10

GESTIÓ I ORGANITZACIÓ D'EMPRESES

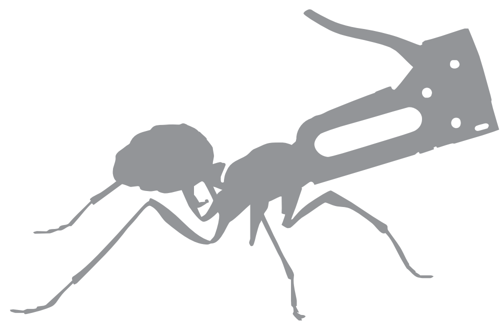
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

UPCPOSTGRAU

Programación dinámica

MÉTODOS CUANTITATIVOS PARA LA TOMA DE DECISIONES II

Joan B. Fonollosa
Josep M. Sallán
Vicenç Fernández
Albert Suñé



UNIVERSITAT POLITÈCNICA
DE CATALUNYA
BARCELONATECH

iniciativa
digital politècnica
Publicacions Acadèmiques UPC

UPCPOSTGRAU

Programación dinámica

MÉTODOS CUANTITATIVOS PARA LA TOMA DE DECISIONES II

Joan B. Fonollosa
Josep M. Sallán
Vicenç Fernández
Albert Suñé

Primera edición: Septiembre de 2016

© Iniciativa Digital Politècnica, 2016
Oficina de Publicacions Acadèmiques Digitals de la UPC
Jordi Girona 31,
Edifici Torre Girona, Planta 1, 08034 Barcelona
Tel.: 934 015 885
www.upc.edu/idp
E-mail: info.idp@upc.edu

Depósito legal: B. 18120-2016
ISBN: 978-84-9880-605-2

Índice

Presentación de la obra

La programación dinámica es una técnica de cálculo recurrente, que permite construir una clase general de modelos matemáticos útiles para resolver problemas que, de otra manera, suelen ser de muy difícil resolución.

Fue inventada en 1953 por el matemático Richard Bellman como una técnica para optimizar problemas complejos que pueden ser discretizados y secuenciados.

El sintagma nominal *programación dinámica* consta de un nombre y un adjetivo, cada uno de los cuales tiene una significación muy determinada, que, si se conoce bien, permite deducir cuáles eran las motivaciones de sus primeros teóricos y, en consecuencia, qué pretendían modelizar con dicha técnica. Sin embargo, como veremos a lo largo de este volumen, la programación dinámica ha desbordado con creces sus estrechos límites originales y puede afirmarse que, en la actualidad, es una de las técnicas más potentes para el estudio y el gobierno de una gran variedad de sistemas.

El nombre *programación* responde a un paralelismo con la técnica más conocida de la investigación de operaciones: la programación lineal. La palabra *programación*, en este contexto, equivale al análisis matemático para hallar la decisión óptima ante una situación determinada. Por tanto, la programación dinámica es una herramienta diseñada esencialmente para la toma de decisiones, pero con dos importantes diferencias:

a) Al revés de lo que ocurre con la programación lineal, que solo considera una única decisión, aquí se puede definir una “regla de decisión condicional”, es decir, unas instrucciones de lo que hay que hacer si en un momento dado nos hallamos en una situación determinada: es lo que se llama una “política”.

b) Por otra parte, las expresiones que relacionan las diferentes variables pueden ser de cualquier naturaleza matemática, mientras que la programación lineal solo admite expresiones lineales.

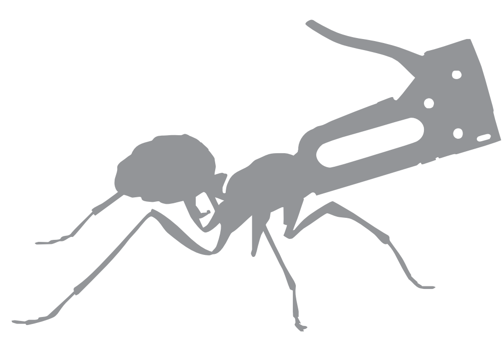
El adjetivo *dinámica* obedece a que originariamente fue diseñada para estudiar el comportamiento de sistemas que evolucionan a lo largo del tiempo. Así pues, la programación dinámica no se enfrenta a una decisión única en que hay que obtener todo el rendimiento posible de la situación actual, sino que la política debe tener en cuenta cómo va a quedar el sistema según lo que hagamos hoy y, por tanto, cómo influirá nuestra decisión actual sobre los rendimientos futuros. Es una situación que el lenguaje popular resume bien con la expresión “pan para hoy, hambre para mañana”. No conviene tanto obtener un buen resultado ahora como alcanzar un resultado global óptimo en el tiempo que va desde hoy hasta al final de la vida del sistema, es decir, en el llamado “horizonte” del problema.

Aunque nació como una técnica independiente, la programación dinámica forma parte de un conjunto de modelos muy general, denominados *modelos polietápicos*, caracterizados por el hecho de que toman en consideración la evolución del sistema objeto de estudio a lo largo del tiempo. Dentro de esta colección, se incluyen otros dos volúmenes dedicados a otros tantos modelos de esta clase:

– La teoría de colas

– Las cadenas de Markov

Los tres modelos comparten determinadas características y, al mismo tiempo, presentan diferencias, algunas notables, por lo cual el presente volumen incluye la sección 2.3, dedicada precisamente a señalar estos paralelismos y las diferencias entre ellos.



→ 1

Introducción a los métodos cuantitativos

Este es un volumen de una serie en que se introducen las técnicas y las aplicaciones de los métodos cuantitativos para la toma de decisiones en el contexto de la organización industrial.

Los volúmenes se corresponden con las diversas asignaturas incluidas en el plan de estudios de las titulaciones impartidas en la Escuela Superior de Ingenierías Industrial, Aeroespacial y Audiovisual de Terrassa (ESEIAAT) sobre métodos cuantitativos para la toma de decisiones.

En este capítulo, se introducen la naturaleza y las especificidades de los métodos cuantitativos en la organización industrial. Esta introducción sigue una lógica de modelos, problemas y técnicas:

a) En muchas ocasiones, en el contexto organizativo nos encontramos (generalmente, en la gestión del sistema productivo, pero también en otros contextos) con situaciones de una cierta complejidad, que plantean problemas susceptibles de ser resueltos elaborando un modelo cuantitativo. Conviene, pues, tener una noción clara de qué es un *modelo* y, en especial, un *modelo matemático o cuantitativo*.

b) Finalmente, necesitamos un conjunto de *técnicas* que nos permitan resolver el modelo, es decir, obtener a un coste asequible y con una precisión aceptable la información que buscábamos al elaborar dicho modelo. Con frecuencia, estas técnicas están relacionadas con problemas tipo, de ahí la importancia de estos.

En definitiva, en los cursos de métodos cuantitativos para la toma de decisiones, se persigue que los estudiantes sean capaces de formular modelos matemáticos a partir de situaciones poco estructuradas. Dichas situaciones pueden tener una cierta relación con problemas tipo a los cuales pueda aplicarse un conjunto bien definido de técnicas para hallar la solución del modelo.

1.1 Concepto de modelo

Una definición clásica de modelo es:

Un objeto M es un modelo de una realidad R para un observador O, si O puede obtener, estudiando M, las respuestas a las preguntas que se hace sobre R.

Así, una misma realidad puede ser representada por diferentes modelos, en función de las preguntas que nos hagamos acerca de esta. Una red eléctrica, por ejemplo, puede ser representada por:

– Un grafo cuyos arcos tienen asociada la distancia entre diferentes puntos, si queremos determinar cuál es el conjunto de conexiones que minimiza la longitud del cable a emplear.
– Un mapa del terreno a escala, para llevar a cabo la obra civil del tendido eléctrico. El mapa tiene más información que el grafo: además de estar a escala, incluye accidentes orográficos (ríos, etc.), curvas de nivel, etc.
– Una red de Kirchhoff, a fin de determinar las intensidades de la línea, pérdidas de energía por transporte, etc.

En definitiva, el *modelo* viene determinado por el *problema* a resolver y por los requerimientos de la *técnica* escogida para resolverlo.

El proceso de construcción y resolución de un modelo puede dividirse en tres partes:

a) Modelización: construcción y elaboración del modelo. Generalmente, se trata de un proceso difícilmente sistematizable, puesto que podemos tener situaciones muy diversas que admiten modelos muy similares (es el caso, por ejemplo, de la programación lineal).

b) Resolución. Diremos que hemos resuelto el modelo cuando hayamos podido responder a las preguntas que nos movieron a elaborarlo. Es decir, cuando hayamos obtenido la información que necesitábamos.

c) Explotación. Una vez obtenidos los resultados, estos han de ser interpretados y se deben analizar las implicaciones para la gestión del sistema afectado. Otra cuestión importante es el mantenimiento del modelo: ver cómo evoluciona la solución cuando los parámetros del sistema evolucionan.

1.2 Tipos de problemas

Los métodos cuantitativos para la toma de decisiones centran su interés en un conjunto de problemas concretos, asociados generalmente a la organización de las interrelaciones entre elementos del sistema que permiten optimizar su comportamiento. En muchas ocasiones, dicha optimización supone la minimización de los recursos empleados o del valor (coste) de esos recursos. En otras, buscaremos la maximización de la utilidad asociada al uso del sistema. Dicha utilidad suele ir asociada, en el contexto empresarial, a la maximización del beneficio que se deriva de la actividad. A continuación, mostramos algunos problemas prototipo. Para cada uno de ellos, definimos tres aspectos:

a) Las *características* del problema.

b) El *objetivo* del problema, esto es, la magnitud a optimizar.

c) Las posibles *variables de decisión*, es decir, las características del sistema que podemos controlar. En general, el valor de la magnitud a optimizar es función de dichas variables de decisión, además de unos *parámetros* sobre los cuales no tendremos control.

1.2.1 Problemas de inventarios (stocks)

Los problemas de inventarios aparecen cuando, por determinadas razones, los flujos de entrada de un recurso determinado en un sistema son diferentes de los flujos de salida. Este hecho exige que dispongamos de una cantidad determinada de dicho recurso, sin más finalidad que ajustar ambos flujos. Esta cantidad son las existencias o *stocks* del sistema.

Un sistema de este tipo tiene un conjunto de costes asociados a la gestión del sistema:

– *Coste de adquisición*. Es posible que el recurso tenga diferentes costes a lo largo del tiempo y que sea interesante adquirir una gran cantidad de recursos en un momento determinado, para su utilización cuando el coste sea elevado. Esta adquisición aumentará el valor de las existencias. En ocasiones, existe un componente de coste independiente de la cantidad adquirida, que suele denominarse *coste de lanzamiento de un pedido*.

– *Coste de mantenimiento o posesión*. Mantener el recurso en *stock* supone unos costes de almacenamiento, financiación, etc., que tienden a limitar el valor del inventario.

– *Coste de rotura*. Puede darse el caso de que, con nuestra programación, no podamos servir el flujo de demanda y provoquemos una falta de suministro. Entonces, tendremos unos costes de rotura de *stocks*. Si estos costes son elevados, podrán hacer aumentar el valor del inventario.

El objetivo del problema será minimizar la suma total de costes de adquisición, mantenimiento y rotura asociados a un determinado flujo de salida. Las variables de decisión serán las cantidades a adquirir en cada instante de tiempo, así como el nivel de inventario entre cada entrada y salida.

La gestión de materiales en una empresa industrial es el ejemplo más evidente de un problema de *stocks*, aunque pueden encontrarse formulaciones similares en otros ámbitos:

- *Gestión de tesorería,* en que el inventario es el saldo de la cuenta bancaria en cada momento.
- *Gestión de un recurso*, por ejemplo, un embalse. En este caso, el inventario es el nivel de agua entre la entrada y la salida del embalse.

1.2.2 Problemas de reparto

Las características generales de un problema de reparto son que disponemos de un conjunto de recursos, con lo cual se realiza un determinado conjunto de actividades. Los parámetros del sistema son:

a) La cantidad de que se dispone de cada recurso.

b) La cantidad de cada recurso necesaria para producir una unidad de cada una de las actividades.

c) El rendimiento o utilidad que se deriva de producir una unidad de cada una de las actividades.

El objetivo es determinar cómo se asignan los recursos a las actividades, de manera que la utilidad o rendimiento total sea máximo.

Una variante de este problema consiste en plantear una situación en que, además de la cantidad máxima de recursos, se deba realizar una cantidad mínima de cada una de las actividades. En este caso, podemos plantearnos determinar si es posible servir esta cantidad mínima o no, encontrar qué actividades deben realizarse primero, etc.

Las variables de decisión serán las cantidades de actividades a realizar, o bien los recursos consumidos en cada actividad. También podemos plantear como variable de decisión qué actividades pueden realizarse y cuáles no.

Algunos problemas asimilables a un problema de reparto son:

- Determinar desde qué almacén debe servirse cada punto de consumo y en qué cantidad, para minimizar los costes de transporte.
- Los problemas de horarios: por ejemplo, qué profesores deben asignarse a cada asignatura.

1.2.3 Problemas de secuencias

En un problema de secuencias, tenemos un conjunto de tareas a realizar, relacionadas entre ellas por algunas restricciones. Por ejemplo, una tarea determinada no puede realizarse hasta que se hayan finalizado otras dos tareas previas. En ocasiones, dichas tareas consumen algunos recursos, de los cuales se dispone en una cantidad limitada.

El objetivo es determinar el momento en que debe iniciarse y acabarse cada tarea, así como el tiempo total de realización de dichas tareas. También puede que sea deseable conocer qué actividades pueden retrasarse sin retrasar la fecha de entrega del conjunto y cuáles no, y la asignación de los recursos limitados a las tareas.

Los problemas de secuencias suelen ir ligados a actividades no repetitivas, como el lanzamiento de un producto al mercado o la construcción de un edificio.

1.2.4 Problemas de colas

Un problema de colas puede formularse del siguiente modo: determinadas unidades llegan a un sistema de forma aleatoria, según una *ley de llegada* conocida. Dichas unidades reciben un servicio determinado en uno o varios servidores. El tiempo de servicio sigue también una determinada *ley de servicio*. Además, el sistema puede tener varios componentes de coste: por ejemplo, el *coste de servicio* (coste asociado al hecho de disponer de un servidor) y el *coste de espera* (coste asociado a esperar a ser servido). Los problemas de colas están condicionados por el carácter aleatorio de las llegadas y del tiempo de servicio.

Lo que se pretende obtener son parámetros asociados al sistema:

a) Tiempo medio de espera

b) Tiempo medio en el sistema (tiempo de espera, más tiempo de servicio)

c) Número medio de unidades en espera

d) Número medio de unidades en el sistema

También pueden plantearse cuestiones relativas al diseño del sistema, como el número de servidores que minimiza la suma de costes de servicio más costes de espera.

Podemos encontrar problemas de colas en cualquier situación en que se produzcan esperas (colas en el supermercado o en el aeropuerto, proceso de matriculación en una universidad, etc.). De forma menos evidente, se pueden encontrar problemas de colas en el diseño de sistemas de mantenimiento, en la asignación de máquinas a operarios, etc.

1.2.5 Problemas de renovación

En un problema de renovación, existe un conjunto de elementos que envejece. Ello puede incrementar los costes de funcionamiento o de reparación, y aumentar las probabilidades de avería. Por otra parte, sustituir el elemento supone un coste, que puede ser superior al de seguir utilizando el elemento en un momento determinado.

El objetivo no es otro que conocer cuándo debe reemplazarse el equipo. En una situación de largo plazo, puede interesar conocer la política de decisión en cada momento.

Los problemas de renovación se encuentran en cualquier sistema cuyos elementos envejezcan y en que el problema de la fiabilidad del conjunto del sistema sea relevante.

1.2.6 Problemas de caminos

Los problemas de caminos son propios de situaciones en que se puede ir de un estado inicial a otro final de varias maneras. Se trata de encontrar cuál de estas es mejor para minimizar el tiempo, los costes o los recursos.

El objetivo es encontrar la política más indicada para hallar el camino más corto entre el inicio y el final del recorrido. Las variables de decisión deben considerar las acciones a realizar, y han de formar parte del camino más corto.

Los problemas de caminos pueden darse en situaciones de encontrar el camino más corto entre dos puntos, o la forma más económica de realizar una actividad con un inicio y un final definidos. Algunos problemas de secuencias pueden formularse en términos de problema de caminos.

1.2.7 Problemas de competencia

Los problemas de competencia estudian situaciones en que diversos actores toman decisiones, de manera que la decisión de un actor afecta los resultados obtenidos por los otros, además de él mismo. Partiendo de una hipótesis determinada, relativa al comportamiento de los actores (usualmente, que se comporten de manera racional), se modeliza el sistema para determinar los resultados alcanzados por cada actor.

El propósito de estos modelos es determinar la *estrategia* óptima para los jugadores, definida como las características de las acciones que cada jugador ha de llevar a cabo para optimizar su utilidad. Estas estrategias son las variables de decisión del modelo.

Cualquier situación en que se dé la interdependencia descrita entre los actores es susceptible de ser analizada como un problema de competencia. Entre los ejemplos más conocidos, podemos citar las decisiones de las empresas de un sector, una subasta o una situación con votaciones sucesivas.

1.2.8 Problemas de búsqueda

Los problemas de búsqueda consisten en situaciones en que, para acceder a una información determinada, se ha de incurrir en un cierto coste de búsqueda. Pueden existir, incluso, situaciones en que no haya la seguridad de que la información exista o esté disponible.

El objetivo de los problemas de búsqueda es determinar una forma de buscar esta información que minimice los costes de búsqueda, y cuándo hay que detener la búsqueda considerando los costes en que se incurre y los beneficios obtenidos con el hallazgo. Una variante de este problema es determinar cómo clasificar la información para facilitar su recuperación con una búsqueda posterior, incurriendo en los mínimos costes de búsqueda.

La variable de decisión es la política de búsqueda, que ha de incluir las reglas para detener la búsqueda. En la segunda versión del problema, las variables definirán la política de clasificación relacionada con la política de búsqueda.

Son ejemplos de situaciones en que se presentan problemas de búsqueda el control de calidad (la búsqueda de defectos), la búsqueda de yacimientos de petróleo, la prospección arqueológica o la búsqueda de información en internet, así como la determinación de las reglas de clasificación y archivo.

1.3 Métodos cuantitativos de tratamiento de los problemas

Los métodos cuantitativos disponibles para resolver los problemas descritos en el apartado anterior pueden clasificarse en dos grandes categorías:

a) *Métodos exactos*. Se caracterizan por que aseguran encontrar la solución óptima, si existe. En ocasiones, hallar esta solución óptima puede suponer costes elevados, en términos de tiempo de cálculo.

b) *Métodos no exactos*. Tienen en común que no aseguran la solución óptima, sino *una solución razonablemente buena en un tiempo razonable*. Existe una cierta probabilidad de encontrar efectivamente el óptimo, pero no podremos asegurar si la solución obtenida es óptima, solo sabremos que es una buena solución.

1.3.1 Métodos exactos

Optimización clásica (análisis matemático)

Engloba diversos procedimientos analíticos que tienen en común la idea de encontrar la función derivada e igualarla a cero para obtener los puntos extremos. Suele requerirse que la función que se quiere optimizar sea continua y derivable. Algunos desarrollos matemáticos, como los multiplicadores de Lagrange, permiten resolver de esta manera problemas en que las variables de decisión están sometidas a determinadas restricciones.

Programación matemática

Los programas matemáticos son formulaciones del tipo:

$$[OPT]\ z = f(x),$$
$$\text{donde } x \in E \subset R^n$$

en que *f(x)* es la *función objetivo* y E es la *región factible*, subconjunto de R^n de soluciones posibles del problema.

El programa matemático será más o menos difícil de resolver según las características de la función objetivo y de la región factible. En este sentido, existen algunos casos especialmente relevantes:

a) *Programación lineal*. La función *f(x)* es lineal y E está determinada por un conjunto de restricciones lineales.

b) *Programación lineal entera, binaria o mixta*. Se trata de programación lineal en que los componentes de *x* son variables enteras (programación lineal entera), binarias (programación lineal binaria), o bien algunas variables son enteras o binarias y otras reales (programación lineal mixta).

c) *Programación no lineal*. Programa matemático que no cumple las condiciones de la programación lineal, tal como se acaba de definir. Los modelos de programación lineal pueden ser de muy difícil resolución, aunque los dos casos que se indican a continuación son especialmente interesantes para la organización industrial:

   – Programación cuadrática: f(x) es un polinomio de segundo grado y las restricciones, lineales o de segundo grado.

   – Programación semilineal: f(x) es un polinomio de grado n y las restricciones son lineales.

La programación lineal es una de las técnicas más utilizadas para resolver los problemas propios de los métodos cuantitativos en el contexto de la organización industrial y se desarrolla en profundidad en este volumen de la colección.

Teoría de grafos

Esta teoría trata de unos objetos matemáticos denominados *grafos*, consistentes en un conjunto de elementos (vértices del grafo) y las relaciones entre ellos (que pueden ser aristas o arcos del grafo).

Programación dinámica

La programación dinámica es una metodología para encontrar políticas óptimas en procesos polietápicos de decisión. Se trata de sistemas que evolucionan en varias etapas, en cada una de las cuales hay que tomar una decisión, que hará que el sistema evolucione a un estado determinado, de manera determinista o siguiendo una distribución de probabilidad.

El objetivo es determinar qué política debemos seguir (decisión a tomar si nos encontramos en un estado determinado, es decir, en una etapa determinada) para optimizar la función objetivo. Dicha función se caracteriza por ser recursiva (eso es, depende de la decisión actual y del estado al cual el sistema puede evolucionar en la etapa siguiente).

Cadenas de Markov

Una cadena de Markov es un sistema que evoluciona a lo largo del tiempo dentro de un conjunto de estados. Se caracteriza por que la probabilidad de que el sistema evolucione a un estado determinado depende exclusivamente del estado en que se encuentra en el momento presente.

Teoría de colas

Se trata de modelos desarrollados específicamente para analizar problemas de colas. Consiste, fundamentalmente, en un conjunto de modelos descriptivos de diversas situaciones, relativas a las leyes de llegada y servicio, y otras características propias de estos sistemas.

Son de especial interés, por su (relativa) sencillez conceptual, los modelos de cola en que las tasas de llegada y de servicio siguen una ley de Poisson.

Teoría de decisión

Se trata de procedimientos que permiten valorar las diferentes alternativas y los posibles modelos de experimentos a realizar para la toma de decisiones en situaciones que pueden describirse mediante un modelo matemático.

Teoría de juegos

Se trata de modelos que estudian, de forma específica, problemas de competencia. Se diferencia de la teoría de la decisión por la existencia de varios actores y por la interdependencia entre las acciones de cada uno de ellos.

Procedimiento de separación y acotación

Es un procedimiento para obtener la solución de problemas en que el conjunto de soluciones es finito. La táctica adoptada consiste, en primer lugar, en dividir el conjunto en dos partes, siguiendo una *regla de separación*. A continuación, se acota el valor de la función objetivo en cada una de las partes, mediante una *regla de acotamiento*. Así, exploramos el conjunto de soluciones cada vez menor hasta encontrar el óptimo.

1.3.2 Métodos no exactos

Métodos heurísticos

Un método heurístico genera soluciones de un determinado problema mediante un método del cual se sabe, bien por experiencia o por razonamiento teórico, que genera *buenas soluciones* con una alta probabilidad. Suele clasificarse en tres grupos:

a) *Algoritmos de un solo paso*. Son métodos que generan una única solución en cada etapa, tomando decisiones sucesivas que ya no son reconsideradas. Cada vez hay menos alternativas y estas están más condicionadas, de forma que las últimas decisiones pueden ser muy malas.

b) *Métodos iterativos*. Son métodos que generan una solución en cada etapa, de manera que puedan reconsiderarse las soluciones obtenidas en etapas anteriores.

c) *Métodos de mejora*. Son métodos que, partiendo de una solución determinada, van mejorándola en las etapas sucesivas.

Simulación

Proceso en que se representa el estado del sistema mediante variables relacionadas por unas reglas determinadas. Una vez establecidas, se observa su evolución a lo largo del tiempo según unas hipótesis y unas reglas de gestión predeterminadas. Permite conocer la solución de problemas difícilmente resolubles mediante métodos exactos.

Modelos descriptivos

Modelos que describen o reproducen, de una manera más o menos simplificada, la realidad que hay que estudiar, con lo cual permiten experimentar y estudiar sus reacciones ante determinadas decisiones, incidencias o políticas.

Algoritmos genéticos

Se trata de procedimientos basados en la selección natural, en que una generación de individuos (conjunto de soluciones) da lugar, mediante el entrecruzamiento entre ellas, a una nueva generación de soluciones que, por término medio, es mejor desde el punto de vista de la calidad del resultado obtenido.

Redes neuronales

Sistemas basados en la simulación de las conexiones neuronales de los seres vivos, capaces de aprender de la experiencia y obtener soluciones cada vez mejores a problemas similares.

Existen varias relaciones entre las diversas técnicas. Por ejemplo, la programación lineal entera utiliza procedimientos de separación y acotación; determinados problemas de programación dinámica utilizan propiedades de las cadenas de Markov, y la teoría de decisión utiliza técnicas de programación lineal, de la teoría de grafos y de simulación, etc.

1.4 Relación entre problemas y técnicas

No existe una relación directa entre los tipos de problemas y las técnicas utilizadas para resolverlos. Además, la mayoría de los problemas no pueden clasificarse de manera unívoca y las técnicas se pueden combinar. La tabla 1 es una aproximación para hallar relaciones entre los problemas (columnas) y las técnicas (filas). Se ha utilizado la notación siguiente:

P: Técnica principal para resolver el problema

S: Técnicas empleadas de forma secundaria

A: Técnicas auxiliares o utilizadas en casos especiales

Tabla 1.1.
Problemas y técnicas de métodos cuantitativos para la toma de decisiones

	Inventarios	Reparto	Colas	Secuencias	Renovación	Caminos	Competencia	Búsqueda
Opt. clásica	P		A		P		S	S
Prog. lineal	S	P		A	A	S	P	A
Prog. no lin.	S	S						
P. dinámica	P	S		S	S	S		P
T. grafos			A	P	S	P	A	A
T. colas			P	S				
P. sep. y acot.		A		P	P	P		A
T. decisión	S		A		S		P	P
T. juegos							P	
Cad. Markov	S	A	S					S
M. heurísticos	S	S		S	S	S	A	A
Simulación	A	S	P	S	P		S	S
M. descript.	A			S			P	P
Alg. genéticos		A		S		P		P
R. neuronales				S	S	P	P	S

1.5 Notación y conocimientos previos

En las expresiones matemáticas, se ha seguido la notación siguiente:

– Las variables escalares se han representado en cursiva, en mayúscula o minúscula, según convenga. Por ejemplo: *n, k, p, N*.

– Los vectores se representan en minúscula. Si no se indica lo contrario, se trata de vectores columna. Por ejemplo: v, g.

