UNIVERSIDAD DE SONORA

División de Ciencias Exactas y Naturales

Departamento de Matemáticas

Procesos de Control de Markov Parcialmente Observables con Aplicaciones a Sistemas de Inventarios

T E S I S

Que para obtener el título de:

Maestro en Ciencias (Matemáticas)

Presenta:

Marla Lizzette Arcega López

Director de Tesis: Dr. Jesús Adolfo Minjárez Sosa

Hermosillo, Sonora, México, 23 de Agosto de 2013

SINODALES

Dr. Agustín Brau Rojas
Universidad de Sonora.

Dr. Carlos Gabriel Pacheco González
CINVESTAV del I.P.N.

Dr. Jesús Adolfo Minjárez Sosa
Universidad de Sonora.

Dr. Oscar Vega Amaya
Universidad de Sonora.

Este trabajo fue financiado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) dentro del proyecto "Aproximación, Estimación y Control en Sistemas Estocásticos" con número de referencia CB2010/154612, bajo la dirección del Dr. Jesús Adolfo Minjárez Sosa.

Dedicatoria

A Dios, eres la principal razón
por la que estoy aquí.

A mi esposo e hijo, Fran y Emanuel,
son mi motivo para salir adelante.

A mis padres, Lourdes y Manuel, que
han estado conmigo siempre.

A mis hermanos, Iván, Selene y Gustavo.

Agradecimientos

A lo largo de todo este tiempo, mientras transcurrían mis estudios de maestría, al igual que durante la realización de este trabajo, estuvieron presente muchas personas mostrándome su cariño y apoyo. Jamás terminaré de agradecérselos.

Primeramente quiero agradecer a Dios, pues nunca nos has dejado solos, a mí y a mi familia, y es gracias a ti que he alcanzado las metas que me he fijado.

Fran, sabes que sin tu compañía, paciencia y apoyo jamás podría haber concluido mis estudios con tanto éxito. Has sido mi soporte durante estos años de arduo trabajo construyendo nuestro futuro. Te agradezco inmensamente por tu amor y por comprender mi pasión hacia mis estudios y las matemáticas. Gracias por ayudarme a alcanzar mis sueños. Emanuel, aunque eres muy pequeño has sido una parte muy importante de todo esto, eres ese motorcito que me impulsa a seguir adelante y dar lo mejor de mí.

Mi agradecimiento es también para mis padres, Lourdes y Manuel. A pesar de estar lejos durante estos años, nunca dejaron de mostrarme su apoyo y amor incondicional cuando más lo necesité. Gracias por escucharme y hacerme sentir que estaban cerca de mí en los momentos difíciles. A mis hermanos Iván, Selene y Gustavo y a sus familias que siempre me han demostrado lo orgullosos que están de mi.

Carolina, te agradezco profundamente por tus consejos y palabras de aliento en los momentos precisos. Que apesar de la distancia siempre te hiciste presente con tu apoyo. Por ser mi amiga y mi hermana y compartir conmigo el amor a las matemáticas. Sé que llegarás muy lejos.

A mi profesor y Director de Tesis, Dr. Adolfo Minjárez. Gracias por tu apoyo y consejos como mentor y amigo y por las horas dedicadas a la preparación y revisión de esta tesis. Has hecho de mí una persona segura de sí misma y de lo que puede logar. Ha sido un honor conocerte y trabajar contigo.

Gracias a mis sinodales Dr. Agustín Brau Rojas, Dr. Oscar Vega Amaya y Dr. Carlos Gabriel Pacheco González, por sus consejos, amistad y el tiempo dedicado a la revisión de este trabajo.

Finalmente, quiero agradecer al Consejo Nacional de Ciencia y Tecnología por el apoyo recibido durante mis estudios de maestría, ya que sin él me hubiera sido imposible realizarlos.

Marla Lizzette Arcega López
Hermosillo, Sonora
Agosto de 2013

Índice

Introducción

El estudio de los Procesos de Control de Markov (PCMs), también conocidos como programación dinámica estocástica, inició en la década de 1960. Su objetivo es modelar y resolver problemas donde se deben de tomar decisiones secuenciales en múltiples períodos bajo un ambiente estocástico, con el fin de optimizar cierta función objetivo. A este problema se le conoce como Problema de Control Óptimo (PCO) y se presenta en áreas donde es requerido controlar sistemas dinámicos que aparecen, por ejemplo, en Ingeniería, Economía, Biología, Investigación de operaciones, entre otras.

Específicamente, en un PCO interviene un sistema dinámico cuyo comportamiento puede ser regulado o influenciado por desiciones que son tomadas por un controlador, las cuales son llamadas variables de control o acción. Las decisiones o acciones que son aplicadas en cada tiempo o etapa del sistema son elegidas de acuerdo a reglas conocidas como políticas de control. Además interviene una función llamada índice de funcionamiento, definida sobre el conjunto de las políticas de control, y evalúa, en algún sentido, la respuesta del sistema con respecto a la política que está siendo utilizada. Entonces el PCO es encontrar una política de control que optimize el índice de funcionamiento.

El estudio del PCO se puede clasificar según 1) el tipo de espacio de estados: numerable o no numerable (espacio de Borel); 2) el criterio de optimalidad, e.g., optimalidad descontada o promedio.

A partir de esta clasificación, se han introducido muchas generalizaciones para resolver problemas mas apegados a la realidad, como por ejemplo:

- PCMs adaptados, en los cuales algunas de las componentes del modelo de control correspondiente no son completamente conocidas;
- PCMs con restricciones, donde además de optimizar el índice de funcionamiento, las decisiones del controlador deben estar dirigidas a que algunas de las componentes del modelo de control satisfagan ciertas restricciones;

- PCMs parcialmente observables, donde el controlador cuenta con información parcial de las variables de estado al momento de tomar sus decisiones.

Además de las combinaciones entre estos procesos.

En el presente trabajo estudiamos el PCO asociado a PCMs Parcialmente Observables (PO) con espacios de Borel y bajo el criterio de optimalidad de costo descontado. Además presentaremos aplicaciones a sistemas de inventarios.

Los PCM-PO ha recibido considerable atención en los últimos años dentro la literatura de los PCMs por sus múltiples aplicaciones (ver, e.g., Bensoussan *et. al.* (2007,2008a,2008b,2010), Borkar (2003), Borkar and Budhirajab (2004), Brooks *et. al.* (2006), Hernández-Lerma and Romera (1999) y Hsu *et. al.* (2006)). En todos estostrabajos se sigue una técnica estándar (ver, e.g., Hernández-Lerma (1989), Bertsekas and Shreve (1978)) que consiste en transformar el PCM-PO en uno Completamente Observable (CO) cuyo espacio de estados es el espacio de todas las medidas de probabilidad sobre el espacio de estados original. Entonces, los dos procesos son equivalentes en el sentido de que ambos tienen el mismo costo óptimo.

Matemáticamente esta técnica es muy elegante y atractiva, pero meramente teórica. En efecto, la transformación está basada en la existencia de una función que define la dinámica del sistema CO en el espacio de medidas. Por lo tanto, el objetivo de muchos de los trabajos antes citados se centra en proporcionar condiciones que garanticen la existencia de políticas óptimas en el nuevo problema lo cual depende fuertemente de dicha función. Por otra parte, otros artículos se centran, especialmente, en obtener explicitamente dicha función en ejemplos concretos, o bajo condiciones particulares en el modelo de control, lo cual constituye precisamente la motivación principal del presente trabajo.

Especificamente, estudiaremos modelos de control de inventarios donde se tiene información parcial de la cantidad de artículos en existencia, es decir, del nivel de inventario.

En la práctica, el hecho de que el nivel de inventario sea la componente parcialmente observada, puede deberse a alguna de las siguientes causas: errores de transacción, mercancía extraviada, merma de productos, robo de mercancía o a la mala calidad de algunos productos. Además, otro hecho que hace que no se tenga información completa del nivel de inventario es cuando la demanda es parcialmente observada. Esta situación es común en los grandes establecimientos ya que en realidad lo que se observa son las ventas, y no la demanda no satisfecha. Entonces, dadas las grandes inversiones que hacen las empresas en sus inventarios, *control de inventarios* a pasado a formar parte de los temas más importantes dentro de las investigaciones en la industria en los últimos años con el objetivo de proponer modelos que consideren todos estos aspectos.

En esta tesis estudiamos dos tipos de modelos. En el primero suponemos que el controlador o el operador del inventario sólo observa cuando el nievel de inventario es cero, al cual se le conoce como *Zero Balance Walk*. En el otro modelo supondremos que el operador del inventario observa sólo las ventas, y por lo tanto la demanda es parcialmente observada. En ambos casos aplicaremos la transformación a un PCM-CO y obtendremos explícitamente la dinámica del nuevo sistema, para después mostrar la existencia de políticas óptimas bajo el criterio de costo descontado.

La tesis está estructurada de la siguiente manera. En el Capítulo 1 se presenta la teoría básica de los PCMs así como resultados en el caso CO. Enseguida, en el Capítulo 2, se formula de forma general el PCO para procesos de control PO, y finalmente, en el Capítulo 3, se presentan las aplicaciones a los sistemas de inventario. Adicionalmente incluimos un Apéndice que contiene resultados que estaremos haciendo referencia durante el desarrollo del trabajo, así como una tabla de notaciones para facilitar la lectura.

Capítulo 1

Procesos de Control de Markov Completamente Observables

1.1. Introducción

En este capítulo introducimos los elementos necesarios para definir el PCO asociado a los PCM-CO, bajo el criterio de costo descontado. Así mismo, daremos las condiciones necesarias que garantizan la existencia de políticas óptimas. Adicionalmente, presentamos algunos resultados que serán utilizados en el Capítulo 3, en el análisis de los problemas de Control de Inventarios

1.2. Procesos de Control de Markov

Definición 1.1. Un Modelo de Control de Markov a tiempo discreto es una colección

$$\mathcal{M} = (X, A, \{A(x) \,|x \in X\}, Q, c)$$

donde

a) X es un espacio de Borel, llamado espacio de estados.

b) A es un espacio de Borel, llamado espacio de controles o acciones.

c) $\{A(x) \,|x \in X\}$ una familia de subconjuntos medibles no vacíos $A(x)$ de A, donde $A(x)$ es el conjunto de controles o acciones admisibles cuando el sistema se encuentra en el estado $x \in X$. El conjunto $\mathbb{K}$ definido por

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

es el conjunto de las parejas admisibles estado-acción, el cual es un subconjunto medible de $X \times A$.

d) $Q(\cdot|x,a)$ es un kérnel estocástico sobre X dado $\mathbb{K}$, llamado ley de transición.

e) $c : \mathbb{K} \to \mathbb{R}$ una función medible que representa el costo por etapa.

En este trabajo se considerará el caso cuando la función c es posiblemente no acotada y no-negativa.

El modelo de control $\mathcal{M}$ representa un sistema estocástico controlado que se observa completamente en los tiempos $t = 0, 1, ...$ Se denotará por x_t y a_t al estado del sistema y la acción aplicada al tiempo t, respectivamente. La evolución del sistema se puede describir de la siguiente manera. Si el sistema se encuentra en el estado $x_t = x \in X$ al tiempo t y el control $a_t = a \in A(x)$ ha sido aplicado, entonces pasan dos cosas: (i) se genera un costo $c(x,a)$, y (ii) el sistema avanza al siguiente estado x_{t+1} el cual es una variable aleatoria definida en X con distribución $Q(\cdot|x,a)$, es decir,

$$Q(B|x,a) := Prob\,(x_{t+1} \in B|x_t = x, a_t = a) \quad B \in \mathcal{B}(X).$$

Una vez que la transición a un nuevo estado haya ocurrido, un nuevo control es elegido y el proceso se repite.

Observación 1.2. Existen situaciones en que la evolución del sistema está determinada por una ecuación en diferencias de la forma

$$x_{t+1} = F(x_t, a_t, \xi_t), \quad t = 0, 1, ...;$$

donde x_0 es conocido, $\{\xi_t\}$ es una sucesión de variables aleatorias independientes e identicamente distribuidas (i.i.d.) que toman valores en un espacio S y con distribución común μ, e independientes del estado inicial x_0. La sucesión $\{\xi_t\}$ es llamada proceso de perturbaciones aleatorias. En este caso la ley de transición $Q(\cdot|x,a)$ está dada por

$$\begin{aligned} Q(B|x,a) &= Prob(F(x_t,a_t,\xi_t) \in B|x_t = x, a_t = a) \\ &= \mu(\{s \in S|F(x,a,s) \in B\}) \\ &= \int_S 1_B[F(x,a,s)]\mu(ds), \quad B \in \mathcal{B}(X). \end{aligned}$$

1.3. Políticas de Control

Las acciones $a \in A(x)$ son seleccionadas de acuerdo a reglas conocidas como políticas de control, las cuales definimos en esta sección.

Definición 1.3. Considere un Modelo de Control de Markov. Una t-historia admisible es un vector de la forma

$$h_t = (x_0, a_0, x_1 ..., x_{t-1}, a_{t-1}, x_t),$$

con $(x_i, a_i) \in \mathbb{K}$ para $i = 0, 1, ..., t-1$ y $x_t \in X$.

Observemos que h_t es un elemento del espacio H_t de las historias admisibles hasta el tiempo t definido de la siguiente manera

$$H_0 := X$$

$$H_t := \mathbb{K}^t \times X = \mathbb{K} \times H_{t-1} \text{ para } t = 1, 2, ...$$

Definición 1.4. Una política de control, es una suceción $\pi = \{\pi_t\}$ $t = 0, 1, ...$ de kérneles estocásticos π_t sobre el espacio de control A dado H_t, que satisfacen

$$\pi_t(A(x_t)|h_t) = 1 \;\; \forall h_t \in H_t, \;\; t = 0, 1, ...$$

El conjunto de todas las políticas se denotará por Π.

Definición 1.5. Denotaremos por Φ al conjunto de todos los kérneles estocásticos $\phi \in \mathcal{P}(A|X)$ tal que $\phi(A(x)|x) = 1$ para toda $x \in X$.

Recuerde que $\mathbb{F}$ denota al conjunto de todos los selectores medibles de X en A, (ver Definición A.12 Apéndice).

Definición 1.6. Una política $\pi = \{\pi_t\} \in \Pi$ se dice ser

a) *de Markov aleatorizada* si existe una sucesión $\{\phi_t\}$ de kérneles estocásticos $\phi_t \in \Phi$ tal que
$$\pi_t(\cdot|h_t) = \phi_t(\cdot|x_t) \;\; \forall h_t \in H_t, \;\; t = 0, 1, ...$$

b) *estacionaria aleatorizada* si existe $\phi \in \Phi$ tal que

$$\pi_t(\cdot|h_t) = \phi(\cdot|x_t) \;\; \forall h_t \in H_t, \;\; t = 0, 1, ...$$

Los conjuntos de las políticas aleatorizadas de Markov y estacionarias se denotan por Π_{RM} y Π_{RS} respectivamente, y nótese que

$$\Pi_{RS} \subset \Pi_{RM} \subset \Pi$$

c) *determinista* si existe una sucesión g_t de funciones medibles $g_t : H_t \to A$ tal que, para toda $h_t \in H_t$ y $t = 0, 1, ...$ $g_t(h_t) \in A(x_t)$ y $\pi_t(\cdot|h_t)$ está concentrada en $g_t(h_t)$, es decir,

$$\pi_t(C|h_t) = I_C[g_t(h_t)] \text{ para toda } C \in \mathcal{B}(A)$$

d) *de Markov determinsta* si existe una sucesion f_t de funciones $f_t \in \mathbb{F}$ tal que $\pi_t(\cdot|h_t)$ está concentrada en $f_t(x_t) \in A(x_t)$ para toda $h_t \in H_t$ y $t = 0, 1, ...$

e) *estacionaria determinista* si existe una función $f \in \mathbb{F}$ tal que $\pi_t(\cdot|h_t)$ está concentrada en $f(x_t) \in A(x_t)$ para toda $h_t \in H_t$ y $t = 0, 1, ...$

   En este caso también tendremos

$$\Pi_{DS} \subset \Pi_{DM} \subset \Pi_D \subset \Pi$$

   donde Π_D, Π_{DM} y Π_{DS} denotan los conjuntos de las políticas deterministas, de Markov deterministas y estacionarias deterministas respectivamente.

Observemos que a las políticas deterministas las definen g_t, f_t o f según sea el caso. De aquí, que para fines prácticos podemos decir que una política determinista es una sucesión de funciones medibles correspondientes.

Notación. Si $\pi \in \Pi$ es una política determinista la denotaremos como $\{g_t\}$, $\{f_t\}$ y f^∞, si π está en Π_D, Π_{DM} y Π_{DS}, respectivamente.

Sea $(\Omega, \mathcal{F})$ el espacio medible que consta de el espacio muestral $\Omega := \bar{H}_\infty = (X \times A)^\infty$, y $\mathcal{F}$ la correspondiente σ-álgebra producto. Los elementos de Ω son sucesiones de la forma $\omega = (x_0, a_0, x_1, a_1, ...)$ con x_t en X y a_t en A para toda $t = 0, 1, ...$ Nótese que Ω contiene al espacio $H_\infty = \mathbb{K}^\infty$ de las historias admisibles $(x_0, a_0, x_1, a_1, ...)$ con $(x_t, a_t) \in \mathbb{K}$ para toda $t = 0, 1, ...$

Sea $\pi = \{\pi_t\}$ una política de control arbitraria y ν, una medida de probabilidad arbitraria sobre X, una distribución inicial. Entonces por el Teorema de Ionescu-Tulcea (ver, Proposición C.10 p178 Hernández-Lerma and Lasserre (1996)), existe una única medida de probabilidad P_ν^π sobre $(\Omega, \mathcal{F})$ concentrada en H_∞, es decir $P_\nu^\pi(H_\infty) = 1$, y además para cada $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(A)$, y $h_t \in H_t$

$$P_\nu^\pi(x_0 \in B) = \nu(B)$$

$$P_\nu^\pi(a_t \in C|h_t) = \pi_t(C|h_t)$$

$$P_\nu^\pi(x_{t+1} \in B|h_t, a_t) = Q(B|x_t, a_t)$$

Observación 1.7. Cuando la distribución inicial está concentrada en un punto, es decir $\nu(\{x\}) = 1$ para algún $x \in X$, la medida de probabilidad P_ν^π se denota por P_x^π, y el operador esperanza asociado a esta medida de probabilidad por E_x^π.

Definición 1.8. Al proceso estocástico $(\Omega, \mathcal{F}, P_\nu^\pi, \{x_t\})$ se le llama proceso de decisión de Markov a tiempo discreto.

