Modelos de Decisión en Ambientes Inciertos

(Apunte de Clases para el Curso Investigación Operativa IN44A)

Departamento de Ingeniería Industrial - Universidad de Chile

René A. Caldentey   Susana V. Mondschein [1]

Enero, 1999

[1]La presente es una versión preliminar de este apunte docente, el cual se encuentra en construcción. Los autores agradecen los comentarios y correcciones de eventuales errores que aún permanezcan en el texto, los cuales pueden ser comunicados a smondsch@dii.uchile.cl, rcaldent@mit.edu o hawad@dii.uchile.cl

Capítulo 2

Introducción a la Programación Dinámica

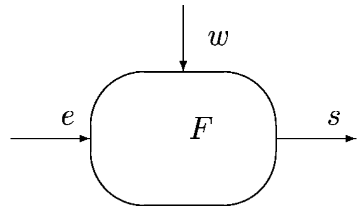
Muchos problemas de decisión de la vida real no corresponden simplemente en determinar la acción a seguir en un instante aislado, sino que se debe tomar una secuencia de decisiones a través del tiempo las cuales interactúan entre sí.

Por ejemplo, consideremos un problema clásico de manejo de inventarios en el que el administrador de una tienda debe decidir cada mes cuántas unidades de cierto producto debe ordenar para enfrentar una demanda aleatoria. Supongamos que el producto en cuestión puede ser vendido por un período máximo de un año y que pasado este período todas las unidades sobrantes pierden la totalidad o la mayor parte de su valor (para fijar ideas piense en agendas, el modelo 99 de un automóvil, artículos con el logo "01/01/00", etc.). Bajo este marco, la decisión de cuánto ordenar un mes dado debe considerar no sólo las ventas de ese mes ya que lo que sobra un período es inventario para el siguiente. De esta forma, la decisión de cuánto ordenar un mes debe ser capaz de incluir el comportamiento de compra de todo el período que resta hasta el final del año.

Para resolver este tipo de problemas se revisará en este capítulo una técnica conocida como *Programación Dinámica* la cual en esencia es una forma ordenada de manejar la información necesaria para la toma de decisión en cada instante.

2.1 Formulación General

Consideremos un sistema dinámico caracterizado por tres conjuntos E, S y W y una función $F : E \times W \to S$. Donde E es el conjunto de variables de entrada (acciones), S el conjunto de variables de salida (estados del sistema), W el conjunto de variables aleatorias y F la función característica del sistema o simplemente función del sistema. De esta forma conocidos un vector de variables de entrada $e \in E$ y una realización de las variables aleatorias $w \in W$ se obtiene un vector de salida $s = F(e, w)$.



Se asumirá que las variables de entrada e son escogidas por el tomador de decisión mientras que las variables aleatorias w son escogidas por la *Naturaleza* de acuerdo a algún mecanismo probabilístico conocido.

Para el ejemplo de manejo de inventarios planteado al comienzo, una entrada $e = (e_1, e_2, ..., e_{12})$ corresponde a las cantidades ordenadas cada uno de los doce meses del año, de forma que el conjunto de entradas posibles, E queda definido por $E = \{e \in I\!N^{12} | e_i \geq 0 \; \forall i = 1, \ldots, 12\}$. Cada elemento del conjunto $W = \{w = (w_1, .., w_{12})\}$ es un vector de 12 componentes, que corresponden a las demandas observadas cada mes. El conjunto $S = \{s = (s_1, ..., s_{12})\}$ contiene los niveles de inventario observado al comienzo de cada mes. Por último, la función característica del sistema para este ejemplo establece la conservación del inventario es decir

$$s = F(e, w) \iff s_{k+1} = \max\{0, s_k + e_k - w_k\}$$

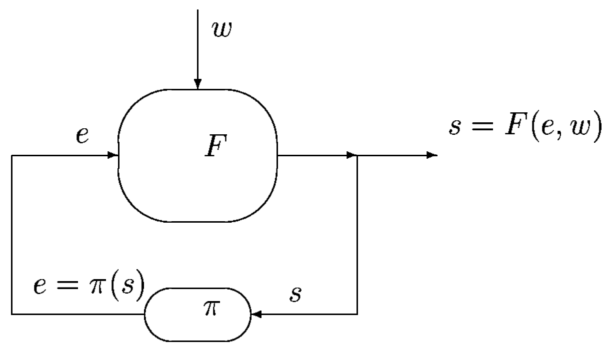
.