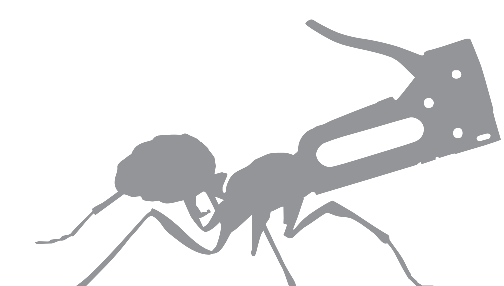
– Las matrices se representan en mayúscula. Por ejemplo: P, A.

En cuanto a los conocimientos previos, se recomienda revisar las materias siguientes:

– Distribuciones de probabilidad, especialmente las propiedades de las distribuciones exponencial y de Poisson.

– Cálculo matricial: producto de matrices, y significado de los valores y vectores propios.

– Obtención de raíces complejas de un número real.



→2

Procesos polietápicos

2.1 Definición

Los denominados *procesos polietápicos* son una clase muy general de modelos matemáticos, que responden a las características siguientes (v. figura 2.1):

1) El sistema representado por el modelo se halla siempre en uno y solo uno de un conjunto de estados: X.

2) El sistema cambia de estado en unos instantes del tiempo, denominados *transiciones* o *etapas*.

3) En cada transición o etapa, puede existir la posibilidad de influir en el comportamiento del sistema mediante una decisión (u_t).

4) Igualmente, en cada transición o etapa, puede existir la posibilidad de obtener una remuneración o rendimiento (r_t). Este nombre no se debe prejuzgar acerca de su naturaleza, ya que tanto puede representar beneficios como costes.

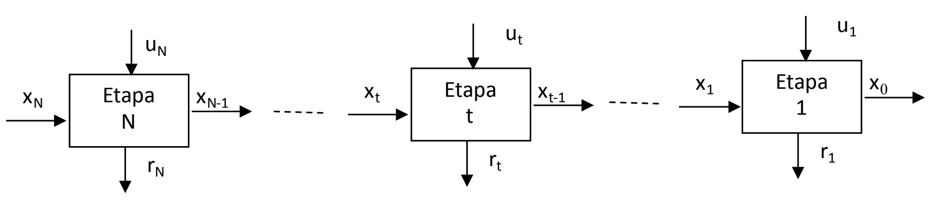


Figura 2.1.
Procesos polietápicos

Algunas observaciones sobre la definición:

- La variable que cuenta el número de etapas (t) recibe el nombre de *variable cronológica*, porque usualmente (pero no siempre) tiene un significado temporal.
- Aunque no es estrictamente obligado, hemos optado por definir esta variable como el número de etapas que faltan para acabar el proceso.
- El cambio de estado del sistema en una transición puede consistir en mantenerse en el mismo estado:

$$x_{t-1} = x_t$$

- La existencia de remuneración es opcional. Pueden existir sistemas sin remuneración.
- En los sistemas con remuneración, usualmente se define un rendimiento total que es función de los rendimientos de cada etapa:

  $$R = R(r_1, r_2, \ldots , r_N)$$

  Esta función R normalmente es la suma, pero puede ser otra función cualquiera monótona creciente con todas las r_t.

- La existencia de una decisión también es opcional. Dicha decisión implica que se desea optimizar el rendimiento total, R, que recibe entonces el nombre de función objetivo o función económica. El sentido de la optimización puede ser de máximo o de mínimo, según el significado del rendimiento.
- La decisión pretende optimizar el rendimiento total R, lo cual puede conllevar que, en determinados casos, la decisión óptima produzca un rendimiento muy pobre en la etapa donde se toma (r_t), pero ello implica dejar el sistema en condiciones mejores para los rendimientos futuros.
- Dado que solo tiene sentido plantear una decisión si hay una función económica, puede haber sistemas con rendimiento y decisión, con rendimiento sin decisión o sin ninguna de las dos cosas, pero no puede haber sistemas con decisión y sin rendimiento.

El número de etapas N definido por la variable cronológica se denomina *horizonte* y puede ser:

a) Continuo. La variable cronológica es continua y, por tanto, las transiciones pueden ocurrir en cualquier momento.

b) Discreto. La variable cronológica es discreta, de modo que las transiciones ocurren en instantes determinados del tiempo (aunque no necesariamente han de estar uniformemente distanciados). En este caso, distinguimos entre:

b.1. Finito. El número de etapas (N) es conocido y finito.

b.2. Indeterminado. El número de etapas (N) es finito pero desconocido.

b.3. Infinito. El número de etapas (N) tiende a infinito numerable.

El caso b3 es, evidentemente, una abstracción; sin embargo, puesto que el comportamiento del sistema en general se estabiliza (bajo determinadas condiciones) al cabo de un cierto número de etapas, se puede plantear como un límite del horizonte finito. Del mismo modo, el caso b2 se tratará como el caso b3 ya que, si se desconoce el número de etapas que faltan, el caso debe tratarse como si el horizonte fuese infinito. De hecho, aunque nos hallemos ante el caso finito, si N es lo bastante grande, se tratará como infinito. En realidad, el caso b1 solamente tiene interés para estudiar el régimen transitorio, o bien cuando el horizonte del problema es lo bastante corto para que no dé tiempo a alcanzar el régimen permanente.

Tanto si el tiempo es continuo como si es discreto, se pueden distinguir dos situaciones diferentes y consecutivas:

- El régimen transitorio (también denominado *de corto plazo*), durante el cual cada etapa es distinta en función de las condiciones iniciales; es decir, el sistema conserva la memoria de la situación de partida, la cual influye en la decisión.
- El régimen permanente (o *de largo plazo*), en que el sistema ya ha alcanzado la estabilidad (aunque puede seguir conservado parte de la información inicial).

El comportamiento del sistema viene definido por dos distribuciones de probabilidad que, en general, pueden definirse así:

- Probabilidad condicional de cambio de estado:

$$Pr_t \{x_{t-1} \mid x_t, x_{t+1}, \ldots, x_{t+k-1}; u_t\}$$

- Probabilidad del rendimiento:

$$Pr_t \{r_t \mid x_t, x_{t+1}; u_t\}$$

La probabilidad condicional del cambio de estado indica que la probabilidad del nuevo estado depende de los k últimos estados. Si el tiempo es continuo, solo es posible $k = 1$; si el tiempo es discreto, puede darse cualquier valor de k. De la misma manera, el rendimiento obtenido en una etapa (si existe) depende de la decisión (si existe) y de los estados anterior y posterior. En este volumen, por las características de la programación dinámica, solo estudiaremos el caso $k = 1$ con decisión.

Como se ha definido, el sistema es aleatorio, pero engloba como caso particular el determinista, ya que basta con considerar que la distribución de probabilidad tiene un valor 1 para una alternativa y el resto son ceros.

Las probabilidades se han indicado con el subíndice t porque pueden ser diferentes para cada etapa, pero ello no necesariamente es así:

Decimos que el sistema es homogéneo en el tiempo (HT) si todas la probabilidades, tanto del nuevo estado como del rendimiento, son iguales para toda t; entonces, se prescinde del subíndice y se escribe Pr{} en lugar de Pr_t{}.

Por el contrario, si el sistema no cumple la condición anterior, decimos que es no homogéneo en el tiempo (NHT).

Dado que en los sistemas NHT debe conocerse el comportamiento de cada etapa, el problema con horizonte infinito o indeterminado solo tiene sentido en el caso HT.

Cuando existe decisión, una política es una regla de decisión que, para cada estado del sistema y para cada etapa, indica qué acción debe elegirse. Obsérvese que fijar una política equivale a convertir un sistema con decisión en uno sin ella. Dicho de otro modo, una vez determinada la política, puede dejarse el control del sistema a un dispositivo automático que la aplique.

Evidentemente, no todas las políticas proporcionan el mismo rendimiento. La política que da un resultado óptimo en un horizonte determinado se denomina *política óptima*.

Normalmente, la optimalidad de una política depende del horizonte que se ha planteado. Cuando el horizonte es finito (N transiciones) existen N políticas óptimas, una para cada etapa; dicho de otro modo, no es lo mismo la política óptima cuando quedan 3 etapas para terminar que cuando quedan 2 o 1. La política óptima depende de la variable cronológica, y entonces se habla de *política óptima a corto plazo*.

En cambio, si se trata de un problema con horizonte infinito o ilimitado, la política óptima no depende de la variable cronológica y se habla de *política óptima a largo plazo*.

2.2 Comparación con la programación lineal

La programación lineal, a la cual se dedica un volumen de esta colección, es probablemente el modelo cuantitativo más conocido en general y, por ello, es interesante compararla con los procesos polietápicos, dado que sus características son muy diferentes y permiten, en conjunto, modelizar una gran cantidad de situaciones distintas. La tabla 2.1 describe las principales diferencias entre los procesos polietápicos y la programación lineal.

	Programación lineal	Procesos polietápicos
Comportamiento del sistema	Determinista	Aleatorio
Tipo de sistema	Estático	Dinámico
Tipo de decisión	Única	Múltiple / no existe
Consideración del tiempo	Ninguna	Discreto / continuo
Distinción corto/largo plazo	No	Sí

Tabla 2.1. Comparación entre la programación lineal y los procesos polietápicos

2.3 Modelos polietápicos existentes

Existen tres tipos de modelos polietápicos a los cuales se dedican tres volúmenes de esta colección:

- Teoría de colas
- Cadenas de Markov
- Programación dinámica

La *teoría de colas* es un modelo polietápico específico para una situación muy concreta, que presenta las características siguientes:

- El tiempo es continuo.
- No existe decisión en cada etapa; únicamente podemos observar el comportamiento del sistema y, por tanto, solo existe la decisión previa del diseño del sistema en función de los resultados que se pretende obtener.
- El estado del sistema se caracteriza por el número de elementos que se hallan en el sistema en cada momento, ya sea esperando a ser servidos o recibiendo un servicio.
- Existe una remuneración, que es el tiempo que cada elemento pasa en el sistema.

Se trata, pues, de un modelo poco general, en el sentido de que solo puede aplicarse a un tipo de problema muy concreto. Sin embargo, su frecuencia de aparición en la realidad es tan grande que vale la pena tomarlo en consideración.

Las *cadenas de Markov* (CM) y la *programación dinámica* (PD) son dos modelos de investigación operativa, de orígenes muy distintos, que responden a conceptos teóricos muy diferentes. Sin embargo, ambos se pueden considerar casos particulares de los procesos polietápicos, tal como se han definido. Por ello, será útil tener en cuenta sus paralelismos y diferencias dentro de este marco general, a fin de facilitar la comprensión de los mismos.

Un efecto de este planteamiento conjunto es que determinados conceptos tienen dos nombres, ya que proceden de las terminologías de la PD y las CM. Como simple indicación, puede establecerse que los términos *transición* y *remuneración* proceden de la literatura sobre CM, mientras que *etapa* y *rendimiento* proceden de la de PD.

Existen tres tipos de cadenas de Markov:

- Las cadenas de Markov simples (CM), sin remuneración ni decisión
- Las cadenas de Markov con remuneración (CMR) pero sin decisión
- Las cadenas de Markov con remuneración y decisión (CMRD)

La relación con el marco general descrito se puede establecer así:

Tabla 2.2. Relaciones entre la PD y las CM

	Con decisión	**Sin decisión**
HT	CMRD, PD	CM, CMR
NHT	PD	(PD)

Así pues, las cadenas de Markov tratan con los sistemas homogéneos en el tiempo (HT), tengan o no decisión o rendimiento, mientras que la programación dinámica exige que exista decisión (y, por tanto, rendimiento), pero admite la posibilidad de sistemas NHT.

De este modo, los problemas HT con decisión pueden tratarse de las dos maneras ya que, como se verá si se comparan las exposiciones en ambos volúmenes, se trata poco más que de un cambio de notación.

Los sistemas NHT sin decisión no entran en el esquema anterior, por lo cual no se abordan explícitamente en esta colección. Se trata de situaciones relativamente poco frecuentes y que no suelen tener gran interés; sin embargo, puede extenderse fácilmente la programación dinámica para tratarlos, si es necesario: basta con considerar que se trata de una decisión con tan solo una alternativa.

2.4 Introducción al cálculo recurrente

En este apartado, hacemos una breve introducción al cálculo recurrente, que es una herramienta fundamental en el tratamiento de los modelos polietápicos, mediante un ejemplo en que no hay decisión.

El problema es el siguiente: Por una carretera muy estrecha donde no se puede avanzar, circulan *n* coches, cada uno de los cuales tiene una velocidad propia, v_i, a la cual circularía si no tuviera nadie delante que se lo impidiera.

Cuando se encuentra con un coche más lento, como no se puede avanzar, tiene que adaptar su velocidad a la de este. Ello quiere decir que se forman "paquetes" de coches, en cada uno de los cuales el primero es el más lento del paquete. Dado que el orden de los coches y sus velocidades propias son magnitudes no correlacionadas (por tanto, estadísticamente independientes) el número de paquetes que se formarán es una magnitud aleatoria y se trata de saber el número esperado de paquetes que se formarán a largo plazo (es decir, en una carretera infinita).

Consideramos el más lento de los coches, es decir, el de menor v_i. Detrás de él, se forma un único paquete, y, si se halla en la posición k, delante suyo habrá k-1 coches, que formarán $p(k\text{-}1)$ paquetes.

Por tanto:

$$p(n / k) = 1 + p(k-1)$$

si el más lento va en posición k. Como el orden es aleatorio:

$$p(n) = \frac{1}{n}\sum_{k=0}^{n}\left[1 + p(k-1)\right]$$

con $p(0) = 0$, evidentemente. Esta expresión permite el cálculo recurrente de $p(n)$. Por tanto, el problema está resuelto, pero con una expresión final no muy satisfactoria.

Consideremos ahora el más rápido. Salvo que vaya el primero, nunca formará un paquete, sino que se añadirá al del delante.

Por tanto:

$$p(n / 1) = 1 + p(n-1)$$
$$p(n/ \neq 1) = p(n-1)$$

Entonces:

$$p(n) = \frac{1}{n}\left[1 + p(n-1)\right] + \frac{n-1}{n}p(n-1) = \frac{1}{n} + p(n-1)$$

Y, dado que $p(0) = 0$, puede deducirse que

$$p(n) = \sum_{k=1}^{n}\frac{1}{n}$$

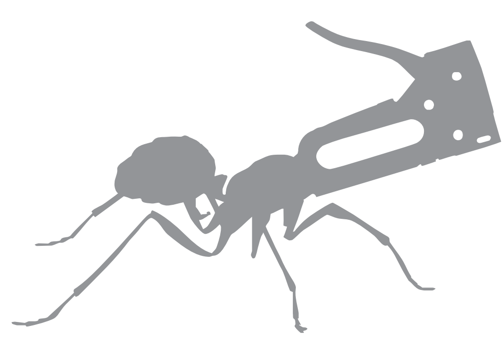
lo cual es una fórmula directa, al tiempo que permite también el cálculo recurrente.

Este problema posibilita ver dos aspectos interesantes:

a) Los fundamentos del cálculo recurrente y la forma de plantear los problemas de este tipo.

b) Cómo uno u otro planteamiento, aunque matemáticamente equivalentes, pueden dar soluciones que permitan calcular los resultados más o menos fácilmente.

Estos son dos detalles a tener muy en cuenta a la hora de utilizar los modelos, tanto en lo referente a los modelos basados en las cadenas de Markov como a los basados en la programación dinámica.



→ 3

Programación dinámica determinista

3.1 Fundamentos teóricos

La programación dinámica se basa en los conceptos teóricos siguientes:

- El problema puede dividirse en dos o más subproblemas más sencillos.
- Dicho proceso puede repetirse tantas veces como sea necesario hasta llegar a un conjunto de problemas triviales.
- Se aplica el principio del óptimo: "Dada una secuencia óptima de decisiones, toda subsecuencia de ella es, a su vez, óptima."

Este principio puede explicarse de manera más simple diciendo que, si el camino óptimo entre A y B pasa por C, los tramos AC y CB son los caminos óptimos entre estas parejas de puntos. Veámoslo con un ejemplo.

3.1.1 Camino mínimo en un grafo

La figura 3.1 representa un mapa simplificado de las carreteras existentes en una zona determinada y, los números que acompañan cada tramo, el tiempo en minutos necesario para recorrerlo. Deseamos ir de A a I por el camino más rápido.

Este "mapa" es, en realidad, un grafo que puede representarse mediante su matriz asociada que, puesto que las carreteras pueden recorrerse en ambas direcciones, será simétrica (ver tabla 3.1).

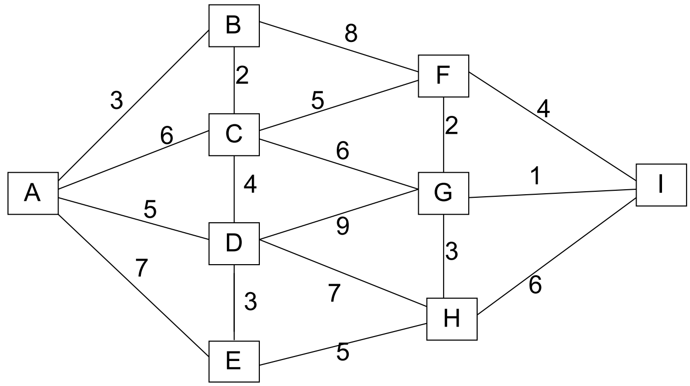


Figura 3.1.
Mapa de carreteras y tiempo necesario para recorrer cada tramo

Tabla 3.1.
Matriz de tiempos del mapa representado en la figura 3.1

	A	B	C	D	E	F	G	H	I
A	0	3	6	5	7				
B	3	0	2			8			
C	6	2	0	4		5	6		
D	5		4	0	3		9	7	
E	7			3	0			5	
F		8	5			0	2		4
G			6	9		2	0	3	1
H				7	5		3	0	6
I						4	1	6	0

Las casillas en blanco corresponden a tramos que no existen y deben considerarse, a efectos de cálculo, infinitos. En la diagonal principal, hay ceros porque el tiempo necesario para ir de un lugar al mismo sitio es cero.

Para hallar el camino óptimo desde A hasta I, veamos si es aplicable el principio del óptimo: por ejemplo, si el camino de A a I pasa por C, ello quiere decir que dicho camino está formado por el camino óptimo A-C y el camino óptimo C-I. Puesto que ello es cierto, podemos dividir el problema en partes: ¿cuál es el camino óptimo desde cualquier punto hasta I?

Empecemos resolviendo un problema trivial: ¿Cuál es el camino hasta I desde cualquier punto en un solo paso? Evidentemente, desde F, G o H, el camino es recorrer el tramo correspondiente, mientras que en los otros casos no existe. Anotemos esta solución en la columna N = 1 de la tabla 3.2.

Tabla 3.2.
Resolución del ejemplo de camino mínimo

	A	B	C	D	E	F	G	H	I	N= 1	N= 2	N= 3	N= 4	N= 5
A	0	3	6	5	7							13;C	12;B	12;B
B	3	0	2			8					12;F	9;C	9;C	9;C
C	6	2	0	4		5	6				7;G	7;G	7;G	7;G
D	5		4	0	3		9	7			10;G	10;G	10;G	10;G
E	7			3	0			5			11;H	9;H	9;H	9;H
F		8	5			0	2		4	4;I	3;G	3;G	3;G	3;G
G			6	9		2	0	3	1	1;I	1;I	1;I	1;I	1;I
H				7	5		3	0	6	6;I	4;G	4;G	4;G	4;G
I						4	1	6	0	0;I	0;I	0;I	0;I	0;I

En la primera iteración (N = 1), solo anotamos soluciones en las filas F, G, H e I, que son los únicos puntos desde donde se puede ir directamente a I. Junto a la longitud del camino correspondiente, anotamos cuál es el punto siguiente a donde ir. Como siempre, las casillas en blanco equivalen a un valor infinito.

Veamos ahora si es posible llegar a I en dos pasos (o menos). Para ello, ha de ser posible ir desde el vértice objeto de estudio a algún otro (fila de la matriz) desde el cual conozcamos ya un camino hasta I (columna N = 1). Observemos que el movimiento de los índices es análogo a multiplicar la matriz por el vector N = 1, salvo que, en lugar de multiplicar y sumar, sumamos y comparamos. Ambas operaciones son un caso particular de otra operación matemática más general denominada *convolución*.

Sea, por ejemplo, el vértice D. Desde él, podemos ir a A, al propio D, a G o a H. Pero, desde A, no hay camino conocido, por lo que la suma es infinita. El camino por G tiene una longitud de 9 + 1 = 10, mientras que por H es de 7 + 6 = 13. Elegimos, pues, el tramo D-G y anotamos la longitud del camino D-I hallado (10). De esta manera, hemos encontrado ya caminos hasta I desde todos los vértices excepto A, como es evidente viendo el dibujo.

Vemos también que, en el caso de H, ocurre algo similar. En lugar de ir directamente a I, resulta más corto el camino por G: 3 + 1 < 6.

Ahora repetimos la operación con la columna N = 2 para hallar el camino en tres pasos o menos (N = 3), donde observamos que por primera vez aparece un camino A-I, de longitud 13, que empieza con el tramo A-C. Veamos lo que ha ocurrido con el vértice E. Habíamos hallado un camino (E-H-I) de longitud 11. Pero la única opción disponible desde E es E-H: 5 + 4 = 9. El primer tramo del camino no ha variado, pero ahora es más corto porque hemos descubierto que se puede ir de H a I en 4 minutos (o km o la unidad que sea) en lugar de 6.

En la cuarta iteración (N = 4), observamos que el camino desde A no solo es más corto que en la iteración anterior, sino que además el primer tramo ha cambiado.

Finalmente, para N = 5, observamos que ya no ha cambiado nada: ni los valores (función económica) ni las decisiones (política), de lo cual se deduce que, aunque hiciésemos más iteraciones, ya no iban a cambiar: hemos alcanzado el estado estable o régimen permanente. El camino más corto desde A hasta I se hace en cinco tramos o menos, y tiene una longitud de 12. Observando, a partir de A, cuál es el vértice siguiente al cual ir (regla de decisión) reconstruimos el camino óptimo A-I, que resulta ser este:

A – B – C – G – I

3.1.2 Generalización

Este ejemplo simple nos permite explicar, de manera sencilla, algunas de las propiedades más importantes de la programación dinámica.

En primer lugar, puede observarse, consultando en el volumen de teoría de grafos de esta misma colección, que el algoritmo utilizado es el que allí se conoce como de Bellman-Kalaba, que será el que utilizaremos principalmente para resolver problemas deterministas como este.

Sin embargo, esto tiene una connotación mucho más general: todo problema de programación dinámica determinista se puede entender como hallar el camino mínimo o máximo de un grafo.

También podemos poner los elementos de este ejemplo en relación con los procesos polietápicos, porque, de hecho, eso es lo que es:

- Cada tramo es una etapa del problema; el número máximo de tramos a recorrer es, pues, la variable cronológica, que en este caso solo tiene un significado de tiempo de una manera muy relativa.
- En cada etapa, el sistema se halla en un estado de un conjunto finito de estados (el vértice donde nos hallamos).
- En cada etapa o transición, hay una decisión que hay que tomar: el vértice a donde ir, que depende de la etapa y del estado.
- Existe un rendimiento (el tiempo total) que debe optimizarse en su conjunto, no en cada paso individual.

Estos conceptos constituyen la base teórica de todo problema de programación dinámica y lo único que necesitamos para resolverlo, al menos mientras sea determinista.

3.2 Modelización en programación dinámica

El ejemplo anterior era un caso de horizonte finito no determinado: el número de pasos ha de ser necesariamente finito, ya que el grafo es finito, pero a priori desconocíamos cuántos íbamos a necesitar; ahora hemos visto que el régimen permanente se alcanza con $N = 4$ y, por tanto, ya no ha sido necesario proseguir. Es decir, el método sirve también para resolver el caso de horizonte finito no determinada y también el de horizonte infinito: basta iterar hasta que el sistema se estabiliza.

Antes de entrar en el detalle del método de cálculo, veamos el proceso de modelización que hemos seguido para llegar al caso que hemos resuelto.

3.2.1 Etapas, estados y decisiones

Un elemento esencial de la programación dinámica es la división del problema en etapas. En el ejemplo anterior, eran los tramos a recorrer, pero puede ser cualquier cosa; qué sean las etapas depende del problema concreto y establecer su naturaleza es el punto esencial para una buena modelización en programación dinámica.

Además del concepto de *etapa*, existen otros dos tan estrechamente relacionados con este que los tres han de definirse simultáneamente; de hecho, definir uno ayuda a definir los otros dos: son los conceptos de *estado* y *decisión*.

En cada etapa, el sistema a modelizar se halla en un solo elemento de un conjunto finito bien definido, denominado *conjunto de estados*, y el elemento en cuestión es el *estado el sistema*. Esta definición, basada en su carácter matemático, tal vez no nos ayude mucho a determinar cuál debe ser este conjunto ante un caso concreto; para ello, en la práctica lo mejor es imaginarse posibles respuestas a la pregunta: ¿Cómo están las cosas? Las respuestas posibles a esta pregunta son los estados posibles.

A la vez que la etapa y el estado, hemos de definir la *decisión*. La característica esencial de la programación dinámica es que en cada etapa podemos decidir ciertas acciones (incluida, por supuesto, la posibilidad de no hacer nada), con lo cual influiremos en el comportamiento futuro del sistema. Así pues, para cada etapa y cada estado, hemos de conocer cuál es el conjunto de acciones posibles entre las cuales debemos elegir para que el sistema pueda pasar a la etapa siguiente.

Al definir las etapas, definimos también la variable que nos cuenta el número de etapas que faltan para terminar, que recibe el nombre de *variable cronológica*.

3.2.2 Rendimiento

La esencia de la decisión es optimizar el rendimiento. Decidimos hacer algo porque queremos obtener un determinado objetivo, que puede ser de máximos o de mínimos, monetario (beneficios o costes) o no (fama o aburrimiento, por ejemplo). Si no hay rendimiento, no hay decisión porque, hagamos lo que hagamos, nos va a dar exactamente lo mismo: cualquier alternativa es tan buena como cualquier otra y no hay ninguna razón para preferir una; por consiguiente, no existe decisión. Si es obligado elegir, podemos hacerlo al azar porque no hay nada que nos motive a preferir una alternativa frente a otra. Por ello, y dado que la programación dinámica es esencialmente un método para optimizar decisiones, hay que definir un rendimiento.

El rendimiento es una magnitud escalar que se obtiene en cada transición y que pretendemos maximizar o minimizar en el conjunto de todas las etapas de que se compone el problema. Dicho rendimiento, en general, depende de los estados anterior y posterior del sistema, y de la decisión adoptada. Evidentemente, puesto que nos hallamos en el caso determinista, el estado posterior es función del estado anterior y de la decisión, por lo cual se puede considerar que el rendimiento, en este caso, solo depende de estos dos factores.

3.2.3 Ecuación de recurrencia

Hemos dicho que el objetivo es optimizar el rendimiento global, es decir, lo que obtendremos en el conjunto de todas las etapas de que conste la vida del sistema. Ello quiere decir que hemos de tener una manera de agregar estos rendimientos, para lo cual la programación dinámica recurre al principio de óptimo, que se expresa, de forma general, con la expresión:

$$f_N^*(i) = \underset{u}{OPT}\left\{r_{ij} \otimes f_{N-1}^*(j)\right\}$$

donde $f^*_N(i)$ representa el rendimiento total óptimo en las próximas N etapas partiendo del estado i. El símbolo * representa que la función es óptima; si este signo se omite, ello quiere decir que no sabemos si dicho resultado es óptimo o no: es solo lo que se obtendría con la decisión que consideramos.