1.4. Criterio de Costo Descontado con Horizonte Infinito

Dado un modelo de control $(X, A, \{A\,(x)\,|x \in X\}, Q, c)$, una política $\pi \in \Pi$ y un estado inicial $x \in X$, el índice de funcionamiento **costo descontado** se define como

$$V\,(\pi, x) := E_x^\pi[\sum_{t=0}^{\infty} \alpha^t c\,(x_t, a_t)], \tag{1.1}$$

donde $\alpha \in (0,1)$ es el factor de descuento. La **función de valor óptimo** está dada por

$$V^*\,(x) := \inf_{\Pi} V\,(\pi, x)\,, \quad x \in X. \tag{1.2}$$

Entonces el **problema de control óptimo** es encontrar una política $\pi^* \in \Pi$ tal que

$$V\,(\pi^*, x) = V^*\,(x)\,, \quad \forall x \in X. \tag{1.3}$$

Observe que si definimos

$$V_n\,(\pi, x) := E_x^\pi\left[\sum_{t=0}^{n-1} \alpha^t c\,(x_t, a_t)\right], \tag{1.4}$$

el cual representa el costo descontado esperado en n-etapas, por el Teorema de Convergencia Monótona, podemos escribir al costo descontado esperado total $V\,(\pi, x)$, en (1.1), como

$$V\,(\pi, x) = \lim_{n\to\infty} V_n\,(\pi, x)\,. \tag{1.5}$$

1.4.1. Ecuación de Optimalidad

Una función medible $v : X \to \mathbb{R}$ se dice ser una solución a la **ecuación de optimalidad** (EO) si satisface

$$v(x) = \min_{a \in A(x)} \left[c(x,a) + \alpha \int_X v(y) Q(dy|x,a) \right], \quad \forall x \in X. \tag{1.6}$$

Para modelos en ecuaciones en diferencias como en la Observación 1.2, la EO toma la forma

$$v(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_S v[F(x,a,s)] \mu(ds) \right\} = \min_{a \in A(x)} \left\{ c(x,a) + \alpha E_x^{\pi} [v(x_1)] \right\} \tag{1.7}$$

Definición 1.9. Para cada $u \in \mathcal{L}(X)$, se definen los operadores $T_a u$ y Tu como

$$T_a u(x) := c(x,a) + \alpha \int_X u(y) Q(dy|x,a), \quad x \in X, a \in A(x), \tag{1.8}$$

y

$$Tu(x) := \min_{a \in A(x)} T_a u(x), \quad x \in X. \tag{1.9}$$

Más adelante se demostrará que la función de valor óptimo, definida anteriormente, es una solución de la EO, es decir

$$V^*(x) = TV^*(x), \quad \forall x \in X. \tag{1.10}$$

Una manera de abordar el problema es utilizando el **Algoritmo de Iteración de Valores** (IV) definido como:

$$v_0(\cdot) \equiv 0$$
$$v_n(x) := Tv_{n-1}(x), \tag{1.11}$$

para toda $x \in X$ y $n = 1, 2, ...,$ y mostrar que

$$V^*(x) = \lim_{n \to \infty} v_n(x) \quad \forall x \in X, \tag{1.12}$$

el cual es un resultado esperado pues v_n es la función de valor óptimo del costo descontado en n-etapas definido en (1.4) con costo terminal cero (ver, Remark 5.2, Hernández-Lerma (1990)), es decir

$$v_n(x) = \inf_{\pi \in \Pi} V_n(\pi, x) \quad \forall x \in X. \tag{1.13}$$

Para mostrar lo anterior se supondran las siguientes condiciones.

Hipótesis 1.10. El modelo de Control es tal que:

a) Para cada $x \in X$, el conjunto $A(x)$ es compacto y la multifunción $x \mapsto A(x)$ es semicontinua superiormente (u.s.c), es decir para cada subconjunto abierto $G \subset A$ el conjunto $\{x|A(x) \subset G\}$ es abierto en X.

b) La función costo-por-etapa c es semicontinua inferiormente (l.s.c.) y no negativa.

c) $T_a u \in \mathcal{L}_+(\mathbb{K})$ para toda $u \in \mathcal{L}_+(X)$.

Observación 1.11. Supóngase que en la Hipótesis 1.10(b), en lugar de $c \geq 0$, tenemos $c \geq -m$, y sea $c'(x,a) := c(x,a) + m \geq 0$. Definimos el nuevo índice de funcionamiento asociado al nuevo costo-por-etapa c' como

$$V^{'}(\pi, x) := E_x^{\pi}[\sum_{t=0}^{\infty} \alpha^t c'(x_t, a_t)].$$

Entonces,es fácil ver que

$$V^{'}(\pi, x) = V(\pi, x) + \frac{m}{1-\alpha},$$

y las funciones de valor óptimo $V^{'*}$ y V^* asociadas a c' y c, respectivamente, satisfacen

$$V^{'*}(x) = V^*(x) + \frac{m}{1-\alpha}.$$

Similarmente podemos demostrar que, v es una solución de la ecuación de optimalidad (1.6) si y sólo si $v' = v + \frac{m}{1-\alpha}$ satisface $v' = T^{'}v'$, donde $T^{'}$ es el operador obtenido cuando c es remplazado por c' en T. Así, si c es acotada inferiormente, es decir $c \geq -m$, para algún $m \in \mathbb{R}^+$, no hay pérdida de generalidad en suponer $c \geq 0$ como el la Hipótesis 1.10(b).

El siguiente resultado proporciona condiciones suficientes que garantizan la Hipótesis 1.10(c).

Lema 1.12. Las siguientes condiciones implican la Hipótesis 1.10(c).

C1 $c \in \mathcal{L}_+(\mathbb{K})$

C2 El kérnel de transición $Q(\cdot|x,a)$ es débilmente continuo, es decir, $v^{'}(x,a) := \int v(y)Q(dy|x,a)$ es continua y acotada en $\mathbb{K}$ para cada función v continua y acotada en X.

Demostración. Sea $v(x) \geq 0$ una función l.s.c en X, entonces por la Proposición A.3(b) del Apéndice sabemos que existe una sucesión de funciones no-negativas $\{v_n\} \subset \mathcal{C}(X)$ tal que $v_n \uparrow v$. Como $v_n(\cdot)$ es continua y acotada para cada $n \in \mathbb{N}$ y el kérnel $Q(\cdot|x,a)$ es débilmente continuo, se sigue que para cada $n \in \mathbb{N}$ la función

$$v_n'(x,a) := \int v_n(y)Q(dy|x,a) \quad (x,a) \in \mathbb{K}$$

es continua, acotada y no-negativa.

Ahora bien, puesto que $v_n(x) \uparrow v(x)$, por el Teorema de Convergencia Monótona

$$v_n'(x,a) \to v'(x,a) := \int v(y)Q(dy|x,a)$$

y de hecho la convergencia es creciente. Así, hemos encontrado una sucesión de funciones continuas y acotadas $v_n'(x,a)$ que convergen crecientemente a $v'(x,a)$ para toda pareja $(x,a) \in \mathbb{K}$, y en consecuencia, por la Proposición A.3(b) del Apéndice $v'(x,a) := \int v(y)Q(dy|x,a)$ es semicontinua inferiormente, acotada por abajo y no-negativa.

Finalmente, de lo anterior y de la Condición C1 se obtiene que

$$c(x,a) + \alpha \int v(y)Q(dy|x,a)$$

pertenece a $\mathcal{L}_+(\mathbb{K})$ para cada $v \in \mathcal{L}_+(X)$. $\square$

De las la Hipótesis 1.10(a)-(c) y la Proposición A.13(d) del Apéndice se obtiene el siguiente lema.

Lema 1.13. *El operador T mapea $\mathcal{L}_+(X)$ en si mismo, es decir, para cada $u \in \mathcal{L}_+(X)$, se tiene que $Tu \in \mathcal{L}_+(X)$, y además existe un selector $f \in \mathbb{F}$ tal que*

$$Tu(x) = c(x,f) + \alpha \int_X u(y)Q(dy|x,f) \quad \forall x \in X. \tag{1.14}$$

Dado que estamos considerando una función de costo-por-etapa c posiblemente no acotada necesitaremos imponer la siguiente condicion

Hipótesis 1.14. Existe una política $\pi \in \Pi$ tal que $V(\pi,x) < \infty$ para toda $x \in X$.

Bajo la Hipótesis 1.14, es claro que la función de valor óptimo V^* es finita, es decir

$$V^*(x) < \infty \text{ para toda } x \in X.$$

Ahora ya estamos en condiciones de enunciar el teorema principal de este capítulo.

Teorema 1.15. *Suponga que se satisfacen las Hipótesis 1.10 y 1.14, y considere las funciones de IV* $\{v_n\}$*. Entonces:*

a) $v_n \uparrow V^*$ *y* V^* *es l.s.c.. Además la función de valor óptimo* V^* *es la mínima solución no-negativa de la EO, es decir*

$$V^*(x) = \min_{A(x)} \left[c(x,a) + \alpha \int_X V^*(y) Q(dy|x,a) \right] \quad \forall x \in X, \tag{1.15}$$

y si u *es cualquier otra solucion de la EO, entonces* $u(\cdot) \geq V^*(\cdot)$*.*

b) Existe un selector $f_* \in \mathbb{F}$*, en el cual se alcanza el mínimo en la ecuación (1.15), es decir,*

$$V^*(x) = c(x, f_*) + \alpha \int_X V^*(y) Q(dy|x, f_*) \ \ \forall x \in X, \tag{1.16}$$

y la política estacionaria determinista $f_*^\infty = \{f_*\}$ *determinada por el selector* f_* *es óptima. Recíprocamente, si* f_*^∞ *es óptima, entonces satisface la ecuación (1.16).*

Demostración. (a) Primero, observemos que T, por las propiedades de la integral, es un operador monótono. Así, el hecho que $c \geq 0$ y el Lema 1.13 implican que las funciones de IV $v_n = Tv_{n-1}$ forman una sucesión no-decreciente de funciones en $\mathcal{L}_+(X)$. Entonces existe una función $v^* \in \mathcal{L}_+(X)$ tal que $v_n \uparrow v^*$. Ahora lo que demostraremos es

(i) v^* satisface (1.15).

(ii) v^* es la mínima solución no-negativa de (1.15).

(iii) $v^* = V^*$.

Definamos las siguientes funciones

$$u_n(x,a) := c(x,a) + \alpha \int v_n(y) Q(dy|x,a),$$
$$u(x,a) := c(x,a) + \alpha \int v^*(y) Q(dy|x,a).$$

Como v^* y v_n $n = 0, 1, \ldots$ son funciones en $\mathcal{L}_+(\mathbb{K})$, entonces por la Hipótesis 1.10(c) las funciones $u(x,a)$ y $u_n(x,a)$ $n = 0, 1, \ldots$ también pertenecen a $\mathcal{L}_+(\mathbb{K})$. Así, por el Teorema de Convergencia Monónota se obtiene que $u_n \uparrow u$. Ahora, de la compacidad de $A(x)$ y dado que $u(x,a)$ y $u_n(x,a)$ $n = 0,1,\ldots$ pertenecen a $\mathcal{L}_+(\mathbb{K})$, los conjuntos $\{a \in A(x)|u(x,a) \leq r\}$ y $\{a \in A(x)|u_n(x,a) \leq r\}$ son compactos para cada $x \in X$ y $r \in \mathbb{R}$. Por lo tanto, por la Proposición A.5 del Apéndice

$$v^* = \lim_{n\to\infty} v_n = \lim_{n\to\infty} Tv_{n-1} = Tv^*,$$

es decir, v^* satisface (1.15).

Para demostrar (ii), supongamos que $u' \in \mathcal{L}_+(X)$ es otra función que satisface $u' = Tu'$. Como $u' \geq v_0 \equiv 0$, la monotonía de T implica que $u' \geq v_n$ para toda $n = 0,1,\ldots,$ y por tanto $u' \geq v^*$, lo cual demuestra (ii).

Finalmente mostremos (iii). Como $v^* = Tv^*$, por el Lema 1.13 existe $f \in \mathbb{F}$ tal que

$$v^*(x) = c(x,f) + \alpha \int_X v^*(y)Q(dy|x,f) \quad \forall x \in X. \tag{1.17}$$

Iterando (1.17) obtenemos, para toda $n \geq 1$ y $x \in X$,

$$\begin{aligned} v^*(x) &= E_x^{f^\infty}\left[\sum_{t=0}^n \alpha^t c(x_t,f)\right] + \alpha^{n+1}E_x^{f^\infty} v^*(x_{n+1}) \\ &\geq E_x^{f^\infty}\left[\sum_{t=0}^n \alpha^t c(x_t,f)\right] \end{aligned} \tag{1.18}$$

pues $v^* \geq 0$. Por tanto, haciendo $n \to \infty$ obtenemos

$$v^*(x) \geq V(f^\infty, x) \geq V^*(x),$$

por definición de V^*. Para obtener la otra desigualdad nótese que de (1.13), (1.1) y de la Hipótesis 1.10(b) $(c \geq 0)$,

$$v_n(x) \leq V_n(\pi,x) \leq V(\pi,x) \quad \forall n,\ \forall \pi \in \Pi,\ \forall x \in X.$$

Luego, como $v_n \uparrow v^*$, tenemos

$$v^*(\cdot) \leq V(\pi,\cdot) \quad \forall \pi \in \Pi,$$

lo cual implica que $v^*(\cdot) \leq V^*(\cdot)$, por la definición de V^*. Esto prueba (iii), completando así la demostración de la parte (a) del teorema.

(b) La existencia de un selector $f_* \in \mathbb{F}$ que satisfaga (1.16) es asegurado por el Lema 1.13. Ahora, iterando (1.16) se tiene [como en (1.18)] que

$$\begin{aligned} V^*(x) &= E_x^{f_*^\infty}\left[\sum_{t=0}^n \alpha^t c(x_t,f_*)\right] + \alpha^{n+1}E_x^{f_*^\infty} V^*(x_{n+1}) \\ &\geq E_x^{f_*^\infty}\left[\sum_{t=0}^n \alpha^t c(x_t,f_*)\right] \quad \forall n \geq 0,\ x \in X. \end{aligned}$$

Esto implica que, haciendo $n \to \infty$, $V^*(x) \geq V(f_*^\infty, x) \quad \forall x \in X$, mientras que de (1.2), $V^*(\cdot) \leq V(f_*^\infty, \cdot)$. Por lo tanto $V^*(\cdot) = V(f_*^\infty, \cdot)$, es decir f_*^∞ es óptima.

Para mostrar el recíproco, notemos primero que para cualquier política estacionaria determinista $f^\infty \in \Pi_{DS}$, el costo descontado $V(f^\infty, \cdot)$ satisface

$$V(f^\infty, x) = c(x, f) + \alpha \int_X V(f^\infty, y) Q(dy|x, f) \;\; \forall x \in X. \tag{1.19}$$

En efecto,

$$\begin{aligned} V(f^\infty, x) &:= E_x^{f^\infty} \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, f) \right] \\ &= E_x^{f^\infty} \left[c(x_0, f) + \sum_{t=1}^{\infty} \alpha^t c(x_t, f) \right] \\ &= c(x, f) + \alpha E_x^{f^\infty} \left[\sum_{t=1}^{\infty} \alpha^{t-1} c(x_t, f) \right]. \end{aligned} \tag{1.20}$$

Ahora, por propiedades de la esperanza condicional y la propiedad de Markov, la suma esperada del lado derecho de (1.20) puede expresarse como

$$\begin{aligned} E_x^{f^\infty} \left[\sum_{t=1}^{\infty} \alpha^{t-1} c(x_t, f) \right] &= \int_X E^{f^\infty} \left[\sum_{t=1}^{\infty} \alpha^{t-1} c(x_t, f) | x_1 = y \right] Q(dy|x, f) \\ &= \int_X V\left(f^\infty, y\right) Q\left(dy|x, f\right). \end{aligned}$$

En particular, si $f_*^\infty \in \Pi_{DS}$ es óptima, entonces $V(f_*^\infty, \cdot) = V^*(\cdot)$, en tal caso (1.19), con $f = f_*$, se reduce a (1.16).

$\square$

Capítulo 2

Procesos de Control de Markov Parcialmente Obsrvables

2.1. Introducción

Los sistemas de control estocástico con información parcial del estado aparecen en muchos problemas de ingeniería, economía, crecimiento de poblaciones, entre muchas otras áreas. La característica de estos sistemas es que, a diferencia de la situación estudiada en el capítulo anterior, el controlador sólo cuenta con observaciones parciales del estado del sistema x_t a través de un proceso de observación y_t. Un modelo típico en estos casos es de la forma:

$$\begin{aligned}
&\text{(a) } x_{t+1} = F(x_t, a_t, \xi_t),\ t = 0, 1, \ldots, \\
&\text{(b) } y_t = G(a_{t-1}, x_t, \eta_t),\ t = 1, 2, \ldots, \\
&\text{(c) } y_0 = G_0(x_0, \eta_0),
\end{aligned} \tag{2.1}$$

donde F y G son funciones conocidas, x_t, a_t y y_t son, respectivamente, el estado, el control y la observación al tiempo t, $\{\xi_t\}$ es una sucesión de v.a. i.i.d. con distribución común μ representando el proceso de perturbaciones en los estados, y $\{\eta_t\}$ son v.a. i.i.d. con distribución λ que representa el ruido aleatorio de la observación. El objetivo del presente capítulo es estudiar PCM-PO modelados por medio de ecuaciones en diferencia de la forma (2.1). Para esto, primero describiremos el Modelo de Control correspondiente, y plantearemos el PCO. Después mediante una transformación, vemos que este problema es equivalente a un PCO-CO, y entonces aplicamos los resultados del capítulo anterior para mostrar la existencia de políticas óptimas.

2.2. Modelo de Control Parcialmente Observable

Definición 2.1. El modelo de control del PCM-PO está definido por $(X, Y, A, Q, K, K_0, \nu, c)$, donde:

(a) X, el espacio de estados, es un espacio de Borel.

(b) Y, el conjunto de observaciones, es un espacio de Borel.

(c) A, el espacio de control, es un espacio de Borel.

(d) $Q(\cdot|x, a)$, la ley de transición, es un kérnel estocástico sobre X dado $X \times A$.

(e) $K(\cdot|a, x)$, el kérnel observación, es un kérnel estocástico sobre Y dado $A \times X$.

(f) $K_0(\cdot|x)$, el kérnel observación inicial, es un kérnel estocástico sobre Y dado X.

(g) $\nu \in \mathbb{P}(X)$, es la distribución inicial (a priori).

(h) c, es la función de costo-por-etapa.

En este capítulo consideraremos que $A(x) = A \; \forall x \in X$ es compacto y la función c posiblemente no acotada.