Defininici'on 2.1 *Se dice que $\pi : S \to E$ es una función de decisión o una política para el sistema si $\forall w \in W$ la ecuación*

$$e = \pi(F(e, w))$$

tiene solución única en e.

Luego, para cada posible realización de las variables aleatorias w, una política π genera una entrada única e y una salida única s.



Para el ejemplo de manejo de inventarios, algunos ejemplos de políticas de decisión son:

1. Ordenar todos los meses una unidad.
2. Ordenar 2 si el inventario vigente es 0 y ordenar 1 si el inventario es 1.
3. Ordenar 1 sólo si el inventario del próximo mes es 0.

No es difícil ver que los tres criterios anteriores definen una única secuencia de acciones a seguir a lo largo del año para niveles fijos de demanda y por tanto son efectivamente políticas de decisión. Sin embargo, el tercer criterio no puede ser aplicado en la práctica, pues con él se toman decisiones basados en información que no está disponible: al momento de decidir cuánto pedir este mes no sabemos cuál va a ser el inventario del próximo mes. Esto motiva la siguiente definición:

Defínici'on 2.2 *Se dice que una política de decisión es* No Anticipativa *si las decisiones tomadas con ella dependen sólo de la información disponible al momento de tomarlas.*
Más formalmente, la política π es no anticipativa ssi se cumple lo siguiente: dados w, $\tilde{w} \in W$, $e, \tilde{e} \in E$ soluciones de $e = \pi(F(e,w))$ y $\tilde{e} = \pi(F(\tilde{e},\tilde{w}))$ respectivamente, si existe $k \geq 1$ tal que $w_1 = \tilde{w}_1, w_2 = \tilde{w}_2, \ldots, w_k = \tilde{w}_k$, entonces $e_1 = \tilde{e}_1, \ldots, e_k = \tilde{e}_k$.

Sólo consideraremos políticas de decisión admisibles aquellas que sean no anticipativas. En lo que sigue designaremos por Π al conjunto de políticas de decisión no anticipativas.

Una política $\pi \in \Pi$ y una realización $w \in W$ generan una única tripleta (e, w, s), donde e es la solución de $e = \pi(F(e,w))$ y $s = F(e,w)$. Lo anterior nos permite garantizar la existencia de una función $f : W \times \Pi \to E \times W \times S$ tal que $(e, w, s) = f(w, \pi)$.

Sea $U : E \times W \times S \to \mathbb{R}$ una función que represente la utilidad del tomador de decisión. Para resolver el problema de decisión se requiere conocer qué tipo de criterio adopta el tomador de decisión frente a la incertidumbre. Por ejemplo para los criterios de *MAXMIN* y de *Valor Esperado* los problemas a resolver son:

- **Criterio MaxMin:**
$$\max_{\pi \in \Pi} \min_{w \in W} \{U(f(w,\pi))\}$$

- **Criterio del Valor Esperado:**
$$\max_{\pi \in \Pi} \{E_w \left[U(f(w,\pi))\right]\}$$

En los dos casos anteriores se tiene que el problema de decisión bajo incertidumbre se reduce a un problema de maximización sobre Π de una función numérica. La dificultad que presentan estos problemas está en que Π es un conjunto de funciones que satisfacen la condición de ser no anticipativas y por tanto se hacen en general inaplicables las técnicas clásicas de optimización como programación matemática. Para resolver este tipo de problema se utiliza una técnica conocida como *Programación Dinámica* (PD), la que consiste básicamente en descomponer el problema de maximización sobre Π en una secuencia de problemas de optimización mucho más simples los que son construidos y resueltos partiendo desde el final del horizonte de planificación y retrocediendo en el tiempo.