OPT es el sentido de optimización, que puede ser MAX o MIN.

u^*_N representa la decisión que tomamos en la N-ésima etapa; igual que en el caso anterior, si se acompaña del *, representa la decisión óptima.

r_{ij} representa el rendimiento obtenido en la primera etapa, al pasar del estado i al estado j.

El símbolo ⊗ representa la operación de agregación de los rendimientos, que usualmente es la suma, aunque puede ser cualquier función monótona creciente.

Por tanto, el rendimiento en N transiciones es la agregación del resultado de la primera etapa y el resultado de las demás, que es otra manera de expresar el principio del óptimo.

3.2.4 Condiciones de contorno

La ecuación de recurrencia nos permite hallar el rendimiento en N etapas si conocemos el rendimiento en N-1. Ello implica que hemos de calcularlo antes, para lo cual necesitamos conocerlo en N-2, etc. Llegamos así a la necesidad de tener una solución inicial, con sus valores correspondientes, que debe ser un problema trivial. Esto es lo que se conoce como *condiciones de contorno*.

Usualmente, las condiciones de contorno se fijan para N = 0 o N = 1, cuyo significado en cuanto a la modelización es el siguiente:

N = 1 significa que, cuando ya solo queda una etapa para terminar, ya sabemos qué hemos de hacer, sea cual sea el estado. Generalmente, se elegirá la decisión que implique mejor r_{ij}.

N = 0 es más usual y, con frecuencia, tiene más interés. La expresión f_0 representa el rendimiento obtenido después de haber terminado la vida del sistema, es decir, el valor residual al final del proceso. En general, se toma 0 ya que, una vez "muerto" el sistema, ya no nos importará lo que vaya a ocurrir, aunque a veces puede ser muy útil para representar el valor residual en casos como los siguientes:

- El valor de una máquina o de un elemento usado que todavía pueda venderse.
- Los costos de eliminación de los residuos finales.
- Si el rendimiento representa la probabilidad de obtener un cierto resultado, expresa si en el estado final se ha obtenido (1) o no (0) dicho resultado.

3.2.5 El caso del camino mínimo

Vamos a ver ahora cómo hemos aplicado todos estos conceptos al caso resuelto en el ejemplo sobre el camino mínimo de un grafo descrito en la sección 3.1.1.

Etapas y estados. El estado es el lugar donde nos hallamos y la etapa es el camino que recorremos hasta el siguiente vértice del grafo.

Decisión. En cada etapa, decidimos a cuál de los vértices posibles nos vamos a dirigir: obsérvese que, en este caso, la decisión consiste precisamente en elegir cuál va ser el próximo estado del sistema.

Rendimiento. En este caso es, claramente, la longitud del camino que falta para llegar al punto I, longitud que puede estar expresada en cualquier unidad, según cuál sea su naturaleza: km (longitud), minutos (tiempo), euros (coste), etc. Lo único importante es que se trata de una magnitud que queremos minimizar.

Ecuación de recurrencia. Dado que nos interesa minimizar la longitud total, la ecuación es obvia: la longitud desde un punto hasta el punto de destino es la suma de la longitud desde dicho punto (i) hasta el siguiente (j) más la longitud desde este al punto final deseado.

Condiciones de contorno. En este caso, hemos tomado como condiciones el caso N = 1, dado que, para los vértices desde donde existe una arista que nos conduzca a I, el resultado es trivial; por otra parte, si dicha arista no existe también lo es: no hay camino.

3.3 Iteración en el espacio de los estados

El procedimiento utilizado para resolver el ejemplo descrito en la sección 3.1.1 se conoce, en el contexto de la programación dinámica, como Iteración en el espacio de los estados y es el que se sigue para resolver los problemas de horizonte finito determinado. Si conocemos el número de iteraciones necesario (N), basta con repetir el cálculo con la ecuación de recurrencia empezando por las condiciones de contorno hasta la N deseada. Veamos el método en detalle con un caso un poco peculiar.

3.3.1 Distribución de recursos

La tabla 3.3 muestra el rendimiento (en cientos de €) que resulta de asignar un determinado número de trabajadores a cuatro tareas.

Tabla 3.3. Rendimientos de asignar trabajadores a tareas en el ejemplo de distribución de recursos

	A	**B**	**C**	**D**
0	0	0	0	0
1	3	4	6	5
2	7	8	9	10
3	9	10	12	11
4	15	13	14	11

Sabiendo que contamos con cuatro trabajadores, ¿cuántos deberemos asignar a cada tarea para que el rendimiento sea óptimo?

Este es un tipo de problema clásico de la investigación operativa: se trata de la asignación óptima de recursos a un cierto conjunto de actividades y es un caso para el cual la programación dinámica está especialmente bien dotada. Veamos cómo definir cada uno de los elementos del modelo.

Rendimiento. Es obvio que se trata de maximizar la suma de los valores de la tabla, pero ¿cuáles son las etapas y cuáles son los estados?

Una primera pista nos la da la decisión: es evidente que lo que hemos de decidir es cuántos trabajadores asignamos a cada tarea y, dado que cada decisión va asociada a una etapa, ello significa que las etapas son las tareas; así pues, se trata de un problema en horizonte finito, con N = 4.

Este es uno de los casos en que la variable cronológica no tiene significado de tiempo, y ello, además, conlleva otra característica peculiar: sabemos cuántas etapas hay, pero su orden no está fijado. Dicho de otra forma: podemos asignar los trabajadores a las tareas en cualquier orden, ya sea respetar el que nos han dado (A, B, C, D), seguir el orden inverso (D, C, B, A) o cualquier otro (por ejemplo, B, C, A, D).

Si ya sabemos qué son las etapas, el rendimiento y las decisiones, solo nos falta conocer los estados. Para ello, podemos formular la pregunta que ya hemos dicho que nos puede ayudar: ¿Cómo están las cosas?

Veamos: hemos dicho que las etapas son las tareas; por tanto, si a mitad del problema nos preguntan eso, ¿qué diremos? Ya sabemos las tareas a las cuales hemos asignado trabajadores y cuántos hemos asignado a cada una, así como el rendimiento que todo ello nos conlleva. ¿Qué nos queda por decir? El número de personas todavía no asignadas (o las ya asignadas, que viene a ser lo mismo) serán, pues, los estados.

Como el rendimiento se acumula sumando y se trata de un máximo, la función de recurrencia será:

$$f_N^*(x) = \underset{0 \le u \le x}{MAX} \left\{ r_N(u) + f_{N-1}^*(x-u) \right\}$$

donde x es el estado (trabajadores no asignados todavía), u es la decisión y f^*_N es el rendimiento obtenido con las asignaciones hechas en las N primeras tareas.

Las condiciones de contorno serán $f^*_N(x) = 0$ para toda x, ya que no nos sirve de nada tener trabajadores disponibles una vez completada la asignación.

Utilicemos todo ello para calcular la política óptima. Recuérdese que podemos definir el orden de las tareas como queramos; lo calcularemos de una manera poco habitual: D A C B.

Veamos las dos primeras tareas:

				u_A						
X	f^*_0	f^*_1	u^*_D	0	1	2	3	4	f^*_2	u^*_A
0	0	0	0	0					0	0
1	0	5	1	5	3				5	0
2	0	10	2	10	8	7			10	0
3	0	11	3	11	13	12	9		13	1
4	0	11	3,4	11	14	17	14	15	17	2

Tabla 3.4.
Cálculo de la política óptima asignando las tareas D y A

Para N = 1 (tarea D), la decisión es trivial (podría haber servido también como condición de contorno), pues asignamos todos los trabajadores disponibles a la tarea; solo en el caso x = 4 da lo mismo asignar 3 o 4, ya que el rendimiento no aumenta.

En N = 2, ya hay que aplicar la fórmula de la ecuación de recurrencia; las columnas encabezadas por una u_A indican el valor de la expresión entre llaves para cada valor de u. Hecho esto, basta elegir el mayor valor (f^*_2) y anotar junto a él la u con que se obtiene (u^*_A). Observemos que si, llegados a este punto, solo disponemos de uno o dos trabajadores, lo mejor será no asignar a ninguno a esta tarea y reservarlos para la D.

Continuemos ahora con las otras dos tareas pendientes:

Tabla 3.5.
Cálculo de la política óptima asignando las 4 tareas

		u_C							u_B						
x	f^*_2	0	1	2	3	4	f^*_3	u^*_C	0	1	2	3	4	f^*_4	u^*_B
0	0	0					0	0	0					0	0
1	5	5	6				6	1	6	4				6	0
2	10	10	11	9			11	1	11	10	8			11	0
3	13	13	16	14	12		16	1	16	15	14	10		16	0
4	17	17	19	19	17	14	19	1,2	19	20	19	16	13	20	1

Para construir la solución (“leer la tabla”), hemos de empezar por N = 4, ya que la tarea B es la última calculada y la primera a la cual asignamos personal. Como disponemos de 4 personas, empezamos con la última fila (x = 4) y ahí leemos que $f^*_4(4) = 20$: el valor de la función económica será 20 cientos de euros, es decir, 2.000 €.

¿Cómo se logra este resultado? Por lo pronto, asignando un solo trabajador a la tarea B, como nos indica la u^*_B. A continuación, nos trasladamos a la etapa siguiente, con x = 3, donde leemos $f^*_3(3) = 16$ y $u^*_3(3) = 1$. Observemos que f^*_N ha disminuido en 4, que es precisamente el rendimiento que ha aportado el trabajador asignado a B. Asignamos, pues, otro trabajador a la tarea C y nos vamos a N = 2, con x = 2, donde vemos que $f^*_2(2) = 10$ y $u^*_A(2) = 0$, lo cual significa que no hemos de asignar a ningún trabajador a esta tarea. Finalmente, $f^*_1(2) = 10$ y $u^*_D(2) = 2$ nos dice que hemos de asignar los dos trabajadores restantes a la tarea D, en que aportarán un rendimiento de 10 cientos de euros.

3.3.2 Multiplicación máxima

Veamos ahora otro ejemplo muy similar al anterior, pero cuya característica es que la función de agregación de rendimientos no es la suma sino la multiplicación: descomponer el número 10 en cuatro sumandos enteros positivos cuyo producto sea máximo.

Está claro que cada sumando es una etapa, que la decisión es el número que asignamos a cada sumando y que el estado es la parte de la suma todavía no asignada.

Las condiciones de contorno, en este caso, son directamente la asignación del último sumando (N = 1), que coincide con x:

$$f_N^*(x) = \underset{u_N=1,2,..x-1}{MAX} \left\{ u_N \times f_{N-1}^*(x-u) \right\}$$

$$f_1^*(x) = x; \quad u_1^*(x) = x$$

3.4 El largo plazo

3.4.1 Paso de régimen transitorio a régimen permanente

Los dos ejemplos que hemos visto hasta ahora tienen en común que el horizonte es finito. En el caso 3.2 de asignación de recursos, además, estaba bien determinado, ya que eran 4 tareas; en el caso del camino mínimo (3.1.1 y 3.2.4), no era tan claro cuál iba a ser el límite, pero era evidente que no podían ser muchas las transiciones: era un caso de horizonte finito indeterminado.

Sin embargo, en muchas ocasiones, ello no es así. El sistema que hemos de gestionar ha de funcionar indefinidamente y, por tanto, hemos de plantear el caso como si el horizonte fuese infinito. El caso del camino mínimo ya nos da una pista de por dónde podemos buscar la solución: si iteramos hasta que el sistema se estabilice, habremos hallado el régimen permanente; aunque no siempre las cosas son tan claras como en este caso.

3.4.2 Procesos homogéneos en el tiempo

Si debe estudiarse el comportamiento de un sistema durante un plazo indefinidamente largo, está claro que ha de existir una regularidad en su comportamiento, lo cual no ocurría en el ejemplo de asignación de recursos que hemos visto anteriormente: en él, cada etapa se comportaba de manera diferente. Era un sistema no homogéneo en el tiempo.

Un sistema es homogéneo en el tiempo (HT) si su comportamiento no depende de la variable cronológica; en caso contrario, es no homogéneo en el tiempo (NHT).

HT significa que todas sus características (estados accesibles, rendimiento, decisiones posibles, ecuación de recurrencia) son las mismas para cualquier etapa en que nos hallemos, lo cual claramente no era el caso de la asignación de recursos.

Veamos un ejemplo distinto, que nos servirá de guía para estudiar el caso.

3.4.3 El catalizador

Un proceso químico determinado mejora notablemente su rendimiento si se introduce en él un catalizador. Dicho catalizador tiene una vida útil máxima de cinco semanas, tras las cuales ha de ser reemplazado de inmediato. También es posible reemplazarlo antes, si las condiciones económicas del proceso así lo aconsejan.

A medida que el catalizador envejece, su rendimiento disminuye y su coste de utilización aumenta, según la tabla 3.6.

Tabla 3.6.
Costes de utilización del catalizador

Semana de funcionamiento	Coste (€ × 100)
1	10
2	12
3	18
4	25
5	33

Cada vez que el catalizador se reemplaza, se incurre en unos costes adicionales de instalación de 20 cientos de euros.

¿Cómo debemos gestionar el catalizador si queremos unos costes óptimos?

Empecemos por crear el modelo, que en este caso no ofrece demasiadas dudas.

Etapas: semanas de funcionamiento del catalizador.

Estado: número de semanas que lleva funcionando el catalizador al principio de la semana (edad del catalizador).

El detalle de decir en qué momento se define el estado del sistema no es trivial: equivale a decir cuándo termina una etapa y comienza la siguiente, lo cual equivale a decir en qué momento tomamos la decisión.

Decisión: evidentemente, hemos de decidir si cambiamos (C) o no (N) el catalizador.

Rendimiento: coste total del sistema (funcionamiento + cambios), que, evidentemente, es de mínimos.

Ecuación de recurrencia:

$$f_N^*(x) = \text{MIN}\,\{20 + c(0) + f_{N-1}^*(1);\, c(x) + f_{N-1}^*(x+1)\}$$

donde la función c(x) representa el coste de funcionamiento durante la semana siguiente si el catalizador la empieza en el estado x. Escribámosla porque tiene una pequeña sutileza:

x	0	1	2	3	4	5
c(x)	10	12	18	25	33	999

Observemos que los valores de x están desplazados una semana con respecto a la forma en que se han indicado en el enunciado. Ello se debe a que si, por ejemplo, el catalizador es nuevo (x = 0), la semana que empieza será su primera semana de vida. Asimismo, para x = 5, hemos puesto un valor muy alto (999), que en realidad debe leerse como infinito; como se verá, ello podría acarrearnos algunas dificultades si alargásemos mucho el cálculo, pero no llegaremos tan lejos.

Condiciones de contorno: a falta de información, tomamos

$$f^*_0(x) = 0$$

Con todo ello, estamos ya en condiciones de iniciar el cálculo:

Tabla 3.7.
Solución del ejemplo del catalizador

x	c(x)	f^*_0	C	NC	f^*_1	u^*_1	C	NC	f^*_2	u^*_2	C	NC	f^*_3	u^*_3
0	10	0	30	10	10	N	42	22	22	N	60	40	40	N
1	12	0	30	12	12	N	42	30	30	N	60	54	54	N
2	18	0	30	18	18	N	42	43	42	C	60	60	60	N,C
3	25	0	30	25	25	N	42	55	42	C	60	67	60	C
4	33	0	30	33	30	C	42	63	42	C	60	75	60	C
5	999	0	30	999	30	C	42	999	42	C	60	999	60	C

Es decir, a tres semanas de finalizar la gestión, debemos sustituir el catalizador si tiene una edad 3 o superior y no cambiarlo si su edad es 0 o 1; será indiferente cambiarlo o no si su edad es 2. Esta será la política a corto plazo. Pero, ¿cómo sería a largo plazo? Podemos iterar un poco más:

Tabla 3.8.
Solución del ejemplo del catalizador (continuación)

x	C	NC	f^*_4	u^*_4	C	NC	f^*_5	u^*_5	C	NC	f^*_6	u^*_6	f^*_7	u^*_7	f^*_8	u^*_3
0	84	64	64	N	102	82	82	N	120	100	100	N	124	N	142	N
1	84	72	72	N	102	90	90	N	120	114	114	N	132	N	150	N
2	84	78	78	N	102	102	102	N,C	120	120	120	N,C	138	N	162	N,C
3	84	85	84	C	102	109	102	C	120	127	120	C	144	C	162	C
4	84	93	84	C	102	117	102	C	120	135	120	C	144	C	162	C
5	84	999	84	C	102	999	102	C	120	999	120	C	144	C	162	C

¿Hemos llegado ya a la estabilidad? No los sabemos. Por lo pronto, la política se parece mucho, pero no se repite exactamente en cada iteración. Además, los valores de f_N^* van aumentando, como es lógico, y amenazan que el 999 con que hemos representado el infinito no sea suficiente. Veamos una técnica que nos ayudará en este punto.

3.4.4 Normalización

Si los rendimientos no son valores de signos distintos, los valores de las f^*_N tenderán a crecer indefinidamente, lo cual puede suponer, como mínimo, un engorro para los cálculos y para la interpretación de los resultados. La normalización permite evitar este inconveniente.

La técnica de la normalización consiste en restar sistemáticamente a los valores obtenidos en cada transición una cantidad determinada y, en lo sucesivo, operar con la diferencia. Así pues, las ecuaciones de recurrencia quedan de la forma siguiente:

$$f_N^*(i) = \underset{u}{OPT}\left\{r_{ij} \otimes f'_{N-1}(j)\right\}$$

$$f'_N(i) = f_N^*(i) - g_N$$

donde g_N es un valor fijado arbitrariamente para cada iteración. Por supuesto, la operación de normalización que hemos representado con la resta ha de ser coherente con la operación de agregación de los rendimientos.

Obsérvese que, una vez efectuados los cálculos con la normalización, el valor del rendimiento es:

$$f_N^{*R}(i) = f'_N(i) + g_N + g_{N-1} + \ldots + g_1$$

Esta técnica tiene dos ventajas: en primer lugar, evita que los valores absolutos de f^*_N crezcan de forma continuada, lo cual facilita indiscutiblemente el cálculo; en segundo lugar, permite observar con más facilidad la llegada del régimen permanente, puesto que cuando en dos transiciones (consecutivas o no) se repitan los valores, tanto de la g_N como de f^*_N, podremos afirmar que se ha superado el régimen transitorio y se ha alcanzado el régimen permanente.

Queda por determinar cómo se fija la g_N, valor que se debe restar a todas las f^*_N: en principio, este valor es arbitrario, y lo importante es que se determine según una regla fija. Las dos más habituales son:

Regla 1:

$$g_N = \underset{i}{MIN}\left\{f_N^*(i)\right\}$$

Regla 2:

$$g_N = f_N^*(k)$$

donde k es un estado elegido previamente.

Ambas reglas conducen, a largo plazo, a los mismos resultados (salvo unas diferencias constantes), pues en régimen permanente el estado de valor mínimo será siempre el mismo. La regla 1 tiene la ventaja de que se asegura que los valores de f'_N son siempre superiores o iguales a cero, lo cual evita errores en los cálculos manuales, mientras que la regla 2 facilita la comparación de los valores de f'_N.

Apliquemos la técnica a nuestro caso:

x	r(x)	f_0	C	N_C	f_1	u_1	f'_1	C	N_C	f_2	u_2	f'_2	C	N_C	f_3	u_3	f'_3
0	10	0	30	10	10	N	0	32	12	12	N	0	38	18	18	N	0
1	12	0	30	12	12	N	2	32	20	20	N	8	38	32	32	N	14
2	18	0	30	18	18	N	8	32	33	32	C	20	38	38	38	N,C	20
3	25	0	30	25	25	N	15	32	45	32	C	20	38	45	38	C	20
4	33	0	30	33	30	C	20	32	53	32	C	20	38	53	38	C	20
5	999	0	30	999	30	C	20	32	999	32	C	20	38	999	38	C	20
g_N					10					12					18		

Tabla 3.9.
Solución del ejemplo del catalizador con normalización

El lector puede comprobar que los resultados son equivalentes con los obtenidos anteriormente, pero con una diferencia: ahora los valores de f'_N ya no crecen en cada iteración y, por tanto, son comparables. La política es la misma de antes y, en consecuencia, presenta los mismos síntomas, pero ahora observamos que también ocurre lo mismo con los valores normalizados.

Un hecho nos permitirá afirmar, cuando se produzca, que hemos llegado al régimen permanente: si se repiten tanto la política como los valores normalizados, es seguro que ello se repetirá indefinidamente en el futuro. Iteremos un poco más:

x	f'_3	C	N_C	f_4	u_4	f'_4	f_5	u_5	f'_5	f_6	u_6	f'_6	f_7	u_7	f'_7	f_8	u_8	f'_8
0	0	44	24	24	N	0	18	N	0	18	N	0	24	N	0	18	N	0
1	14	44	32	32	N	8	26	N	8	32	N	14	32	N	8	26	N	8
2	20	44	38	38	C	14	38	N,C	20	38	N,C	20	38	C	14	38	N,C	20
3	20	44	45	44	C	20	38	C	20	38	C	20	44	C	20	38	C	20
4	20	44	53	44	C	20	38	C	20	38	C	20	44	C	20	38	C	20
5	20	44	999	44	C	20	38	C	20	38	C	20	44	C	20	38	C	20
g_N				24			18			18			24			18		

Tabla 3.10.
Solución del ejemplo del catalizador con normalización (continuación)

En la tabla 3.10, observamos que, en realidad, ya hay repetición desde la tercera iteración: las columnas u y f' de N = 3 se repiten para N = 6. Lo mismo ocurre con las parejas 4-7 y 5-8, lo cual hace ocioso proseguir con los cálculos: a largo plazo, el sistema tiene un ciclo de longitud 3 y se va a repetir indefinidamente. Ya hemos llegado al régimen permanente.

¿Cuál es, pues, la política a seguir si hemos de seguir gestionando el catalizador indefinidamente? Dado que, en el estado 2, a veces hay que cambiarlo y a veces da igual hacerlo o no, la política a seguir es cambiar el catalizador si tiene edad 2 o superior, es decir, {N, N, C, C, C, C}, con lo cual, en la práctica, se va a repetir el ciclo 1-2-3 (recordemos que este es el estado antes de decidir si cambiamos o no), es decir, un ciclo de tres semanas. Si supiéramos cuántas semanas nos quedan hasta terminar (módulo 3), antes de acabar sabríamos cuándo nos podemos permitir el lujo de no cambiar un catalizador de edad 2, aunque ello es irrelevante, puesto que no tiene influencia en el coste.

Nos preguntamos ahora cuál va ser el coste a largo plazo con esta política. Está claro que la pregunta no tiene respuesta “en euros” o, mejor dicho, su respuesta es “infinito”, porque hay infinitas semanas: la respuesta solo tiene sentido en euros/semana, lo cual es el valor de g_N.

Ahora bien, g_N tiene valores distintos según la semana. Ello no es problema, porque sabemos que hay un ciclo de tres semanas, así que el valor medio será:

$$g = (18+18+24)/3=20 \text{ cientos €/semana}$$

3.4.5 Iteración en el espacio de las políticas

No siempre vamos a tener tanta suerte. El procedimiento que hemos visto para hallar la solución óptima a largo plazo no suele hallarla tan rápido. De hecho, es probable que, aunque las políticas se empiecen a repetir relativamente pronto, los f’ no, e incluso sucede con frecuencia que solo tienden hacia los valores finales de manera asintótica, de forma que siempre vamos a tener “bailando” el último decimal. Nos interesa, pues, hallar un procedimiento que nos permita hallar la solución óptima a largo plazo directamente.

Para ello, tengamos en cuenta que, a largo plazo, se puede escribir:

$$f_N^*(i) \xrightarrow[N\to\infty]{} Ng + w_i$$

O sea, el rendimiento total será el rendimiento medio por semana por el número de semanas y por un factor de corrección, que dependerá del estado inicial. Huelga decir que la g será el límite de las g_N y que las w_i lo serán de las f’(i).

Supongamos ahora que hemos determinado una política a largo plazo, óptima o no; en cada estado i, tenemos una decisión que nos conducirá a un estado j = u(i).

Si sustituimos la expresión anterior en la ecuación de recurrencia:

$$Ng + w_i = r_{ij} + (N-1)g + w_j$$

con lo cual obtenemos:

$$g + w_i = r_{ij} + w_j$$

que es la expresión básica para la iteración en el espacio de las políticas, lo cual es un procedimiento que permite hallar la política óptima para el sistema a largo plazo; es decir, para cada uno de los n estados posibles indica la acción a realizar si el sistema está lo bastante alejado de su final para considerar que se halla en régimen permanente.

La expresión anterior nos proporciona un sistema de tantas ecuaciones como estados y tantas incógnitas como estados (las w_i) más una (g). Tenemos, pues, un grado de libertad para conocer estos valores. Sin embargo, dadas las características del sistema, este grado de libertad afecta las w_i, no la g, que está determinada. De hecho, también lo están las diferencias $w_i - w_j$ para todo par i, j.

En esencia, el procedimiento consiste en:

1. Fijar una política inicial cualquiera.
2. Determinar la g y las w_i para la política actual.
3. Comprobar si los valores hallados cumplen la condición de óptimo.
4. Si cumplen la condición, fin del algoritmo.
5. Si no la cumplen, sustituir la política actual por otra mejor y regresar al punto 2.

Vamos a examinar cada paso con más detalle.

Paso 1: Fijar una política inicial

En principio, no existe ninguna restricción para fijar esta política; sin embargo, como es lógico, cuanto mejor sea, menos iteraciones serán necesarias y menos trabajo habrá que realizar para hallar el óptimo. Por tanto, podemos fijar algunas reglas prácticas:

- Si se ha calculado mediante la iteración en el espacio de los estados la política óptima a corto plazo, la mejor opción es tomar la política para la mayor N calculada; de esta manera, se podrá confirmar si los resultados obtenidos mediante la iteración en los estados son ya óptimos en régimen permanente.
- A falta de mejor información, una buena elección es escoger para cada estado aquella acción cuya $r_i(u)$ sea mejor.

Paso 2: Hallar los valores de g y wi

Se trata de resolver para la política fijada el sistema ya conocido:

$$g + w_i = r_{ij} + w_j$$

Recordemos que hemos de fijar una de las w_i a un valor arbitrario, usualmente $w_i = 0$.

Paso 3: Comprobar si cumplen la condición de óptimo

Si la política es óptima, para cada estado debe cumplirse:

$$w_i = OPT_u \{ r_i(u) + w_j \} - g$$

Así pues, se trata de comprobar si, en cada estado, la acción definida por la política cumple o no dicha condición.

Paso 4: Definición de la nueva política o fin del algoritmo

Si, en el paso anterior, se ha cumplido la condición en todos los estados, la política es óptima y se termina el algoritmo.