El PCM-PO evoluciona de la siguiente manera. Al tiempo $t = 0$, el estado inicial (no-observable) x_0 tiene una distribución inicial a priori ν, y la observación inicial y_0 es generada de acuerdo al kérnel observación inicial $K_0(\cdot|x_0)$. Tomando en cuenta esta observación, el controlador elige una acción $a_0 \in A$, y entonces 1) se genera un costo $c(x_0, a_0)$; y 2) el sistema se mueve a un nuevo estado x_1 de acuerdo a la ley de transición $Q(\cdot|x_0, a_0)$. De aquí, una nueva observacioń y_1 es obtenida de acuerdo al kérnel $K(\cdot|a_0, x_1)$ y tomando en cuenta esta observación se selecciona el control a_1, y el proceso se repite.

Observación 2.2. Consideremos un sistema como en (2.1), donde x_t, y_t y a_t toman sus respectivos valores en los espacios de Borel X, Y y A. Supongamos también que $\{\xi_t\}$ y $\{\eta_t\}$ son sucesiones de elementos aleatorios mutuamente independientes, i.i.d. y con valores en los espacios de Borel S y N, con distribuciones $\mu \in \mathbb{P}(S)$ y $\lambda \in \mathbb{P}(N)$, respectivamente; F, G y G_0 son funciones medibles dadas, y x_0 es independiente de $\{\xi_t\}$ y $\{\eta_t\}$. Entonces la ley de transición está dada por

$$Q(B|x, a) = \int_S 1_B[F(x, a, s)]\mu(ds) \;\; B \in \mathcal{B}(X), \; x \in X \text{ y } a \in A. \tag{2.2}$$

Similarmente, si $a_t = a$ y $x_{t+1} = x$, el kérnel de observación está dado por

$$K(C|a,x) = \int_N 1_C[G(a,x,n)]\lambda(dn) \ \ \forall C \in \mathcal{B}(Y), \tag{2.3}$$

y si $x_0 = x$,

$$K_0(C|x) = \int_N 1_C[G_0(x,n)]\lambda(dn) \ \ \forall C \in \mathcal{B}(Y).$$

Una realización del sistema Parcialmente Observable (PO) toma la forma

$$(x_0, y_0, a_0, x_1, y_1, a_1, \dots) \in \Omega := (X \times Y \times A)^\infty,$$

donde x_0 tiene una distribución dada $\nu \in \mathbb{P}(X)$, y $\{a_t\}$ es una sucesión de controles en A determinada por un política de control definida más adelante.

Observemos que en este caso las decisiones del controlador dependen de la historia observada definida de la siguiente manera

$$\begin{aligned} h_0 &:= (\nu, y_0) \in H_0 \\ h_t &:= (\nu, y_0, a_0, \dots, y_{t-1}, a_{t-1}, y_t) \in H_t \ \text{ para } \ t \geq 1, \end{aligned}$$

donde $H_0 := \mathbf{Z} \times Y$ y $H_t := H_{t-1} \times A \times Y$ si $t \geq 1$ con $\mathbf{Z} := \mathbb{P}(X)$. Entonces una política para el PCM-PO está definida como una sucesión $\pi = \{\pi_t\}$, tal que, para cada t, π_t es un kérnel estocástico sobre A dado H_t. Además como en el Capítulo anterior una política de control determinista $\{\pi_t\}$ puede ser definida como una sucesión de funciones H_t-medibles que toman valores en A. De nuevo, al conjunto de todas las políticas lo denotaremos por Π, tomando en cuenta el presente contexto.

El problema de control PO se plantea similarmente considerando el conjunto de políticas Π, y suponiendo que x_0 tiene distribución $\nu \in \mathbf{Z}$.

Sea

$$J(\pi, \nu) := E_\nu^\pi \sum_{t=0}^\infty \alpha^t c(x_t, a_t)$$

el costo descontado total esperado bajo la política $\pi \in \Pi$ y distribución inicial $\nu \in \mathbf{Z}$, donde $\alpha \in (0,1)$ es el factor de descuento dado. Entonces el problema de control PO es encontrar un política $\pi^* \in \Pi$ tal que

$$J(\pi^*, \nu) = J^*(\nu) \ \text{ para toda } \nu \in \mathbf{Z},$$

donde

$$J^*(\nu) := \inf_{\pi \in \Pi} J(\pi, \nu), \ \ \nu \in \mathbf{Z},$$

es la función de costo óptimo.

2.3. Transformación a un Problema de Control Completamente Observable

Existe una metodología estandar para estudiar modelos de control PO la cual consiste en transformar el problema correspondiente en uno completamente observable. Esta transformación se hace introduciendo un proceso de filtrado para el proceso de estados $\{x_t\}$ con lo cual obtenemos un nuevo proceso $\{z_t\} \subset \mathbf{Z} = \mathbb{P}(X)$ que evoluciona recursivamente de acuerdo a una ecuación en diferencias de la forma

$$\begin{aligned} z_0 &= H_0(\nu, y_0) \\ z_{t+1} &= H(z_t, a_t, y_{t+1}), \quad t = 0, 1, \ldots, \end{aligned} \tag{2.4}$$

el cual constituirá el nuevo proceso completamente observable.

En (2.4) H es una función conocida, $\{a_t\}$ y $\{y_t\}$ son las sucesiones de controles y observaciones respectivamente del PCM-PO original, y $\nu \in \mathbf{Z}$ es la distribución inicial a priori del estado incial x_0. Por lo tanto, el problema principal al que nos enfrentamos es cómo obtener la función H. Para esto seguiremos las ideas de Hernández-Lerma (1989) las cuales estan basadas en la Proposición A.10 del Apéndice. En efecto, tomemos X y Y como los espacios de estados y de observaciones del PCM-PO, respectivamente, y $W := \mathbf{Z} \times A$, donde A es el espacio de control y $\mathbf{Z} = \mathbb{P}(X)$ como antes. Sea $R(d(x,y)|z,a)$ el kérnel estocástico sobre $X \times Y$ dado $W = \mathbf{Z} \times A$ tal que para toda $B \in \mathcal{B}(X)$ y $C \in \mathcal{B}(Y)$,

$$R(B \times C|z,a) := \int_X \int_B K(C|a,x')Q(dx'|x,a)z(dx), \tag{2.5}$$

donde $Q(\cdot|x,a)$ y $K(\cdot|a,x)$ son, respectivamente, la ley de transición de los estados y el kérnel observación en el PCM-PO. Entonces por la Proposición A.10 del Apéndice, existe un kérnel estocástico $H'(dx|z,a,y)$ sobre X dado $W \times Y = \mathbf{Z} \times A \times Y$ tal que, para toda $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(Y)$ y $(z,a) \in W$,

$$R(B \times C|z,a) = \int_C H'(B|z,a,y)R'(dy|z,a), \tag{2.6}$$

donde $R'(C|z,a) := R(X \times C|z,a)$ es la marginal de $R(\cdot|z,a)$ sobre Y, la cual toma la forma

$$R'(C|z,a) = \int_X \int_X K(C|a,x')Q(dx'|x,a)z(dx),$$

para todo $C \in \mathcal{B}(Y)$ y $(z,a) \in W$.

Entonces se sigue, por las propiedades de los kérneles estocásticos, que la función $H : \mathbf{Z} \times A \times Y \to \mathbf{Z}$ definida por

$$H(z, a, y) := H'(\cdot|z, a, y) \tag{2.7}$$

es medible, y por tanto

$$k(D|z, a) := \int_Y 1_D[H(z, a, y)]R'(dy|z, a), \tag{2.8}$$

donde $D \in \mathcal{B}(\mathbf{Z})$ y $(z, a) \in W = \mathbf{Z} \times A$, define un kérnel estocástico sobre $\mathbf{Z}$ dado $\mathbf{Z} \times A$.

Un argumento similar muestra que si $K_0(\cdot|x)$ es el kérnel de observación inicial en el PCM-PO, entonces existe un kérnel estocástico $H_0^{'}(dx|\nu, y)$ sobre X dado $\mathbf{Z} \times Y$ tal que, para todo $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(Y)$ y $\nu \in \mathbf{Z}$, el kérnel estocástico

$$R_0(B \times C|\nu) := \int_B K_0(C|x)\nu(dx)$$

sobre $X \times Y$ dado $\mathbf{Z}$, puede ser descompuesto como

$$R_0(B \times C|\nu) = \int_C H_0^{'}(B|\nu, y)R_0^{'}(dy|\nu),$$

donde

$$R_0^{'}(C|\nu) = \int_X K_0(C|x)\nu(dx) \;\; \forall C \in \mathcal{B}(Y) \;\; y \;\; \nu \in \mathbf{Z}.$$

Así, la función $H_0 : \mathbf{Z} \times Y \to \mathbf{Z}$ definada por

$$H_0(\nu, y) := H_0^{'}(\cdot|\nu, y) \;\text{ para toda }\; (\nu, y) \in \mathbf{Z} \times Y \tag{2.9}$$

es medible, y por tanto,

$$k_0(D|\nu) := \int_Y 1_D[H_0(\nu, y)]R_0^{'}(dy|\nu), \;\; D \in \mathcal{B}(\mathbf{Z}), \; \nu \in \mathbf{Z}, \tag{2.10}$$

define un kérnel estocástico sobre $\mathbf{Z}$ dado $\mathbf{Z}$.

Observación 2.3. Las funciones H y H_0 en (2.4) están definidas por (2.7) y (2.9). Siendo así, el elemento aleatorio z_t en el proceso de filtrado (2.4) puede ser interpretado como la distribución a posteriori del estado no-observable x_t dada la historia observable h_t. Esto es, para alguna política π, una distribución inicial a priori $\nu_0 = \nu \in \mathbf{Z}$ y $B \in \mathcal{B}(X)$,

$$P_\nu^\pi(x_0 \in B|h_0) = H_0(h_0)(B) = z_0(B) \;\; P_\nu^\pi - c.s., \tag{2.11}$$

donde $h_0 = (\nu, y_0)$, y para toda $t \geq 0$ e historia observable $h_{t+1} = (h_t, a_t, y_{t+1})$

$$P_\nu^\pi(x_{t+1} \in B | h_{t+1}) = H(z_t, a_t, y_{t+1})(B) = z_{t+1}(B) \ \ P_\nu^\pi - c.s.$$

Observemos que por definición de política, $\sigma(y_0, y_1 \dots y_t) = \sigma(h_t)$. Entonces denotando $\mathcal{Y}_t = \sigma(y_0, y_1, \dots y_t)$ la σ-álgebra de la historia observada, tenemos que

$$z_{t+1}(B) = P_\nu^\pi[x_{t+1} \in B | \mathcal{Y}_t]. \tag{2.12}$$

[Para la demostración véase, por ejemplo, Bertsekas and Shreve (1978), Sección 10.3.1; Striebel (1975), Sección 2.2; o Sawaragi y Yoshikawa (1970) cuando X y Y son numerables.]

2.4. I-Políticas

Ahora introducimos el conjunto de políticas correspondiente al sistema controlado completamente observado (2.4). Definimos los vectores información i_t hasta el tiempo t como

$$i_t = (z_0, a_0, \dots, z_{t-1}, a_{t-1}, z_t) \in I_t,$$

donde $I_t := \mathbf{Z} \times (A \times \mathbf{Z})^t$ para toda $t = 1, 2, \dots$, con $I_0 := \mathbf{Z}$. Entonces, de acuerdo a la definición de política en el caso completamente observable vista en el capítulo anterior, una política de información o I-política es una sucesión $\delta = \{\delta_t\}$ tal que, para cada t, $\delta_t(da|i_t)$ es un kérnel estocástico sobre A dado I_t. Denotaremos por Δ al conjunto de todas las I-políticas.

Una sucesión $\{f_t\}$ de funciones medibles $f_t : \mathbf{Z} \to A$ es llamada una I-política de Markov en el sentido usual, y nótese que el conjunto de todas las I-políticas de Markov es un subconjunto de Δ. También como es usual, una I-política de Markov $\{f_t\}$ en la cual $f_t = f$ independiente de t es llamada una I-política estacionaria y nos referiremos a ella como la I-política estacionaria f^∞.

Consideraremos Δ como un subconjunto de Π; esto es, consideraremos una I-política $\delta \in \Delta$ como una política $\pi \in \Pi$. Podemos hacer esto porque cualquier I-política $\delta = \{\delta_t\}$ define una política $\pi^\delta = \{\pi_t^\delta\}$ en Π dada por

$$\pi_t^\delta(\cdot|h_t) := \delta_t(\cdot|i_t(h_t)) \ \ \forall \, h_t \in H_t, \ \ t \geq 0,$$

donde $i_t(h_t) \in I_t$ es el vector información determinado por la historia observable h_t por medio de (2.4). Asi δ y π^δ son equivalentes en el sentido que, para cada $t \geq 0$, π_t^δ asigna la misma probabilidad condicional sobre A que la que es asignada por δ_t

para cada historia observable h_t. De hecho se tiene el siguiente resultado [Ver, por ejemplo, Hernandez-Lerma (1989), Yushkevich (1976)].

Proposición 2.4. Δ *es completo. Esto es, para cualquier política* $\pi \in \Pi$ *existe una I-política* $\delta \in \Delta$ *tal que*

$$J(\delta, \nu) = J(\pi, \nu) \ \ \forall \nu \in \mathbf{Z}.$$

2.5. Modelo de Control CO

Dado el Modelo de Control PO $(X, Y, A, Q, K, K_0, \nu, c)$ de la Definición 2.1, considere el Modelo de Control CO $(\mathbf{Z}, A, k, k_0, \widetilde{c})$ con espacio de estados $\mathbf{Z} = \mathbb{P}(X)$, espacio de control A, ley de transición $k(\cdot|z,a)$ definida en (2.8), distribución inicial $k_0(\cdot|\nu)$ como en (2.10) cuando la distribución a priori del estado inicial no-observable x_0 es $\nu \in \mathbf{Z}$, y función de costo $\widetilde{c} : \mathbf{Z} \times A \to \mathbb{R}$ definida por

$$\widetilde{c}(z, a) := \int_X c(x, a) z(dx). \tag{2.13}$$

En este caso, el proceso de control CO está definido por medio de la ecuación en diferencias (2.4), similar a la situación que se presentó en la Observación 1.2.

El conjunto de políticas del PCM-CO es el conjunto Δ de las I-políticas. Como en el capítulo anterior, una I-política $\delta \in \Delta$ y una distribución a priori $\nu \in \mathbf{Z}$ de x_0 (junto con los kérneles estocásticos k y $k_0(\cdot|\nu)$) definen una medida de probabilidad $P_\nu^{\prime\delta}$ sobre el espacio $(\mathbf{Z} \times A)^\infty$ y el operador esperanza correspondiente es denotado por $E_\nu^{\prime\delta}$. Entonces el costo descontado total esperado del PCM-CO está dado por

$$V(\delta, \nu) := E_\nu^{\prime\delta} \sum_{t=0}^{\infty} \alpha^t \widetilde{c}(z_t, a_t) \ \text{ para } \ \delta \in \Delta \ \text{ y } \ \nu \in \mathbf{Z},$$

y (como en (1.1)) el problema de control CO es determinar una I-política δ^* tal que

$$V(\delta^*, \nu) = V^*(\nu) \ \text{ para toda } \ \nu \in \mathbf{Z},$$

donde

$$V^*(\nu) := \inf_{\delta \in \Delta} V(\delta, \nu). \tag{2.14}$$

Este problema de control y el problema de control PO original son equivalentes en el sentido del siguiente resultado [Ver, por ejemplo, Hernandez-Lerma (1989), Yushkevich (1976)].

Proposición 2.5. $V(\delta, \nu) = J(\delta, \nu)$ *para toda* $\delta \in \Delta$ *y* $\nu \in \mathbf{Z}$

De las Proposiciones 2.4 y 2.5 se sigue que una I-política es óptima para el problema de control CO si, y sólo si, es óptima para el problema PO original. En otras palabras, los resultados para el problema CO pueden ser aplicadas para obtener la solución del problema PO, remplazando políticas por I-políticas. Seguiremos este enfoque en la siguiente sección para obtener políticas óptimas para el problema de control PO.

2.6. I-Políticas Óptimas

El objetivo principal en esta sección es dar condiciones bajo las cuales los resultados de optimalidad para los problemas bajo el criterio de costo descontado obtenidos en el Capítulo 1 se cumplan para el Modelo de Control CO $(\mathbf{Z}, A, k, k_0, \widetilde{c})$. Para hacer esto consideraremos el Modelo $(X, Y, A, Q, K, K_0, \nu, c)$ de la Definición 2.1 y supongamos que satisface las siguientes condiciones.

Hipótesis 2.6. El Modelo de Control es tal que:

(a) $A(x) = A$ es compacto $\forall x \in X$.

(b) La función de costo por etapa $c(x, a)$ es continua y no-negativa en $X \times A$.

(c) La ley de transición $Q(\cdot|x, a)$ y el kérnel observación $K(\cdot|a, x)$ son kérneles estocásticos débilmente continuos.

(d) La función $H(z, a, y)$ en (2.7) es continua en $\mathbf{Z} \times A \times Y$

Estas condiciones, serán brevemente discutidas en la Observación 2.9 al final del capítulo.

Lema 2.7. *La Hipótesis 2.6 implica que el Modelo de Control CO* $(\mathbf{Z}, A, k, k_0, \widetilde{c})$ *satisface:*

(a) La función de costo por etapa $\widetilde{c}(z, a)$ *es l.s.c. y no-neg en* $\mathbf{Z} \times A$.

(b) El kérnel estocástico $k(dz'|z, a)$ *es débilmente continuo.*

Demostración. (a) Sea (z_n, a_n) una sucesión en $\mathbf{Z} \times A$ que converge a $(z, a) \in \mathbf{Z} \times A$. Queremos mostrar que

$$\liminf_{n\to\infty} \widetilde{c}(z_n, a_n) \geq \widetilde{c}(z, a),$$

es decir, por (2.13),

$$\liminf_{n\to\infty} \int_X c(x,a_n) z_n(dx) \geq \int_X c(x,a) z(dx).$$

Tomando $v_n(x) := c(x,a_n)$ y $v(x) := c(x,a)$ en la Proposición A.6 de Apéndice se sigue el resultado.

(b) El objetivo es mostrar que para cada función $v \in C(\mathbf{Z})$, la función

$$\widetilde{v}(z,a) := \int_{\mathbf{Z}} v(z^{'}) k(dz^{'}|z,a)$$

es continua y acotada en $\mathbf{Z} \times A$. Por la definición de k en (2.8), podemos escribir

$$\begin{aligned}\widetilde{v}(z,a) &= \int_Y v[H(z,a,y)] R^{'}(dy|z,a) \\ &= \int_X \int_X \int_Y v[H(z,a,y)] K(dy|a,x^{'}) Q(dx^{'}|x,a) z(dx).\end{aligned}$$

Entonces por las Hipótesis 2.6(c)-(d), y la repetida aplicación de la Proposición A.9(b) del Apéndice, obtenemos que $\widetilde{v}$ es continua. Esto completa la demostración.

