacionaria, entonces

$$\pi_y = \sum_x \pi_x p_{xy}^n \quad \forall\, y \in S.$$

Definición. 5.15. Si $y \in S$ es recurrente positivo ($m_{yy} < \infty$), definimos

$$\rho_x(y) := \sum_{n=0}^{\infty} \Pr\left[X_n = x,\ T_{yy} > n \mid X_0 = y\right].$$

Notemos que $\rho_x(y)$ es el número promedio de visitas de la cadena al estado x entre dos visitas consecutivas al estado y.

Lema 5.16. Si y es un estado positivo de una cadena irreducible recurrente, entonces existe una distribución estacionaria $\Pi = \{\pi_x,\ x \in S\}$ tal que

$$\pi_x = \frac{\rho_x(y)}{m_{yy}}.$$

Lema 5.17. Si $\Pi = \{\pi_x,\ x \in S\}$ es distribución estacionaria de una cadena irreducible, entonces $\pi_x > 0 \quad \forall\ x \in S$.

Teorema 5.18. Una cadena de Markov irreducible tiene una distribución estacionaria Π si y sólo si todos los estados son recurrentes positivos, en cuyo caso Π es la única distribución estacionaria y esta dada por

$$\pi_x = \frac{1}{m_{xx}} \quad \forall\ x \in S.$$

Nota: Las demostraciones de estos resultados pueden consultarse en cualquier libro de Cadenas de Markov; en particular en los que enunciamos en la bibliografia. ■

6. Operadores de Contracción.

Definición 6.1. Un espacio métrico es una pareja (M, d), donde M es un conjunto no vacio y d es una función de $M \times M$ en $\mathbf{R}$ que satisface las propiedades siguientes, cualesquiera que sean los puntos x, y, z de M:

1. $d(x,x) = 0$
2. $d(x,y) > 0$ si $x \neq y$
3. $d(x,y) = d(y,x)$
4. $d(x,y) \leq d(x,z) + d(z,y)$

Definición 6.2. Sea (M,d) un espacio métrico. Se dice que (M,d) es un espacio métrico completo si cualquier sucesión de cauchy en M converge en M.

Definición 6.3. Sea (M,d) un espacio métrico, un operador $T : M \to M$ se dice de $\alpha-$contracción $(0 < \alpha < 1)$ si $d(Tx,Ty) \leq \alpha d(x,y)$ para todo $x, y \in M$.

Teorema 6.4. (Teorema de Punto Fijo para Operadores de Contracción). Si (M,d) es un espacio métrico completo y T es un operador de contracción, entonces
(i) Existe un único $x \in M$ tal que $Tx = x$.
(ii) $\lim_{n\to\infty} T^n y = x$ para cada $y \in M$.

Demostración. Probaremos en primer lugar la unicidad del inciso (i).
Sean x, y tales que $x = Tx$ y $y = Ty$ con $x \neq y$. Entonces

$$d(x,y) = d(Tx,Ty) \tag{6.1}$$

y como T es de contracción

$$d(Tx,Ty) \leq \alpha d(x,y), \tag{6.2}$$

de (6.1) y (6.2) tenemos que $1 \leq \alpha$ lo cual es una contradicción, así $x = y$.
Sea ahora $y \in M$, puesto que T es completo:
$\{T^n y\}$ es convergente si y sólo si $d(T^n y, T^m y) \to 0$.
Por lo tanto supongamos que $m \geq n$ $(m = k + n)$, entonces puesto que T es de contracción:

$$d\left(T^{n+k}y, T^m y\right) \leq \alpha d\left(T^{n+k-1}y, T^{m-1}y\right) \leq ... \leq \alpha^n d\left(T^k y, y\right)$$

y por la propiedad 4 de la def. 6.1

$$d\left(T^k y, y\right) \leq d\left(T^k y, T^{k-1}y\right) + ... + d(Ty,y)$$

así,

$$\begin{aligned} d\left(T^{n+k}y, T^m y\right) &\leq \alpha^n \left[d\left(T^k y, T^{k-1}y\right) + ... + d(Ty,y)\right] \\ &\leq \alpha^n \left[\alpha^{k-1} + ... + 1\right] d(Ty,y), \end{aligned}$$

tomando límite cuando $n \to \infty$ tenemos

$$d(T^m y, T^m y) \to 0$$

de donde $\{T^n y\}$ es convergente.
Sea entonces $x = \lim_{n\to\infty} T^n y$, como $T^{n+1}y \to x$ y $T^{n+1}y = T(T^n y)$ converge a Tx entonces $Tx = x$. ■