En caso contrario, habrá al menos un estado para el cual exista por lo menos una decisión (u) para la cual el valor $r_i(u) + w_j - g$ será mejor (mayor si maximizamos, menor si minimizamos) que el que corresponda a la política ensayada. Cambiando esta decisión, se garantizará que la política definida por este cambio ya será mejor que la anterior. Evidentemente, es aconsejable hacer en cada iteración todos los cambios necesarios en todos los estados, eligiendo sistemáticamente aquella u cuyo valor sea mejor. De esta manera, se obtiene una nueva política, que se debe analizar regresando al paso 2 del algoritmo.

Obsérvese que el hecho de que en una iteración se haya sustituido una decisión concreta por otra no garantiza que dicha decisión no vuelva a aparecer más adelante en otra iteración, ya que su optimalidad no depende solo de ella misma, sino del conjunto de la política.

Apliquémoslo a nuestro catalizador. Hemos de empezar fijando una política cualquiera, óptima o no; por supuesto, podemos empezar por cualquiera, pero como ya conocemos políticas a corto es recomendable empezar por ellas. Elijamos, por ejemplo la siguiente: {N, N, C, C, C, C}, es decir, cambiar el catalizador si tiene edad 2 o superior.

El sistema correspondiente será:

$$\begin{aligned} w_0 + g &= 10 + w_1 \\ w_1 + g &= 12 + w_2 \\ w_2 + g &= 30 + w_1 \\ w_3 + g &= 30 + w_1 \\ w_4 + g &= 30 + w_1 \\ w_5 + g &= 30 + w_1 \end{aligned}$$

Hemos de fijar una w_i. Por ejemplo, $w_1 = 0$.

Resolviendo el sistema, obtenemos:

$$g=21;\ w_0=-11\ ;\ w_1=0\ ;\ w_2=w_3=w_4=w_5=9$$

Comprobemos si es óptimo:

w_0) MIN {20+10 + w_1; 10 + w_1} – g = MIN {30; 10} – 21 = - 11 = $w_0 \Rightarrow$ OK
w_1) MIN {20+10 + w_1; 12 + w_2} – g = MIN {30; 21} – 21 = 0 = $w_1 \Rightarrow$ OK
w_2) MIN {20+10 + w_1; 18 + w_3} – g = MIN {30; 27} – 21 = 6 < $w_2 \Rightarrow$ cambiar
w_3) MIN {20+10 + w_1; 25 + w_4} – g = MIN {30; 34} – 21 = 8 = $w_3 \Rightarrow$ OK
w_4) MIN {20+10 + w_1; 33 + w_5} – g = MIN {30; 43} – 21 = 9 = $w_4 \Rightarrow$ OK
w_5) MIN {20+10 + w_1; 999} – g = MIN {30; 999} – 21 = 9 = $w_5 \Rightarrow$ OK

Así pues, la política {N, N, N, C, C, C} es mejor (como ya sabíamos).

Comprobémosla:

$$w_0 + g = 10 + w_1$$
$$w_1 + g = 12 + w_2$$
$$w_2 + g = 18 + w_3$$
$$w_3 + g = 30 + w_1$$
$$w_4 + g = 30 + w_1$$
$$w_5 + g = 30 + w_1$$

Supongamos, como antes, que $w_1 = 0$. Resolviendo el sistema, obtenemos:

$$g=20;\ w_0=-10\ ;\ w_1=0\ ;\ w_2=8;\ \ w_3=w_4=w_5=10$$

Comprobemos si es óptimo:

w_0) MIN {20+10 + w_1; 10 + w_1} – g = MIN {30; 10} – 20 = - 10 = $w_0 \Rightarrow$ OK
w_1) MIN {20+10 + w_1; 12 + w_2} – g = MIN {30; 20} – 20 = 0 = $w_1 \Rightarrow$ OK
w_2) MIN {20+10 + w_1; 18 + w_3} – g = MIN {30; 28} – 20 = 8 = $w_2 \Rightarrow$ OK
w_3) MIN {20+10 + w_1; 25 + w_4} – g = MIN {30; 35} – 20 = 10 = $w_3 \Rightarrow$ OK
w_4) MIN {20+10 + w_1; 33 + w_5} – g = MIN {30; 43} – 20 = 10 = $w_4 \Rightarrow$ OK
w_5) MIN {20+10 + w_1; 999} – g = MIN {30; 999} – 20 = 10 = $w_5 \Rightarrow$ OK

La política es óptima a largo plazo, como ya sabíamos.

Comparemos estos resultados con los obtenidos iterando en el espacio de los estados. Ya hemos visto que g, el coste medio por semana, es el mismo (20 cientos de euros), pero los valores de w_i, que deberían ser el límite de las f'_N, no se parecen en nada a ellos. ¿Qué ocurre?

Lo que ocurre es efecto del grado de libertad del sistema de ecuaciones. Recordemos que hemos de tomar la media de tres semanas para obtener el valor medio a largo plazo y comparar esta media con la w_i correspondiente:

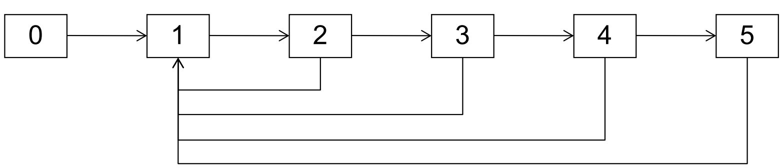
i	media f'(i)	wi	dif.
0	0	-10	10
1	(8 + 8 + 14)/3 = 10	0	10
2	(14 + 20 + 20)/3 = 18	8	10
3	20	10	10
4	20	10	10
5	20	10	10

Es decir, lo único que ha ocurrido es que hemos desplazado el nivel de referencia del estado 0 al estado 1. Estos valores tan solo nos dicen que, si empezamos con un catalizador nuevo ya montado, nos ahorraremos 10 cientos de euros frente a si empezamos con uno de edad 1; que si empezamos con uno de edad 2, ello nos costará 800 euros más, y que si es de edad 3 o superior, nos costará 1.000 euros más. Por supuesto, a largo plazo, estas diferencias son irrelevantes frente a los 2.000 euros por semana que nos cuesta el sistema.

Un último detalle antes de terminar con este ejemplo. Hemos dicho en 3.1.2 que todo problema de programación dinámica determinista se puede entender como hallar el camino mínimo o máximo de un grafo.

El grafo en cuestión se caracteriza porque tiene los vértices equivalentes a los estados y los arcos, a las transiciones posibles cuya elección es la decisión. Una transición equivale a recorrer un arco y el rendimiento es un valor asociado al mismo. En nuestro caso, el grafo está representado en la figura 3.2 y el problema consiste en hallar un camino infinito de longitud mínima, que es dar vueltas al circuito 1-2-3-1.

Figura 3.2.
Grafo del ejemplo del catalizador.



3.5 Técnicas adicionales de modelización

En este apartado, vemos dos técnicas de modelización, que suelen ser útiles, ya que se dan en situaciones bastante habituales: la programación dinámica separable y la actualización de rendimientos.

3.5.1 Programación dinámica separable

El ejemplo del catalizador visto en 3.4.3 es un caso al cual es aplicable esta técnica y nos será útil para exponer el concepto antes de presentarlo de forma completa.

El problema del catalizador se reduce, esencialmente, a una decisión que debe tomarse en un momento determinado, al inicio de la semana, que tendrá consecuencias durante la transición hasta la decisión siguiente. Por tanto, la transición entre dos etapas puede dividirse en dos pasos, que denominaremos *decisión* y *evolución*. La cuestión es que se puede definir un estado del sistema después de la decisión y antes de la evolución, y escribir la ecuación de recurrencia en dos pasos (v. figura 3.3.)

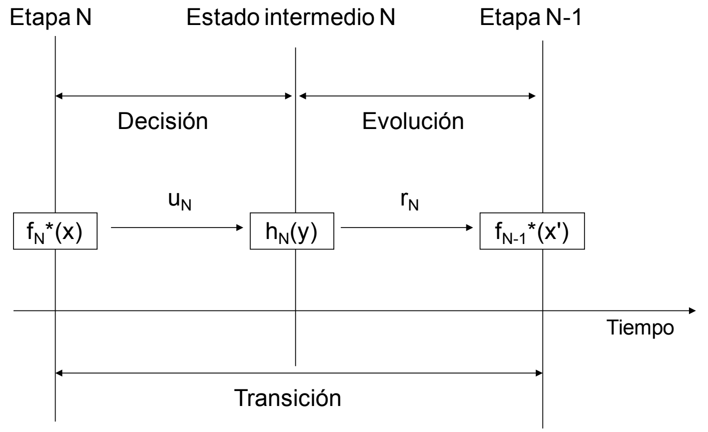


Figura 3.3. Representación de los dos pasos de una transición (decisión y evolución)

Partimos de la situación cuando faltan N etapas para el final. En este momento, hemos definido la función objetivo $f_N^*(x)$ como habitualmente (tal vez normalizada, aunque eso no tiene importancia ahora). En este momento, tomamos una decisión u_N, con lo cual el sistema evolucionará hacia un cierto estado, que puede o no ser uno de los que hemos definido para $f_N^*(x)$; en este momento, el rendimiento tendrá un cierto valor, que será el del $f_N^*(x)$, más una posible modificación debida a la decisión tomada. En nuestro caso, si la decisión es cambiar, habremos añadido 20 al coste, y el catalizador estará en el estado 0 (nuevo). Si no hemos realizado el cambio, ni el estado del catalizador ni el coste habrán cambiado.

A continuación, viene la evolución, en la cual no tenemos ninguna intervención: el sistema evoluciona del estado intermedio *y* a un nuevo estado *x'*, con el cual empezamos una nueva etapa. En el caso del catalizador, será siempre $x' = y + 1$, y la nueva $f_{N-1}^*(x)$ será el coste en el punto intermedio menos el coste debido a la antigüedad del catalizador.

En definitiva, podemos escribir la ecuación de recurrencia de esta forma:

$$f_N^*(x) = \text{MIN}\,\{20 + h_N(0); h_N(x)\}$$
$$h_N(x) = c(x) + f_{N-1}^*(x+1)$$

En este caso, los estados intermedios *y* son los mismos que hemos definido para las *x*, pero esto no tiene por qué suceder en general.

Veamos cómo sería el cálculo con este método:

Tabla 3.11.
Solución del ejemplo del catalizador con PD separable.

x	c(x)	f^*_0	h_1	f^*_1	u_1	f'_1	h_2	f^*_2	u_2	f'_2	h_3	f^*_3	u_3	f'_3
0	10	0	10	10	N	0	12	12	N	0	18	18	N	0
1	12	0	12	12	N	2	20	20	N	8	32	32	N	14
2	18	0	18	18	N	8	33	32	C	20	38	38	N,C	20
3	25	0	25	25	N	15	45	32	C	20	45	38	C	20
4	33	0	33	30	C	20	53	32	C	20	53	38	C	20
5	999	0	999	30	C	20	999	32	C	20	999	38	C	20
				10				12				18		

Hemos aplicado la normalización, aunque ello no tiene por qué ser así. Desde luego, los resultados coinciden con los que habíamos encontrado.

Para iterar en el espacio de las políticas, basta con tener en cuenta que la función h(y) interviene en los dos tramos de la ecuación:

$$Ng + w_X = r_1(x,u) + h(y)$$
$$h(y) = r_2(y) + (N-1)g + w_{x'}$$

donde $r_1(x,u)$ y $r_2(y)$ representan los rendimientos existentes en la primera y en la segunda parte de la transición, que, como es lógico, dependen del estado inicial y de la decisión (la primera) y del estado intermedio (la segunda).

Para resolver el sistema, hemos de eliminar la N, que, como sabemos, es infinito; para ello, basta con sustituir la h(y) de la segunda ecuación en la primera, con lo cual nos queda la misma expresión que si no hubiésemos separado la transición en dos partes y, por tanto, el método de resolución es el mismo. Debe tenerse en cuenta que las funciones h(y) son valores totales, es decir, que no tiene sentido intentar determinarlos en un horizonte infinito.

Veamos ahora el concepto de forma general:

Se dice que un problema es tratable mediante programación dinámica separable si es posible dividir la unidad de la variable cronológica que transcurre entre dos etapas en dos partes, una que corresponde a la decisión y otra a la evolución del sistema sobre la cual no tenemos capacidad alguna de actuación; dicha división puede hacerse de forma que la primera corresponda a la decisión y la segunda, a la evolución (separación D/E), o bien al revés (separación E/D). En todo caso, se define un conjunto de estados intermedios del sistema {y}, que puede ser o no el mismo de los estados en el punto de cambio de etapa, y en ellos se define una función del rendimiento en las condiciones correspondientes a dicho estado, h(y), que se relaciona con las dos $f_N^*(x)$ anterior y posterior mediante las expresiones correspondientes.

La figura 3.3 muestra un sistema D/E y la figura 3.4, una separación E/D:

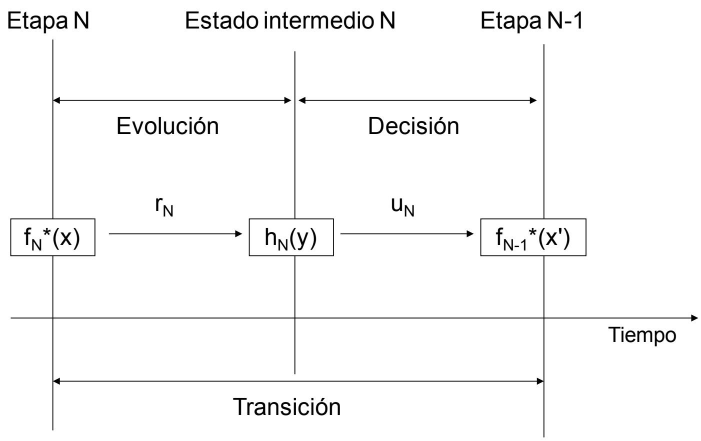


Figura 3.4. Representación de una separación E/D (evolución-decisión)

Esta técnica es utilizable en la mayoría de los casos y suele ser muy útil si el modelo que requiere el problema es complejo, lo cual ocurre principalmente en programación dinámica aleatoria, como veremos en el capítulo siguiente. Usualmente, en programación dinámica determinista, no vale la pena utilizarlo. El caso del catalizador está bastante "en el límite".

3.5.2 Actualización

La actualización es un caso en que la función de agregación del rendimiento no es simplemente la suma, sino una suma afectada por un coeficiente de actualización α, que casi siempre es positivo e inferior a la unidad: $0 < \alpha < 1$.

Ello sucede, sobre todo, en problemas de tipo financiero. Como es sabido, el valor del dinero disminuye con el paso del tiempo, de modo que los resultados obtenidos en una etapa no pueden sumarse sin más a los de la etapa anterior y deben afectarse del coeficiente de actualización. Así, si deseamos tener en cuenta una tasa de interés r, el valor del coeficiente de actualización ha de ser $\alpha = 1 - r$.

Desde el punto de vista del modelo, ello solo significa que en la expresión de la ecuación de recurrencia debe introducirse dicho coeficiente:

$$f_N^*(x) = \underset{u}{OPT}\left\{r_N(u) + a f_{N-1}^*(x - u)\right\}$$

lo cual permite resolver el problema iterando en el espacio de los estados. Si el rendimiento tiene un significado monetario, la función objetivo f^*_N representa el valor actual de los flujos financieros estudiados.

La utilización de la técnica de normalización presenta la característica que, para calcular el valor real, han de afectarse las distintas g_N del coeficiente de actualización correspondiente:

$$f_N^{*R}(i) = f_N^{'}(i) + g_N + \alpha g_{N-1} + \alpha^2 g_{N-2} + \ldots + \alpha^{N-1} g_1$$

Si deseamos estudiar el largo plazo, hemos de tener en cuenta que los valores de las f^*_N tienden hacia un valor finito, en cuyo caso es recomendable utilizar esta expresión:

$$f^*(x) = \underset{u}{OPT}\left\{r(u) + \alpha f^*(x')\right\}$$

en que se ha prescindido del subíndice que indica la etapa, dado que dicho valor no afecta a largo plazo. También puede utilizarse esta versión de la fórmula usual:

$$g + w_i = r_{ij} + \alpha w_j$$

si bien, en este caso, cabe tener en cuenta que la g, que estrictamente representa el rendimiento medio actualizado por iteración, presenta algunas dificultades de interpretación. El método permite igualmente hallar la política óptima a largo plazo, pero el valor resultante no es claro.

3.5.3 Renovación de maquinaria

Para terminar con la programación dinámica determinista, veamos un ejemplo completo, con todas las alternativas que hemos visto.

Deseamos estudiar la política de renovación de una máquina. Según la edad de la máquina, al inicio del año considerado tiene los valores siguientes (en miles de euros)

Tabla 3.12. Datos de partida del problema de renovación de maquinaria en miles de euros

Edad	**x**	0	1	2	3	4	5	6	7
Coste funcionamiento	**c(x)**	10	13	20	40	70	100	200	400
Descuento	**d(x)**		32	21	11	5	0	0	0
Valor residual	**s(x)**		25	17	8	0	0	0	0

El coste de una máquina nueva es de 50.000 euros, pero el proveedor nos hará un descuento si le entregamos la vieja a cambio, siempre que su edad no supere los 4 años. De la misma forma, al final de la gestión, podemos desprendernos de la máquina resultante en el mercado de segunda mano y obtener el valor residual que se indica.

Para plantear el modelo, podemos utilizar la técnica de la programación dinámica separable: una vez tomada la decisión, tenemos unos costes hasta el año siguiente.

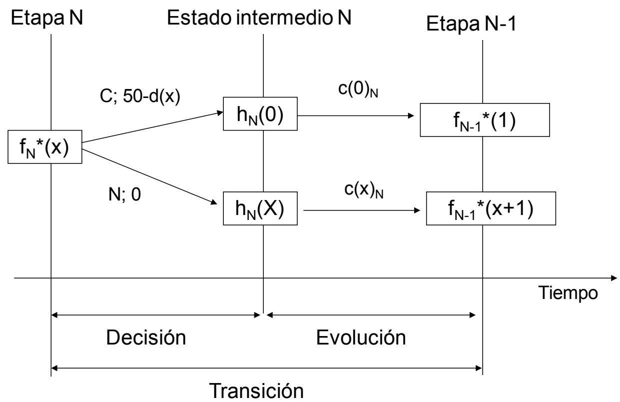


Figura 3.5. Representación de una transición en el problema de renovación de maquinaria

La parte de la transición entre el momento de la decisión y el inicio de la evolución tiene un coste que depende de lo que hayamos decidido: si cambiamos, hay el precio de la máquina menos el descuento y, si no cambiamos, el coste es cero. También la evolución es diferente: si hemos cambiado, tenemos una máquina nueva y, si no, tenemos la misma que en el momento de la decisión (ver figura 3.5).

En definitiva, las ecuaciones de recurrencia son:

$$f_N^*(x) = MIN\{50 - d(x) + h(0); h(x)\}$$

$$h(x) = c(x) + af_{N-1}^*(x)$$

$$f_0^*(x) = -s(x)$$

donde hemos indicado un coeficiente de actualización α, que de momento establecemos igual a uno. Obsérvese también que las condiciones de contorno ya no son cero, como es habitual: ahora tenemos un valor residual, que es un ingreso (por tanto, coste negativo) que se produce después de la última etapa.

En estas condiciones, veamos cómo se comporta la decisión en un horizonte de cinco años:

x	f0	h1	C	N	u1	f1	f'1	h2	C	N	u2	f2	f'2
0	0	-15		-15	N	-15	-11	10		10	N	10	-18
1	-25	-4	3	-4	N	-4	0	29	28	29	C	28	0
2	-17	12	14	12	N	12	16	48	39	48	C	39	11
3	-8	40	24	40	C	24	28	74	49	74	C	49	21
4	0	70	30	70	C	30	34	109	55	109	C	55	27
5	0	100	35	100	C	35	39	139	60	139	C	60	32
6	0	200	35	200	C	35	39	239	60	239	C	60	32
7	0	400	35	400	C	35	39	400	60	400	C	60	32
g						-4						28	

Tabla 3.13a. Solución del ejemplo renovación de maquinaria sin actualización del capital.

Tabla 3.13b. Solución del ejemplo renovación de maquinaria sin actualización del capital.

x	h3	C	N	u3	f3	f'3	h4	C	N	u4	f4	f'4	h5	C	N	u5	f5	f'5
0	10		10	N	10	-14	10		10	N	10	-18	10		10	N	10	-14
1	24	28	24	N	24	0	28	28	28	N,C	28	0	24	28	24	N	24	0
2	41	39	41	C	39	15	45	39	45	C	39	11	41	39	41	C	39	15
3	67	49	67	C	49	25	71	49	71	C	49	21	67	49	67	C	49	25
4	102	55	102	C	55	31	106	55	106	C	55	27	102	55	102	C	55	31
5	132	60	132	C	60	36	136	60	136	C	60	32	132	60	132	C	60	36
6	232	60	232	C	60	36	236	60	236	C	60	32	232	60	232	C	60	36
7	400	60	400	C	60	36	400	60	400	C	60	32	400	60	400	C	60	36
g					24						28						24	

Parece, pues, que hemos de cambiar la máquina cuando tiene una antigüedad de 2 o más años; dicho de otro modo, cada dos años, excepto en el último año de gestión, en que no la cambiaremos aunque tenga una antigüedad 2 (sí cambiaremos si tiene 3 o más años). Observemos también que el estado x = 0 es perfectamente viable como estado intermedio, pero no como estado en el momento de la decisión (salvo, quizás, como estado inicial).

Como vemos, la política se repite ya desde N = 5, que es igual a N = 3. Aunque ya sabemos que es óptima a largo plazo, lo comprobaremos con la iteración en el espacio de las políticas. Establezcamos las ecuaciones para la política {N, N, C, C, C, C, C, C}. Como ya hemos indicado, lo mejor es eliminar las expresiones de h(y):

$$
\begin{aligned}
g + w(0) &= 10 + \alpha w(1) \\
g + w(1) &= 13 + \alpha w(2) \\
g + w(2) &= 50 - 21 + 10 + \alpha w(1) \\
g + w(3) &= 50 - 11 + 10 + \alpha w(1) \\
g + w(4) &= 50 - 5 + 10 + \alpha w(1) \\
g + w(5) &= 50 + 10 + \alpha w(1) \\
g + w(6) &= 50 + 10 + \alpha w(1) \\
g + w(7) &= 50 + 10 + \alpha w(1)
\end{aligned}
$$

donde, como siempre, disponemos de un grado de libertad; haciendo w(1) = 0, tenemos:

$$g = 26;\ w(0) = -16;\ w(2) = 13;\ w(3) = 23;\ w(4) = 29;\ w(5) = w(6) = w(7) = 34$$

El lector comprobará que, efectivamente, la política es óptima y los valores coinciden con los ya hallados, teniendo en cuenta que existe un ciclo de 2 años.

Supongamos ahora que hay actualización: el valor de un año se deprecia un 10 % para el siguiente; ello quiere decir que $\alpha = 0{,}9$. Las expresiones anteriores nos valen igual simplemente cambiando el valor de α.

x	f0	h1	C	N	u1	f1	f'1	h2	C	N	u2	f2	f'2
0	0	-13		-13	N	-13	-10	10		10	N	10	-17
1	-25	-2,3	5,5	-2,3	N	-2,3	0	26,6	28	26,6	N	26,6	0
2	-17	12,8	16,5	12,8	N	12,8	15,1	45,9	39	45,9	C	39	12,4
3	-8	40	26,5	40	C	26,5	28,8	71,3	49	71,3	C	49	22,4
4	0	70	32,5	70	C	32,5	34,8	106	55	106	C	55	28,4
5	0	100	37,5	100	C	37,5	39,8	136	60	136	C	60	33,4
6	0	200	37,5	200	C	37,5	39,8	236	60	236	C	60	33,4
7	0	400	37,5	400	C	37,5	39,8	400	60	400	C	60	33,4
g						-2,3						26,6	

Tabla 3.14a. Solución del ejemplo de renovación de maquinaria con actualización

x	h3	C	N	u3	f3	f'3	h4	C	N	u4	f4	f'4	h5	C	N	u5	f5	f'5
0	10		10	N	10	-14	10		10	N	10	-16	10		10	N	10	-14
1	24,2	28	24,2	N	24,2	0	26,3	28	26,3	N	26,3	0	24,4	28	24,4	N	24,4	0
2	40,2	39	40,2	C	39	14,8	42,3	39	42,3	C	39	12,7	40,4	39	40,4	C	39	14,6
3	65,6	49	65,6	C	49	24,8	67,7	49	67,7	C	49	22,7	65,8	49	65,8	C	49	24,6
4	100	55	100	C	55	30,8	102	55	102	C	55	28,7	100	55	100	C	55	30,6
5	130	60	130	C	60	35,8	132	60	132	C	60	33,7	130	60	130	C	60	35,6
6	230	60	230	C	60	35,8	232	60	232	C	60	33,7	230	60	230	C	60	35,6
7	400	60	400	C	60	35,8	400	60	400	C	60	33,7	400	60	400	C	60	35,6
g					24,2						26,3						24,4	

Tabla 3.14b. Solución del ejemplo renovación de maquinaria con actualización (continuación).

En primer lugar, debe observarse que, aunque parezca lo contrario, todavía no hemos alcanzado la estabilidad: desde N = 3, se repite la política y parece que se repiten las $f_N'^*$, pero ello es debido a que mostramos un solo decimal. En rigor, pues, no debemos dar la política por buena, aunque podemos apostar que muy probablemente sí lo será.

Sobre el largo plazo, vamos a ver las dos versiones de la iteración en el espacio de las políticas.

Empezaremos utilizando la versión con la función sin normalizar, que, como veremos, resulta mucho más conveniente:

$$f(0) = 10 + \alpha\, f(1)$$
$$f(1) = 13 + \alpha\, f(2)$$
$$f(2) = 39 + \alpha\, f(1)$$
$$f(3) = 49 + \alpha\, f(1)$$
$$f(4) = 55 + \alpha\, f(1)$$
$$f(5) = 60 + \alpha\, f(1)$$
$$f(6) = 60 + \alpha\, f(1)$$
$$f(7) = 60 + \alpha\, f(1)$$

Este sistema no tiene ningún grado de libertad, de modo que la solución es única:

$$f^*(0) = 237,84;\ f^*(1) = 235,15;\ f^*(2) = 266,84;\ f^*(3) = 276,84;\ f^*(4) = 282,84;$$
$$f^*(5) = f^*(6) = f^*(7) = 287,84$$

que son los valores actualizados según el estado inicial de una gestión con horizonte infinito.

Veamos ahora la versión con actualización de las mismas expresiones que ya hemos visto anteriormente, simplemente utilizando el valor de $\alpha = 0,9$. Igual que antes, tenemos un grado de libertad que no afecta la g, así que haciendo w(1) = 0 obtenemos:

$$g = 25,32;\ w(0) = -15,32;\ w(2) = 13,67;\ w(3) = 23,67;\ w(4) = 29,67;$$
$$w(5) = w(6) = w(7) = 34,67$$

valores que se puede comprobar que son óptimos. Estos valores son muy similares, pero no idénticos, a los obtenidos iterando en el espacio de los estados y, de hecho, es a los que tienden si alargamos indefinidamente la iteración.