□

Así de acuerdo a los Lemas 1.12 y 2.7, bajo la Hipótesis 2.6, el PCM-CO $(\mathbf{Z}, A, k, k_0, \widetilde{c})$ satisface la Hipótesis 1.10 del Capítulo 1. Por lo tanto, por el Teorema 1.15 del capítulo anterior tenemos el siguiente resultado.

Teorema 2.8. *Supongamos que la Hipótesis 2.6 se satisface. Entonces:*

(a) La función de costo óptimo $V^* : \mathbf{Z} \to \mathbb{R}$ en (2.14) es la mínima solución en $\mathcal{L}_+(\mathbf{Z})$ de la ecuación de programación dinámica

$$\begin{aligned}V^*(z) &= \min_{a\in A} \left\{ \widetilde{c}(z,a) + \alpha \int_{\mathbf{Z}} V^*(z') k(dz'|z,a) \right\}, \quad z \in \mathbf{Z}, \\ &=: TV^*(z),\end{aligned}$$

donde T es el operador definido como

$$Tv(z) := \min_{a\in A} \left\{ \widetilde{c}(z,a) + \alpha \int_{\mathbf{Z}} v(z') k(dz'|z,a) \right\}.$$

(b) Existe una I-política estacionaria $f_*^\infty = \{f_*\}$ óptima, tal que

$$v^*(z) = \widetilde{c}(z, f_*) + \alpha \int_{\mathbf{Z}} v^*(z')k(dz'|z, f_*), \quad \forall z \in \mathbf{Z}. \tag{2.15}$$

Recíprocamente si f_*^∞ es óptima, entonces satisface (2.15).

Observación 2.9. Las Hipótesis 2.6(a)-(b) son estandar en el estudio de modelos de control PO. Los aspectos técnicos que se deben implementar hacen que dichas condiciones sean más restrictivas que en el caso CO, específicamente al considerar $A(x) = A\ \forall x \in X$ y función de costo continuo.

Para sistemas de ecuaciones en diferencias como (2.1), la Hipótesis 2.6(c) puede ser expresada en términos de las funciones $F(x, a, s)$ y $G(a, x, n)$. Por ejemplo, de la ecuación (2.2), podemos ver que la continuidad de $F(x, a, s)$ implica la continuidad de $Q(\cdot|x, a)$, ya que

$$\int_X h(x^{'})Q(dx^{'}|x, a) = \int_S h[F(x, a, s)]\mu(ds)$$

es continua en (x, a) para toda $h \in C(X)$ si F es continua. Similarmente usando (2.3), la continuidad de $G(a, x, n)$ implica la continuidad de $K(\cdot|a, x)$.

A diferencia de las Hipótesis 2.6(a)-(c), que establecen condiciones sobre PCM-PO original, la Hipótesis 2.6(d) es sobre el PCM-CO. Condiciones suficientes para esta hipótesis se tienen para casos muy particulares, por ejemplo, para sistemas finitos o numerables (Striebel (1975)), para sistemas con ruido aditivo (Hernández-Lerma and Romera (1999)), o para el caso específico de sistemas de inventario (Bensoussan *et. al.* (2007, 2008a, 2008b, 2010)), las cuales estudiaremos en el siguiente capítulo.

Capítulo 3

Sistemas de Inventarios Parcialmente Observables

3.1. Introducción

En muchas situaciones, los niveles de inventario en un establecimiento comercial son sólo parcialmente observados. Esto puede deberse a alguno de los siguientes factores:

- Errores de transacción. Estos pueden ser, un mal conteo del inventario o al momento de darle salida a los productos en la caja registradora, puesto que se suelen marcar unos productos por otros si dichos artículos son parecidos y del mismo precio.
- Mercancía extraviada. Frcuentemente, cuando esto pasa, la mercancía no es recuperda inmediatamente lo cual provoca que no sea considerada en el conteo del inventario.
- Merma de productos. Algunos productos, como los medicamentos o comestibles, tienen un tiempo limitado de vida y si éste llega a su fin sin que sea observado, entonces el inventario real es menor al registrado y así parcialmente observado.
- Calidad o funcionamiento de los productos. Cuando un producto de mala calidad o defectuoso es aceptado en el almacén, el nivel de inventario real es desconocido pues frecuentemente la avería del producto no es inmediatamente identificada, lo cual implica que el nivel de inventario, de los artículos viables, es parcialmente observado.
- Robo de mercancía. El más común y difícil de detectar es el robo continuo, pues no siempre es observado sin una inspección del inventario, así si ya ocurrió un robo y no fue detectado se está considerando un nivel de inventario que no es el real.

En los ultimos años, el problema de control de inventario se ha convertido en uno de los temas más importantes dentro de las investigaciones en la industria debido a las grandes inversiones que hacen las empresas en sus inventarios. En este cápitulo abordaremos dos problemas de Inventarios Parcialmente Observados a los cuales nos referiremos como *Zero-Balance Walk* y *Demanda Parcialmente Observada.*

3.2. Descripción de un Sistema de Inventarios

Un Sistema de Producción-Inventario, o Sistema de Inventario, modela la dinámica de un establecimiento comercial con el fin de optimizar sus costos bajo cierto criterio de rendimiento siendo el costo descontado el más utilizado en este tipo de sistemas. En cada periodo t, $t = 1, 2, \ldots$, antes de la apertura de la tienda a los clientes, se mide el nivel de inventario x_t (la variables de estado), y se ordena cierta cantidad a_t (la variable de control), la cual es provista inmediatamente. De esta manera, al momento de la apertura del establecimeinto el nivel de inventario es $x_t + a_t$ y la demanda que se registra durante ese período es ξ_t, el proceso de perturbaciones aleatorias, las cuales, comúnmente, se suponen v.a. i.i.d. e independientes del nivel de stock inicial x_1. Por lo tanto, el nivel de inventario al inicio del siguiente período, si se supone que la demanda no satisfecha al final de cada período se pierde, puede ser expresado como $x_{t+1} = (x_t + a_t - \xi_t)^+, \;\; t = 1, 2, \ldots$ la cual es llamada ecuación del sistema.

Por lo tanto, para los dos problemas que se estudian en este capítulo, consideremos un sistema que evoluciona de acuerdo a la ecuación en diferencias

$$x_{t+1} = (x_t + a_t - \xi_t)^+ \;\; t = 1, 2, ... \tag{3.1}$$

donde

x_t: es el nivel de inventario al inicio del periodo t;

a_t: es la cantidad a ordenar al inicio del periodo t;

$\xi_t \geq 0$: es la demanda durante el periodo t.

En ambos casos supondremos que $A(x) = A = [0, \overline{a}]$, $x \in X$, para algún $\overline{a} \in \mathbb{R}^+$.

3.3. El Modelo Zero Balance Walk

Cuando el nivel de invetario es cero o negativo, una de las siguientes situaciones pudo haber ocurrido: errores de transacción, mercancía extraviada, robos, etc. Muchas compañías ponen especial atención al momento en que los niveles de inventario llegan a cero. En estas compañías hay empleados, denominados Inventory Managers (IM), que recorren las estanterías e identifican los elementos cuyo nivel de existencia es cero, este proceso es llamado *Zero-Balance Walk* (ZBW). En esta sección se construye un modelo basado en este proceso asumiendo que los niveles de inventario son completamente observados cuando estos son cero.

Específicamente, el controlador observa el evento $[x_t = 0]$ cuando el nivel de inventario es cero, pero no observa el nivel de inventario cuando éste es positivo $[x_t > 0]$.

Para formular y analizar el modelo ZBW haremos las siguientes suposiciones:

- $\{\xi_t\}$ es una sucesión de v.a. i.i.d. con densidad f y una función de distribución F, y donde $\overline{F} = 1 - F$;
- el nivel de inventario inicial x_1 es cero, por lo tanto es observado, o tiene una densidad de probabilidad $\kappa(\cdot)$;
- el estado no observado x_t (i.e., cuando $x_t > 0$), tiene densidad condicional dada la historia observada $\kappa_t(\cdot)$;
- la función costo por etapa $c(x, a)$ es continua y no-negativa, y además la función

$$(a, \kappa) \mapsto \int_0^\infty c(x, a)\kappa(x)dx \tag{3.2}$$

  es continua en $(a, \kappa) \in A \times \mathbb{D}$, (ver Apéndice A).

Bajo estas condiciones, el proceso de observación $\{y_t\}$ está definido como

$$y_t := 1_{[x_t=0]} \quad t = 1, 2, \ldots$$

Observe que $\{y_t\}$ es una cadena de Markov a tiempo discreto con espacio de estados $\{0, 1\}$, pues el hecho de que haya o no productos en el inventario en cierto periodo sólo depende de si hubo o no en el período anterior.

Entonces las condiciones iniciales del proceso parcialmente observado son parejas $(y_1, \kappa_1) := (y, \kappa) \in \{0, 1\} \times \mathbb{D}$ pues

- Si $y_1 = 1$ entonces $x_1 = 0$.
- Si $y_1 = 0$ entonces $x_1 > 0$ y $\kappa(\cdot)$ es su densidad.

Por lo tanto, podemos definir la distribución inicial $\nu \in \mathbb{P}(X)$ del nivel de inventario x_1 como

$$\nu(B) := y\delta_0(x_1) + (1-y)\int_B \kappa(w)dw \quad B \in \mathcal{B}(X),$$

donde δ_0 es la función delta de Dirac. Observe que la distribución ν está determinada por las condiciones iniciales $(y_1, \kappa_1) = (y, \kappa)$.

Sea $\mathcal{Y}_t := \sigma(y_1, y_2, ..., y_t)$ para $t = 1, 2, ...$, la σ-álgebra generada por las observaciones. Así, $\mathcal{Y}_t$ es la historia observable para el IM hasta el período t.

Para una fácil referencia, plantearemos el problema de control para el modelo ZBW bajo el criterio de costo descontado. Para alguna política $\pi \in \Pi$ y distribución inicial $\nu \in \mathbb{P}(X)$, sean

$$J(\nu, \pi) = \sum_{t=1}^{\infty} \alpha^{t-1} E_\nu^\pi \left[c(x_t, a_t)\right]$$

el costo descontado total y

$$J^*(\nu) = \inf_{\pi \in \Pi} J(\nu, \pi) \quad \nu \in \mathbb{P}(X)$$

el costo óptimo. Entonces el Problema de Inventario Parcialmente Observado para el modelo ZBW es encontrar una política óptima π^*, es decir un política π^* que satisfaga

$$J(\nu, \pi^*) = J^*(\nu) \quad \forall \nu \in \mathbb{P}(X).$$

Ahora que hemos planteado el Problema de Inventario PO, procederemos de acuerdo al Capítulo 2. Recordemos cuál fue el esquema que se siguió:

- Primero se expuso el problema de control PO.
- Luego se transformó a uno CO mediante un proceso de filtrado $\{z_t\}$, el cual evoluciona recursivamente de acuerdo a una ecuación en diferencias, definido en (2.4).
- Después se plantea el problema de control CO.
- Finalmente se demuestra que el modelo de control CO satisface las condiciones de Capítulo 1 para asegurar la existencia de políticas óptimas y que la función de valor óptimo es la mínima solución a la EO.

3.3.1. Evolución de las densidades

Si $\kappa_t(\cdot)$ es la densidad condicional de x_t dado $\mathcal{Y}_{t-1}$ y $x_t > 0$, para cada política $\pi \in \Pi$ y distribución inicial $\nu \in \mathbb{P}(X)$, definamos el proceso de filtrado $\{z_t\}$ (ver (2.12)) como

$$\begin{aligned}
&z_1(B) := P_\nu^\pi[x_1 \in B] = \nu(B)\\
&z_t(B) := P_\nu^\pi[x_t \in B|\mathcal{Y}_{t-1}] = y_t\delta_0(x_t) + (1-y_t)\int\limits_B \kappa_t(w)dw \; t > 1,\ B \in \mathcal{B}(X).
\end{aligned}$$

De aquí, la distribución z_t queda determinada por el par (y_t, κ_t), y por lo tanto para determinar una relación recursiva para z_t es suficiente encontrar la relación entre las densidades κ_t. Como el evento $[x_t = 0]$ es observado, las densidades de probabilidad son necesarias sólo cuando $[x_t > 0]$, es decir cuando $y_t = 0$. Por lo tanto es suficiente considerar el proceso $\{\kappa_t\} \in \mathbb{D}$.

Teorema 3.1. *El proceso $\{\kappa_t\}$ evoluciona de acuerdo a la siguiente ecuación recursiva*

$$\begin{aligned}
\kappa_t(s) = 1_{[x_{t-1}=0]} &\left\{\frac{f(a_{t-1}-s)1_{[s\le a_{t-1}]}}{F(a_{t-1})}\right\}\\
+ 1_{[x_{t-1}>0]} &\left\{\frac{\int_{(s-a_{t-1})^+}^\infty f(w+a_{t-1}-s)\kappa_{t-1}(w)dw}{\int_0^\infty F(w+a_{t-1})\kappa_{t-1}(w)dw}\right\}
\end{aligned} \tag{3.3}$$

donde $\kappa_1 = \kappa \in \mathbb{D}$ está dada.

El Teorema 3.1 es consecuencia de los siguientes dos lemas. En el primero de ellos mostraremos que $E[\varphi(x_t)|\mathcal{Y}_t]$ se puede escribir en términos de $\mathcal{Y}_{t-1}$, lo cual será la clave en el Lema 3.3, para poder expresar esta esperanza condicional en términos de la densidad y función de distribución de la demanda.

Para facilitar la notación, para $\pi \in \Pi$ y $\nu \in \mathbb{P}(X)$, denotaremos por E al operador esperanza E_ν^π y P a la medida de probabilidad P_ν^π inducida por π y ν.

Lema 3.2. *Para cualquier función real y acotada φ*

$$\begin{aligned}
E[\varphi(x_t)|\mathcal{Y}_t] &= 1_{[x_t=0]}\varphi(0) + 1_{[x_t>0]}\frac{E\left[\varphi(x_t)1_{[x_t>0]}|\mathcal{Y}_{t-1}\right]}{P[x_t>0|\mathcal{Y}_{t-1}]}\\
&= 1_{[x_t=0]}\varphi(0) + 1_{[x_t>0]}E[\varphi(x_t)|\mathcal{Y}_{t-1}, x_t>0]\,.
\end{aligned} \tag{3.4}$$

Lema 3.3.

$$E\left[\varphi(x_t)|\mathcal{Y}_t\right]1_{[x_t>0]} = 1_{[x_{t-1}=0]}\left\{\frac{\int_0^\infty \varphi(s)f(a_{t-1}-s)1_{[s\leq a_{t-1}]}ds}{F(a_{t-1})}\right\} + 1_{[x_{t-1}>0]}\left\{\frac{\int_0^\infty \varphi(s)\int_{(s-a_{t-1})^+}^\infty f(w+a_{t-1}-s)\kappa_{t-1}(w)dwds}{\int_0^\infty F(w+a_{t-1})\kappa_{t-1}(w)dw}\right\} \tag{3.5}$$

Para no desviarnos de nuestro objetivo, las demostraciones de los Lemas 3.2 y 3.3 se presentan al final de este capítulo.

Demostración (del Teorema 3.1). Primero observemos que la esperanza condicional en el Lema 3.2, puede ser expresada en términos de la densidad κ_t, de la siguiente manera:

$$E\left[\varphi(x_t)|\mathcal{Y}_t\right] = 1_{[x_t=0]}\varphi(0) + 1_{[x_t>0]}\int_0^\infty \varphi(s)\kappa_t(s)ds. \tag{3.6}$$

Entonces, comparando las ecuaciones (3.6) y (3.5) obtenemos que

$$\kappa_t(s) = 1_{[x_{t-1}=0]}\left\{\frac{f(a_{t-1}-s)1_{[s\leq a_{t-1}]}}{F(a_{t-1})}\right\} + 1_{[x_{t-1}>0]}\left\{\frac{\int_{(s-a_{t-1})^+}^\infty f(w+a_{t-1}-s)\kappa_{t-1}(w)dw}{\int_0^\infty F(w+a_{t-1})\kappa_{t-1}(w)dw}\right\}.$$

El cual es el resultado deseado.

□

Podemos obtener una expresión más corta para $\kappa_t(s)$ de la siguiente manera. Sean p y Φ, funciones reales, y $a\in A$. Definamos

$$\langle p,\Phi\rangle = \int_0^\infty p(s)\Phi(s)ds$$

y

$$\rho(a,p)(s) = \int_{(s-a)^+}^\infty f(w+a-s)p(w)dw.$$

Observemos que

$$\rho(a,\delta)(s) = \int\limits_{(s-a)^+}^{\infty} f(w+a-s)\delta(w)dw = f(a-s)1_{[s\leq a]}.$$

Entonces, $\rho(0,\delta)(s) = 0$, donde δ es la delta de Dirac.

Definamos el operador no lineal θ como

$$\theta(a,p)(s) = \frac{\rho(a,p)(s)}{\langle \rho(a,p),1\rangle}.$$

Entonces

$$\langle \theta(a,p),1\rangle = \int\limits_{0}^{\infty} \frac{\rho(a,p)}{\langle \rho(a,p),1\rangle} = 1,$$

y además $\theta(a,\lambda p) = \theta(a,p)$ para cada $\lambda \in \mathbb{R}$.