Nuestro interés se concentrará en un caso especial de problemas, aquellos en que se puede actuar sobre el sistema (tomar decisiones) sólo en un conjunto discreto de instantes, de forma que la evolución del sistema puede ser modelada con un eje de tiempo discreto.

2.2 Modelo de Decisión Secuencial en Tiempo Discreto

Consideremos un sistema que evoluciona a través del tiempo durante N períodos. Al comienzo de cada período se observa el estado del sistema, se puede decidir actuar sobre el sistema de alguna forma, y una vez tomada la decisión se observa la realización de alguna variable aleatoria. Llamando s_k el estado del sistema en el período k, e_k acciones tomadas en el período k y w_k el comportamiento de la naturaleza en el período k, se tiene que el estado del sistema en el período $k+1$ está determinado por s_k, e_k y w_k de acuerdo a la ecuación de recurrencia:
$$s_{k+1} = f_k(s_k, e_k, w_k) \quad k = 0, 1, .., N-1$$

2.2.1 Supuestos

- Se asume que w_k se selecciona de acuerdo a algún mecanismo probabilístico que en principio puede depender de s_k y e_k pero que no depende de $w_0, w_1, .., w_{k-1}$ (si dependiera de ellas entonces $w_0, w_1, .., w_{k-1}$ debieran estar incluidas en la información que define s_k).

- Existe *Información Perfecta*, es decir, el tomador de decisión escoge e_k conociendo perfectamente el estado del sistema s_k.
- La función de utilidad es *Aditiva*, es decir,

$$U(e, w, s) = \sum_{k=0}^{N-1} U_k(e_k, w_k, s_k) + U_N(s_N)$$

  El término $U_N(s_N)$ representa el beneficio residual que se obtiene por terminar la evolución del sistema en el estado s_N.
- Se utiliza como criterio de decisión la maximización del valor esperado de la utilidad.

2.2.2 Resolución

De acuerdo a los supuestos establecidos y a las características discretas del modelo una política de decisión π toma la forma:

$$\pi = (\pi_0(s_0), \pi_1(s_1), ..., \pi_{N-1}(s_{N-1})),$$

tal que la acción a seguir en el período k es $e_k = \pi_k(s_k)$. Vale decir es una regla que indica qué acción tomar en el período k si el estado del sistema es s_k, para todo período k y estado posible s_k.

El problema a resolver consiste entonces en determinar la función $\pi^* = (\pi_0^*, \pi_1^*, ..., \pi_{N-1}^*)$ que maximice

$$E_w \left\{ \sum_{k=0}^{N-1} U_k(\pi_k(s_k), w_k, s_k) + U_N(s_N) \right\},$$

sujeto a la restricción $s_{k+1} = f_k(s_k, e_k, w_k) \quad k = 0, 1, .., N-1$.

Para resolver este problema se construyen en forma secuencial y retrocediendo en el tiempo las funciones π_k^*, de la forma que se explica a continuación.

1. Cálculo de π_{N-1}^*. Sea s_{N-1} el estado del sistema en el período $N-1$. Entonces si se escoge una acción e a seguir la utilidad esperada que se obtendrá desde el período $N-1$ hasta el final $(V_{N-1}(e, s_{n-1}))$ será:

$$V_{N-1}(e, s_{N-1}) = E_{w_{N-1}} \{U_{N-1}(e, w_{N-1}, s_{N-1}) + U_N(f_{N-1}(s_{N-1}, e, w_{N-1}))\}$$

   Luego $\pi_{N-1}^*(s_{N-1}) = \arg\max_e \{V_{N-1}(e, s_{N-1})\}$. Al valor de dicho máximo lo denotamos $V_{N-1}^*(s_{N-1}) \equiv V_{N-1}(\pi_{N-1}^*(s_{N-1}), s_{N-1})$.