Sin embargo, el verdadero significado de estos valores no es fácil de explicar: el valor de g es el del rendimiento medio en cada anualidad en términos de la propia anualidad, pero cabe recordar que dicho valor no es constante cada año, sino que es el promedio de varios años (dos, en este caso). Por lo que respecta al significado de las w(x), todavía es menos claro. Desde luego, son diferencias motivadas por el diferente estado inicial pero, para entenderlo, hay que hacer algunos cálculos.

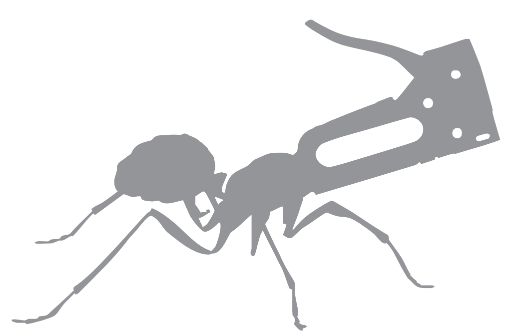
El rendimiento medio anual es $g = 25,32$. El valor actualizado de este rendimiento anual durante infinitos años es:

$$f^* = g\ /(1-\alpha) = 253,2$$

Restando este valor de las $f^*(x)$ obtenidas antes, obtenemos los valores de las

$$w(x) = f^*(x) - f^*$$

que es el significado exacto de las w(x), que, aparte de decirnos cuál es el estadio inicial más conveniente (suponiendo que se pueda elegir), es de difícil interpretación.



→ 4

Programación dinámica aleatoria

4.1 Conceptos básicos

La extensión de los conceptos vistos en programación dinámica determinista al caso aleatorio no presenta grandes dificultades conceptuales, aunque sí operativas. Esencialmente, es el mismo caso, pero ahora ni el nuevo estado ni el resultado obtenido están determinados, sino que tan solo conocemos su distribución de probabilidad:

Probabilidad del nuevo estado:

$$\mathrm{Pr}_N\left\{x_{N-1} \mid x_N, u_N\right\}$$

Rendimiento:

$$r_N = r_N(x_N, u_N, x_{N-1})$$

Igual que en el caso determinista, puede suceder que las expresiones sean las mismas o no para cada etapa, es decir, se puede tratar de un caso HT (si son iguales para todas) o NHT (si no son todas iguales).

Observemos que, en general, el rendimiento depende no solo del estado anterior y de la decisión, sino también del nuevo estado, que no sabemos cuál será. En consecuencia, el rendimiento obtenido en todas las etapas no será un valor conocido que se pueda optimizar. Lo único que podemos hacer es optimizar su valor esperado. Por tanto, la expresión general para la iteración en el espacio de las políticas será:

$$f_N^*(x) = \underset{u}{OPT} \quad EM\left\{r(x,u) \otimes f_{N-1}^*(x')\right\}$$

e igualmente por lo que respecta a la iteración en el espacio de la políticas:

$$g + w_i = \underset{u}{OPT} \quad EM\left\{r_{ij} + w_j\right\}$$

Por supuesto aquí también se pueden utilizar las técnicas que ya hemos visto, como la normalización, la programación dinámica separable y la actualización. Basta con recordar, a efectos de la interpretación de los resultados, que lo que se obtiene no son valores reales, sino valores esperados, es decir, la esperanza matemática del rendimiento.

Mención especial merece la técnica de la programación dinámica separable. Por supuesto, si es posible utilizarla en un problema dado, es optativo hacerlo o no. Sin embargo, al revés de lo que ocurre en la programación dinámica determinista, casi siempre resulta muy útil para plantar la ecuación de recurrencia.

Veamos, a través de algunos ejemplos, cómo se adapta lo que hemos visto en el caso determinista al caso aleatorio.

4.2 Planificación de inventarios

El precio de venta de un artículo determinado es de 10 euros/unidad y la demanda diaria sigue la probabilidad expresada en la tabla 4.1.

Tabla 4.1.
Ley de probabilidad de la demanda

Demanda	0	1	2	3
Probabilidad	0,4	0,3	0,2	0,1

El coste de posesión de cada unidad en *stock* al final del día es de 1 euro. Al final del día, tenemos la posibilidad de pedir una cantidad cualquiera a un proveedor, que nos la sirve al día siguiente, a primera hora de la mañana, con un coste de 2 euros/unidad, más 3 euros por costes de transporte. ¿Cuántas unidades hemos de comprar cada día para maximizar el beneficio esperado?

En primer lugar, hemos de definir los elementos del modelo de programación dinámica.

Etapas del sistema. Dadas las condiciones del problema, cada día en que se sirve la demanda constituye una etapa en la evolución del proceso.

Estados del sistema. La variable de estado del sistema ha de contener la información necesaria para describir la situación del sistema en la etapa considerada. La única circunstancia relevante a la hora de tomar la decisión de compra es saber de cuántas unidades disponemos, una vez se ha servido la demanda de la etapa anterior. En definitiva, la variable de estado del sistema es el nivel de inventario en el almacén antes de comprar las unidades adicionales y servir la demanda.

En principio, este nivel de inventario puede tomar cualquier valor entero no negativo. Sin embargo, más adelante veremos que, en la práctica, existe una cota superior para dicho nivel de inventario.

Variables de decisión. Claramente, la decisión que hay que tomar es la cantidad de unidades a comprar. Entre las unidades compradas y las unidades disponibles (inventario inicial), se intentará cubrir la demanda. En general, tenemos la expresión clásica de los *stocks*:

Stock inicial + Unidades compradas = Demanda satisfecha + *Stock* final

En este punto, cabe realizar varias observaciones:

– En la ecuación, se habla de demanda satisfecha y no de demanda total. Si, entre el inventario inicial y las unidades compradas, no podemos cubrir la demanda, quedará demanda por satisfacer y se perderán ingresos que podrían haberse obtenido de comprar más cantidad.
– La máxima demanda es de tres unidades, por lo cual puede parecer razonable plantear que inventario inicial + unidades compradas ≤ 3. Sin embargo, no lo es, ya que existen unos costes fijos por aprovisionar y puede ser interesante soportar un coste de posesión del *stock* a cambio de no tener que aprovisionar tan frecuentemente. En este caso, no es así porque el coste fijo es relativamente pequeño.

Puesto que la demanda sigue una ley de probabilidad, la evolución del sistema es aleatoria. El inventario final, que representará el estado hacia el que evolucionará el sistema, también seguirá una evolución aleatoria. Formalmente, como puede quedar demanda sin cubrir, el estado al cual evolucionará el sistema vendrá dado por:

Inventario final = MAX {Inventario inicial + Unidades compradas – Demanda; 0}

Función de recurrencia. La función de recurrencia debe expresar el beneficio esperado. En este caso, la función es de tipo aditivo:

Beneficio esperado en las N últimas etapas antes del final = Beneficio esperado en la etapa N + Beneficio esperado en las N–1 últimas etapas

Para plantear correctamente el modelo, será útil dividir cada etapa en dos partes (programación dinámica separable):

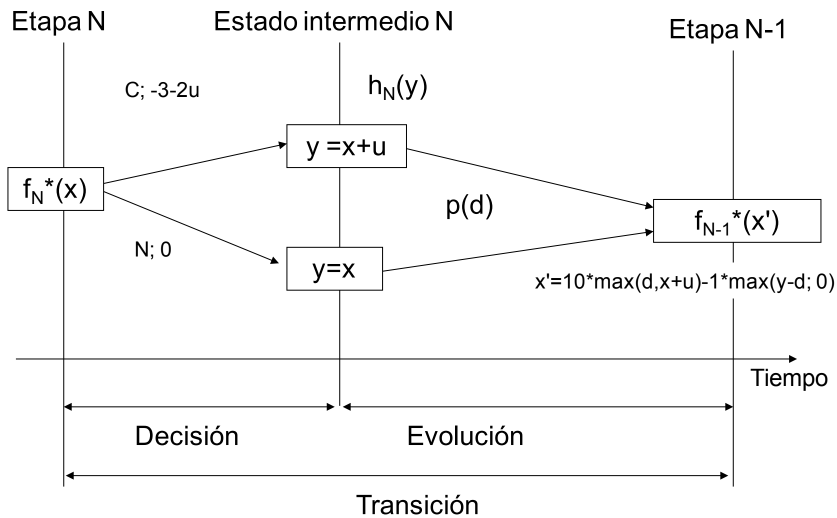


Figura 4.1. Representación de una transición en el problema de planificación de inventarios

con lo cual podemos plantear el sistema siguiente:

$$f_N^*(x) = MAX\left\{h_N(y); \max_u\left[-3-2u+h(x+u)\right]\right\}$$

$$h(y) = \sum_{d=0}^{y} p(d)\left[10d - 1(y-d) + f_{N-1}^*(y-d)\right] + \sum_{d=y+1}^{\infty} p(d)\left[10y + f_{N-1}^*(0)\right]$$

donde las dos alternativas en la expresión de f_N*(x) indican la decisión de no comprar o sí y, en este caso, cuánto (u). Una vez tomada la decisión, nos encontramos con la situación representada por la función del estado intermedio h(y), que es el valor esperado después de la decisión de comprar o no (y, en su caso, cuánto), tras la cual el sistema evoluciona hacia la nueva transición. Por supuesto, los dos sumatorios indican que, si la demanda es superior al *stock* disponible, solo podremos vender lo que tenemos y el *stock* final será cero.

En este caso, se trata de un problema de programación dinámica homogénea. Por esta razón, podemos plantear el problema tanto para un número determinado de semanas (horizonte finito) como para el largo plazo (horizonte infinito). Primero, plantearemos el caso del horizonte finito.

La tabla correspondiente a los cálculos será indefinidamente larga ya que, en teoría, podríamos tener cualquier cantidad de unidades en el almacén; mostraremos hasta x = 8 solamente, lo que debe ser suficiente para ver el comportamiento del sistema para valores mayores de x en el corto plazo:

		N=1			u1					N=2			u2				
x	f*0	h1	0	1	2	3	4	f*1	f'*1	h2	0	1	2	3	4	f*2	f'*2
0	0	0	0	0,6	**0,9**	-1	-4	0,9	0	0	0	2,48	**5,11**	4,88	2,44	5,11	0
1	0	5,6	**5,6**	2,9	1	-2	-5	5,6	4,7	7,48	**7,48**	7,11	6,88	4,44	0,99	7,48	2,37
2	0	7,9	**7,9**	3	0	-3	-6	7,9	7	12,1	**12,1**	8,88	6,44	2,99	-0,9	12,1	7
3	0	8	**8**	2	-1	-4	-7	8	7,1	13,9	**13,9**	8,44	4,99	1,1	-2,9	13,9	8,77
4	0	7	**7**	1	-2	-5	-8	7	6,1	13,4	**13,4**	6,99	3,1	-0,9	-4,9	13,4	8,33
5	0	6	**6**	0	-3	-6	-9	6	5,1	12	**12**	5,1	1,1	-2,9	-6,9	12	6,88
6	0	5	**5**	-1	-4	-7	-10	5	4,1	10,1	**10,1**	3,1	-0,9	-4,9	-8,9	10,1	4,99
7	0	4	**4**	-2	-5	-8	-11	4	3,1	8,1	**8,1**	1,1	-2,9	-6,9	-11	8,1	2,99
8	0	3	**3**	-3	-6	-9	-12	3	2,1	6,1	**6,1**	-0,9	-4,9	-8,9	-13	6,1	0,99
g								0,9								5,11	

	N=3			u_3					N=4			u_4				
x	h_3	0	1	2	3	4	f^*_3	f'^*_3	h_4	0	1	2	3	4	f^*_4	f'^*_4
0	0	0	1,55	4,41	**5,08**	3,6	5,08	0	0	0	1,4	4,03	**4,9**	3,97	4,9	0
1	6,55	6,55	6,41	**7,08**	5,6	2,71	7,08	2	6,4	6,4	6,03	**6,9**	5,97	3,74	6,9	2
2	11,4	**11,4**	9,08	7,6	4,71	0,6	11,4	6,33	11	**11**	8,9	7,97	5,74	2	11	6,13
3	14,1	**14,1**	9,6	6,71	2,6	-2,1	14,1	9	13,9	**13,9**	9,97	7,74	4	-0,8	13,9	9
4	14,6	**14,6**	8,71	4,6	-0,1	-5	14,6	9,52	15	**15**	9,74	6	1,16	-4,3	15	10,1
5	13,7	**13,7**	6,6	1,9	-3	-8	13,7	8,62	14,7	**14,7**	8	3,16	-2,3	-8,2	14,7	9,84
6	11,6	**11,6**	3,9	-1	-6	-11	11,6	6,52	13	**13**	5,16	-0,3	-6,2	-12	13	8,1
7	8,9	**8,9**	0,98	-4	-9	-14	8,9	3,82	10,2	**10,2**	1,67	-4,2	-10	-16	10,2	5,26
8	5,98	**5,98**	-2	-7	-12	-17	5,98	0,9	6,67	**6,67**	-2,2	-8,1	-14	-20	6,67	1,77
g							5,08								4,9	

Como puede observarse, superado el estado transitorio de las dos primeras semanas, se repite la política de "pedir hasta 3 si hay 1 o menos", un tipo de política muy habitual en la gestión de *stocks*. El beneficio medio por semana, sin embargo, está aún lejos de estabilizarse, de modo que podemos decir poco de su valor a largo plazo.

Por supuesto, ahora ya no podemos decir cuál será la evolución del sistema una vez haya empezado ni cuál será el beneficio real, dado que el valor es aleatorio. Solamente podemos decir que la política indicada es óptima para N = 3 o N = 4 (es decir, cuando faltan 3 o 4 semanas para terminar), que en N = 1 y N = 2 (las dos últimas semanas de gestión) la política es pedir 2 unidades, si nos hemos quedado sin *stock*, y no pedir nada en caso contrario, y que, con todo ello, si, a 4 semanas del final, el estado inicial es de 5 unidades, por ejemplo, el beneficio esperado será:

$$f_4{}^{*R}(5) = 9{,}84 + 4{,}90 + 5{,}08 + 5{,}11 + 0{,}90 = 25{,}83$$

Veamos si la política hallada es óptima a largo plazo. Planteemos el sistema correspondiente para iterar en el espacio de las políticas. Al revés de lo que ocurre en programación dinámica determinista, ahora es conveniente mantener en el sistema las funciones del estado intermedio:

w(0) + g = -9 + h(3)	h(0) = w(0)
w(1) + g = -7 + h(3)	h(1) = 0,4 [-1 +w(1)]+0,6[10+w(0)]
w(2) + g = h(2)	h(2) = 0,4 [-2 +w(2)]+0,3[9+w(1)]+0,3[20+w(0)]
w(3) + g = h(3)	h(3) = 0,4 [-3 +w(3)]+0,3[8+w(2)]+0,2[19+w(1)]+0,1[30+w(0)]
w(4) + g = h(3)	h(4) = 0,4 [-4 +w(5)]+0,3[7+w(3)]+0,2[18+w(3)]+0,1[29+w(1)]
w(5) + g = h(5)	h(5) = 0,4 [-5 +w(5)]+0,3[6+w(4)]+0,2[17+w(3)]+0,1[28+w(2)]
etc.	etc.

Como siempre, podemos fijar una w(x), por ejemplo, w(0) = 0, con lo cual obtenemos:

$$g = 4{,}83$$

$$w(0) = 0;\ w(1) = 2;\ w(2) = 6{,}11;\ w(3) = 9;\ w(4) = 10{,}5;\ w(5) = 11{,}2$$

$$h(0) = 0;\ h(1) = 6{,}4;\ h(2) = 10{,}9;\ h(3) = 13{,}8;\ h(4) = 15{,}3;\ h(5) = 16$$

valores que, aunque con ciertas dificultades de cálculo, debidas a los decimales que aparecen, podemos comprobar que son óptimos.

En este punto, es importante recordar que los problemas homogéneos en el tiempo, como en este caso, pueden tratarse también como cadenas de Markov con remuneración y decisión (CMRD), que se estudian en otro volumen de esta colección, al cual remitimos al lector.

Las CMRD permiten un tratamiento numérico mucho más cómodo pero, por el contrario, la facilidad de planteamiento que permite la programación dinámica separable hace que la creación del modelo sea mucho más simple. Por ello, lo más recomendable es este tipo de planteamiento, especialmente a largo plazo.

4.3 El caso del camionero

Un camionero que hace el transporte entre dos ciudades, A y B, tiene tres clientes fijos: P y Q en A, y R en B. El viaje de una ciudad a otra tiene un coste de 40 euros y consume un día completo, por lo cual ha de quedarse a dormir en la ciudad de destino. Al día siguiente, se pone en contacto con sus clientes de la ciudad donde se encuentra para ver si tienen carga para él. Estadísticamente, sabe que el cliente P tiene carga el 25 % de los días, mientras que tanto el cliente Q como el cliente R la tienen el 40 %. Según los contratos que tiene con ellos, la facturación por viaje es de 180 euros al cliente P, 100 euros al cliente Q y 120 euros al cliente R.

Cada día, tanto si hay carga como si no, el camionero puede decidir si hace el viaje o no. Por supuesto, si decide viajar y hay carga, la llevará, pero puede también decidir viajar de vacío si no hay carga o quedarse aunque haya carga. Si decide quedarse en la ciudad, el coste es de 10 euros si se halla en A y de 20 euros si se halla en B.

En este problema, se pone de manifiesto la importancia de definir cuidadosamente el momento en que cambiamos de etapa para una buena modelización. Veámoslo con detalle. Evidentemente, cada etapa es un día en que puede hacer un viaje o no (la decisión). Pero la decisión la toma sabiendo si hay carga o no, lo cual quiere decir que la división de la etapa en dos partes tiene dos alternativas:

- El cambio de etapa se define a última hora de la noche, antes de saber si mañana habrá carga o no, con lo cual se trata de un caso de evolución/decisión.
- El cambio de etapa se define al día siguiente por la mañana, cuando ya sabe si hay carga o no, y toma la decisión de viajar o no; en este caso, el modelo será de decisión/evolución.

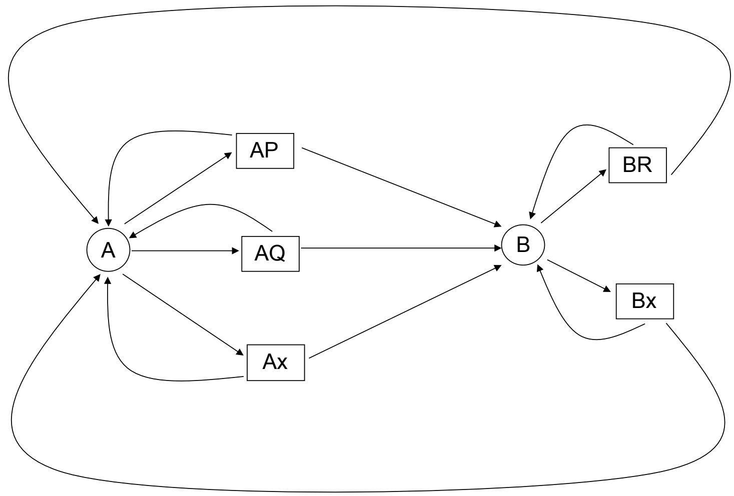
Antes de decidirnos por uno u otro, representemos gráficamente la situación en la figura 4.2.

Hemos representado con un círculo los estados A y B donde se inicia la evolución, que es aleatoria; dichos estados representan la situación por la noche, cuando el camionero solo sabe dónde se halla, pero no si mañana habrá carga o no.

La evolución llevará a los estados posibles en esa ciudad.

- AP representa que el camionero se halla en A y hay carga en el cliente P. La transición A – AP tiene, pues, una probabilidad de 0,25.
- AQ representa que el camionero se halla en A y hay carga en el cliente Q y no la hay en P: en efecto, si hay carga en los dos, preferirá P, puesto que la facturación es mayor; la transición A – AQ tendrá, pues, una probabilidad de (1 – 0,25)* 0,40 = 0,30.
- Ax representa que el camionero se halla en A y no hay carga ni en el cliente P ni el Q; la transición tiene una probabilidad de 1 – 0,25 – 0,30 = 0,45.
- BR representa que el camionero se halla en B y hay carga en el cliente R. La transición B – BR tiene, pues, una probabilidad de 0,40.
- Bx representa que el camionero se halla en B y no hay carga en el cliente R. La transición B – Bx tiene, pues, una probabilidad de 1 – 0,40 = 0,60.

Figura 4.2.
Evolución de los estados posibles del caso del camionero



La función objetivo es, claramente, maximizar el beneficio, de manera que, una vez tomada la decisión de viajar (V) o no (N), las transiciones tendrán las remuneraciones que se indican en la tabla 4.2.

Tabla 4.2.
Remuneraciones del caso del camionero

De \ a	A	B
AP	-20 (N)	140 (V)
AQ	-20 (N)	60 (V)
Ax	-20 (N)	- 40 (V)
BR	80 (V)	-10 (N)
Bx	- 40 (V)	-10 (N)

Podemos preguntarnos si tienen sentido las decisiones de viajar de vacío, aunque el coste de permanecer donde esté sea menor. Más allá del debate, es importante observar que, al crear el modelo, no deben menospreciarse las opciones que no sean muy claramente absurdas puesto que, en ciertas condiciones (como en este caso, según veremos), resultan óptimas; en todo caso, ya las eliminará el propio modelo si realmente no son buenas.

Observemos que, según definamos el modelo E/D o D/E, los estados de cambio de etapa e intermedio serán distintos (v. tabla 4.3).

Tabla 4.3.
Estados de cambio de etapa e intermedios en el caso del camionero

	Estados	
Modelo	**Cambio de etapa**	**Intermedios**
D / E	AP, AQ, Ax, BR, Bx	A B
E / D	A B	AP, AQ, Ax, BR, Bx

Es hora de decidir qué modelo queremos. Cualquiera de los dos es bueno y los dos dan idénticos resultados: nos decidimos por un modelo E/D. En estas condiciones, las ecuaciones de recurrencia serán:

$$
\begin{aligned}
&f_N(A) = 0{,}25h_N^*(AP) + 0{,}30h_N^*(AQ) + 0{,}45h_N^*(Ax)\\
&f_N(B) = 0{,}40h(BR) + 0{,}60h(Bx)\\
&h_N^*(AP) = MAX\left\{140 + f_{N-1}(B); -10 + f_{N-1}(A)\right\}\\
&h_N^*(AQ) = MAX\left\{60 + f_{N-1}(B); -10 + f_{N-1}(A)\right\}\\
&h_N^*(Ax) = MAX\left\{-40 + f_{N-1}(B); -10 + f_{N-1}(A)\right\}\\
&h_N^*(BR) = MAX\left\{80 + f_{N-1}(A); -20 + f_{N-1}(B)\right\}\\
&h_N^*(Bx) = MAX\left\{-40 + f_{N-1}(A); -20 + f_{N-1}(B)\right\}
\end{aligned}
$$

Observemos que ahora hemos indicado el superíndice * no en las funciones f_N, sino en las h_N, ya que es aquí donde se toma la decisión, aunque ello sea a mitad de la transición.

Vamos a resolver el problema. Empezaremos suponiendo que el camionero tiene familia en A y, por tanto, el viernes por la noche debe terminar su actividad en esa ciudad para pasar el fin de semana con la familia. Esto significa un horizonte limitado con N = 5 (los cinco días laborables) y que la decisión del viernes ya está tomada: si se halla en B, viajará con carga o sin ella y, si se halla en A, no viajará en ningún caso. Iteremos en el espacio de los estados utilizando la normalización para hallar la política óptima Y:

		Viernes				Jueves				Miércoles			
x	y	f_1	u^*_1	h^*_1	h'^*_1	f_2	u^*_2	h^*_2	h'^*_2	f_3	u^*_3	h^*_3	h'^*_3
A	AP	0	N	-10	30	30	V	188	168	68,4	V	180,8	152,4
	AQ		N	-10	30		V	100	80		V	100,8	72,4
	Ax		N	-10	30		N	20	0		N	58,8	30
B	BR	0	V	80	120	48	V	110	90	40,8	V	148,4	120
	Bx		V	-40	0		V	28	8		V	28,4	0
	g				-40				20				28,4

Tabla 4.4.a. Solución del caso del camionero iterando en el espacio de los estados

		Martes				Lunes				Dom
X	y	f_4	u^*_4	h^*_4	h'^*_4	f_5	u^*_5	h^*_5	h'^*_5	f_6
A	AP	73,32	V	188	154,68	74,57	V	188	153,43	73,89
	AQ		V	108	74,68		V	108	73,43	
	Ax		N	63,32	30		N	64,57	30	
B	BR	48	V	153,32	120	48	V	157,57	120	48
	Bx		V	33,32	0		V	34,57	0	
	g				33,32				34,57	

Tabla 4.4.b. Solución del caso del camionero iterando en el espacio de los estados (continuación)

Por tanto, con esta política, y dado que el camionero pasa el fin de semana en A con la familia, al acostarse el domingo tendrá una ganancia esperada para la semana próxima de:

$$\text{Ganancia esperada} = 73.89+34,57+33,32+28,4+20-40 = 150,18\ €$$

El lunes por la mañana, su beneficio semanal esperado subirá unos 80 euros, si hay carga en P; se mantendrá, si no hay carga en P pero sí en Q, o disminuirá en 43 euros, si ese lunes no hay carga en ninguno de ambos clientes.

En todo caso, la política óptima (excepto el viernes) es viajar, salvo que esté en A y no haya carga.

Veamos ahora qué ocurriría si no hubiese la condición de pasar el fin de semana en A. Ello significa que al camionero le es indiferente donde pase el fin de semana (cuyo coste supondremos nulo) y, por tanto, se transforma en un problema con horizonte ilimitado.

La iteración en el espacio de los estados ya nos sugiere que la política, en este caso, será {V, V, N, V, V}. Veámoslo.