Con esta nueva notación, podemos escribir a (3.3) de la siguiente manera

$$\begin{aligned} \kappa_1 &= \kappa \\ \kappa_t(s) &= y_{t-1}\theta(a_{t-1},\delta)(s) + (1-y_{t-1})\theta(a_{t-1},\kappa_{t-1})(s) \quad t>1 \end{aligned}$$

En efecto:

$$\begin{aligned} & y_{t-1}\theta(a_{t-1},\delta)(s) + (1-y_{t-1})\theta(a_{t-1},\kappa_{t-1})(s) \\ &= y_{t-1}\frac{f(a_{t-1}-s)1_{[s\leq a_{t-1}]}}{\int_0^\infty f(a_{t-1}-s)1_{[s\leq a_{t-1}]}ds} + (1-y_{t-1})\frac{\int_{(s-a_{t-1})^+}^\infty f(w+a_{t-1}-s)\kappa_{t-1}(w)dw}{\int_0^\infty \int_{(s-a_{t-1})^+}^\infty f(w+a_{t-1}-s)\kappa_{t-1}(w)dwds} \\ &= y_{t-1}\frac{f(a_{t-1}-s)1_{[s\leq a_{t-1}]}}{\int_0^{a_{t-1}} f(a_{t-1}-s)ds} + (1-y_{t-1})\frac{\int_{(s-a_{t-1})^+}^\infty f(w+a_{t-1}-s)\kappa_{t-1}(w)dw}{\int_0^\infty \left[\int_0^{w+a_{t-1}} f(w+a_{t-1}-s)ds\right]\kappa_{t-1}(w)dw} \\ &= 1_{[x_{t-1}=0]}\frac{f(a_{t-1}-s)1_{[s\leq a_{t-1}]}}{F(a_{t-1})} + 1_{[x_{t-1}>0]}\frac{\int_{(s-a_{t-1})^+}^\infty f(w+a_{t-1}-s)\kappa_{t-1}(w)dw}{\int_0^\infty F(w+a_{t-1})\kappa_{t-1}(w)dw}, \end{aligned}$$

donde la primera igualdad se obtiene de aplicar la definición de θ; la segunda se obtiene de aplicar el Teorema de Fubini en el segundo sumando; y la tercera se sigue aplicando los cambios de varibales $z' := a_{t-1} - s$ y $z'' := w + a_{t-1} - s$ en los denominadores del primer y segundo sumandos, respectivamente. Así, recordando que $y_{t-1} = 1_{[x_{t-1}=0]}$, obtenemos finalmente (3.3).

3.3.2. Problema de Control CO

De acuerdo al esquema que estamos siguiendo, el siguiente paso es plantear el problema de control completamente observable, donde ahora el proceso de estados es $\{z_t\}$, el cual es determinado por (y_t, κ_t).

Definamos la función de costo-por-etapa

$$\begin{aligned}\widetilde{c}(z_t, a_t) = \widetilde{c}(y_t, \kappa_t, a_t) &= \int c(x, a_t) z_t(dx) \\ &= y_t c(0, a_t) + (1 - y_t) \int c(x, a_t)\kappa_t(x)dx \\ &= y_t c(0, a_t) + (1 - y_t) \langle c(\cdot, a_t), \kappa_t(\cdot)\rangle\end{aligned}$$

donde

$$z_t(B) = y_t\delta(x_t) + (1 - y_t) \int_B \kappa_t(s)ds \;\; B \in \mathcal{B}(X).$$

Como z_t queda totalmente determinada por y_t y κ_t la nueva función de costo-por-etapa es

$$\widetilde{c}(y_t, \kappa_t, a_t) = y_t c(0, a_t) + (1 - y_t) \langle c(\cdot, a_t), \kappa_t(\cdot)\rangle \, . \tag{3.7}$$

Si $(y, \kappa) \in \{0, 1\} \times \mathbb{D}$ es la condición inical, para cada $\pi \in \Pi$

$$V(y, \kappa, \pi) = \sum_{t=1}^{\infty} \alpha^{t-1} E^{\pi}_{y,\kappa}[\widetilde{c}(y_t, \kappa_t, a_t)]$$

es el costo descontado total y

$$V^*(y, \kappa) = \inf_{\pi\in\Pi} V(y, \kappa, \pi) \tag{3.8}$$

la función de valor óptimo. Por lo tanto el problema de control CO es encontrar una política π^* tal que

$$V^*(y, \kappa) = V(y, \kappa, \pi^*) \;\; \forall(y, \kappa) \in \{0, 1\} \times \mathbb{D}.$$

Como ya se mencionó en el Capítulo 2, este problema es equivalente al Problema PO en el sentido que una política óptima para el modelo CO es óptima para el PO.

3.3.3. Ecuación de Optimalidad

A continuación presentamos la ecuación de optimalidad para dar explícitamente la forma del operador T_a.

Similar a (1.7) una función $\mathcal{U} : \{0,1\} \times \mathbb{D} \to \mathbb{R}$ satisface la EO si

$$\mathcal{U}(y,\kappa) = \min_{a\in A} \left\{ yc(0,a) + (1-y) \langle c(\cdot,a), \kappa(\cdot)\rangle + \alpha E^{\pi}_{y,\kappa}[\mathcal{U}(y_2,\kappa_2)] \right\}, \tag{3.9}$$

donde el tercer sumando, dada la manera en que evoluciona el proceso $\{\kappa_t\}$, se puede expresar de la siguiente manera

$$\begin{aligned} E^{\pi}_{y,\kappa}[\mathcal{U}(y_2,\kappa_2)] &= E^{\pi}_{y,\kappa}[\mathcal{U}(y_2, y\theta(a,\delta)(x) + (1-y)\theta(a,\kappa)(x))] \\ &= \mathcal{U}(1, y\theta(a,\delta)(x) + (1-y)\theta(a,\kappa)(x)) P^{\pi}_{y,\kappa}[x_2 = 0] \\ &+ \mathcal{U}(0, y\theta(a,\delta)(x) + (1-y)\theta(a,\kappa)(x)) P^{\pi}_{y,\kappa}[x_2 > 0]. \end{aligned} \tag{3.10}$$

Ahora usando propiedades de la esperanza condicional se obtiene

$$\begin{aligned} P^{\pi}_{y,\kappa}[x_2 = 0] = P^{\pi}_{y,\kappa}[\xi \geq x + a] &= 1 - P^{\pi}_{y,\kappa}[\xi < x + a] \\ &= 1 - E^{\pi}_{y,\kappa}[E^{\pi}_{y,\kappa}[1_{[\xi < x+a]} | x]] \\ &= 1 - \int P^{\pi}_{y,\kappa}[\xi < w + a] \kappa(w) dw \\ &= 1 - \int F(w+a)\kappa(w) dw \\ &= \int \overline{F}(w+a)\kappa(w) dw \end{aligned} \tag{3.11}$$

y

$$P^{\pi}_{y,\kappa}[x_2 > 0] = P^{\pi}_{y,\kappa}[\xi < x + a] = \int F(w+a)\kappa(w) dw. \tag{3.12}$$

Así, sustituyendo (3.12) y (3.11) en (3.10) y después en (3.9) obtenemos la Ecuación de Optimalidad

$$\begin{aligned} \mathcal{U}(y,\kappa) = \min_{a\in A} \{ & yc(0,a) + (1-y) \langle c(\cdot,a), \kappa(\cdot)\rangle \\ &+ \alpha \mathcal{U}(1, y\theta(a,\delta)(x) + (1-y)\theta(a,\kappa)(x)) \int \overline{F}(w+a)\kappa(w) dw \\ &+ \alpha \mathcal{U}(0, y\theta(a,\delta)(x) + (1-y)\theta(a,\kappa)(x)) \int F(w+a)\kappa(w) dw \}. \end{aligned} \tag{3.13}$$

De aquí, definimos los operadores $T_a\mathcal{U}$ y $T\mathcal{U}$ como

$$\begin{aligned} T_a\mathcal{U}(y,\kappa) := {} & yc(0,a) + (1-y) \langle c(\cdot,a), \kappa(\cdot)\rangle \\ &+ \alpha \mathcal{U}(1, y\theta(a,\delta)(x) + (1-y)\theta(a,\kappa)(x)) \int \overline{F}(w+a)\kappa(w) dw \\ &+ \alpha \mathcal{U}(0, y\theta(a,\delta)(x) + (1-y)\theta(a,\kappa)(x)) \int F(w+a)\kappa(w) dw \end{aligned} \tag{3.14}$$

y

$$TU(y,\kappa) := \min_{a\in A} T_a U(y,\kappa).$$

Finalmente, para demostrar que la función de valor óptimo (3.8) es la mínima solución en $\mathcal{L}(\{0,1\}\times\mathbb{D})$ de la Ecuación de Optimalidad (3.13) y garantizar la existencia de una política óptima para este problema, se requiere verificar las condiciones del Teorema 1.15, es decir

a) $A(x)$ es compacto $\forall x \in X$ y el mapeo $x \mapsto A(x)$ es u.s.c.

b) La función costo-por-etapa es l.s.c. y no negativa en $\{0,1\}\times\mathbb{D}\times A$

c) $T_a U \in \mathcal{L}_+(\{0,1\}\times\mathbb{D}\times A)$ para cada $U \in \mathcal{L}_+(\{0,1\}\times\mathbb{D})$.

La condición (a) se satisface pues $A(x) = A$ para toda $x \in X$, y $A = [0,\overline{a}]$. Para verificar b), consideremos la función de costo-por-etapa

$$\widetilde{c}(y_t,\kappa_t,a_t) = y_t c(0,a_t) + (1-y_t)\left\langle c(\cdot,a_t),\kappa_t(\cdot)\right\rangle,$$

donde

$$\langle c(\cdot,a_t),\kappa_t(\cdot)\rangle = \int c(x,a_t)\kappa_t(x)dx,$$

entonces por la suposición hecha en (3.2), y dado que la función c es continua y no-negativa, la función $\widetilde{c}$ es continua y no negativa.

Por último, la demostración de (c) la obtendremos del Lema 3.4 y Observación 3.5, dados a continuación, acerca de las propiedades de los operadores θ y ρ.

Lema 3.4. *La función*

$$(a,p) \mapsto \int f(w+a-x)p(w)dw \quad (a,p)\in A\times\mathcal{H}^+$$

es continua.

Demostración. Sea $\{(a_t,p_t)\}$ una sucesión en $A\times\mathcal{H}^+$ que converge a $(a,p)\in A\times\mathcal{H}^+$. Sumando y restando $\int f(w+a_t-x)p(w)dw$ se tiene

$$\begin{aligned}
&\left|\int f(w+a_t-x)p_t(w)dw - \int f(w+a-x)p(w)dw\right| \leq \\
&\left|\int f(w+a_t-x)(p_t(w)-p(w))dw\right| + \left|\int (f(w+a_t-x)-f(w+a-x))p(w)dw\right| \\
&\leq \int f(w+a_t-x)\left|p_t(w)-p(w)\right|dw + \int \left|f(w+a_t-x)-f(w+a-x)\right|p(w)dw.
\end{aligned}$$

Como f está acotada, digamos por $M \in \mathbb{R}$, haciendo $t \to \infty$ y por el Teorema de Convergencia Dominada obtenemos

$$\begin{aligned}
&\lim_{t\to\infty} \left| \int f(w + a_t - x)p_t(w)dw - \int f(w + a - x)p(w)dw \right| \leq \\
\lim_{t\to\infty} M \int |p_t(w) - p(w)|\, dw &+ \int \lim_{t\to\infty} |f(w + a_t - x) - f(w + a - x)|\, p(w)dw \\
&= 0
\end{aligned}$$

donde la última igualdad se tiene por la definición de la convergencia en $\mathcal{H}^+$ (ver Apéndice) y la continuidad de f. Por lo tanto $(a, p) \mapsto \int f(w + a - x)p(w)dw$ es continua en $(a, p) \in A \times \mathcal{H}^+$.

$\square$

Observación 3.5. Siguiendo argumentos similares se puede demostrar la continuidad de las siguientes funciones

$$\begin{aligned}
(a, p) &\mapsto \int F(w + a)p(w)dw \\
(a, p) &\mapsto \int \overline{F}(w + a)p(w)dw \\
(a, p) &\mapsto \rho(a, p)(x) = \int_{(x-a)^+}^{\infty} f(w + a - x)p(w)dw.
\end{aligned}$$

Además la función
$$(a, p) \mapsto \int \int f(w + a - x)p(w)dwdx$$
es continua en $A \times \mathcal{H}^+$. De esta manera, para cada $x \geq 0$, la función
$$\theta(a, p)(x) = \frac{\int_{(x-a)^+}^{\infty} f(w + a - x)p(w)dw}{\int_0^{\infty} \int_{(x-a)^+}^{\infty} f(w + a - x)p(w)dwdx}$$
es continua en (a, p).

Por lo tanto, de acuerdo al Lema 3.4, la Observación 3.5 y (3.14), tenemos que

$$\begin{aligned}
T_a U(y, \kappa) &= yc(0, a) + (1 - y) \langle c(\cdot, a), \kappa(\cdot) \rangle \\
&\quad + \alpha U(1, y\theta(a, \delta)(x) + (1 - y)\theta(a, \kappa)(x)) \int \overline{F}(w + a)\kappa(w)dw \\
&\quad + \alpha U(0, y\theta(a, \delta)(x) + (1 - y)\theta(a, \kappa)(x)) \int F(w + a)\kappa(w)dw
\end{aligned}$$

pertenece a $\mathcal{L}_+(\{0, 1\} \times \mathbb{D} \times A)$ para cada $U \in \mathcal{L}_+(\{0, 1\} \times \mathbb{D})$.

3.4. Demanda Parcialmente Observada

Otro modelo de inventario PO que en la práctica resulta importante es cuando las observaciones parciales son hechas sobre la demanda y no directamente en el nivel de inventario.

La situación particular que analizaremos es cuando la demanda no satisfecha no es observada por el controlador, y la demanda satisfecha es observada a través de las ventas. Esto se presenta cada vez con mayor frecuencia debido a la competitividad del mercado de hoy en día. Por ejemplo, si un cliente acude a una tienda departamental o supermercado en busca de un producto específico y no lo encuentra acude a algún otro establecimiento a adquirirlo, lo cual representa una venta perdida para la primera tienda, la cual fue no observada.

Para el análisis de este modelo haremos las siguientes suposiciones:

- la demanda es completamente observada cuando es menor o igual al nivel de inventario disponible, y si es mayor puede ser representada por una densidad condicional dada las observaciones pasadas, $\kappa_n(\cdot)$.

- la demanda $\xi_n \in [0, \infty)$, durante el periodo n, es un proceso de Markov con probabilidades de transición

$$P[\xi_{n+1} \in B | \xi_n = s] = \int\limits_B p(x|s)dx$$

  donde $p(\cdot|\cdot)$ es una densidad condicional, ξ_1 dada y $E\xi_1 < \infty$.

- la función de costo-por-etapa es de la forma

$$c(x, \xi, a) = \begin{cases} pa + h(x + a - \xi) & \text{si } \xi \leq x + a \\ pa + b(\xi - x - a) & \text{si } \xi > x + a \end{cases}, \tag{3.15}$$

  donde

  p: costo por artículo al orednar,

  h: costo de almacenamiento,

  b: costo por escacés de artículo,

  las cuales satisfacen $0 < p < b$ y $h > 0$.

Por lo anterior, el proceso de observación y_n está definido como

$$y_n := \min\{\xi_n, x_n + a_n\}$$

el cual representa la venta durante el periodo n, para $n = 1, 2, ...$

Es decir, si $\xi_n \leq x_n + a_n$, entonces la demanda fue satisfecha por completo y por lo tanto observada. Por otra parte si $\xi_n > x_n + a_n$, el inventario no fue suficiente para satisfacer toda la demanda durante ese período y por consiguiente parcialmente observada, en este caso, la venta fue $x_n + a_n$ y $\xi_n - (x_n + a_n)$ la demanda no satisfecha.

Definamos $\mathcal{Y}_n := \sigma(y_1, y_2, ..., y_n)$, la σ-álgebra generada por las observaciones hasta el periodo n.

Para una política $\pi \in \Pi$ y un nivel de invenatrio inicial $x_1 = x \in X$, sean

$$J(x, \pi) = E_x^\pi \sum_{t=1}^{\infty} \alpha^{t-1} c(x_t, \xi_t, a_t)$$

el costo descontado total esperado, con $\alpha \in (0, 1)$ el factor de descuento dado, y

$$J^*(x) = \inf_{\pi \in \Pi} J(x, \pi)$$

el costo óptimo. Entonces el Problema de Inventario PO es encontrar una política $\pi^* \in \Pi$ tal que

$$J(x, \pi^*) = J^*(x) \;\; \forall x \in X.$$

3.4.1. Evolución de las Densidades

Al igual que en el ejemplo anterior, básicamente el problema es encontrar una relación recursiva para el proceso $\{\kappa_n\}$, donde κ_n es la densidad condicional de la demanda ξ_n dada la historia observada, es decir

$$P[\xi_n \in B | \mathcal{Y}_{n-1}] = \int_B \kappa(s) ds \;\; B \in \mathcal{B}([0, \infty)),$$

A partir de aquí, plantearemos el problema de control completamente observable.

Teorema 3.6. *El proceso $\{\kappa_n\}$ evoluciona de acuerdo a la siguiente ecuación recursiva*

$$\kappa_{n+1}(x) = 1_{[y_n = x_n + a_n]} \frac{\int_{x_n+a_n}^{\infty} \kappa_n(s) p(x|s) ds}{\int_{x_n+a_n}^{\infty} \kappa_n(s) ds} + 1_{[y_n < x_n + a_n]} p(x|y_n) \tag{3.16}$$

donde $\kappa_1 \in \mathbb{D}$ es conocido.

Nótese que el primero y segundo terminos de la derecha corresponden, respectivamente, a los eventos $[\xi_n \geq x_n + a_n]$ y $[\xi_n < x_n + a_n]$. En el primer evento, la demanda es mayor o igual al nivel de invenatio después de ordenar y por tanto no observada. En el segundo evento la demanda es observada con el valor $\xi_n = y_n$ y por tanto $\kappa_{n+1}(x) = p(x|y_n)$.

La demostración del Teorema 3.6 se deriva de los siguientes dos Lemas cuyas demostraciones serán presentadas al final de este capítulo.

Sean $\pi \in \Pi$, $x \in X$ y $\kappa \in \mathbb{D}$ fijos. Para simplificar notación escribiremos E en lugar de $E^{\pi}_{x,\kappa}$.

Lema 3.7. *Para toda función integrable ψ y $n \geq 1$*

$$E[\psi(\xi_n)|\mathcal{Y}_n] = 1_{[y_n = x_n + a_n]} \frac{\int_{x_n+a_n}^{\infty} \psi(s)\kappa_n(s)ds}{\int_{x_n+a_n}^{\infty} \kappa_n(s)ds} + 1_{[y_n < x_n + a_n]}\psi(y_n). \tag{3.17}$$

Lema 3.8. *Sean η_n una v.a. $\mathcal{Y}_n$-medible y G una función integrable arbitraria. Entonces*

$$E[G(\xi_{n+1})\eta_n 1_{[\xi_n \geq x_n + a_n]}] = E\left[\eta_n 1_{[y_n = x_n + a_n]} \frac{\int_{x_n+a_n}^{\infty} \kappa_n(x)[\int G(s)p(s|x)ds]dx}{\int_{x_n+a_n}^{\infty} \kappa_n(x)dx}\right] \tag{3.18}$$

y

$$E[G(\xi_{n+1})\eta_n 1_{[\xi_n < x_n + a_n]}] = E\left[\eta_n 1_{[y_n < x_n + a_n]} \int G(s)p(s|y_n)ds\right] \tag{3.19}$$

Ahora, con los Lemas 3.7 y 3.8 establecidos podemos demostrar el Teorema 3.6.