2. Cálculo de π_k^*, $k = 0, 1, .., N-1$. En forma análoga consideramos un estado del sistema s_k en el período k y buscamos la acción que maximice la utilidad esperada acumulada del período k hasta el final $(V_k(e, s_k))$. En forma recursiva $V_k(e, s_k)$ se construye como:

$$V_k(e, s_k) = E_{w_k} \left\{ U_k(e, w_k, s_k) + V_{k+1}^*(f_k(s_k, e, w_k)) \right\}$$

De donde $\pi_k^*(s_k) = \arg\max_e \{V_k(e, s_k)\}$ y $V_k^*(s_k) = \max_e \{V_k(e, s_k)\} = V_k(\pi_k^*(s_k), s_k)$.

El procedimiento anterior permite encontrar la estrategia óptima de acción a lo largo de los N períodos de tal forma de maximizar la utilidad esperada. Para ello se apoya en la construcción de funciones auxiliares V_k^* que representan la utilidad máxima en términos esperados que se puede obtener desde el período k hasta el final. La idea básica en su formulación es la siguiente: si actualmente el sistema se encuentra en el período k, todo lo que ha pasado antes de k está definido y sólo se puede optimizar la evolución futura del sistema.

La principal desventaja de este método está en la cantidad de subproblemas de optimización que deben ser resueltos: en efecto, para conocer V_k^* se debe determinar la acción óptima $\pi_k^*(s_k)\ \forall s_k \in S$, es decir, en cada período se deben resolver tantos problemas de maximización como estados posibles tenga el sistema.

2.3 Ejemplos

2.3.1 Problema de Inventarios

Retomemos el problema de manejo de inventarios expuesto al comienzo de esta sección. Supongamos que restan 3 meses para el final del período de comercialización y que actualmente existe una unidad disponible en inventario. La demanda mensual es estocástica con las siguientes características: con probabilidad $p_0 = 0.2$ no existen compradores interesados por el producto, con probabilidad $p_1 = 0.5$ existe un comprador interesado por el producto y con probabilidad $p_2 = 0.3$ existen dos clientes interesados en el producto. El precio de venta del producto es $\$P = 10$, el costo del producto es $\$\ c = 5$, el costo de poner una orden es $\$K = 10$ y el costo de rechazar una venta (no vender el producto a un cliente que lo demanda) es $\$I = 8$. ¿Cuál es la mejor estrategia para ordenar producto durante los próximos tres meses?

Sea s_k el nivel de inventario al comienzo del período k antes de realizar la orden (variable de estado), e_k la cantidad de producto a ordenar en el período k (variable de decisión) y w_k la demanda observada en el período k.

De esta forma la relación recursiva que satisface el inventario $s_{k+1} = f(s_k, e_k, w_k)$ toma la forma:

$$s_{k+1} = max\{0, s_k + e_k - w_k\}$$

La función de utilidad para un mes k cualquiera $U_k(e_k, w_k, s_k)$ viene dada por:

$$U_k(e_k, w_k, s_k) = P \cdot \min(w_k, e_k + s_k) - c \cdot e_k - K \cdot r_k - I \cdot \max(0, w_k - e_k - s_k),$$

donde $r_k = 1$ si se ordena producto en el período k y $r_k = 0$ si no.

Las funciones de beneficio esperado acumulado toman la forma:

$$V_k(e_k, s_k) = E_{w_k}\{U_k(e_k, w_k, s_k) + V^*_{k+1}(s_{k+1})\}$$

Ahora bien para resolver este problema debemos construir en forma recursiva las funciones V^*_k. Para ello aprovecharemos el carácter discreto del conjunto de acciones posibles para representar matricialmente el problema. Si consideramos que a lo más pueden venderse 2 unidades por mes y que sólo restan tres meses de venta y se dispone de una unidad de inventario inicial, a lo más se pueden ordenar 5 unidades al comienzo, 4 unidades después del primer mes y sólo 2 en el último mes.[1] Esto limita a 6 las acciones posibles en el primer período, a 5 en el segundo y a 3 en el tercero.