Para iterar en el espacio de las políticas, hemos de situarnos en el momento de la decisión, es decir, en el estadio intermedio de la transición. En estas condiciones, la política indicada implica que las ecuaciones serán:

$$\begin{aligned}
w(AP) + g &= 140 + 0,4\ w(BR) + 0,6\ w(Bx)\\
w(AQ) + g &= 60 + 0,4\ w(BR) + 0,6\ w(Bx)\\
w(Ax) + g &= -10 + 0,25\ w(AP) + 0,30\ w(AQ) + 0,45\ w(Ax)\\
w(BR) &= 80 + 0,25\ w(AP) + 0,30\ w(AQ) + 0,45\ w(Ax)\\
w(Bx) &= -40 + 0,25\ w(AP) + 0,30\ w(AQ) + 0,45\ w(Ax)
\end{aligned}$$

cuya solución haciendo $w(Bx) = 0$ es:

$$g = 34,13;\ w(AP)=153,87;\ w(AQ)=73,87;\ w(Ax)=30;\ w(BR)=120$$

Comprobemos si es óptima:

$AP \Rightarrow \text{MAX}\{140+0,4w(BR)+0,6w(Bx); -10+0,25w(AP)+0,30\ w(AQ)+0,45\ w(Ax)\} - g$

$= \text{MAX}\ (153,87; 30) = 153,87 = w(AP) \Rightarrow \text{OK}$

$AQ \Rightarrow \text{MAX}\ \{60+0,4w(BR)+0,6w(Bx); -10+0,25\ w(AP)+0,30w(AQ)+0,45w(Ax)\} - g$

$= \text{MAX}\ (73,87; 30) = 73,87 = w(AQ) \Rightarrow \text{OK}$

Ax $\Rightarrow$ MAX {-40+0,4w(BR)+0,6w(Bx); -10+0,25w(AP)+0,30w(AQ)+0,45w(Ax)} – g

= MAX (-26,13; 30) = 30= w(Ax) $\Rightarrow$ OK

BR $\Rightarrow$ MAX {80+0,25w(AP)+0,30w(AQ)+0,45w(Ax); -20+0,4w(BR)+0,6w(Bx)} – g

= MAX (120; - 6,13) = 120 = w(BR) $\Rightarrow$ OK

Bx $\Rightarrow$ MAX {-40+0,25w(AP)+0,30w(AQ)+0,45w(Ax); -20+0,4w(BR)+0,6 w(Bx)} – g

= MAX (0; - 6,13) = 0 = w(Bx) $\Rightarrow$ OK

La política es óptima a largo plazo y el beneficio medio diario es de 34,13 euros/día.

4.4 El caso del tahúr

Un tahúr afirma que tiene un método para ganar en la ruleta del casino que le permite, con una probabilidad de 2/3, doblar la apuesta y, de 1/3, de perder lo apostado. Como tiene deudas con determinadas organizaciones poco recomendables, ha hecho con ellas el trato siguiente: ya que su capital actual es de 3.000 euros, irá al casino, comprará tres fichas de 1.000 euros y si, al cabo de solo tres jugadas, es capaz de entregar al menos 5.000 euros en fichas, la deuda quedará saldada; pero si no logra obtener este objetivo, la deuda se saldará mediante un baño del tahúr en el río, convenientemente equipado con unos zapatos de cemento. ¿Qué debe hacer el tahúr?

Para resolver el problema mediante programación dinámica, supondremos que huir no es una opción y que es cierto lo que afirma el tahúr. Está claro que se trata de un problema con horizonte finito (N = 3), pero ahora, contra lo que es usual, el dinero no es el rendimiento: no se trata de ganar muchas fichas, sino de salvar la vida. El tahúr quiere maximizar la probabilidad de poder saldar la deuda, y lo conseguirá si al final del proceso logra tener 5 fichas o más. Por tanto, el número de fichas que tenga en la mano no será el rendimiento, sino el estado del sistema. Y también están claras las condiciones de contorno: si tiene 5 fichas o más gana y si tiene 4 o menos pierde; por tanto, la probabilidad de salvar la vida será 1 si el estado final es 5 o más, y 0 en caso contrario.

La ecuación de recurrencia también será un tanto peculiar, ya que se trata de probabilidades: ahora la función con que se agregan no es la suma, sino la multiplicación, es decir:

$$f_n(x) = MAX_u \{2/3\ f_{n-1}\ (x+u) + 1/3\ f_{n-1}\ (x-u)\}$$

Claramente, la decisión será el número de fichas que tiene que apostar, que puede ser 0 (no apostar) o un número inferior o igual al estado del sistema. Es evidente que, si antes de llegar a la tercera jugada ya tiene 5 fichas o más, la estrategia óptima es no apostar y podemos asimilar todos los valores del 5 en adelante como un único estado.

Calculemos las dos últimas transiciones iterando en el espacio de los estados:

Tabla 4.5. Solución del caso del tahúr iterando en el espacio de los estados

x	f^*_0	u=0	u=1	u=2	u=3	u=4	f^*_1	u=0	u=1	u=2	u=3	u=4	f^*_2
0	0	0					0	0					0
1	0	0	0				0	0	0				0
2	0	0	0	0			0	0	4/9	4/9			4/9
3	0	0	0	2/3	2/3		2/3	2/3	4/9	2/3	2/3		2/3
4	0	0	2/3	2/3	2/3	2/3	2/3	2/3	8/9	2/3	2/3	2/3	8/9
5	1	1	<1	<1	<1	<1	1	1	<1	<1	<1	<1	1

Para N = 3, basta con calcular el caso x = 3, ya que es así como empezaremos:

x	u = 0	u = 1	u = 2	u = 3	u = 4	f*3
3	2/3	20/27	2/3	2/3		20/27

Observemos algunas peculiaridades de la solución:

El tahúr tiene una probabilidad bastante alta (20/27 ≈ 74 %) de salir bien librado, pero ha de seguir una estrategia muy concreta:

- En la primera tirada (N = 3), ha de apostar solamente una ficha ya que, si apuesta más y falla, ya no podrá recuperarse.
- Si en la primera tirada gana, su probabilidad de sobrevivir habrá aumentado hasta el 89 % y deberá volver a jugar una sola ficha. Por el contrario, si pierde, la probabilidad habrá disminuido hasta el 44 %, y tanto da que apueste una o las dos fichas que tiene.
- Por fin, en la última tirada (N = 1), solo deberá apostar si tiene 3 o 4 fichas (de lo contrario, ya sabrá si ha ganado o ha perdido), y en ambos casos tendrá diversas opciones equivalentes: apostar 2 o 3 fichas si tiene 3, y cualquier número si tiene 4.

Obsérvese que la situación (y, por tanto, la estrategia) sería muy distinta si el objetivo fuese maximizar la ganancia, o sea, el número de fichas.

Veamos ahora qué ocurre si el tahúr consigue convencer a sus acreedores de que le dejen jugar tantas veces como haga falta hasta que se quede sin fichas (y, por tanto, se vaya directo al río) o bien consiga reunir 5 o más fichas (y, por tanto, pueda saldar su deuda). Ello quiere decir que el problema se ha transformado en uno con horizonte ilimitado, cuya estrategia para $x \geq 5$ es no jugar, con $f^*(x) = 1$, y para $x = 0$ también es no jugar, pero con $f^*(0) = 0$.

Examinemos los otros casos con la política $u(x) = 1$ para $x = 1, 2, 3, 4$ sugerida por la solución a corto plazo:

$$\begin{array}{lcl} f^*(1) = 2/3\, f^*(2) & & f^*(1) = 16/13 = 0{,}52 \\ f^*(2) = 1/3\, f^*(1) + 2/3\, f^*(3) & & f^*(2) = 24/31 = 0{,}77 \\ f^*(3) = 1/3\, f^*(2) + 2/3\, f^*(4) & \Rightarrow & f^*(3) = 28/31 = 0{,}90 \\ f^*(4) = 1/3\, f^*(3) + 2/3 & & f^*(4) = 30/31 = 0{,}97 \end{array}$$

Veamos si es óptima, descartando la opción de no jugar:

$$f^*(1) = \text{MAX}\ \{2/3\, f^*(2)\} = 0{,}52 \Rightarrow \text{OK}$$

$$f^*(2) = \text{MAX}\ \{\ 1/3\, f^*(1)+2/3\, f^*(3);\ 2/3\, f^*(4)\ \} = \text{MAX}\ (0{,}77;\ 0{,}65) = 0{,}77 \Rightarrow \text{OK}$$

$$f^*(3) = \text{MAX}\ \{1/3\, f^*(2)+2/3\, f^*(4);\ 1/3\, f^*(1)+2/3\, f^*(4);\ 2/3\} \\ = \text{MAX}\ (0{,}90;\ 0{,}84;\ 0{,}67) = 0{,}90 \Rightarrow \text{OK}$$

$$f^*(4) = \text{MAX}\ \{1/3\, f^*(3) +2/3;\ 1/3\, f^*(2) +2/3;\ 1/3\, f^*(1) +2/3; 12/3\ \} \\ = \text{MAX}\ (0{,}97;\ 0{,}92;\ 0{,}67;\ 0{,}67) = 0{,}97 \Rightarrow \text{OK}$$

Por tanto, la política es óptima y la probabilidad de que el tahúr pueda saldar su deuda si empieza con tres fichas es del 90 %. Se puede observar que, en este caso, mientras tenga alguna ficha, su probabilidad de ganar la apuesta se mantendrá siempre por encima del 50 %.

4.5 Gestión de stocks

4.5.1 Definición

En este apartado, vamos a tratar un modelo muy general de gestión de *stocks* de indudable interés práctico. Empecemos definiendo la notación, que al mismo tiempo nos servirá para enunciar el caso.

I_N = *stock* al final del mes N (para simplificar, hablaremos de meses, pero el modelo sirve igual si son días, semanas o cualquier otro período)

CL = coste fijo de aprovisionamiento (euros/lote)

CA = coste unitario de aprovisionamiento (euros/unidad)

CS = coste de posesión del *stock*, cargado sobre la cantidad I_N, si esta es positiva (euros/unidad)

F_N^* (I) = coste de la gestión del *stock* en los últimos N meses, antes de acabar

S_N = nivel de *stock* con que iniciamos el mes N. La decisión será, pues, no aprovisionar (así que $S_N = I_N$) o bien aprovisionar la diferencia ($S_N - I_N > 0$).

$p_N(x)$ = probabilidad de la demanda (x) en el mes N

$H_N(S)$ = función de coste en el estado intermedio, una vez tomada la decisión.

α = coeficiente de actualización, si procede ($0 < \alpha \leq 1$)

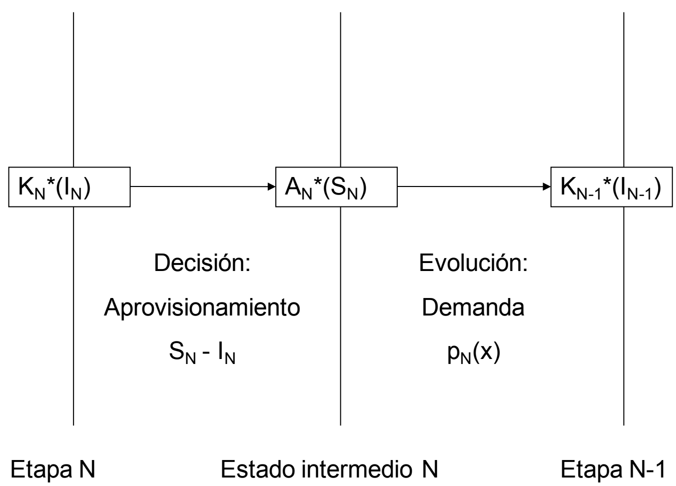
Distinguiremos dos casos:

1. Demanda insatisfecha perdida: si la demanda es superior al *stock*, se pierde, con un coste de rotura (CR) (euros/unidad). Por tanto, siempre será $I_N \geq 0$.
2. Demanda insatisfecha diferida: si la demanda es superior al *stock*, queda el material pendiente de servir, lo cual se hará con el material que llegue al inicio del mes siguiente; dicho retraso tiene un CR (euros/unidad). Como acaso debamos material a fin de mes, I_N puede ser positivo, nulo o negativo.

4.5.2 Modelos

Con estas definiciones, está claro que se trata de un modelo separable D/E no homogéneo en el tiempo.

Figura 4.3.
Evolución de los estados en el caso de gestión de *stocks*



Para escribir la ecuación de recurrencia, empezamos definiendo una función auxiliar G(S):

$$G(S_N) = CS\sum_{x=0}^{S_N} p_N(x)\left(S_N - x\right) + CR \sum_{x=S_N+1}^{\infty} p_N(x)\left(x - S_N\right)$$

que nos permite escribir fácilmente la relación entre la función intermedia $H_N(S)$ y la función objetivo del mes siguiente, que será distinta según el tratamiento de la ruptura. En caso de demanda perdida:

$$H_N(S_N) = G(S_N) + \alpha \sum_{x=0}^{S_N} p_N(x) F_{N-1}^*(S_N - x) + \alpha \sum_{x=S_N+1}^{\infty} p_N(x) F_{N-1}^*(0)$$

Y, en caso de demanda diferida:

$$H_N(S_N) = G(S_N) + \alpha \sum_{x=0}^{\infty} p_N(x) F_{N-1}^*(S_N - x)$$

con lo cual estamos en condiciones de establecer la recurrencia:

$$F_N^*(I_N) = MIN\left\{ H_N(I_N); \min_{S_N > I_N} \left[CL + CS(S_N - I_N) + H_N(S_N) \right] \right\}$$

Dado que se trata de un modelo no homogéneo, no se puede, en general, plantear el caso de horizonte ilimitado. Si existe un valor residual de *stock*, debe usarse para establecer la F_N* A fin de simplificar el cálculo y la interpretación de los resultados, suele ser recomendable normalizar con F_N*(0) = 0, salvo que las condiciones de contorno F_0* sean negativas (existe valor residual), en cuyo caso es más recomendable normalizar por el valor mínimo.

Vamos a aplicar el modelo al caso de la demanda perdida, que se expone a continuación.

4.5.3 Caso de demanda perdida

El almacén central de distribución debe gestionar el *stock* de un cierto producto de temporada durante los próximos tres meses de verano. Dado que se trata de aprovisionar al por mayor, la unidad logística es el palé. El coste de adquisición de un palé es de 200 euros, más 500 euros por lote de coste de transporte, independientemente de la cantidad de palés transportados. Los palés que queden en el almacén al final de temporada se podrán saldar a un precio de 50 euros/palé. Puesto que solamente se puede aprovisionar a principio de cada mes, la demanda insatisfecha se pierde, lo cual se valora con un coste de oportunidad de 1.000 euros/palé.

La demanda estimada para cada mes se indica en la tabla 4.6. Para cada cantidad (entre 0 y 8 palés), en la tabla se expresa la probabilidad de que esta sea la demanda durante el mes.

Tabla 4.6.
Datos de demanda estimada mensual en el caso de gestión de *stocks* con demanda perdida

Demanda	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Junio	0,40	0,30	0,20	0,10	0,00	0,00	0,00	0,00	0,00
Julio	0,05	0,10	0,10	0,15	0,20	0,15	0,10	0,10	0,05
Agosto	0,20	0,20	0,30	0,10	0,10	0,10	0,00	0,00	0,00

Dado que, como es habitual, numeramos las etapas en función de las que quedan antes de acabar, N = 1 corresponderá al mes de agosto; julio será N = 2, y junio será N = 3.

Calculemos, ante todo, la función auxiliar G(S), para lo cual es útil la tabla 4.7.

Tabla 4.7.
Función G(S)

				S											
Dem.	**N= 1**	**N= 2**	**N= 3**	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	
0	0,20	0,05	0,40	0	20	40	60	80	100	120	140	160	180	200	
1	0,20	0,10	0,30	1.000	0	20	40	60	80	100	120	140	160	180	
2	0,30	0,10	0,20	2.000	1.000	0	20	40	60	80	100	120	140	160	
3	0,10	0,15	0,10	3.000	2.000	1.000	0	20	40	60	80	100	120	140	
4	0,10	0,20	0,00	4.000	3.000	2.000	1.000	0	20	40	60	80	100	120	
5	0,10	0,15	0,00	5.000	4.000	3.000	2.000	1.000	0	20	40	60	80	100	
6	0,00	0,10	0,00	6.000	5.000	4.000	3.000	2.000	1.000	0	20	40	60	80	
7	0,00	0,10	0,00	7.000	6.000	5.000	4.000	3.000	2.000	1.000	0	20	40	60	
8	0,00	0,05	0,00	8.000	7.000	6.000	5.000	4.000	3.000	2.000	1.000	0	20	40	
9	0,00	0,00	0,00	9.000	8.000	7.000	6.000	5.000	4.000	3.000	2.000	1.000	0	20	
10	0,00	0,00	0,00	10.000	9.000	8.000	7.000	6.000	5.000	4.000	3.000	2.000	1.000	0	
D.esp	**2,00**	**4,00**	**1,00**	**2.000**	**1.204**	**612**	**326**	**142**	**60**	**80**	**100**	**120**	**140**	**160**	**G_1**
				4.000	**3.051**	**2.204**	**1.459**	**867**	**479**	**244**	**111**	**80**	**100**	**120**	**G_2**
				1.000	**408**	**122**	**40**	**60**	**80**	**100**	**120**	**140**	**160**	**180**	**G_3**

Dado que existe un valor residual, será un coste negativo que aplicaremos al *stock* sobrante al final de agosto. Por tanto, y puesto que se trata de un caso de demanda perdida, la gestión correspondiente a agosto será:

		N= 1	S										-240= g	
I/S	**F*0**	**H1**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**F*1**	**F'*1**
0	0	2000	1894	1482	1361	**1300**	2461	2637	2910	3280	3650	4020	1300	1540
1	-50	1194		1282	1161	**1137**	1210	1380	1550	1720	1890	2060	1137	1377
2	-100	**582**			961	937	1010	1180	1350	1520	1690	1860	582	822
3	-150	**261**				737	810	980	1150	1320	1490	1660	261	501
4	-200	**37**					610	780	950	1120	1290	1460	37	277
5	-250	**-90**						580	750	920	1090	1260	-90	150
6	-300	**-120**							550	720	890	1060	-120	120
7	-350	**-150**								520	690	860	-150	90
8	-400	**-180**									490	660	-180	60
9	-450	**-210**										460	-210	30
10	-500	**-240**											-240	0

Tabla 4.8. Decisiones que hay que tomar en el mes de agosto

Es decir, en agosto pediremos hasta el nivel de 4 palés si tenemos uno o ninguno, y no pediremos nada en caso contrario.

La gestión de julio será:

	N= 2	S										307,9= g	
I/S	**H2**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**F*2**	**F'*2**
0	5540	5283	4592	3959	3444	3079	2824	2676	**2654**	2707	2808	2653,95	2346,05
1	4582,85		4392	3759	3244	2879	2624	2476	**2454**	2507	2608	2453,95	2146,05
2	3691,8			3559	3044	2679	2424	2276	**2254**	2307	2408	2253,95	1946,05
3	2858,95				2844	2479	2224	2076	**2054**	2107	2208	2053,95	1746,05
4	2143,7					2279	2024	1876	**1854**	1907	2008	1853,95	1546,05
5	**1579**						1824	1676	1654	1707	1808	1579	1271,1
6	**1123,8**							1476	1454	1507	1608	1123,8	815,9
7	**776,25**								1254	1454	1654	776,25	468,35
8	**553,95**									1107	1208	553,95	246,05
9	**407,2**										1008	407,2	99,3
10	**307,9**											307,9	0

Tabla 4.9. Decisiones a tomar en el mes de julio.

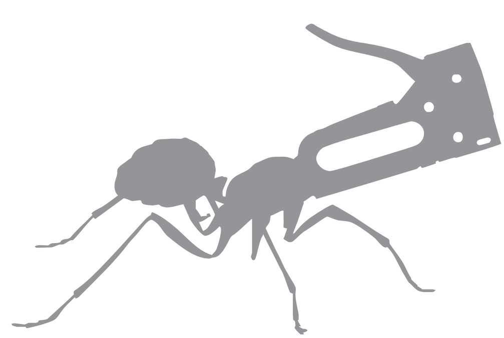
El mes de julio, pediremos lo que falte hasta llegar a 8 palés si el *stock* es de 4 o menos, y no pediremos nada en caso contrario.

Finalmente, en junio:

Tabla 4.10. Decisiones a tomar en el mes de junio

	N= 3	**S**										**306= g**		**373,74**
I/S	**H3**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**F*3**	**F'*3**	**FR***
0	3346,05	2846	2846	2846	2846	2771	2516	2368	**2346**	2399	2500	2346	2040,215	**2.413,95**
1	2674,05		2646	2646	2646	2571	2316	2168	**2146**	2199	2300	2146	1840,215	2.213,95
2	2248,05			2446	2446	2371	2116	1968	**1946**	1999	2100	1946	1640,215	2.013,95
3	1986,05				2246	2171	1916	1768	**1746**	1799	1900	1746	1440,215	1.813,95
4	1806,05					1971	1716	1568	**1546**	1599	1700	1546	1240,215	1.613,95
5	1596,07						1516	1368	**1346**	1399	1500	1346	1040,215	1.413,95
6	1291,505							1168	**1146**	1199	1300	1146	840,215	1.213,95
7	960,935								**946**	1146	1346	946	640,215	1.013,95
8	**669,215**									799	900	669	363,38	737,12
9	**448,795**										700	449	142,96	516,70
10	**305,835**											306	0	373,74

Si, como es de suponer, no hay *stock* al empezar la temporada, aprovisionaremos 8 palés con un coste esperado en toda la temporada de 2.413,95 euros. Obsérvese que la demanda esperada total en los tres meses es de solo 7 palés, por lo cual, de entrada, ya se pedirá un palé más de los que se espera que se necesitarán. Este palé extra cuesta 200 euros y estará (en valor esperado) tres meses en el almacén, a 20 euros/mes, pero se puede vender al final por 50 euros. Luego, el coste total esperado para este palé será de 200+3×20–50 = 210 euros, muy inferior al coste de rotura, que es de 1.000 euros/palé.



→ 5

Programación dinámica continua

5.1 Paso de programación dinámica discreta a continua

Los modelos y los métodos descritos en los capítulos anteriores suponen siempre que, tanto el espacio de los estados como el de las soluciones, así como la variable cronológica, son discretos. Es decir, finitos o infinito numerables.

Pero todo ello admite el paso a dimensiones continuas, sin grandes cambios conceptuales.

Dicho paso puede ser:

- Espacio de estados continuo
- Espacio de decisiones continuo
- Tiempo continuo

que pueden darse de manera simultánea o no.

Ahora bien, que el cambio sea simple conceptualmente no quiere decir que lo sea operativamente. Desde el punto de vista operativo, los cambios en las dos operaciones fundamentales que deberemos hacer siempre consisten básicamente en:

Operación	**Caso discreto**	**Caso continuo**
Composición de rendimientos	Suma	Integral
Optimización	Elección en una tabla	Derivada igual a cero

Tabla 5.1. Correspondencia entre operaciones discretas y continuas

lo cual, en general, introduce una notable complejidad operativa, que no siempre es fácil de gestionar.

Veamos, a título de ejemplo, un caso relativamente simple: el caso del barniz volátil.

5.2 El caso del barniz volátil

Hemos de gestionar la producción de un barniz muy volátil, de manera que el 10 % del *stock* inicial y de la producción del mes no se halla disponible para el mes siguiente. El *stock* inicial es de 200 litros y la demanda prevista para los próximos dos meses es de 80 y 60 litros, que debe servirse obligatoriamente.

El coste de producción es de 1 euro/litro, más 50 euros de costes fijos de preparación. Aunque no vamos a planificar más de tres meses, valoraremos el *stock* final a 1 euro/litro, ya que lo podremos aprovechar para servir la demanda de los meses siguientes.

Dado que numeramos las etapas como el tiempo que falta para acabar, haremos N = 1 para el segundo mes y N = 2 para el primero.

Empecemos fijando las condiciones de contorno: el valor de la $f^*_0(x)$ será el resultado que habrá que obtener en los meses siguientes con el *stock* final; por tanto, dado que valoramos a 1 euro/litro dicho *stock*, eso será un coste negativo:

$$f^*_0(x) = -1 . x_0$$

Veamos ahora la decisión que hay que tomar en N = 1. Si el *stock* inicial es inferior a la demanda, deberemos fabricar obligatoriamente:

$$f^*_1(x) = \mathrm{MIN}\{ 50+u+f^*_0[0{,}9(x+u-60)]\} = \mathrm{MIN}\{50+u-0{,}9x-0{,}9u+54\} =$$

$$= \mathrm{MIN}\ \{114 + 0{,}1\ u - 0{,}9\ x\}$$

Dado que $x + u \geq 60$, el mínimo de la expresión anterior es:

$$u^*_1 = 60-x_1$$

$$f^*_1(x_1) = 120-x_1$$

Veamos qué ocurre si $x \geq 60$. Dado que fabricar tiene un coste fijo y el coste variable es de 1 euro/litro, del cual solo aprovecharemos el 90 %, no interesa fabricar. Por tanto, una vez descontados los 60 litros a servir, el resto pasará a N = 0.

$$f^*_1(x_1) = f^*_0[0{,}9(x_1-60)] = -0{,}9(x_1-60) = 54-0{,}9\ x_1$$

En resumen, la decisión para N = 1 es:

- Si $x_1<60 \Rightarrow$ fabricar $u^*_1=60-x_1 \Rightarrow f^*_1(x_1)=120-x_1$
- Si $x_1\geq 60 \Rightarrow$ no fabricar $u^*_1=0 \Rightarrow f^*_1(x)=54-0{,}9x_1$

Veamos ahora el primer mes (N = 2). Hay tres casos, según sea el *stock* inicial:

a) Si $x_2 > 80$, ya tenemos material para servir la demanda del mes; puesto que, si fabricamos ahora, se perderá un 10 % del material, es obvio que no debemos fabricar; por tanto, la única diferencia es si dicho *stock* inicial es suficiente para atender la demanda de los dos meses o no, lo cual se produce cuando:

$$x_2 > 80 + 60/0{,}9 = 146{,}67 \text{ litros}$$

Por tanto:

a1) Si $x_2 \geq 146{,}67$, entonces:

$$u^*_2=0; \;\; f^*_2\,(x_2) = f^*_1[0{,}9(x_2–80)] = 54–0{,}9^2(x_2–80) = 118{,}8–0{,}81x_2$$

a2) Si $146{,}67 > x_2 \geq 80$, entonces:

$$u^*_2=0; \;\; f^*_2\,(x_2) = f^*_1\,[0{,}9(x_2–80)] = 120–0{,}9(x_2–80) = 192–0{,}9x_2$$

b) Si $x_2 < 80$, es obligatorio fabricar:

$$f^*_2(x_2) = \text{MIN}\,\{\,50+u_2+f^*_1[0{,}9(x_2+u_2–80)]\}$$

Dado que, del sobrante, solo el 90 % se aprovechará para el mes siguiente, tenemos dos casos:

b1) Fabricar para los dos meses o más:

$0{,}9(x_2+u_2–80)\geq 60 \Rightarrow x_2+u_2\geq 146{,}67$, entonces:

$f^*_1(x_2) = 54–0{,}9^2(x_2+u_2–80)$, con lo cual:

$f_2(x) = \text{MIN}\,\{50+u+54–0{,}9^2(x+u–80)\} = 168{,}8+0{,}19u–0{,}81x$

Como conviene que u sea el menor posible, se fabricará:

$u^*_2 = 146{,}66–x_2$; $f^*_2(x) = 196{,}67–x_2$

b2) Fabricar para el mes actual:

$u_2=80–x_2$, entonces el stock final del mes es 0 y $f^*_1(0)=120$, es decir:

$f_2(x_2) = 50+80–x_2+120 = 250–x_2$, que es peor que la opción b1.