7. Teorema Tauberiano.

Lema 7.1. Sean $\alpha \in (0,1)$ y $\{C_t\}_{t=0}$ una sucesión de números reales no negativos y definamos $S_n = \sum_{t=0}^{n-1} C_t$ para $n = 1, 2, \ldots$ y $S_0 = 0$. Entonces si $\limsup_{n\to\infty} \frac{S_n}{n} < \infty$,

$$\sum_{t=0}^{\infty} \alpha^t C_t = (1-\alpha) \sum_{t=0}^{\infty} \alpha^t S_{t+1}.$$

Demostración. Observemos que para $t \geq 0$ se tiene que $S_{t+1} - S_t = C_t$.

Sea N un entero positivo:

$$\begin{aligned}
\sum_{t=0}^{N} \alpha^t C_t &= \sum_{t=0}^{N} \alpha^t (S_{t+1} - S_t) \\
&= \sum_{t=0}^{N} \alpha^t S_{t+1} - \sum_{t=0}^{N} \alpha^t S_t \\
&= \sum_{t=0}^{N-1} \alpha^t S_{t+1} + \alpha^N S_{N+1} - \sum_{t=0}^{N} \alpha^t S_t \\
&= \sum_{t=0}^{N-1} \alpha^t S_{t+1} + \alpha^N S_{N+1} - \sum_{t=1}^{N} \alpha^t S_t \\
&= \sum_{t=0}^{N-1} \alpha^t S_{t+1} - \alpha \sum_{t=1}^{N} \alpha^{t-1} S_t + \alpha^N S_{N+1} \\
&= \sum_{t=0}^{N-1} \alpha^t S_{t+1} - \alpha \sum_{t=1}^{N-1} \alpha^t S_{t+1} + \alpha^N S_{N+1}
\end{aligned}$$

así,

$$\sum_{t=0}^{N} \alpha^t C_t = (1-\alpha) \sum_{t=0}^{N-1} \alpha^t S_{t+1} + \alpha^N S_{N+1}$$

de donde

$$\lim_{N\to\infty}\sum_{t=0}^{N}\alpha^t C_t = \lim_{N\to\infty}\left[(1-\alpha)\sum_{t=0}^{N-1}\alpha^t S_{t+1} + \alpha^N S_{N+1}\right].$$

Por otro lado, como

$$\alpha^N S_{N+1} = (N+1)\,\alpha^N \frac{S_{N+1}}{N+1} \to 0 \text{ cuando } N \to \infty$$

puesto que la serie $\sum_{t=0}^{\infty}(N+1)\,\alpha^N$ es convergente si $|\alpha| < 1$ y por hipótesis $\limsup_{n\to\infty}\frac{S_n}{n} < \infty$. Así, se tiene que

$$\sum_{t=0}^{\infty}\alpha^t C_t = (1-\alpha)\sum_{t=0}^{\infty}\alpha^t S_{t+1}. \blacksquare$$

Teorema 7.2.(Teorema Tauberiano). Sea $\{C_t\}_{t=0}$ una sucesión de números reales no negativos y definamos $S_0 = 0$ y $S_n = \sum_{t=0}^{n-1} C_t$ para $n = 1, 2, ...$Entonces si $\limsup_{n\to\infty}\dfrac{S_n}{n} < \infty$,

$$\begin{aligned}\liminf_{n\to\infty}\frac{S_n}{n} &\leq \liminf_{\alpha\uparrow 1^-}(1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t \\ &\leq \limsup_{\alpha\uparrow 1^-}(1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t \\ &\leq \limsup_{n\to\infty}\frac{S_n}{n}\end{aligned}$$