Demostración (del Teorema 3.6). Para cualquier v.a. η_n $\mathcal{Y}_n$-medible y función integrable G, se tiene

$$\begin{aligned} E[G(\xi_{n+1})\eta_n] &= E[\eta_n G(\xi_{n+1})(1_{[\xi_n \geq x_n + a_n]} + 1_{[\xi_n < x_n + a_n]})] \\ &= E[\eta_n G(\xi_{n+1})1_{[\xi_n \geq x_n + a_n]}] + E[\eta_n G(\xi_{n+1})1_{[\xi_n < x_n + a_n]}]. \end{aligned}$$

De aquí y las ecuaciones (3.18) y (3.19) obtenemos

$$\begin{aligned} &E[G(\xi_{n+1})\eta_n] = \\ &\qquad E\{\eta_n[1_{[y_n = x_n + a_n]} \frac{\int_{x_n+a_n}^{\infty} \kappa_n(x)[\int G(s)p(s|x)ds]dx}{\int_{x_n+a_n}^{\infty} \kappa_n(x)dx} \\ &\qquad\qquad + 1_{[y_n < x_n + a_n]} \int G(s)p(s|y_n)ds]\}. \end{aligned}$$

Como η_n es una variable aleatoria $\mathcal{Y}_n$-medible arbitraria y lo que está entre corchetes también es $\mathcal{Y}_n$-medible, obtenemos, por la definición de esperanza condicional

$$\begin{aligned}
&\int G(x)\kappa_{n+1}(x)dx = E[G(\xi_{n+1})|\mathcal{Y}_n] = \\
&1_{[y_n=x_n+a_n]}\frac{\int_{x_n+a_n}^{\infty}\kappa_n(x)[\int G(s)p(s|x)ds]dx}{\int_{x_n+a_n}^{\infty}\kappa_n(x)dx} + 1_{[y_n<x_n+a_n]}\int G(s)p(s|y_n)ds \\
&= 1_{[y_n=x_n+a_n]}\frac{\int G(s)[\int_{x_n+a_n}^{\infty}\kappa_n(x)p(s|x)ds]dx}{\int_{x_n+a_n}^{\infty}\kappa_n(x)dx} + 1_{[y_n<x_n+a_n]}\int G(s)p(s|y_n)ds
\end{aligned}$$

donde la última igualdad se da por el Teorema de Fubini.

Por lo tanto, como $G(x)$ es arbitraria

$$\kappa_{n+1}(x) = 1_{[y_n=x_n+a_n]}\frac{\int_{x_n+a_n}^{\infty}\kappa_n(s)p(x|s)ds}{\int_{x_n+a_n}^{\infty}\kappa_n(s)ds} + 1_{[y_n<x_n+a_n]}p(x|y_n).$$

□

3.4.2. Problema de Control CO

Ahora introduciremos el modelo de control completamente observable. Definamos para cada $n \in \mathbb{N}$, la función de costo-por-etapa

$$\widetilde{c}(x_n,\kappa_n,a_n) = \int c(x_n,s,a_n)\kappa_n(s)ds \quad x_n \in X,\ a_n \in A. \tag{3.20}$$

Dado $x_1 = x$ y $\kappa_1 = \kappa$, para alguna política $\pi \in \Pi$ sean

$$V(x,\kappa,\pi) = \sum_{n=1}^{\infty}\alpha^{n-1}E\widetilde{c}(x_n,\kappa_n,a_n),$$

el costo descontado total esperado, y

$$V^*(x,\kappa) = \min_{\pi\in\Pi} V(x,\kappa,\pi) \tag{3.21}$$

la función de valor óptimo.

El PCO correspondiente es encontrar una política $\pi^* \in \Pi$ tal que para cada nivel de inventario inicial $x \in X$ y densidad inicial $\kappa \in \mathbb{D}$ de la demanda,

$$V^*(x,\kappa) = V(x,\kappa,\pi^*).$$

3.4.3. Ecuación de Optimalidad

Una fución $\mathcal{U} : X \times \mathbb{D} \to \mathbb{R}$ satisface la EO si

$$\mathcal{U}(x,\kappa) = \min_{a\in A} \left\{\widetilde{c}(x,\kappa,a) + \alpha E[\mathcal{U}(x_2,\kappa_2)]\right\}.$$

Entonces, por (3.1) y (3.20), podemos expresar la ecuación de programación dinámica como

$$\begin{aligned}\mathcal{U}(x,\kappa) &= \min_{a\in A}\left\{\int c(x,s,a)\kappa(s)ds + \alpha E[\mathcal{U}(x_2,\kappa_2)]\right\}\\ &= \min_{a\in A}\left\{\int c(x,s,a)\kappa(s)ds + \alpha \int \mathcal{U}[(x+a-s)^+,\kappa_2]\kappa(s)ds\right\}\\ &= \min_{a\in A}\left\{\int c(x,s,a)\kappa(s)ds + \alpha\mathcal{U}(0,\kappa_2)\int_{x+a}^{\infty}\kappa(s)ds + \alpha\int_0^{x+a}\mathcal{U}[x+a-s,\kappa_2]\kappa(s)ds\right\}.\end{aligned}$$

De lo anterior y la ecuación (3.16) obtenemos la Ecuación de Optimalidad

$$\begin{aligned}\mathcal{U}(x,\kappa) = \min_{a\in A}\{&\int c(x,s,a)\kappa(s)ds + \alpha\mathcal{U}(0,\frac{\int_{x+a}^{\infty}p(\cdot|s)\kappa(s)ds}{\int_{x+a}^{\infty}\kappa(s)ds})\int_{x+a}^{\infty}\kappa(s)ds\\ &+\alpha\int_0^{x+a}\mathcal{U}[x+a-s,p(\cdot|s)]\kappa(s)ds\}.\end{aligned} \tag{3.22}$$

Para $\mathcal{U} : X \times \mathbb{D} \to \mathbb{R}$ los operadores $T_a\mathcal{U}$ y $T\mathcal{U}$ son definidos como

$$\begin{aligned}T_a\mathcal{U}(x,\kappa) = &\int c(x,s,a)\kappa(s)ds + \alpha\mathcal{U}(0,\frac{\int_{x+a}^{\infty}p(\cdot|s)\kappa(s)ds}{\int_{x+a}^{\infty}\kappa(s)ds})\int_{x+a}^{\infty}\kappa(s)ds\\ &+\alpha\int_0^{x+a}\mathcal{U}[x+a-s,p(\cdot|s)]\kappa(s)ds.\end{aligned} \tag{3.23}$$

y

$$T\mathcal{U}(x,\kappa) = \min_{a\in A} T_a\mathcal{U}(x,\kappa).$$

Ahora como en el caso anterior, para asegurar que la función de valor óptimo (3.21) es la mínima solución de la EO (3.22) y asegurar la existencia de una política óptima se requiere verificar

a) $A(x)$ es compacto $\forall x \in X$ y $x \mapsto A(x)$ es semicontinuo superiormente.

b) La función costo-por-etapa $\widetilde{c}$ es l.s.c. y no negativa en $X \times \mathbb{D} \times A$.

c) El operador $T_a U \in \mathcal{L}_+(X \times \mathbb{D} \times A)$ para cada $U \in \mathcal{L}_+(X \times \mathbb{D})$.

La condición (a) se satisface pues $A(x) = A = [0, \overline{a}]$ para toda $x \in X$. De (3.15) se sigue que la función de costo $\widetilde{c}$ definida en (3.20) es continua y no-negativa, lo cual implica (b). Para mostrar (c) veamos los siguientes lemas.

Lema 3.9. *Para* $(x, \rho, a) \in X \times \mathcal{H}^+ \times A$ *la función*

$$(x, \rho, a) \mapsto \int \phi(x + a - s, p(\cdot|s))\rho(s)ds$$

es continua para cada función continua y acotada ϕ *en* $X \times \mathcal{H}^+$.

Demostración. Sea ϕ una función continua y acotada por alguna constante M y $\{(x_n, \rho_n, a_n)\}$ una sucesión en $X \times \mathcal{H}^+ \times A$ que converge a $(x, \rho, a) \in X \times \mathcal{H}^+ \times A$. Entonces sumando y restando el término $\int \phi(x_n + a_n - s, p(\cdot|s))\rho(s)ds$, tenemos

$$\begin{aligned} \left| \int \phi(x_n + a_n - s, p(\cdot|s))\rho_n(s)ds - \int \phi(x + a - s, p(\cdot|s))\rho(s)ds \right| \\ \leq \int |\phi(x_n + a_n - s, p(\cdot|s))[\rho_n(s) - \rho(s)]|ds \\ + \int |\phi(x_n + a_n - s, p(\cdot|s)) - \phi(x + a - s, p(\cdot|s))|\rho(s)ds. \end{aligned} \tag{3.24}$$

Ahora, como ϕ es acotada y $\rho_n \to \rho$ en $\mathcal{H}^+$

$$\int |\phi(x_n + a_n - s, p(\cdot|s))[\rho_n(s) - \rho(s)]|ds \to 0 \text{ cuando } n \to \infty. \tag{3.25}$$

Además, dado que ϕ es continua en $X \times \mathcal{H}^+$, por el Teorema de Convergencia Dominada, se tiene

$$\int |\phi(x_n + a_n - s, p(\cdot|s)) - \phi(x + a - s, p(\cdot|s))|\rho(s)ds \to 0 \text{ cuando } n \to \infty. \tag{3.26}$$

Así, de (3.24), (3.25) y (3.26), haciendo $n \to \infty$, se obtiene

$$\left| \int \phi(x_n + a_n - s, p(\cdot|s))\rho_n(s)ds - \int \phi(x + a - s, p(\cdot|s))\rho(s)ds \right| \to 0.$$

Por lo tanto

$$(x, \rho, a) \mapsto \int \phi(x + a - s, p(\cdot|s))\rho(s)ds$$

es continua para cada función continua y acotada ϕ en $X \times \mathcal{H}^+$. □

Lema 3.10. *Para cada función $\phi \in \mathcal{L}_+(X \times \mathcal{H}^+)$, la función*

$$(x, \rho, a) \mapsto \int \phi(x + a - s, p(\cdot|s))\rho(s)ds \tag{3.27}$$

pertenece a $\mathcal{L}_+(X \times \mathcal{H}^+)$.

Demostración. Sea $\{(x_n, \rho_n, a_n)\}$ una sucesión en $X \times \mathcal{H}^+ \times A$ que converge a $(x, \rho, a) \in X \times \mathcal{H}^+ \times A$, y sea ϕ una función en $\mathcal{L}_+(X \times \mathcal{H}^+)$, entonces, por la Proposición A.3(b) del Apéndice, existe una sucesión $\{\phi_k\}$ de funciones continuas, acotadas y no-negativas en $X \times \mathcal{H}^+$ tal que $\phi_k \uparrow \phi$. Así del Lema 3.9, para cada $k \in \mathbb{N}$,

$$\liminf_{n\to\infty} \int \phi_k[x_n + a_n - s, p(\cdot|s)]\rho_n(s)ds = \int \phi_k[x + a - s, p(\cdot|s)]\rho(s)ds.$$

Además, como $\phi \geq \phi_k$ para toda $k \in \mathbb{N}$

$$\begin{aligned}\liminf_{n\to\infty} \int \phi[x_n + a_n - s, p(\cdot|s)]\rho_n(s)ds &\geq \liminf_{n\to} \int \phi_k[x_n + a_n - s, p(\cdot|s)]\rho_n(s)ds \\ &= \int \phi_k[x + a - s, p(\cdot|s)]\rho(s)ds.\end{aligned}$$

Haciendo $k \to \infty$ y usando el Lema de Fatou, obtenemos

$$\liminf_{n\to\infty} \int \phi[x_n + a_n - s, p(\cdot|s)]\rho_n(s)ds \geq \int \phi[x + a - s, p(\cdot|s)]\rho(s)ds.$$

De lo anterior y la Proposición A.3(a) del Apéndice, se obtiene que la función (3.27) pertenece a $\mathcal{L}_+(X \times \mathcal{H}^+)$. □

Observación 3.11. Similarmente se muestra que las funciones

$$(x, \rho, a) \mapsto \int_0^{x+a} \phi[x + a - s, p(\cdot|s)]\rho(s)ds,$$

$$(x, \rho, a) \mapsto \int_{x+a}^{\infty} p(\cdot|s)\rho(s)ds$$

y por tanto

$$(x, \rho, a) \mapsto \frac{\int_{x+a}^{\infty} p(\cdot|s)\rho(s)ds}{\int_{x+a}^{\infty} \rho(s)ds}$$

son l.s.c y no-negativas, donde ϕ es cualquier función l.s.c. y no-negativa.

Finalmente, del Lema 3.10, la Observación 3.11 y dado que $\widetilde{c}$ es continua obtenemos que el operador $T_a\mathcal{U}(x,\kappa)$ en (3.23) es l.s.c. y no-negativo para cada $\mathcal{U} \in \mathcal{L}_+(X \times \mathbb{D})$, lo cual demuestra (c).

Concluimos este capítulo presentando las demostraciones de los lemas en los cuales se basan los resultados obtenidos.

3.5. Demostraciones

3.5.1. Demostración del Lema 3.2

Como $\kappa_t(\cdot)$ es la densidad condicional de x_t dado $\mathcal{Y}_{t-1}$ y $x_t > 0$, por definición

$$\int_0^x \kappa_t(w)dw = P\left[x_t \leq x | x_t > 0, \mathcal{Y}_{t-1}\right].$$

Observemos que

$$\begin{aligned} E\left[\varphi(x_t)|\mathcal{Y}_t\right] &= E\left[\varphi(x_t)(1_{[x_t=0]} + 1_{[x_t>0]})|\mathcal{Y}_t\right] \\ &= E\left[\varphi(x_t)1_{[x_t=0]}|\mathcal{Y}_t\right] + E\left[\varphi(x_t)1_{[x_t>0]}|\mathcal{Y}_t\right] \\ &= 1_{[x_t=0]}\varphi(0) + E\left[\varphi(x_t)1_{[x_t>0]}|\mathcal{Y}_t\right]. \end{aligned} \tag{3.28}$$

Por otro lado, por definición de esperanza condicional, existe una función medible Ψ tal que $E\left[\varphi(x_t)|\mathcal{Y}_t\right] = \Psi(y_1, y_2, ...y_{t-1}, y_t)$. Entonces, el último término en (3.28) lo podemos expresar como

$$\begin{aligned} E\left[\varphi(x_t)1_{[x_t>0]}|\mathcal{Y}_t\right] &= 1_{[x_t>0]}E\left[\varphi(x_t)|\mathcal{Y}_t\right] \\ &= 1_{[x_t>0]}\Psi(y_1, y_2, ...y_{t-1}, y_t) \\ &= 1_{[x_t>0]}\Psi(y_1, y_2, ...y_{t-1}, 0) \end{aligned} \tag{3.29}$$

donde la primera igualdad se obtiene por la $\mathcal{Y}_t$-medibilidad del evento $[x_t > 0]$, y la última se sigue del hecho que $x_t > 0$ si y sólo si $y_t = 0$.

Ahora tomando esperanza con respecto a $\mathcal{Y}_{t-1}$ en (3.29), dado que $\mathcal{Y}_{t-1} \subseteq \mathcal{Y}_t$ y $\Psi(y_1, y_2, ...y_{t-1}, 0)$ es $\mathcal{Y}_{t-1}$-medible, obtenemos

$$E\left[\varphi(x_t)1_{[x_t>0]}|\mathcal{Y}_{t-1}\right] = \Psi(y_1, y_2, ...y_{t-1}, 0)P\left[x_t > 0|\mathcal{Y}_{t-1}\right] \tag{3.30}$$

o bien

$$\Psi(y_1, y_2, ...y_{t-1}, 0) = \frac{E\left[\varphi(x_t)1_{[x_t>0]}|\mathcal{Y}_{t-1}\right]}{P\left[x_t > 0|\mathcal{Y}_{t-1}\right]}. \tag{3.31}$$

Finalmente sustituyendo (3.31) en (3.29) y despues en (3.28) obtenemos que

$$E\left[\varphi(x_t)|\mathcal{Y}_t\right] = 1_{[x_t=0]}\varphi(0) + 1_{[x_t>0]}\frac{E\left[\varphi(x_t)1_{[x_t>0]}|\mathcal{Y}_{t-1}\right]}{P\left[x_t>0|\mathcal{Y}_{t-1}\right]} \tag{3.32}$$

la cual es la primera igualdad en (3.4).