En definitiva, la política para N = 2 es:

Si $x_2 < 80 \Rightarrow$ fabricar $u^*_2=146{,}66–x_2 \Rightarrow f^*_2(x_2) = 196{,}67–x_2$

Si $146{,}67 > x_2 \geq 80 \Rightarrow$ no fabricar $u^*_2=0 \Rightarrow f^*_2(x_2) = 192–0{,}9x_2$

Si $x_2 \geq 146{,}67 \Rightarrow$ no fabricar $u^*_2=0 \Rightarrow f^*_2(x_2) = 118{,}8–0{,}81x_2$

→6

Ejercicios

6.1 Ejercicios resueltos

6.1.1 Sustitución de equipos

Un sistema productivo incluye dos elementos muy delicados (A y B), que influyen decisivamente en el rendimiento de la producción. Ambos pueden sustituirse de manera independiente entre dos procesos cualesquiera, y el coste de sustitución es de 9 y 3 miles de euros, respectivamente.

El elemento A puede utilizarse en tres procesos como máximo, mientras que el elemento B solo puede utilizarse dos veces. El rendimiento del sistema (en miles de euros por proceso) depende de la edad de ambos elementos (veces que ha sido utilizado anteriormente), según la tabla 6.1.

		Edad B	
		0	**1**
Edad A	**0**	14	9
	1	10	7
	2	5	3

Tabla 6.1. Rendimiento del sistema en función de A y B

Se trata de hallar la política óptima de renovación de los elementos A y B en los casos siguientes:

a) Hemos de gestionar el sistema cuyo estado es (2,1) durante cinco procesos, al final de los cuales venderemos los elementos residuales con un ingreso que se indica en la tabla 6.2.

Edad	**0**	**1**	**≥2**
A	6	3	0
B	2	1	0

Tabla 6.2. Ingreso residual por la venta de A y B.

b) A largo plazo.

c) Estudiar el corto plazo con un coeficiente de actualización de 0,9.

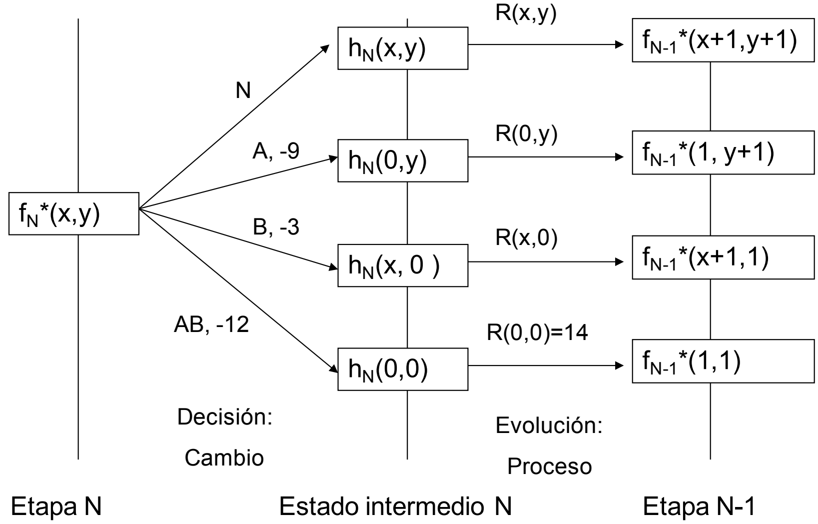
Solución

Empecemos por definir los estados del sistema. Para ello, nos será útil la técnica de la programación dinámica separable. Definiremos el instante de cambio de transición como aquel en que se toma la decisión de cambiar o no un elemento.

En principio, hay cuatro alternativas: no cambiar ninguno (N), cambiar el elemento A (A), cambiar el elemento B (B) o cambiar los dos elementos (AB).

El estado del sistema será la edad de ambos equipos, así que lo definiremos por un par de números (x, y), donde x = 0, 1, 2, 3 e y = 0, 1, 2. En estas condiciones, las transiciones posibles se pueden representar gráficamente según la figura 6.1.

Figura 6.1. Transiciones en el ejercicio de sustitución de equipos.



La ecuación de recurrencia según este esquema es:

$$f_N^*(x,y) = MAX\left\{h(x,y); -9 + h(0,y); -3 + h(x,0); -12 + h(0,0)\right\}$$

$$h(x,y) = R(x,y) + f_{N-1}^*(x+1, y+1)$$

Sin embargo, dada la simplicidad del sistema, podemos ahorrarnos escribir explícitamente las funciones h(x, y). Además, hay algunas decisiones evidentemente absurdas, que descartamos de entrada:

– No se cambia un elemento nuevo (edad 0).

– Un elemento gastado (x = 3 o y = 2) ha de cambiarse obligatoriamente.

Así pues, las transiciones posibles son las que se describen en la tabla 6.3.

Tabla 6.3. Transiciones posibles en el ejercicio de sustitución de equipos

	0; 0	**0; 1**	**0; 2**	**1; 0**	**1; 1**	**1; 2**	**2;0**	**2;1**	**2;2**	**3;0**	**3;1**	**3;2**
0; 0					14							
0; 1					11	9						
0; 2					11							
1; 0					5			10				
1; 1					2	1		7	7			
1; 2					2			7				
2;0					5						5	
2;1					2	0					2	3
2;2					2						2	
3;0					5						2	
3;1					2	0						
3;2					2							

y la correspondiente ecuación de recurrencia:

$$f_N^*(x,y) = MAX\left\{R(x,y) + f_{N-1}^*(x+1,y+1); -9 + R(0,y) + f_{N-1}^*(1,y+1); -3 + R(x,0) + f_{N-1}^*(x+1,1); 2 + f_{N-1}^*(1,1)\right\}$$

expresión que utilizaremos en esta resolución del problema.

a) Hemos de gestionar el sistema cuyo estado es (2,1) durante cinco procesos, al final de los cuales venderemos los elementos residuales con un ingreso que se indica en la tabla 6.2.

Se trata de un caso con N = 5 y valor residual, que hemos de considerar como componente de la f^*_0.

Aplicando la iteración en el espacio de los estados:

Tabla 6.4.a. Matriz de transiciones del ejercicio de sustitución de equipos

x,y	f*0	N	A	B	AB	f*1	f'*1	N	A	B	AB	f*2	f'*2	N	A	B	AB	f*3	f'*3
0; 0	8	18				18	12	16				16	12	17				17	12
0; 1	7	12		15		15	9	11		13		13	9	12		14		14	9
0; 2	6			15		15	9			13		13	9			14		14	9
1; 0	5	11		9		11	5	10		7		10	6	10		8		10	5
1; 1	4	7	4	8	6	8	2	7	3	7	4	7	3	7	4	7	5	7	2
1; 2	3			8	6	8	2			7	4	7	3			7	5	7	2
2;0	2	6	9			9	3	5	7			7	3	5	8			8	3
2;1	1	3	3	3	6	6	0	3	2	2	4	4	0	3	3	2	5	5	0
2;2	0			3	6	6	0			2	4	4	0			2	5	5	0
3;0	2		9			9	3		7			7	3		8			8	3
3;1	1		3		6	6	0		2		4	4	0		3		5	5	0
3;2	0				6	6	0				4	4	0				5	5	0
	g						6						4						5

Tabla 6.4.b. Matriz de transiciones del ejercicio de sustitución de equipos (continuación)

x,y	N	A	B	AB	f*4	f'*4	N	A	B	AB	f*5	f'*5
0; 0	16				16	12	17				17	12
0; 1	11		13		13	9	12		14		14	9
0; 2			13		13	9			14		14	9
1; 0	10		7		10	6	10		8		10	5
1; 1	7	3	7	4	7	3	7	4	7	5	7	2
1; 2			7	4	7	3			7	5	7	2
2;0	5	7			7	3	5	8			8	3
2;1	3	2	2	4	4	0	3	3	2	5	5	0
2;2			2	4	4	0			2	5	5	0
3;0		7			7	3		8			8	3
3;1		2		4	4	0		3		5	5	0
3;2				4	4	0				5	5	0
						4						5

Así pues, si el estado es (2,1), la política a seguir será (marcado en fondo gris en las tablas 6.4):

Tabla 6.5. Decisiones a tomar, considerando la solución obtenida

N	5	4	3	2	1	0
Estado	2,1	1,1	2,2	1,1	2,1	1,1
Decisión	AB	N	AB	B	AB	-

cuyo beneficio será 0+5+4+5+4+6 = 24 miles de euros, como puede comprobarse sumando el valor asociado a cada transición en la matriz de transiciones.

Observemos también que en N = 4 y N = 2 nos hallamos en el estado 1,1, en el cual hay dos decisiones igualmente óptimas: N y B. En la solución dada, hemos elegido arbitrariamente una en cada ocasión. El resultado total habría sido el mismo si hubiésemos realizado otra elección.

Aunque el enunciado no nos lo pide, veamos qué ocurriría si no hubiese valor residual:

x,y	f*0	N	A	B	AB	f*1	f'*1	N	A	B	AB	f*2	f'*2	N	A	B	AB	f*3	f'*3
0; 0	0	14				14	12	19				19	12	15				15	12
0; 1	0	9		11		11	9	14		16		16	9	10		12		12	9
0; 2	0			11		11	9			16		16	9			12		12	9
1; 0	0	10		5		10	8	11		10		11	4	10		6		10	7
1; 1	0	7	1	7	2	7	5	7	6	8	7	8	1	7	2	7	3	7	4
1; 2	0			7	2	7	5			8	7	8	1			7	3	7	4
2;0	0	5	5			5	3	5	10			10	3	5	6			6	3
2;1	0	3	0	2	2	3	1	3	5	2	7	7	0	3	1	2	3	3	0
2;2	0			2	2	2	0			2	7	7	0			2	3	3	0
3;0	0		5			5	3		10			10	3		6			6	3
3;1	0		0		2	2	0		5		7	7	0		1		3	3	0
3;2	0				2	2	0				7	7	0				3	3	0
	g						2						7						3

Tabla 6.6.a.
Matriz de transiciones sin valor residual de los elementos

x,y	N	A	B	AB	f*4	f'*4	N	A	B	AB	f*5	f'*5
0; 0	18				18	12	15				15	12
0; 1	13		15		15	9	10		12		12	9
0; 2			15		15	9			12		12	9
1; 0	10		9		10	4	10		6		10	7
1; 1	7	5	7	6	7	1	7	2	7	3	7	4
1; 2			7	6	7	1			7	3	7	4
2;0	5	9			9	3	5	6			6	3
2;1	3	4	2	6	6	0	3	1	2	3	3	0
2;2			2	6	6	0			2	3	3	0
3;0		9			9	3		6			6	3
3;1		4		6	6	0		1		3	3	0
3;2				6	6	0				3	3	0
						6						3

Tabla 6.6.b.
Matriz de transiciones sin valor residual de los elementos (continuación)

La política es ligeramente distinta y el valor total es 0+3+6+3+7+2 = 21 miles de euros. Este menor valor es debido a que no hay valor residual, pero se puede observar que la diferencia (24–21 = 3) es menor que el valor de la f^*_0

del primer caso (4), lo cual indica que la política a corto plazo está fuertemente influida por lo que ocurre al final del proceso. Se observa también que las g de normalización son distintas: así como en el primer caso se alternaban los valores 5 y 4, ahora lo hacen 6 y 3. Sin embargo, la media (4,5) es la misma, y lo mismo ocurre con las f'^*_N. Todo ello nos servirá para estudiar el significado en el caso a largo plazo.

b) A largo plazo.

Una respuesta sencilla, en este caso, es observar que las tablas de N = 3 y N = 5 son idénticas, y lo mismo ocurre con N = 2 y N = 4, lo cual indica que ya se ha alcanzado la estabilidad. Pero se va a proceder ignorando este detalle, con el fin de ilustrar al lector.

Dado que ya se conoce una política a corto plazo, se puede analizar si esta es también óptima a largo plazo. De las tablas anteriores, se puede deducir que la mejor política a ensayar es {N, B, B, N, B, B, A, AB, AB, A, AB, AB}, que es óptima para N = 5 en ambas situaciones.

Para comprobar si también lo es a largo plazo, se aplica la iteración en el espacio de las políticas, para lo cual se empieza resolviendo el sistema:

$$
\begin{array}{llcllll}
w_{00} + g = 14 + w_{11} & w_{20} + g = 5 + w_{11} & & & g = 4{,}5 & \\
w_{01} + g = 11 + w_{11} & w_{21} + g = 2 + w_{11} & & w_{00} = 12 & w_{11} = 2{,}5 & w_{22} = 0 \\
w_{02} + g = 11 + w_{11} & w_{22} + g = 2 + w_{11} & \Rightarrow & w_{01} = 9 & w_{12} = 2{,}5 & w_{30} = 3 \\
w_{10} + g = 10 + w_{12} & w_{30} + g = 2 + w_{31} & & w_{02} = 9 & w_{20} = 3 & w_{31} = 0 \\
w_{11} + g = 7 + w_{21} & w_{31} + g = 2 + w_{11} & & w_{10} = 5{,}5 & w_{21} = 0 & w_{32} = 0 \\
w_{12} + g = 7 + w_{21} & w_{32} + g = 2 + w_{11} & & & &
\end{array}
$$

En la resolución, se ha asignado $w_{32} = 0$, por coherencia con la tabla anterior. El lector comprobará que esta política es óptima a largo plazo.

A continuación, se deben ligar los valores que se han obtenido con los obtenidos en la iteración en el espacio de los estados. En primer lugar, se puede observar que el valor de g es la misma media que se había obtenido antes, con o sin valores residuales: 4,5; el hecho de que los valores que se alternaban como normalización sean distintos (4 y 5 o 6 y 3) es un detalle de cálculo sin importancia práctica. Lo mismo ocurre con las w_{xy}: los valores obtenidos son la media de los que se alternaban en las iteraciones pares e impares, tanto en un caso como en otro.

Ello es debido a que la política ensayada conduce a un ciclo en que se alternan los estados (1,1) y (2,1), cuya ganancia es, respectivamente, 7 y 2, lo cual arroja un promedio de 4,5 = g.

Por lo que respecta a las w_{xy}, la alternancia de valores se produce porque el método de cálculo en el espacio de los estados “sabe” siempre si el número de transiciones que faltan para terminar es par o impar, lo cual se produce también en las políticas: así, en el estado (2,1), se puede elegir entre las decisiones N o AB si N es impar, mientras que debe elegirse AB si N es par; por supuesto, ello no ocurre en el espacio de las políticas, porque ahora no se conoce esta paridad.

c) Estudiar el corto plazo con un coeficiente de actualización de 0,9.

A continuación, se resuelve el ejercicio a corto plazo con actualización.

Las tablas 6.7 presentan las transiciones con actualización:

Tabla 6.7.a. Matriz de transiciones con actualización

x,y	**f*0**	N	A	B	AB	**f*1**	**f'*1**	N	A	B	AB	**f*2**	**f'*2**	N	A	B	AB	**f*3**	**f'*3**
0; 0	**8**	**18**				**18**	**12,4**	**16,3**				**16,3**	**12,2**	**16,9**				**16,9**	**12,3**
0; 1	**7**	11,7		**14,6**		**14,6**	**9**	11,1		**13,1**		**13,1**	**9**	11,6		**13,6**		**13,6**	**9**
0; 2	**6**			**14,6**		**14,6**	**9**			**13,1**		**13,1**	**9**			**13,6**		**13,6**	**9**
1; 0	**5**	**10,9**		8,6		**10,9**	**5,3**	**10**		7,07		**10**	**5,93**	**10**		7,64		**10**	**5,36**
1; 1	**4**	7	3,7	**7,9**	5,6	**7,9**	**2,3**	**7**	3,07	**7**	4,07	**7**	**2,93**	**7**	3,64	**7**	4,64	**7**	**2,36**
1; 2	**3**			**7,9**	5,6	**7,9**	**2,3**			**7**	4,07	**7**	**2,93**			**7**	4,64	**7**	**2,36**
2;0	**2**	5,9	**8,6**			**8,6**	**3**	5	**7,07**			**7,07**	**3**	5	**7,64**			**7,64**	**3**
2;1	**1**	3	2,7	2,9	**5,6**	**5,6**	**0**	3	2,07	2	**4,07**	**4,07**	**0**	3	2,64	2	**4,64**	**4,64**	**0**
2;2	**0**			2,9	**5,6**	**5,6**	**0**			2	**4,07**	**4,07**	**0**			2	**4,64**	**4,64**	**0**
3;0	**2**		**8,6**			**8,6**	**3**		**7,07**			**7,07**	**3**		**7,64**			**7,64**	**3**
3;1	**1**		2,7		**5,6**	**5,6**	**0**		2,07		**4,07**	**4,07**	**0**		2,64		**4,64**	**4,64**	**0**
3;2	**0**				**5,6**	**5,6**	**0**				**4,07**	**4,07**	**0**				**4,64**	**4,64**	**0**
	g						**5,6**						**4,07**						**4,64**

Tabla 6.7.b. Matriz de transiciones con actualización (continuación)

x,y	N	A	B	AB	f*4	f'*4	N	A	B	AB	f*5	f'*5
0; 0	**16,4**				**16,4**	**12,2**	**16,9**				**16,9**	**12,3**
0; 1	11,1		**13,1**		**13,1**	**9**	11,6		**13,6**		**13,6**	**9**
0; 2			**13,1**		**13,1**	**9**			**13,6**		**13,6**	**9**
1; 0	**10**		7,13		**10**	**5,87**	**10**		7,59		**10**	**5,41**
1; 1	**7**	3,13	**7**	4,13	**7**	**2,87**	**7**	3,59	**7**	4,59	**7**	**2,41**
1; 2			**7**	4,13	**7**	**2,87**			**7**	4,59	**7**	**2,41**
2;0	5	**7,13**			**7,13**	**3**	5	**7,59**			**7,59**	**3**
2;1	3	2,13	2	**4,13**	**4,13**	**0**	3	2,59	2	**4,59**	**4,59**	**0**
2;2			2	**4,13**	**4,13**	**0**			2	**4,59**	**4,59**	**0**
3;0		**7,13**			**7,13**	**3**		**7,59**			**7,59**	**3**
3;1		2,13		**4,13**	**4,13**	**0**		2,59		**4,59**	**4,59**	**0**
3;2				**4,13**	**4,13**	**0**				**4,59**	**4,59**	**0**
						4,13						**4,59**

En las tablas 6.7, se observa que, aunque la política es la misma, los valores de normalización no se repiten exactamente, de modo que, a diferencia del caso sin actualización, sería preciso confirmar que la política hallada es óptima a largo plazo.

6.1.2 Un sistema convencional

Un sistema evoluciona entre tres estados, denominados convencionalmente X, Y y Z, y a cada cambio de estado se puede emprender una acción (p, q, r, s, t). Las ganancias obtenidas en cada cambio de estado y las probabilidades de transición se resumen en la tabla 6.8.

Tabla 6.8. Rendimientos y probabilidades del sistema

		Probabilidades			Rendimientos		
Estado	**Acción**	**X**	**Y**	**Z**	**X**	**Y**	**Z**
X	**p**	0,3	0,4	0,3	2	9	6
X	**q**	0	1	0	-	2	-
X	**r**	0	0,5	0,5	-	7	3
Y	**p**	0,3	0,4	0,3	4	4	14
Y	**s**	0,8	0,2	0	1	6	-
Y	**t**	0	1	0	-	3	-
Z	**q**	0	0	1	-	-	2
Z	**r**	0,5	0,5	0	4	6	-
Z	**s**	0,2	0,2	0,6	1	7	4
Z	**t**	0	0	1	-	-	3

a) Cuando se acaba la gestión del sistema, hay unos costes de eliminación de los residuos, valorados en 5, 7 o 9 unidades, según si el sistema ha quedado en el estado X, Y o Z, respectivamente. Determina la política óptima en las tres últimas transiciones y la ganancia esperada si el estado inicial es Z.

b) Sin hacer más cálculos, indica razonadamente si la política hallada para la 3ª transición antes del final es óptima a largo plazo; si no lo es, o si no se puede afirmar nada con seguridad; en este caso, indica qué es más probable.

c) Determina la política óptima a largo plazo y la ganancia media por transición correspondiente.

Solución:

a) Cuando se acaba la gestión del sistema, hay unos costes de eliminación de los residuos, valorados en 5, 7 o 9 unidades, según si el sistema ha quedado en el estado X, Y o Z, respectivamente. Determina la política óptima en las tres últimas transiciones y la ganancia esperada si el estado inicial es Z.

$$f_N^*(x) = \underset{u}{MAX} \sum_y \left[r_u(x,y) + f_{N-1}^*(y) \right] p_u(y \mid x) = \underset{u}{MAX} \left\{ q(x,u) + \sum_y f_{N-1}^*(y) p_u(y \mid x) \right\}$$

Calculamos la expresión q(x,u) en la tabla 6.9, que será de ayuda para cálculos posteriores.

Tabla 6.9. Cálculo de la función q(x, u)

		Probabilidades			Rendimientos			
Estado	**Acción**	**X**	**Y**	**Z**	**X**	**Y**	**Z**	**q**
X	**p**	0,3	0,4	0,3	2	9	6	6
X	**q**	0	1	0	-	2	-	2
X	**r**	0	0,5	0,5	-	7	3	5
Y	**p**	0,3	0,4	0,3	4	4	14	7
Y	**s**	0,8	0,2	0	1	6	-	2
Y	**t**	0	1	0	-	3	-	3
Z	**q**	0	0	1	-	-	2	2
Z	**r**	0,5	0,5	0	4	6	-	5
Z	**s**	0,2	0,2	0,6	1	7	4	4
Z	**t**	0	0	1	-	-	3	3

De acuerdo con los resultados obtenidos en la tabla 6.10, la política óptima es hacer p si el sistema está en X o Y, y hacer r si está en Z, en todas las iteraciones.

Tabla 6.10.
Matriz de transiciones

	$f^*_0(x)$	$f^*_1(x)$	$u^*_1(x)$	$f^n_1(x)$	$f^*_2(x)$	$u^*_2(x)$	$f^n_2(x)$	$f^*_3(x)$	$u^*_3(x)$	$f^n_3(x)$
X	-5	-1	p	0	6,4	p	0,9	7,41	p	1,01
Y	-7	0	p	1	7,4	p	1,9	8,41	p	2,01
Z	-9	-1	r	0	5,5	r	0	6,40	r	0
g				-1			5,5			6,40

Empezando en Z, en tres iteraciones la ganancia esperada es:

$$-1 + 5{,}5 + 6{,}40 + 0 = 10{,}9$$

b) Sin hacer más cálculos, indica razonadamente si la política hallada para la 3ª transición antes del final es óptima a largo plazo; si no lo es, o si no se puede afirmar nada con seguridad; en este caso, indica qué es más probable.

No se sabe si la política {p, p, r} es óptima o no porque, aunque $u^*_2(x) = u^*_3(x)$, la otra condición no se cumple, ya que $f^n_2(x) \neq f^n_3(x)$. Por tanto, no se puede afirmar nada con seguridad.

Ahora bien, el hecho de que se repita la política y que las $f^n_N(x)$ sean bastante parecidas hace creer, casi con toda seguridad, que efectivamente sí lo debe de ser.

c) Determina la política óptima a largo plazo y la ganancia media por transición correspondiente.

Hay que comprobar la política {p, p, r}:

$$\begin{aligned} w_x + g &= 6 + 0{,}3w_x + 0{,}4w_y + 0{,}3w_z \\ w_y + g &= 7 + 0{,}3w_x + 0{,}4w_y + 0{,}3w_z \\ w_z + g &= 5 + 0{,}5w_x + 0{,}5w_y \end{aligned} \quad \Rightarrow \quad \begin{aligned} & g = 161/26 \approx 6{,}19 \\ & w_z = 0 \\ & w_x = 9/13;\ w_y = 22/13 \end{aligned}$$

$W_x = \text{MAX}\{6+0{,}3w_x+0{,}4w_y+0{,}3w_z;\ 2+w_y;\ 5+0{,}5w_x+0{,}5w_y\} - g = 9/13 = w_x \Rightarrow p$✓

$W_y = \text{MAX}\{7+0{,}3w_x+0{,}4w_y+0{,}3w_z;\ 2+0{,}8w_x+0{,}2w_y;\ 3+w_y\}- g = 22/13=w_y \Rightarrow p$✓

$W_z = \text{MAX}\{2+w_z;\ 5+0{,}5w_x+0{,}5w_y\ ;\ 4+0{,}2w_x+0{,}2w_y+0{,}6w_z;\ 3+w_z\}-g=0=w_z \Rightarrow r$✓

Es decir, la política es efectivamente óptima. Obsérvese cómo los valores reales de g y de las w son relativamente similares a los hallados en la última iteración. Este es un fenómeno habitual en la programación dinámica aleatoria: los valores tanto de las g como de las f*N tienden muy lentamente hacia sus valores límite, básicamente por el efecto de la aleatoriedad.

6.1.3 Mantenimiento

Una línea automática de mecanizado trabaja ininterrumpidamente cada semana, de lunes a viernes, y se aprovecha el sábado para hacer diversas operaciones de mantenimiento, una de las cuales es revisar una herramienta sometida a un gran desgaste.

Para describir el desgaste de la herramienta, se han definido los siguientes niveles: nueva (N), seminueva (S), usada (U), desgastada (D) y rota (R). A fin de evaluar las probabilidades de cambio de estado, se dispone de una estadística, que indica el número de veces que se ha observado el cambio indicado en una semana:

Estado el sábado →		N	S	U	D	R
Estado el lunes por la mañana →	N	20	15	15		
	S		16	20	4	
	U			3	5	2
	D				1	1

Tabla 6.11. Datos históricos sobre el cambio de estado entre el lunes y el sábado

Una herramienta rota (R) siempre se queda R, naturalmente.

Si la herramienta no es N, hay unos costes por pérdida de eficiencia que dependen del estado inicial y final de la herramienta; en concreto, los costes estimados si la herramienta acaba la semana igual que la ha empezado son:

Estado	N	S	U	D	R
Coste (€)	0	200	500	1.000	60.000

Tabla 6.12. Costes por pérdida de eficiencia

Si el estado final no es igual al inicial, se considera un coste que es la semisuma de los correspondientes a ambos estados: por ejemplo, si empieza N y acaba U, el coste es 250 euros.

El problema que se debe estudiar es cuándo y cómo hay que cambiar la herramienta, porque una herramienta nueva es muy cara (2.500 €), si bien también existe la posibilidad de poner una de segunda mano, en estado S, que tan solo cuesta 1.000 euros.

a) Desde tiempo inmemorial, hay en la empresa una polémica sobre si es mejor poner una herramienta nueva cuando se halla S o peor, o bien es preferible esperar a que esté U o peor, y entonces poner una seminueva. ¿Cuál de las dos propuestas es mejor?

b) Comprueba, para la mejor política conocida, si es óptima o no.

c) ¿Cada cuántas semanas se cambiará la herramienta con la política óptima?

Solución

a) Desde tiempo inmemorial, existe en la empresa una polémica sobre si es mejor poner una herramienta nueva cuando se halla S o peor, o bien es preferible esperar a que esté U o peor, y entonces poner una seminueva. ¿Cuál de las dos propuestas es mejor?

Una iteración es lo que ocurre en una semana, y definimos el momento de cambio de transición como la observación que se hace el sábado. Los estados del sistema son N, S, U, D, R, es decir, los estados de la herramienta. Las probabilidades de cambio de estado durante la semana se obtienen dividiendo el número de observaciones por el total de la fila, y la remuneración son los costes totales durante la semana, que incluyen la herramienta repuesta (N o S) y los costes de pérdida de eficiencia.