Demostración. Por propiedades del liminf y limsup se tiene que siempre se cumple

$$\liminf_{\alpha\uparrow 1^-}(1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t \leq \limsup_{\alpha\uparrow 1^-}(1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t$$

por lo que sólo se tiene que probar las siguientes desigualdades:

(i) $\limsup_{\alpha\uparrow 1^-}(1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t \leq \limsup_{n\to\infty}\frac{S_n}{n}$

(ii) $\liminf_{n\to\infty}\frac{S_n}{n} \leq \liminf_{\alpha\uparrow 1^-}(1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t.$

Demostremos la parte en (i).

Del lema anterior sabemos que para $\alpha \in (0,1)$ se cumple

$$\begin{aligned}\sum_{t=0}^{\infty}\alpha^t C_t &= (1-\alpha)\sum_{t=0}^{\infty}\alpha^t S_{t+1}\\ &= (1-\alpha)\sum_{t=0}^{N-1}\alpha^t S_{t+1} + (1-\alpha)\sum_{t=N}^{\infty}\alpha^t S_{t+1}\\ &= (1-\alpha)\sum_{t=0}^{N-1}\alpha^t S_{t+1} + (1-\alpha)\sum_{t=N}^{\infty}(t+1)\,\alpha^t\frac{S_{t+1}}{t+1}\\ &= (1-\alpha)\sum_{t=0}^{N-1}\alpha^t S_{t+1} + (1-\alpha)\left[\sup_{t\geq N+1}\frac{S_t}{t}\right]\sum_{t=N}^{\infty}(t+1)\,\alpha^t\end{aligned}$$

así,

$$\sum_{t=0}^{\infty}\alpha^t C_t = (1-\alpha)\sum_{t=0}^{N-1}\alpha^t S_{t+1} + (1-\alpha)\left[\sup_{t\geq N+1}\frac{S_t}{t}\right]\sum_{t=N}^{\infty}(t+1)\,\alpha^t.$$

Ahora recuerdese los siguientes hechos:

(a) $\sum_{t=0}^{\infty}\alpha^t = \frac{1}{1-\alpha}$ para $|\alpha|<1$

(b) $\sum_{t=0}^{\infty}t\alpha^{t-1} = \sum_{t=0}^{\infty}(t+1)\,\alpha^t = \frac{1}{(1-\alpha)^2}$ si $|\alpha|<1$.

Entonces

$$(1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t \leq (1-\alpha)^2\sum_{t=0}^{N-1}\alpha^t S_{t+1} + \sup_{t\geq N+1}\frac{S_t}{t}$$

puesto que $\lim_{\alpha\uparrow 1^-}(1-\alpha)^2\sum_{t=0}^{N-1}\alpha^t S_{t+1} = 0$ se concluye que

$$\limsup_{\alpha\uparrow 1^-}(1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t \leq \sup_{t\geq N+1}\frac{S_t}{t}.$$

De la última desiguadad, tomando límite cuando $N\to\infty$ se concluye que

$$\limsup_{\alpha\uparrow 1^-}(1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t \leq \limsup_{n\to\infty}\frac{S_n}{n}.$$

Ahora demostraremos la parte en (ii):
De nuevo tenemos la relación

$$(1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t = (1-\alpha)^2\sum_{t=0}^{N-1}\alpha^t S_{t+1} + (1-\alpha)^2\sum_{t=N}^{\infty}\alpha^t S_{t+1}$$

y así,

$$(1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t \geq (1-\alpha)^2\sum_{t=0}^{N-1}\alpha^t S_{t+1} + (1-\alpha)^2\left[\inf_{t\geq N+1}\frac{S_t}{t}\right]\sum_{t=N}^{\infty}(t+1)\,\alpha^t.$$

De nuevo, usando propiedades de la serie geométrica

$$\begin{aligned}\sum_{t=N}^{\infty}(t+1)\,\alpha^t &= \frac{d}{d\alpha}\left[\frac{\alpha^{N+1}}{1-\alpha}\right]\\ &= \frac{N\alpha^N+\alpha^N-N\alpha^{N+1}}{(1-\alpha)^2}\end{aligned}$$