Ahora, por el Teorema Condicional de Bayes (ver Elliott (1995)) tenemos

$$\begin{aligned}\int_0^\infty \varphi(s)\kappa_t(s)ds &= E\left[\varphi(x_t)|\mathcal{Y}_{t-1}, x_t>0\right] \\ &= \frac{E\left[\varphi(x_t)1_{[x_t>0]}|\mathcal{Y}_{t-1}\right]}{E\left[1_{[x_t>0]}|\mathcal{Y}_{t-1}\right]} \\ &= \frac{E\left[\varphi(x_t)1_{[x_t>0]}|\mathcal{Y}_{t-1}\right]}{P\left[x_t>0|\mathcal{Y}_{t-1}\right]}. \end{aligned} \tag{3.33}$$

Finalmente sustituyendo (3.33) en el segundo término de (3.32) obtenemos la segunda igualdad

$$E\left[\varphi(x_t)|\mathcal{Y}_t\right] = 1_{[x_t=0]}\varphi(0) + 1_{[x_t>0]}E\left[\varphi(x_t)|\mathcal{Y}_{t-1}, x_t>0\right].$$

3.5.2. Demostración del Lema 3.3

Consideremos por ahora el numerador del segundo sumando del lado derecho en (3.4). Como $\mathcal{Y}_{t-1} = \sigma(y_1,...,y_{t-1}) \subseteq \sigma(y_1,...,y_{t-1},x_{t-1})$, aplicando propiedades de la esperanza condicional obtenemos

$$\begin{aligned}E\left[\varphi(x_t)1_{[x_t>0]}|\mathcal{Y}_{t-1}\right] &= E\left[\varphi(x_{t-1}+a_{t-1}-\xi_{t-1})1_{[x_{t-1}+a_{t-1}-\xi_{t-1}>0]}|\mathcal{Y}_{t-1}\right] \\ &= E\left[E\left[\varphi(x_{t-1}+a_{t-1}-\xi_{t-1})1_{[x_{t-1}+a_{t-1}-\xi_{t-1}>0]}|\mathcal{Y}_{t-1},x_{t-1}\right]|\mathcal{Y}_{t-1}\right] \\ &= E\left[\int_0^\infty \varphi(x_{t-1}+a_{t-1}-w)1_{[x_{t-1}+a_{t-1}-w>0]}f(w)dw|\mathcal{Y}_{t-1}\right] \\ &= E\left[\int_0^{x_{t-1}+a_{t-1}} \varphi(x_{t-1}+a_{t-1}-w)f(w)dw|\mathcal{Y}_{t-1}\right].\end{aligned}$$

Ahora, haciendo el cambio de variable $s := x_{t-1} + a_{t-1} - w$, $ds = -dw$ y utilizando propiedades de la integral definda se obtiene

$$
\begin{aligned}
E\left[\varphi(x_t)1_{[x_t>0]}|\mathcal{Y}_{t-1}\right] &= E\left[-\int_{x_{t-1}+a_{t-1}}^{0} \varphi(s)f(x_{t-1}+a_{t-1}-s)ds|\mathcal{Y}_{t-1}\right] \\
&= E\left[\int_{0}^{x_{t-1}+a_{t-1}} \varphi(s)f(x_{t-1}+a_{t-1}-s)ds|\mathcal{Y}_{t-1}\right] \\
&= E\left[\int_{0}^{\infty} \varphi(s)f(x_{t-1}+a_{t-1}-s)1_{[x_{t-1}+a_{t-1}-s\geq 0]}ds|\mathcal{Y}_{t-1}\right] \\
&= \int_{0}^{\infty} \varphi(s)E\left[f(x_{t-1}+a_{t-1}-s)1_{[x_{t-1}+a_{t-1}-s\geq 0]}|\mathcal{Y}_{t-1}\right]ds.
\end{aligned}
\tag{3.34}
$$

Usando (3.6) con el índice $t-1$ en lugar de t y remplazando $\varphi(x_{t-1})$ por

$$f(x_{t-1}+a_{t-1}-s)1_{[x_{t-1}+a_{t-1}-s\geq 0]},$$

lo cual se puede hacer ya que $\varphi(\cdot)$ es cualquier función real y acotada, se obtiene

$$
\begin{aligned}
E\left[f(x_{t-1}+a_{t-1}-s)1_{[x_{t-1}+a_{t-1}-s\geq 0]}|\mathcal{Y}_{t-1}\right] &= 1_{[x_{t-1}=0]}f(a_{t-1}-s)1_{[a_{t-1}-s\geq 0]} \\
&+ 1_{[x_{t-1}>0]}\int_{0}^{\infty} f(w+a_{t-1}-s)1_{[w+a_{t-1}-s\geq 0]}\kappa_{t-1}(w)dw.
\end{aligned}
\tag{3.35}
$$

Ahora sustituyendo (3.35) en (3.34) se obtiene

$$
\begin{aligned}
E\left[\varphi(x_t)1_{[x_t>0]}|\mathcal{Y}_{t-1}\right] &= 1_{[x_{t-1}=0]}\int_{0}^{\infty} \varphi(s)f(a_{t-1}-s)1_{[s\leq a_{t-1}]}ds \\
&+ 1_{[x_{t-1}>0]}\int_{0}^{\infty} \varphi(s)\int_{(s-a_{t-1})^+}^{\infty} f(w+a_{t-1}-s)\kappa_{t-1}(w)dwds.
\end{aligned}
\tag{3.36}
$$

Por otro lado, el denominador del segundo término del lado derecho de (3.4) lo podemos expresar como

$$
\begin{aligned}
P\left[x_t > 0 | \mathcal{Y}_{t-1}\right] &= E\left[1_{[x_{t-1}+a_{t-1}-\xi_{t-1}>0]} | \mathcal{Y}_{t-1}\right] \\
&= E\left[E\left[1_{[x_{t-1}+a_{t-1}-\xi_{t-1}>0]} | \mathcal{Y}_{t-1}, x_{t-1}\right] | \mathcal{Y}_{t-1}\right] \\
&= E\left[E\left[1_{[\xi_{t-1}\leq x_{t-1}+a_{t-1}]} | \mathcal{Y}_{t-1}, x_{t-1}\right] | \mathcal{Y}_{t-1}\right] \\
&= E\left[P\left[\xi_{t-1} \leq x_{t-1}+a_{t-1} | \mathcal{Y}_{t-1}, x_{t-1}\right] | \mathcal{Y}_{t-1}\right] \\
&= E\left[F(x_{t-1}+a_{t-1}) | \mathcal{Y}_{t-1}\right] \\
&= 1_{[x_{t-1}=0]} F(a_{t-1}) + 1_{[x_{t-1}>0]} \int_0^\infty F(w+a_{t-1})\kappa_{t-1}(w)dw
\end{aligned} \tag{3.37}
$$

donde la penúltima igualdad es debido a la independencia entre ξ_{t-1} y $\mathcal{Y}_{t-1}$ y la última igualdad se obtiene de aplicar la ecuación (3.6). Así, insertando (3.36) y (3.37) en (3.4), despues multiplicando en ambos lados de la igualdad por $1_{[x_t>0]}$ y utilizando propiedades de las funciones indicadoras obtenemos

$$
\begin{aligned}
E\left[\varphi(x_t)|\mathcal{Y}_t\right] 1_{[x_t>0]} &= 1_{[x_{t-1}=0]} \left\{ \frac{\int_0^\infty \varphi(s) f(a_{t-1}-s) 1_{[s\leq a_{t-1}]} ds}{F(a_{t-1})} \right\} \\
&+ 1_{[x_{t-1}>0]} \left\{ \frac{\int_0^\infty \varphi(s) \int_{(s-a_{t-1})^+}^\infty f(w+a_{t-1}-s)\kappa_{t-1}(w)dwds}{\int_0^\infty F(w+a_{t-1})\kappa_{t-1}(w)dw} \right\},
\end{aligned}
$$

lo cual demuestra el Lema 3.3.

3.5.3. Demostración del Lema 3.7

Como el controlador sólo observa las ventas, la demanda ξ_n no es $\mathcal{Y}_n$-medible. Recuerde también que los eventos $[y_n \geq x_n + a_n]$ y $[y_n < x_n + a_n]$ son equivalentes respectivamente a $[y_n = x_n + a_n]$ y $[y_n = \xi_n]$. Entonces

$$
\begin{aligned}
E[\psi(\xi_n)|\mathcal{Y}_n] &= E[\psi(\xi_n)1_{[y_n=x_n+a_n]}|\mathcal{Y}_n] + E[\psi(\xi_n)1_{[y_n<x_n+a_n]}|\mathcal{Y}_n] \\
&= 1_{[y_n=x_n+a_n]}E[\psi(\xi_n)|\mathcal{Y}_n] + 1_{[y_n<x_n+a_n]}E[\psi(\xi_n)|\mathcal{Y}_n] \\
&= 1_{[y_n=x_n+a_n]}E[\psi(\xi_n)|\mathcal{Y}_n] + 1_{[y_n<x_n+a_n]}\psi(y_n).
\end{aligned} \tag{3.38}
$$

Por otra parte como x_n+a_n es $\mathcal{Y}_{n-1}$-medible, entonces en el evento $[y_n = x_n+a_n]$

podemos escribir

$$\begin{aligned} 1_{[y_n=x_n+a_n]}E[\psi(\xi_n)|\mathcal{Y}_n] &= E[\psi(\xi_n)1_{[y_n=x_n+a_n]}|\mathcal{Y}_n] \\ &= E[\psi(\xi_n)1_{[y_n=x_n+a_n]}|\mathcal{Y}_{n-1}] \\ &= 1_{[y_n=x_n+a_n]}E[\psi(\xi_n)|\mathcal{Y}_{n-1}] \\ &= 1_{[y_n=x_n+a_n]}\varphi(y_1,y_2,\ldots,y_{n-1}) \end{aligned} \tag{3.39}$$

para alguna función φ medible. Pero cuando $[y_n = x_n + a_n]$ también sabemos que $[\xi_n \geq x_n + a_n]$ y por tanto

$$\begin{aligned} \varphi(y_1,y_2,\ldots,y_{n-1})1_{[y_n=x_n+a_n]} &= E[\psi(\xi_n)1_{[y_n=x_n+a_n]}|\mathcal{Y}_n] \\ &= E[\psi(\xi_n)1_{[\xi_n\geq x_n+a_n]}|\mathcal{Y}_n], \end{aligned}$$

como $\varphi(y_1,y_2,\ldots,y_{n-1})$ es $\mathcal{Y}_{n-1}$-medible y $\mathcal{Y}_{n-1}\subseteq\mathcal{Y}$, condicionando con respecto a $\mathcal{Y}_{n-1}$ tenemos

$$\varphi(y_1,y_2,\ldots,y_{n-1})E[1_{[y_n=x_n+a_n]}|\mathcal{Y}_{n-1}] = E[\psi(\xi_n)1_{[\xi_n\geq x_n+a_n]}|\mathcal{Y}_{n-1}].$$

Por lo tanto

$$\begin{aligned} \varphi(y_1,y_2,\ldots,y_{n-1}) &= \frac{E[\psi(\xi_n)1_{[\xi_n\geq x_n+a_n]}|\mathcal{Y}_{n-1}]}{E[1_{[y_n=x_n+a_n]}|\mathcal{Y}_{n-1}]} \\ &= \frac{E[\psi(\xi_n)1_{[\xi_n\geq x_n+a_n]}|\mathcal{Y}_{n-1}]}{E[1_{[\xi_n\geq x_n+a_n]}|\mathcal{Y}_{n-1}]} \\ &= \frac{\int_{x_n+a_n}^{\infty}\psi(s)\kappa_n(s)ds}{\int_{x_n+a_n}^{\infty}\kappa_n(s)ds}. \end{aligned} \tag{3.40}$$

Finalmente sustituyendo (3.40) en (3.39), y (3.39) en (3.38) obtenemos

$$E[\psi(\xi_n)|\mathcal{Y}_n] = 1_{[y_n=x_n+a_n]}\frac{\int_{x_n+a_n}^{\infty}\psi(s)\kappa_n(s)ds}{\int_{x_n+a_n}^{\infty}\kappa_n(s)ds} + 1_{[y_n<x_n+a_n]\psi(y_n)}$$

que es el resultado deseado.

3.5.4. Demostración del Lema 3.8

Nótese primero que, como $x_n + a_n$ es $\mathcal{Y}_{n-1}$-medible, $y_n = \min\{\xi_n, x_n + a_n\}$ es $(\mathcal{Y}_{n-1},\xi_n)$-medible y por tanto $1_{[\xi_n\geq x_n+a_n]}$ también lo es. Así, $\eta_n 1_{[\xi_n\geq x_n+a_n]}$ es $(\mathcal{Y}_{n-1},\xi_n)$-medible, pues la venta y_n depende de la demanda ξ_n, y η_n es una v.a. $\mathcal{Y}_n$-medible.

Ahora, para cualquier $G(\cdot)$ integrable, se tiene

$$\begin{aligned}
E[G(\xi_{n+1})\eta_n 1_{[\xi_n \geq x_n+a_n]}|\mathcal{Y}_{n-1},\xi_n] &= \eta_n 1_{[\xi_n \geq x_n+a_n]} E[G(\xi_{n+1})|\mathcal{Y}_{n-1},\xi_n] \\
&= \eta_n 1_{[\xi_n \geq x_n+a_n]} E[G(\xi_{n+1})|\xi_n] \\
&= \eta_n 1_{[\xi_n \geq x_n+a_n]} \int G(s)p(s|\xi_n)ds
\end{aligned}$$

donde la segunda igualdad es debido a la propiedad de Markov del proceso $\{\xi_n\}$. Ahora nótese que

$$\begin{aligned}
E[G(\xi_{n+1})\eta_n 1_{[\xi_n \geq x_n+a_n]}] &= E\left\{E[G(\xi_{n+1})\eta_n 1_{[\xi_n \geq x_n+a_n]}|\mathcal{Y}_{n-1},\xi_n]\right\} \\
&= E\left[\eta_n 1_{[\xi_n \geq x_n+a_n]} \int G(s)p(s|\xi_n)ds\right] \\
&= E\left\{E\left[\eta_n 1_{[\xi_n \geq x_n+a_n]} \int G(s)p(s|\xi_n)ds|\mathcal{Y}_n\right]\right\} \qquad (3.41) \\
&= E\left\{\eta_n 1_{[\xi_n \geq x_n+a_n]} E\left[\int G(s)p(s|\xi_n)ds|\mathcal{Y}_n\right]\right\},
\end{aligned}$$

donde la última igualdad se obtiene porque η_n es $\mathcal{Y}_n$-medible y $[\xi_n \geq x_n+a_n] = [y_n = x_n + a_n] \in \mathcal{Y}_n$. Ahora, de (3.41), tomando $\int G(s)p(s|x)ds = \psi(x)$ y sustituyendo en (3.17), obtenemos

$$\begin{aligned}
E[G(\xi_{n+1})\eta_n 1_{[\xi_n \geq x_n+a_n]}] &= E\left\{\eta_n 1_{[\xi_n \geq x_n+a_n]} E\left[\int G(s)p(s|\xi_n)ds|\mathcal{Y}_n\right]\right\} \\
= E\left\{\eta_n 1_{[\xi_n \geq x_n+a_n]} \left[1_{[y_n = x_n+a_n]} \frac{\int_{x_n+a_n}^{\infty} \kappa_n(x) \int G(s)p(s|x)dsdx}{\int_{x_n+a_n}^{\infty} \kappa_n(x)dx} + 1_{[y_n < x_n+a_n]} \int G(s)p(s|y_n)ds\right]\right\} & \\
&= E\left[\eta_n 1_{[y_n = x_n+a_n]} \frac{\int_{x_n+a_n}^{\infty} \kappa_n(x) \int G(s)p(s|x)dsdx}{\int_{x_n+a_n}^{\infty} \kappa_n(x)dx}\right]
\end{aligned}$$

la cual es precisamente la ecuación (3.18). Ahora para demostrar (3.19), como $\eta_n 1_{[\xi_n < x_n+a_n]}$ es $(\mathcal{Y}_{n-1}, \xi_n)$-medible tenemos

$$\begin{aligned}
E[G(\xi_{n+1})\eta_n 1_{[\xi_n < x_n+a_n]}|\mathcal{Y}_{n-1},\xi_n] &= \eta_n 1_{[\xi_n < x_n+a_n]} E[G(\xi_{n+1})|\mathcal{Y}_{n-1},\xi_n] \\
&= \eta_n 1_{[\xi_n < x_n+a_n]} E[G(\xi_{n+1})|\xi_n] \\
&= \eta_n 1_{[\xi_n < x_n+a_n]} \int G(s)p(s|\xi_n)ds \\
&= \eta_n 1_{[\xi_n < x_n+a_n]} \int G(s)p(s|y_n)ds,
\end{aligned}$$

la segunda igualdad se obtiene por la propiedad de Markov del proceso $\{\xi_n\}$ y la última es puesto que en el evento $[\xi_n < x_n + a_n]$, tenemos que $\xi_n = y_n$. Finalmente tomando esperanza en ambos lados obtenemos

$$E[G(\xi_{n+1})\eta_n 1_{[\xi_n < x_n + a_n]}] = E\left[\eta_n 1_{[\xi_n < x_n + a_n]} \int G(s)p(s|y_n)ds\right],$$

que es la ecuación (3.19)

Conclusiones

En este trabajo se estudió la teoría general de los procesos de control de Markov parcialmente observables y se presentaron aplicaciones a sistemas de inventario. Para esto se aplicó un técnica estandar que consiste en transformar el problema de control óptimo parcialmente observable a uno completamente observable el cual está definido en el espacio de medidas de probabilidad sobre el espacio de estados del problema original. Esta transformación está basada en la existencia de una función que determina la dinámica del nuevo sistema completamente observable, y por lo tanto es meramente teórica.

Muchos de los trabajos en el área de los procesos de control parcialmente observables se centran en establecer condiciones que garanticen la existencia de políticas ótimas en el problema completamente observable, a partir de la existencia de dicha función. Sin embargo otros se centran en proporcionar explícitamente la transformación, lo cual es posible bajo condiciones específicas sobre el modelo de control o en ejemplos muy particulares. En nuestro caso nos centramos en este último aspecto presentando dos modelos de inventario parcialmente observable, el *Zero Balance Walk* y el de *Demanda Parcialmente Observable.*

Existen otros trabajos relacionados con sistemas de inventarios parcialmente observables que podemos considerar una extensión de nuestros modelos, a los cuales se les conoce en la literatura como *Rain Checks* y *Two Distributions*, ver Bensoussan, et.al. (2008b, 2010). Estos modelos permiten acumular la demanda no satisfecha, y por tanto la dinámica del sistema toma la forma

$$x_{t+1} = x_t + a_t - \xi_t, \quad t = 0, 1, ...$$

Específicamente en le modelo Rain Checks se considera que el operador del inventario sólo observa cuando el inventario es negativo, mientras que en el segundo modelo sólo se observa cuando el inventario es cero.

Un aspecto que es importante remarcar es que, bajo el esquema estandar, la solución a un problema de control parcialmente observable se obtiene a través de resolver un problema de control completamente observable definido en espacio de medidas. Este hecho hace muy dificil implementar algoritmos computacionales que nos permitan obtener o aproximar la solución al problema, por lo cual constituye un problema que se puede estudiar a futuro.

Por otra parte, considerando que sólo nos centramos en analizar el criterio de optimalidad de costo descontado para procesos de Markov, creemos importante estudiar los modelos de inventario parcialmente observables bajo otros índices de funcionamiento, digamos el de costo promedio, y considerar además procesos semi markovianos. Mas aún, se pueden explorar extensiones a otros problemas como procesos de control adaptados, control minimax y juegos estocásticos parcialmente observables.

Apéndice

Apéndice A

En este apéndice describimos la terminología usada en la tesis, así como algunos resultados auxiliares para el desarrollo del trabajo.

A.1. Espacios y Funciones

Dado un espacio topológico X, su σ-álgebra de Borel es denotada por $\mathcal{B}(X)$. Medibilidad de conjuntos y funciones, significa Borel-medibles.

Un espacio de Borel es un subconjunto de Borel de un espacio métrico completo y separable. Será denotado por $\mathbb{P}(X)$ al espacio de todas las medidas de probabilidad en X.

Proposición A.1. *Si X es un espacio de Borel, entonces $\mathbb{P}(X)$ también es un espacio de Borel.*

Demostración. Ver, por ejemplo Hinderer (1970) p.91. $\square$

Si X es un espacio topológico , definimos los espacios $M(X) \supset B(X) \supset \mathcal{C}(X)$ de funciones de X en $\mathbb{R}$ como: $M(X)$ el espacio de las funciones medibles en X, y $B(X)$ y $\mathcal{C}(X)$ los subespacios de las funciones medibles y acotadas y continuas y acotadas, respectivamente. $B(X)$ y $\mathcal{C}(X)$ son espacios de Banach bajo la norma del supremo $\|v\| := \sup_x |v(x)|$.