Las acciones posibles son poner una herramienta nueva (N), poner una seminueva (S) o no cambiar (NC).

Ante estas condiciones, las transiciones posibles en una semana serían las expresadas en la tabla 6.13.a. (poniendo solo las opciones que tienen sentido y calculando el vector q de costes esperados en la semana):

Tabla 6.13.a. Matriz de transiciones en una semana

		Probabilidades (P)					Costes (R)					
Est.	Dec	N	S	U	D	R	N	S	U	D	R	**Q**
N	NC	0,4	0,3	0,3	0	0	0	100	250	500	30.000	105
S	N	0,4	0,3	0,3	0	0	2.500	2.600	2.750	3.000	32.500	2.605
S	NC	0	0,4	0,5	0,1	0	0	200	350	600	30.100	315
U	N	0,4	0,3	0,3	0	0	2.500	2.600	2.750	3.000	32.500	2.605
U	S	0	0,4	0,5	0,1	0	1.000	1.200	1.350	1.600	31.100	1.315
U	NC	0	0	0,3	0,5	0,2	0	0	500	750	30.250	6.575
D	N	0,4	0,3	0,3	0	0	2.500	2.600	2.750	3.000	32.500	2.605
D	S	0	0,4	0,5	0,1	0	1.000	1.200	1.350	1.600	31.100	1.315
D	NC	0	0	0	0,5	0,5	0	0	0	1.000	30.500	15.750
R	N	0,4	0,3	0,3	0	0	2.500	2.600	2.750	3.000	32.500	2.605
R	S	0	0,4	0,5	0,1	0	1.000	1.200	1.350	1.600	31.100	1.315

La política de cambiar por N cuando está en S o peor es [NC,N,N,N,N] y lleva a:

Tabla 6.13.b. Resumen de la matriz de transiciones

		Probs.(**P**)					Costes (**R**)				
Est.	Dec	N	S	U	D	R	N	S	U	D	R
N	NC	0,4	0,3	0,3	0	0	0	100	250	500	30.000
S	N	0,4	0,3	0,3	0	0	2.500	2.600	2.750	3.000	32.500
U	N	0,4	0,3	0,3	0	0	2.500	2.600	2.750	3.000	32.500
D	N	0,4	0,3	0,3	0	0	2.500	2.600	2.750	3.000	32.500
R	N	0,4	0,3	0,3	0	0	2.500	2.600	2.750	3.000	32.500

$w_n + g = 105 + 0{,}4w_n + 0{,}3w_s + 0{,}3w_u$

$w_s + g = 2.605 + 0{,}4w_n + 0{,}3w_s + 0{,}3w_u$

$w_u + g = 2.605 + 0{,}4w_n + 0{,}3w_s + 0{,}3w_u$

$w_d + g = 2.605 + 0{,}4w_n + 0{,}3w_s + 0{,}3w_u$

$w_r + g = 2.605 + 0{,}4\ w_n + 0{,}3w_s + 0{,}3w_u$

Igualando w_n=0, se obtiene:

$w_s = w_u = w_d = w_r = 2.500$; $g = 1.605$

Comprobemos ahora la optimalidad:

w_n no hace falta comprobar porque solo hay una opción.

w_s: MIN{$2.605+0{,}4w_n+0{,}3w_s+0{,}3w_u$;

$\mathbf{315+0{,}4w_s+0{,}5w_u+0{,}1w_d}$}$-1.605=1.210 < w_s \rightarrow$ NC

w_u: MIN{$2.605+0{,}4w_n+0{,}3w_s+0{,}3w_u$;

$1.315+0{,}4w_s+0{,}5w_u+0{,}1w_d$;

$6.557+0{,}3w_u+0{,}5w_d\ +0{,}2\ w_r$ } $- 1.605= 2.210 < w_u \rightarrow$ S

w_d: MIN{$2.605+0{,}4w_n+0{,}3w_s+0{,}3w_u$;

$\mathbf{1.315+0{,}4\ w_s+0{,}5w_u+0{,}1w_d}$;

$15.750+0{,}5w_d +0{,}5w_r$} $- 1.605= 2.210 < w_d \rightarrow$ S

w_r: MIN{$2.605+0{,}4w_n+0{,}3w_s+0{,}3w_u$;

$\mathbf{1.315+0{,}4w_s+0{,}5w_u+0{,}1w_d}$}$-1.605 = 2.210 < w_r \rightarrow$S

Resulta que es preferible la política [NC,NC,S,S,S], es decir, esperar a que la herramienta esté U o peor, y entonces cambiarla por una seminueva.

b) Comprueba, para la mejor política conocida, si es óptima o no.

Veamos ahora si la política [NC,NC,S,S,S] es óptima. Resolvemos el sistema:

$$w_n + g = 105 + 0{,}4w_n + 0{,}3w_s + 0{,}3w_u$$

$$w_s + g = 315 + 0{,}4w_s + 0{,}5w_u + 0{,}1w_d$$

$$w_u + g = 1315 + 0{,}4w_s + 0{,}5w_u + 0{,}1w_d$$

$$w_d + g = 1315 + 0{,}4w_s + 0{,}5w_u + 0{,}1w_d$$

$$w_r + g = 1315 + 0{,}4w_s + 0{,}5w_u + 0{,}1w_d$$

y, haciendo $w_s = 0$, obtenemos:

$$w_u = w_d = w_r = 1000;\ \ w_n = -850;\ g = 915$$

Comprobemos ahora la optimalidad:

w_n no hace falta comprobar porque solo hay una opción.

w_s: $\text{MIN}\{2.605+0{,}4w_n+0{,}3w_s+0{,}3w_u;$

$315+0{,}4w_s+0{,}5w_u+0{,}1w_d\}-915=0= w_s \rightarrow$ OK

w_u: $\text{MIN}\{2.605+0{,}4w_n+0{,}3w_s+0{,}3w_u;$

$1.315+0{,}4w_s+0{,}5w_u+0{,}1w_d;$

$6.557+0{,}3w_u+0{,}5w_d +0{,}2w_t\}-915=1.000=w_u \rightarrow$ OK

w_d: $\text{MIN}\{2.605+0{,}4w_n+0{,}3w_s+0{,}3w_u;$

$1.315+0{,}4w_s+0{,}5w_u+0{,}1w_d;$

$15.750+0{,}5w_d +0{,}5w_t\}-915=1.000=w_d \rightarrow$ OK

w_r: $\text{MIN}\{2.605+0{,}4w_n+0{,}3w_s+0{,}3w_u;$

$1.315+0{,}4w_s+0{,}5w_u+0{,}1w_d\}-915=1.000 = w_t \rightarrow$ OK

Por tanto, la política es óptima.

c) ¿Cada cuántas semanas se cambiará la herramienta con la política óptima?

Para saber cada cuántas semanas se cambiará la herramienta, hemos de definir una remuneración que cuente las transiciones en que efectivamente se haga el cambio. Para ello, basta con definir un vector $q'_1 = [0\ 1\ 1\ 1\ 1]^T$, en el caso de la primera política, o bien $q'_2 = [0\ 0\ 1\ 1\ 1]^T$, en el caso de la segunda.

En estas condiciones, el valor de g nos dirá cada cuántas semanas de media se cambia la herramienta.

Las w nos indican la diferencia de tiempo esperado hasta que se produzca el primer cambio:

$$w_n + g = 0 + 0{,}4w_n + 0{,}3w_s + 0{,}3w_u$$

$$w_s + g = 0 + 0{,}4w_s + 0{,}5w_u + 0{,}1w_d$$

$$w_u + g = 1 + 0{,}4w_s + 0{,}5w_u + 0{,}1w_d$$

$$w_d + g = 1 + 0{,}4w_s + 0{,}5w_u + 0{,}1w_d$$

$$w_r + g = 1 + 0{,}4w_s + 0{,}5w_u + 0{,}1w_d$$

y, haciendo $w_s = 0$, obtenemos:

$$w_u = w_d = w_r = 1; \quad w_n = -0{,}5; \; g = 0{,}6$$

Según los resultados obtenidos, la herramienta se cambiará el 60 % de las semanas, es decir, cada 1/0,6 ≈ 1,67 semanas de media.

Si consideramos como base de referencia el tiempo que se tarda en cambiar una herramienta en estado S, en el estado N se tardará 0,5 semanas más, mientras que en el resto de los estados se tardará una semana menos (siempre en valores esperados, por supuesto).

NOTA: Este problema, como todos los de programación dinámica aleatoria homogénea en el tiempo, puede resolverse también mediante cadenas de Markov. En el volumen sobre cadenas de Markov de esta misma colección, se puede hallar la resolución correspondiente.

6.1.4 Cultivos rotativos

Una ONG que regenta una explotación agraria en un país del tercer mundo se encuentra cada año con el problema de decidir qué cultivo plantar, entre los tres posibles: bambú, café o tabaco. Los resultados que se obtengan dependerán del tipo de cultivo que se decida, del cultivo anterior y de si las circunstancias climatológicas y económicas son favorables o desfavorables.

En concreto, los beneficios esperados (en miles de euros) son los descritos en la tabla 6.14.

Tabla 6.14. Beneficios esperados en miles de euros

Si el año anterior se ha plantado…	… y este año se planta…	en año favorable, se obtendrá:	en año desfavorable, se obtendrá:
Bambú	Bambú	20	5
Bambú	Café	60	20
Bambú	Tabaco	100	30
Café	Bambú	85	30
Café	Café	10	10
Café	Tabaco	50	40
Tabaco	Bambú	55	20
Tabaco	Café	90	15
Tabaco	Tabaco	30	10

Se ha observado que, estadísticamente, la alternancia de años favorables y desfavorables sigue una regularidad notable: después de un año favorable, solo el 30 % de las veces habrá otro también favorable; en cambio, después de un año desfavorable, el siguiente es favorable un 60 % de las veces.

Halla la política óptima a largo plazo.

Solución

El problema es aleatorio, de horizonte indeterminado y homogéneo en el tiempo. El planteamiento por programación dinámica, si se subdivide la transición en dos pasos (decisión / evolución), es el siguiente:

Estados al inicio de la transición (decisión): [Función f: 6 estados]

{Año anterior (F/D); cultivo anterior (B/C/T)}

Estados después de tomar la decisión (evolución): [Función φ: 18 estados]

{Año anterior (F/D); cultivo anterior (B/C/T); cultivo actual (B/C/T)}

Se toman los símbolos X, Q, y Q' para representar estas magnitudes de forma genérica.

Condiciones de contorno: $f^*_0 (X, Q) = 0 \quad \forall X, Q$

Ecuaciones de recurrencia:

En la decisión:

$$f_N^*(F,Q) = \underset{Q'}{MAX}\{\phi_N(F,Q,Q')\};$$

$$f_N^*(D,Q) = \underset{Q'}{MAX}\{\phi_N(D,Q,Q')\}$$

y en la evolución, puesta explícitamente:

$$\phi_N\left(F,B,B\right)=0{,}3\left[20+f^*_{N-1}(F,B)\right]+0{,}7\left[5+f^*_{N-1}(D,B)\right]$$
$$\phi_N\left(F,B,C\right)=0{,}3\left[60+f^*_{N-1}(F,C)\right]+0{,}7\left[20+f^*_{N-1}(D,C)\right]$$
$$\phi_N\left(F,B,T\right)=0{,}3\left[100+f^*_{N-1}(F,T)\right]+0{,}7\left[30+f^*_{N-1}(D,T)\right]$$

$$\phi_N\left(F,B,B\right)=0{,}6\left[20+f^*_{N-1}(F,B)\right]+0{,}4\left[5+f^*_{N-1}(D,B)\right]$$
$$\phi_N\left(F,B,C\right)=0{,}6\left[60+f^*_{N-1}(F,C)\right]+0{,}4\left[20+f^*_{N-1}(D,C)\right]$$
$$\phi_N\left(F,B,T\right)=0{,}6\left[100+f^*_{N-1}(F,T)\right]+0{,}4\left[30+f^*_{N-1}(D,T)\right]$$

$$\phi_N\left(F,C,B\right)=0{,}3\left[85+f^*_{N-1}(F,B)\right]+0{,}7\left[30+f^*_{N-1}(D,B)\right]$$
$$\phi_N\left(F,C,C\right)=0{,}3\left[10+f^*_{N-1}(F,C)\right]+0{,}7\left[10+f^*_{N-1}(D,C)\right]$$
$$\phi_N\left(F,C,T\right)=0{,}3\left[50+f^*_{N-1}(F,T)\right]+0{,}7\left[40+f^*_{N-1}(D,T)\right]$$

$$\phi_N\left(F,C,B\right)=0{,}6\left[85+f^*_{N-1}(F,B)\right]+0{,}4\left[30+f^*_{N-1}(D,B)\right]$$
$$\phi_N\left(F,C,C\right)=0{,}6\left[10+f^*_{N-1}(F,C)\right]+0{,}4\left[10+f^*_{N-1}(D,C)\right]$$
$$\phi_N\left(F,C,T\right)=0{,}6\left[50+f^*_{N-1}(F,T)\right]+0{,}4\left[40+f^*_{N-1}(D,T)\right]$$

$$\phi_N\left(F,T,B\right)=0{,}3\left[55+f^*_{N-1}(F,B)\right]+0{,}7\left[20+f^*_{N-1}(D,B)\right]$$
$$\phi_N\left(F,T,C\right)=0{,}3\left[90+f^*_{N-1}(F,C)\right]+0{,}7\left[15+f^*_{N-1}(D,C)\right]$$
$$\phi_N\left(F,T,T\right)=0{,}3\left[30+f^*_{N-1}(F,T)\right]+0{,}7\left[10+f^*_{N-1}(D,T)\right]$$

$$\phi_N\left(F,T,B\right)=0{,}6\left[55+f^*_{N-1}(F,B)\right]+0{,}4\left[20+f^*_{N-1}(D,B)\right]$$
$$\phi_N\left(F,T,C\right)=0{,}6\left[90+f^*_{N-1}(F,C)\right]+0{,}4\left[15+f^*_{N-1}(D,C)\right]$$
$$\phi_N\left(F,T,T\right)=0{,}6\left[30+f^*_{N-1}(F,T)\right]+0{,}4\left[10+f^*_{N-1}(D,T)\right]$$

Si resumimos las dos partes de la transición en una sola tabla, obtenemos la tabla 6.15.

Tabla 6.15.
Matriz de transiciones del problema de cultivos rotativos

		Matriz P (probabilidades)						**Matriz R (remuneraciones)**					
Est.	**Dec.**	**BF**	**BD**	**CF**	**CD**	**TF**	**TD**	**BF**	**BD**	**CF**	**CD**	**TF**	**TD**
BF	**B**	0,3	0,7					20	5				
	C			0,3	0,7					60	20		
	T					0,3	0,7					100	30
BD	**B**	0,6	0,4					20	5				
	C			0,6	0,4					60	20		
	T					0,6	0,4					100	30
CF	**B**	0,3	0,7					85	30				
	C			0,3	0,7					10	10		
	T					0,3	0,7					50	40
CD	**B**	0,6	0,4					85	30				
	C			0,6	0,4					10	10		
	T					0,6	0,4					50	40
TF	**B**	0,3	0,7					55	20				
	C			0,3	0,7					90	15		
	T					0,3	0,7					30	10
TD	**B**	0,6	0,4					55	20				
	C			0,6	0,4					90	15		
	T					0,6	0,4					30	10

y el vector q correspondiente es:

$$q^T = \left|\, 0 \,\middle|\, 32 \,\middle|\, 51 \,\middle|\, 14 \,\middle|\, 44 \,\middle|\, 72 \,\middle|\, 46{,}5 \,\middle|\, 10 \,\middle|\, 43 \,\middle|\, 63 \,\middle|\, 10 \,\middle|\, 46 \,\middle|\, 31 \,\middle|\, 38 \,\middle|\, 16 \,\middle|\, 41 \,\middle|\, 60 \,\middle|\, 22 \,\right|$$

Dado que nos interesa la solución óptima a largo plazo, hemos de establecer una política inicial para iterar en el espacio de las políticas. A falta de mejor información, tomaremos la que mejor ganancia da en el primer año, es decir, [T, T, B, B, C, C], lo cual implica un cultivo rotativo entre las tres posibilidades: B·T·C·B, de lo cual resulta:

Obsérvese que esta política es cíclica de período 3 (hacemos un mismo cultivo cada tres años), lo cual implica que el comportamiento de la función iterada en el espacio de los estados se comportará también cíclicamente.

Para ver si la política es óptima, planteamos el correspondiente sistema y, haciendo $w_{TD} = 0$, resulta:

$$g = 55{,}8;\ w_{BF} = -10{,}45;\ w_{BD} = 4{,}87;\ w_{CF} = -8{,}99;\ w_{CD} = 2{,}91;\ w_{TF} = -18{,}93$$

El lector podrá comprobar que esta política es óptima.

NOTA: Este problema puede también resolverse mediante una cadena de Markov con remuneración y decisión. En el volumen correspondiente de esta misma colección, se puede hallar la resolución correspondiente.

6.2 Ejercicios propuestos

6.2.1 El problema de la mochila

Un turista está planificando una excursión corta por el campo y para ello cuenta con tres recursos: comida (recurso A), bebida (recurso B) y libros para leer a la sombra de un árbol (recurso C). La utilidad que le reporta llevar cada uno de estos recursos aumenta con las unidades que se lleva, según las tablas 6.16 a 6.18.

Tabla 6.16. Utilidad aportada por el recurso A

Cantidad de recurso (unidades)	Utilidad total
0	0
1	10
2	15
3	18

Peso del recurso A: 3 kg por unidad

La columna de la utilidad total significa que, por ejemplo, tres unidades del recurso A proporcionan una utilidad de 18 en su conjunto.

Tabla 6.17. Utilidad aportada por el recurso B

Cantidad de recurso (unidades)	Utilidad total
0	0
1	14
2	21

Peso del recurso B: 4 kg por unidad

Tabla 6.18. Utilidad aportada por el recurso C

Cantidad de recurso (unidades)	Utilidad total
0	0
1	6
2	9
3	11
4	12
5	13

Peso del recurso C: 2 kg por unidad

Sabiendo que el turista puede llevar en su mochila una cantidad máxima de 10 kg, se trata de saber cuántas unidades de cada producto debe llevar en la mochila para maximizar su utilidad total.

Se pide:

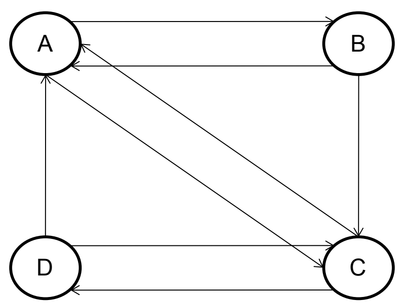
a) Definir la naturaleza de las etapas, los estados y las variables de decisión para plantear este problema mediante la programación dinámica.

b) Resolver el problema utilizando la programación dinámica.

6.2.2 El problema de la ruta

Un transportista tiene una serie de clientes en cuatro ciudades, que denominaremos A, B, C y D. Las rutas posibles son tales que, desde cada una de las ciudades, tiene dos rutas disponibles con carga, tal como aparece en el gráfico. La tabla 6.19 muestra las distancias entre las cuatro ciudades, y la tabla 6.20 muestra la remuneración obtenida por transportar mercancías desde la ciudad i a la ciudad j.

Figura 6.2. Posibles rutas entre ciudades



Los costes del transportista son proporcionales a la distancia y, en estos momentos, son de 10 euros por kilómetro.

Tabla 6.19. Distancias entre ciudades (en km)

	A	**B**	**C**	**D**
A	-	100	200	500
B	100	-	600	800
C	200	600	-	300
D	500	800	300	-

Tabla 6.20. Ingresos por transporte entre ciudades (en €)

	A	**B**	**C**	**D**
A	-	4.000	8.000	-
B	2.500	-	8.000	-
C	2.500	-	-	4.000
D	5.500	-	5.000	-

Con estos datos, se pide:

a) Plantear un modelo de programación dinámica que permita obtener la política a seguir para maximizar el beneficio del transportista, en caso de que deba llevar carga en cada viaje. (Este hecho elimina, por ejemplo, la posibilidad de viajar de A a D.)

b) Determinar qué ruta debe seguir el transportista en los tres próximos viajes, si ahora se encuentra en la ciudad A. Determinar también cuál será el beneficio obtenido para esta ruta óptima.

c) ¿Cómo debería modificarse el modelo si se considerara la posibilidad de volver de vacío? (No es necesario determinar las expresiones de la función de recurrencia para este caso.)

6.2.3 El problema de las máquinas en paralelo

Un sistema productivo consta de dos máquinas en paralelo, que realizan la misma función. La empresa se ha planteado cuál debe ser la política de mantenimiento que maximice el valor esperado para la producción horaria a largo plazo. Para ello, dispone de la información siguiente acerca del funcionamiento del sistema y de los diferentes casos que pueden darse:

Primer caso. Cuando funcionan las dos máquinas en régimen normal, cada una de ellas tiene una productividad de 30 piezas por hora. Sin embargo, cada una de ellas tiene una probabilidad de avería igual a 0,2. En esta situación, podría plantearse la realización de un programa de mantenimiento preventivo al principio de cada hora de funcionamiento, que asegurase el funcionamiento correcto del sistema al final de la hora. Las interrupciones por mantenimiento supondrían que cada máquina tendría un rendimiento de 25 piezas por hora.

Segundo caso. Si solo funciona una máquina, puede optarse por seguir en esta situación o por reparar la otra máquina. De seguir la primera opción, la máquina que funciona tendría que trabajar a 40 piezas por hora y su probabilidad de avería sería de 0,4. Si se opta por reparar la máquina, el sistema permanecería parado durante una hora. Por cuestiones de planificación, la decisión de realizar la reparación debería tomarse al principio de cada hora de funcionamiento.

Tercer caso. Si se estropean las dos máquinas a la vez (en el primer caso) o se estropea la única máquina que funciona (en el segundo caso), la reparación exigiría dos horas para que el sistema estuviera a pleno rendimiento. Por razones similares a la situación anterior, esta reparación debería iniciarse al principio de la hora siguiente a la(s) avería(s).

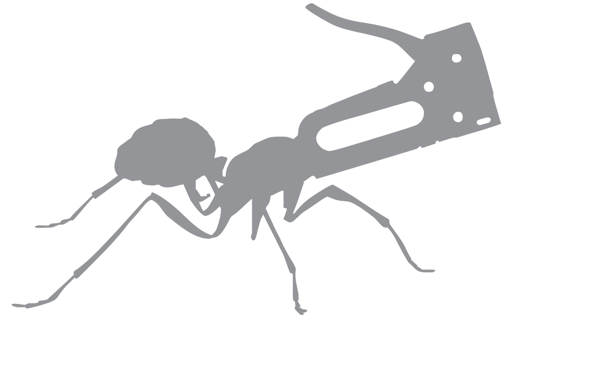
Cuando se produce una avería, el rendimiento es la media entre la situación de antes y de después de la avería. Por ejemplo, si se estropeara una de las dos máquinas, el rendimiento resultante sería de (60+40)/2 = 50 piezas por hora.

Se pide:

a) Definir los parámetros del sistema que permiten su modelización como un modelo de programación dinámica aleatoria.

b) Determinar la política óptima para tres horas de funcionamiento del sistema, si empezamos con las dos máquinas en funcionamiento.

c) Si, a largo plazo, se opta por no realizar el mantenimiento preventivo y por reparar las máquinas inmediatamente después de que se averíen, determinar cuál sería el valor esperado de la producción horaria para esta política.



Bibliografía

Ackoff, R. L.; Sasieni, M. W. (1968). *Fundamentals of Operations Research*. John Wiley & Sons.

Bellman, R. (2003). *Dynamic Programming.* Dover Publications, Inc., Nueva York.

Bierman, H.; Bonini, C. P.; Hausman, W. H. (1969). *Quantitative Analysis for Business Decisions*. 3ª ed. Richard Irwin Inc., Illinois.

Denardo, E. V. (2003). *Dynamic Programming: Models and applications.* Prentice-Hall, Inc., Nueva Jersey.

Dreyfus, S. E.; Law, A. M. (1977). *The Art and Theory of Dynamic Programming.* Academic Press, Nueva York.

Hamdy, A. T. (2004). *Investigación de Operaciones*. 7ª ed. Prentice-Hall.

Hillier, F.; Lieberman, G. (2010). *Introducción a la investigación de operaciones.* 9ª ed. McGraw Hill.

Jiménez, G. (2009). *Optimización*. Universidad Nacional de Colombia.

Lew, A.; Mauch, H. (2007). *Dynamic Programming. A computational tool.* Springer.

Ross, S. M. (1995). *Introduction to Stochastic Dynamic Programming.* Academic Press, Inc., Nueva York.

Puterman, M. (1978). *Dynamic Programming and Its Applications.* Academic Press, Inc., Nueva York.

Wagner, H. M. (1969). *Principles of Operations Research with Applications to Managerial Decisions.* Prentice-Hall, Nueva Jersey.

Glosario

Estado

Variable descriptiva de la información necesaria para conocer la evolución del sistema a partir de ese momento. Dicha evolución dependerá de los valores del estado y de los valores de la variable de decisión.

Función de recurrencia

Función que se desea optimizar al resolver un modelo de programación dinámica. Para que la programación dinámica sea operativa, dicha función ha de ser recursiva.

Homogénea, programación dinámica

Modelo de programación dinámica en que la definición de sus elementos (estados, variable de decisión y función de recurrencia) es la misma para todas las etapas del sistema.

No homogénea, programación dinámica

Cualquier modelo de programación dinámica que no cumpla la propiedad que caracteriza la programación dinámica homogénea: la homogeneidad a lo largo de las etapas de los estados, las variables de decisión y la función de recurrencia.

Política

La política es una regla de decisión para una etapa determinada. Son los valores de las variables de decisión que optimizan el comportamiento del sistema a partir de esa etapa, para cada uno de los estados del sistema.

Proceso polietápico

Proceso que puede dividirse en etapas, representativas de la evolución temporal del mismo. Dicho proceso va evolucionando (de manera determinista o aleatoria) entre un conjunto de estados posibles. Algunos procesos polietápicos prevén la posibilidad de tomar una decisión al principio de cada etapa. Así, la evolución del sistema en el futuro dependerá de la decisión que se tome y del estado en que se halle el sistema en la etapa anterior.

Programación dinámica

Estrategia de resolución de problemas que pueden representarse como procesos polietápicos de decisión. Cabe distinguir dos tipos de programación dinámica: si se puede conocer la evolución del sistema una vez tomada la decisión, se tiene un modelo de programación dinámica determinista. Si la evolución del sistema no está determinada, sino que puede evolucionar a diferentes estados según una ley de probabilidad conocida, el modelo será de programación dinámica aleatoria.

Recursividad

Una función descriptiva del comportamiento de un proceso polietápico es recursiva cuando sus valores para una etapa determinada dependen de los valores que pueda tomar la función en las etapas siguientes.

Variable de decisión

Variable descriptiva de las decisiones que pueden tomarse en un estado determinado. En función del modelo, el rango de valores que puede tomar la variable de decisión puede depender del estado en que se halle el sistema.