Así,

$$\begin{aligned}(1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t &\geq (1-\alpha)^2\sum_{t=0}^{N-1}\alpha^t S_{t+1} + (1-\alpha)^2\inf_{t\geq N+1}\frac{S_t}{t}\left[\frac{N\alpha^N+\alpha^N-N\alpha^{N+1}}{(1-\alpha)^2}\right]\\ &= (1-\alpha)^2\sum_{t=0}^{N-1}\alpha^t S_{t+1} + \inf_{t\geq N+1}\frac{S_t}{t}\left[N\alpha^N+\alpha^N-N\alpha^{N+1}\right].\end{aligned}$$

Tomando liminf cuando $\alpha \uparrow 1^-$ se obtiene que

$$\liminf_{\alpha\uparrow 1^-} (1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t \geq \inf_{t\geq N+1}\frac{S_t}{t}.$$

Ahora, haciendo $N \to \infty$ se concluye que

$$\liminf_{\alpha\uparrow 1^-} (1-\alpha)\sum_{t=0}^{\infty}\alpha^t C_t \geq \liminf_{n\to\infty}\frac{S_n}{n}. \blacksquare$$

CONCLUSIONES.

En el primer capítulo de este trabajo planteamos el Problema de Control Óptimo, para lo cual definimos:
— modelo de control,
— políticas de control admisibles, y
— índice de funcionamiento.
Antes de dar las definiciones de tales elementos, dimos un ejemplo(un sistema de producción/inventario) con el cual intentamos introducir de manera natural estos elementos e ilustrar el tipo de problemas que desarrollamos en este trabajo. En el capítulo 2 probamos el Algoritmo de la Programación Dinámica y dimos dos ejemplos, en los cuales para su solución aplicamos este teorema. En el capítulo 3 estudiamos el problema de control estocástico en tiempo discreto con espacio de estados numerable y horizonte infinito tanto para el caso de costos acotados como para el caso de costos no acotados, en ambos casos, dimos condiciones bajo las cuales existen políticas óptimas y formas de conocer tales políticas; para esto, establecimos la Ecuación de Optimalidad y la existencia de soluciones a la ecuación de optimalidad, asi como sus relaciones con tal ecuación; mostramos como se caracterizan las políticas óptimas y estudiamos un método para obtener la *función de valor óptimo*. En el capítulo 4 estudiamos el Criterio en Costo Promedio para el caso en Costos Acotados y Costos no Acotados. En ambos casos establecimos condiciones bajo las cuales existe una política estacionaria óptima y dimos un ejemplo(un modelo de colas).

Los primeros 3 capítulos y la primera parte del último fueron estudiados principalmente de M. Ross(1983) y el resto del capítulo 4 se basó en el artículo de Sennott(1987). En el apéndice desarrolla mos el material necesario para este trabajo.

BIBLIOGRAFIA.

[1] Ross S. M., "Introduction to Stochastic Dynamic Programming". Academic Press, Inc., San Diego, 1983.

[2] Bertsekas D. P., "Dynamic Programming: Determinisc and Stochastic Models". Prentice Hall, New Jersey, 1976.

[3] Hernández Lerma O., "Lecture Notes on Discrete-Time Markov Control Processes". Depto. de Matemáticas Centro de Investigación I.P.N., 1990.

[4] Sennott L. I., "A New Condition for the Existence de Optimal Stationary Policies in Average Cost Markov Decision Processes", 1986.

[5] Sennott L. I., "Average Cost Optimal Stationary Policies in Infinite State Markov Decision Processes whit Unbonded Costs", 1987.

[6] Stirzaker D., "Elementary Probability". Cambridge University Press, 1994.

[7] Karlin & Taylor, "A First Course in Stochastic Processes". Academic Press, Inc., 1975.

[8] Issacson & Madsen, "Markov Chains". John Wiley & Sons, Inc., New York, 1976.