Definición A.2. Sea X un espacio topológico y v una función de X en los reales extendidos. La función v se dice ser *semicontinua inferiormente* (l.s.c.) si el conjunto $\{x \in X | v(x) \leq r\}$ es cerrado en X para toda $r \in \mathbb{R}$, y *semicontinua superiormente* (u.s.c) si $\{x \in X | v(x) \geq r\}$ es cerrado para cada $r \in \mathbb{R}$.

Observe que v es l.s.c. si y sólo si $-v$ es u.s.c.; además, v es continua si y sólo si v es l.s.c. y u.s.c.

Denotaremos por $\mathcal{L}(X)$ el conjunto de todas las funciones l.s.c. definidas en X que son acotadas inferiormente y por $\mathcal{L}_+(X)$ al subconjunto de las funciones no-negativas.

Proposición A.3. *Sea X un espacio métrico y v una función de X en los reales extendidos. Entonces:*

(a) v es l.s.c. si y sólo si para cada sucesión $\{x_n\} \subset X$ tal que $x_n \to x$, $x \in X$, se tiene que $\liminf_{n\to\infty} v(x_n) \geq v(x)$.

(b) $v \in \mathcal{L}(X)$ si y sólo si existe una sucesión de funciones $v_n \in \mathcal{C}(X)$ tal que $v_n \uparrow v$.

Por la dualidad entre las funciones l.s.c. y u.s.c. (v es l.s.c. si y sólo si $-v$ es u.s.c.), un resultado similar a la Proposición A.3 puede ser obtenido para las funciones u.s.c.. Por ejemplo, de la Proposición A.3(a) obtenemos: v es u.s.c. si y sólo si para cada sucesión $\{x_n\}$ de X tal que $x_n \to x$, $x \in X$, se tiene que

$$\limsup_{n\to\infty} v(x_n) \leq v(x).$$

Ahora algunas propiedades elementales, pero útiles, de las funciones l.s.c. y acotadas inferiormente.

Proposición A.4. *Si $v, v_1, \ldots, v_n$ pertenecen a $\mathcal{L}(X)$, entonces:*

(a) $\alpha v(\cdot) \in \mathcal{L}(X)$ para toda $\alpha > 0$.

(b) $v_1 + \cdots + v_n$ y $\min_i v_i$ pertenecen a $\mathcal{L}(X)$.

(c) Si X es compacto, entonces v alcanza el ínfimo, es decir, existe $x^* \in X$ tal que $v(x^*) = \inf_x v(x)$

Para las demostraciones de las Proposiciones A.3 y A.4 ver, por ejemplo, Ash (1972), Bertsekas and Shreve (1978).

Proposición A.5. *Sean u y u_n, $n = 1, 2, \ldots$ funciones l.s.c. y acotadas inferiormente tales que, para cada $x \in X$ y $r \in \mathbb{R}$, los conjuntos $\{a \in A(x) | u(x,a) \leq r\}$ y $\{a \in A(x) | u_n(x,a) \leq r\}$ son compactos. Si $u_n \uparrow u$, entonces*

$$\lim_{n\to} \min_{a\in A(x)} u_n(x,a) = \min_{a\in A(x)} u(x,a) \;\; \forall x \in X.$$

Demostración. Ver, Lema 4.2.4 de Hernández-Lerma and Lasserre (1996) □

Proposición A.6. *Sea X un espacio de Borel arbitrario. Supóngase que $\{z_n\}$ es una sucesión en $\mathbb{P}(X)$ que converge débilmente a $z \in \mathbb{P}(X)$, y $\{v_n\}$ una sucesión de funciones no-negativas y l.s.c. en X tal que*

$$\liminf_{n\to\infty} v_n(x) \geq v(x) \;\; \forall x \in X.$$

Entonces

$$\liminf_{n\to\infty} \int\limits_X v_n(x)z_n(dx) \geq \int\limits_X v(x)z(dx).$$

Demostración. Ver, Lema 3.1 en p.11 de Hernández-Lerma and Romera (1999). □

Denotamos por $\mathbb{D}$ el conjunto de las funciones de densidad κ definidas en $[0,\infty)$ tales que $\int_0^\infty x\kappa(x)dx < \infty$. Definamos los siguientes espacios

$$\mathcal{H} := \left\{ \rho \in L^1(\mathbb{R}^+) : \int\limits_0^\infty x|\rho(x)|dx < \infty \right\},$$

y

$$\mathcal{H}^+ = \{\rho \in \mathcal{H} | \rho \geq 0\},$$

donde $L^1(\mathbb{R}^+)$ es el espacio de las funciones integrables cuyo dominio es el conjunto de los números reales no-negativos. Note que $\mathbb{D} \subset \mathcal{H}^+$ y además $\mathcal{H}^+$ es un subconjunto cerrado de $\mathcal{H}$ con la norma

$$\|\rho\| = \int\limits_0^\infty |\rho(x)|dx + \int\limits_0^\infty x|\rho(x)|dx. \tag{A.1}$$

Sea $\{\rho_t\}$ una sucesión en $\mathcal{H}^+$ y $\rho \in \mathcal{H}^+$, se dice que $\rho_t \to \rho$ si $\|\rho_t - \rho\| \to 0$ cuando $t \to \infty$, lo cual es equivalente a

$$\int\limits_0^\infty |\rho_t(x) - \rho(x)|dx \to 0 \;\text{ y }\; \int\limits_0^\infty x|\rho_t(x) - \rho(x)|dx \to 0.$$

Por las propiedades de la integral y el valor absoluto se verifica que en efecto el operador en (A.1) define una norma.

A.2. Kérneles Estocásticos

A lo largo de esta sección X y Z denotan espacios de Borel.

Definición A.7. Un kérnel estocástico sobre X dado Z es una función $Q(\cdot|\cdot)$, o $Q(dx|z)$, tal que:

(a) $Q(\cdot|z)$ es una medida de probabilidad en X para cada $z \in Z$, y

(b) $Q(B|\cdot)$ es una función medible en Z para cada $B \in \mathcal{B}(X)$.

Equivalentemente, una medida de probabilidad $Q(dx|z)$ en X, para cada $z \in Z$, es un kérnel estocástico si y sólo si la función $h : Z \to \mathbb{P}(X)$ definida por

$$h(z) := Q(\cdot|z) \tag{A.2}$$

es medible.

La familia de todos los kérneles estocásticos sobre X dado Z es denotado por $\mathcal{P}(X|Z)$.

Definición A.8. Sea $Q(dx|z)$ un kérnel estocástico sobre X dado Z. Se dice que

(a) Q es *fuertemente continuo* si la función

$$z \mapsto \int v(x)Q(dx|z) \tag{A.3}$$

es continua y acotada para cada función $v \in B(X)$.

(b) Q es *débilmente continuo* si la función en (A.3) es continua y acotada para cada función $v \in \mathcal{C}(X)$

(c) Q es l.s.c. si la función en (A.3) es l.s.c. para cada $v \in \mathcal{L}(X)$.

Proposición A.9. *Sea $Q(dx|y)$ un kérnel estocástico sobre X dado Y; sea $f(x,y)$ una función real medible sobre $X \times Y$, y sea $f^{'} : Y \to \mathbb{R}$ la función definida por*

$$f^{'}(y) := \int f(x,y)Q(dx|y)$$

cuando la integral existe.

(a) Si $f \in B(X \times Y)$, entonces $f^{'} \in B(Y)$.

(b) Si $Q(dx|y)$ es continuo and $f \in \mathcal{C}(X \times Y)$, entonces $f^{'} \in \mathcal{C}(Y)$.

Demostración. Ver, por ejemplo, Bertsekas and Shreve (1978). □

Proposición A.10. *Sean X, Y y W espacios de Borel y sea $R(d(x,y)|w)$ un kérnel estocástico sobre $X \times Y$ dado W. Entonces existen kérneles estocásticos $H'(dx|w,y)$ y $R'(dy|w)$ sobre X dado $W \times Y$ y sobre Y dado W respectivamente tales que*

$$R(B \times C|w) = \int_C H'(B|w,y)R'(dy|w)$$

para toda $B \in \mathcal{B}(X)$, $C \in \mathcal{B}(Y)$ y $w \in W$, donde $R'(dy|w)$ es la marginal de $R(d(x,y)|w)$ sobre Y, es decir,

$$R'(C|w) := R(X \times C|w), \quad C \in \mathcal{B}(Y). \tag{A.4}$$

Demostración. Bertsekas and Shreve (1978), Corollary 7.27.1, p. 139; o Dynkin and Yushkevich (1979), p. 215; o Striebel (1975), Appendix A.1, etc. □

A.3. Multifunciones y Selectores Medibles

Consideraremos los espacios de Borel (no-vacíos) X y A.

Una *multifunción* φ de X en A es una función cuyos valores $\varphi(x)$, para cada $x \in X$, son subconjuntos no-vacíos de A. La *gráfica* de φ es el subconjunto $Gr(\varphi)$ de $X \times A$ definido como

$$Gr(\varphi) := \{(x,a)|x \in X, a \in \varphi(x)\},$$

y si B es un subconjunto no-vacío de A, definimos $\varphi^{-1}[B] := \{x \in X|\varphi(x) \cap B \neq\}$. (En el texto, se escribe $A(x)$ en lugar de $\varphi(x)$, y $Gr(\varphi)$ como $\mathbb{K}$).

Una multifunción φ de X en A se dice ser compacta (cerrada) si $\varphi(x)$ es un subconjunto compacto (cerrado) de A para cada $x \in X$.

Definición A.11. Sea φ una multifunción de X en A. Se dice que:

(a) φ es *Borel-medible* si $\varphi^{-1}[F]$ es un subconjunto de Borel de X para cada subconjunto cerrado F de A.

(b) φ es *semicontinua superiormente* (u.s.c.) si para cada subconjunto abierto $G \subset A$, el conjunto $\{x \in X|\varphi(x) \subset G\}$ es abierto en X.

(c) φ es *semicontinua inferiormente* (l.s.c.) si para cada subconjunto cerrado $F \subset A$, el conjunto $\{x \in X|\varphi(x) \subset F\}$ es cerrado en X.

(d) φ es *continua* si es u.s.c. y l.s.c.

Definición A.12. Sea φ una multifunción medible de X en A. Entonces una función medible $f : X \to A$ tal que $f(x) \in \varphi(x)$ para cada $x \in X$ es llamado un *selector medible* de φ. Se denota por $\mathbb{F}$ al conjunto de todos los selectores medibles de φ.

Proposición A.13. *Sea φ una multifunción de X en A Borel-medible y compacta, y sea $v : Gr(\varphi) \to \mathbb{R}$ una función medible.*

(a) Si $v(x,\cdot)$ es u.s.c. sobre $\varphi(x)$ para cada $x \in X$, entonces existe un selector medible f^* de φ tal que

$$v(x, f^*(x)) = \max_{a \in \varphi(x)} v(x,a) =: v^*(x) \quad \forall x \in X, \tag{A.5}$$

y la función v^* es medible.

(b) Si $v(x,\cdot)$ es l.s.c. sobre $\varphi(x)$ para cada $x \in X$, entonces existe un selector medible f_* de φ tal que

$$v(x, f_*(x)) = \min_{a \in \varphi(x)} v(x,a) =: v_*(x) \quad \forall x \in X, \tag{A.6}$$

y v_* es medible.

(c) Si φ es u.s.c. y v es u.s.c. sobre $Gr(\varphi)$ y acotada superiormente, entonces existe un selector medible f^* de φ tal que

$$v(x, f^*(x)) = \max_{a \in \varphi(x)} v(x,a) =: v^*(x) \quad \forall x \in X,$$

y la función v^* es u.s.c. y acotada superiormente.

(d) Si φ es u.s.c. y v es l.s.c. sobre $Gr(\varphi)$ y acotada inferiormente, entonces existe un selector medible f_* de φ tal que

$$v(x, f_*(x)) = \min_{a \in \varphi(x)} v(x,a) =: v_*(x) \quad \forall x \in X,$$

y la función $v_*(x)$ es l.s.c. y acotada inferiormente.

Demostración. Himmelberg *et al.* (1976); Schäl (1975). □

Apéndice B

Tabla de Notaciones

B.1. Abreviaturas

- CO: Completamente Observale
- EO: Ecuación de Optimalidad
- l.s.c.: semicontinua inferiormente
- PCMs: Procesos de Control de Markov
- PCO: Problema de Control Óptimo
- PO: Parcialmente Observable
- u.s.c.: semicontinua superiormente
- ZBW: Zero Balance Walk

B.2. Conjuntos y Espacios

- A: espacio de acciones, espacio de Borel
- $\mathcal{B}(X)$: σ-álgebra de Borel de X
- $\mathbb{F}$: conjunto de todos los selectores medibles
- H_t: espacio de historias admisibles
- $\mathbb{K}$: conjunto de parejas admisibles estado-acción
- $\mathcal{P}(W|Z)$: familia de todos los kérneles estocásticos sobre W dado Z
- $\mathbb{P}(X)$: espacio de todas las medidas de probabilidad en X

- X: espacio de estados, espacio de Borel
- Y: espacio de las observaciones, espacio de Borel
- $\mathcal{Y}_t$: $\sigma-$álgebra generada por las observaciones hasta el tiempo t
- $\mathbf{Z}$:=$\mathbb{P}(X)$
- Π: conjunto de todas las políticas
- Δ: conjunto de todas las I-políticas

B.3. Espacios de Funciones

- $B(X)$: espacio de las funciones medibles y acotadas en X
- $\mathcal{C}(X)$: espacio de las funciones continuas y acotadas en X
- $\mathbb{D}$: conjunto de las funciones de densidad κ definidas en $[0, \infty)$ tales que

$$\int_0^\infty x\kappa(x)dx < \infty$$

- $\mathcal{H}$:=$\left\{\rho \in L^1(\mathbb{R}^+) : \int_0^\infty x|\rho(x)|dx < \infty\right\}$
- $\mathcal{H}^+$:=$\{\rho \in \mathcal{H} | \rho \geq 0\}$
- $\mathcal{L}_+(X)$: conjunto de las funciones l.s.c. y no-negativas
- $M(X)$: espacio de las funciones medibles en X

B.4. Kérnesles Estocásticos

- K: kérnel estocástico sobre Y dado $A \times X$
- K_0: kérnel estocástico sobre Y dado X
- Q: kérnel estocástico sobre X dado $\mathbb{K}$

B.5. Notación

- E_ν^π: operador esperanza asociada a la medida de probabilidad P_ν^π
- f^∞: política estacionaria determinista
- f: función de densidad de ξ_t
- F: función de distribución de ξ_t
- $\overline{F}$:=$1 - F$
- P_ν^π: medida de probabilidad inducida por la distribución inicial ν y política π
- $\{z_n\} \subset \mathbf{Z}$: proceso de estados en el PCM-CO.

Bibliografía

[1] Ash, R.B., *Real Analysis and Probability*, Academic Press, (1972).

[2] Bensoussan, A., Cakanyildirim, M., and Sethi, S.P., *Partially observed inventory systems: the case of zero-balnce walk*, SIAM J. Control Optim. **46**, No. 1, 176 – 209, (2007).

[3] Bensoussan, A., Cakanyildirim, M., Minjárez-Sosa, J.A., and Sethi, S.P., *Inventory problems with partially observed demands and lost sales*, J. Optim. Theory Appl. **136**, 321 – 340, (2008a).

[4] Bensoussan, A., Cakanyildirim, M., Minjárez-Sosa, J.A., Sethi, S.P., and Shi, R., *Partially observed inventory systems: the case of rain checks*, SIAM J. Control Optim. **47**, No. 5, 2490 – 2519, (2008b).

[5] Bensoussan, A., Cakanyildirim, M., Minjárez-Sosa, J.A., Sethi, S.P., and Shi, R., *An incomplete information inventory model with presence of inventories or backorders as only observations*, J. Optim. Theory Appl. **146**, 544580, (2010).

[6] Bertsekas, D.P., and Shreve, S.E., *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, (1978).

[7] Borkar, V.S., *Dynamic programming for ergodic control with partial observations*, Stoch. Proc. Appl. **103**, 293 – 310, (2003).

[8] Borkar, V.S., and Budhirajab, A., *A further remark on dynamic programming for pPartially observed Markov processes*, Stoch. Proc. Appl. **112**, 79 – 93, (2004).

[9] Brooks, A., Makarenko, A., Williams, S., and Durrant-Whyte, H., *Parametric POMDPs for planning in continuous state Spaces*, Robotics and Autonomous Systems **54**, 887 – 897, (2006).

[10] Dynkin, E.B., and Yushkevich, A.A., *Controlled Markov Processes*, Springer-Verlag, (1979).

[11] Elliott, R.J., Aggoun, L., and Moore, J.B., *Hidden Markov Models: Estimation and Control*, Springer-Verlag, (1995).

[12] Hernández-Lerma, O., *Adaptive Markov Control Processes*, Springer-Verlag, (1989).

[13] Hernández-Lerma, O., *Lecture Notes on Discrete-Time Markov Control Processes*, Departamenento de Matemáticas, Centro de Investigación del IPN, (1990).

[14] Hernández-Lerma, O., and Lasserre, J.B., *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer, (1996).

[15] Hernández-Lerma, O., and Romera, R., *Limiting discounted-cost control of partially bbservable stochastic system*, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, (1999).

[16] Himmelberg, C.J., Parthasarathy, T., and Van Vleck, F.S., *Optimal plans for dynamic programming problems*, Math. Oper. Res. **1**, 390 – 394, (1976).

[17] Hinderer, K., *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*, Lecture Notes Oper. Res. **33**, Springer-Verlag, (1970).

[18] Hsu, S.P., Chuang, D.M., and Arapostathisa, A., *On the existence of stationary optimal policies for partially observed MDPs under the long-run average cost criterion*, Systems & Control Letters **55**, 165 – 173, (2006).

[19] Sawaragi, Y., and Yoshikawa, T., *Discrete-time markovian decision processes with incomplete state observation*, Ann. Math. Statist. **41**, 78 – 86, (1970).

[20] Schäl, M., *Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal*, Z. Wahrs. verw. Gerb. **32**, 179 – 196, (1975).

[21] Striebel, C., *Optimal Control of Discrete Time Stochastic Systems*, Lecture Notes Econ. Math. Syst. **110**, Springer-Verlag, (1975).

[22] Yushkevich, A.A., *Reduction of a controlled Markov model with incomplete data to a problem with complete information in the case of borel state and control spaces*, Theory Probab. Appl. **21**, 153 – 158, (1976).