- **período 4**
  Al comenzar el cuarto mes ya no es posible vender más unidades y por tanto se tiene $V^*_4(s) = 0 \quad \forall s \geq 0$.

- **período 3**

$$V^*_3(s_3) = \max_{e_3} V_3(e_3, s_3) \;=\; \max_{e_3} \{E_{w_3}[U_3(e_3, w_3, s_3) + V^*_4(s_4)]\}$$

	$V_3(e_3, s_3)$			$V^*_3(s_3)$	e^*_3
s_3	$e_3 = 0$	$e_3 = 1$	$e_3 = 2$		
0	-8.8	-9.4	-9	-8.8	0
1	5.6	-4	-9	5.6	0
2	11	-4	-9	11	0
3	11	-4	-9	11	0
4	11	-4	-9	11	0
5	11	-4	-9	11	0
6	11	-4	-9	11	0

En la tabla anterior la primera columna representa los posibles estados (inventario disponible) al comenzar el último mes (período 3). Las siguientes 3 columnas presentan, para cada estado posible, el beneficio esperado si se escoge la acción no ordenar, ordenar 1 unidad u ordenar 2 unidades respectivamente. La penúltima columna presenta V^*_3, el máximo beneficio esperado que se puede alcanzar para cada estado, mientras que la última columna indica cuál es la mejor acción posible para cada estado. Se puede ver que independiente de s_3 la acción óptima a seguir es no ordenar.

Por ejemplo, el elemento asociado a $s_3 = 1$ y $e_3 = 1$ se calcula como

$$\begin{aligned}
V_3(1,1) &= E_{w_3}\left[U_3(1, w_3, 1) + \overbrace{V^*_4(s_4)}^{0}\right] \\
&= -K - c + p_0 \cdot 0 + p_1 \cdot P + p_2 \cdot 2P \\
&= -4
\end{aligned}$$

[1] Esto supone un tomador de decisiones racional.

- **período 2**

$$V_2^*(s_2) = \max_{e_2}\{E_{w_2}[U_2(e_2,w_2,s_2)+V_3^*(s_3)]\}$$

s_2	$V_2(e_2,s_2)$ $e_2=0$	$e_2=1$	$e_2=2$	$e_2=3$	$e_2=4$	$V_2^*(s_2)$	e_2^*
0	-17.6	-15.32	-6.64	-4.62	-8	-4.62	3
1	-0.32	-1.64	0.38	-3	-8	0.38	2
2	13.36	5.38	2	-3	-8	13.36	0
3	20.38	7	2	-3	-8	20.38	0
4	22	7	2	-3	-8	22	0
5	22	7	2	-3	-8	22	0
6	22	7	2	-3	-8	22	0

Por ejemplo, el elemento asociado a $s_2=1$ y $e_2=0$ se calcula como

$$\begin{aligned} V_2(0,1) &= E_{w_2}[U_2(0,w_2,1)+V_3^*(s_3)] \\ &= p_0\cdot(0P+\underbrace{V_3^*(1)}_{5.6})+p_1\cdot(P+\underbrace{V_3^*(0)}_{-8.8})+p_2\cdot(P-I+\underbrace{V_3^*(0)}_{-8.8}) \\ &= -0.32 \end{aligned}$$

Para el segundo período conviene ordenar 3 unidades si el inventario inicial en este período es 0, conviene ordenar 2 si el inventario inicial es 1 y conviene no ordenar si el inventario inicial es superior o igual a 2 unidades.

- **período 1**

$$V_1^*(s_1) = \max_{e_1}\{E_{w_1}[U_1(e_1,w_1,s_1)+V_2^*(s_2)]\}$$

s_1	$V_1(e_1,s_1)$ $e_1=0$	$e_1=1$	$e_1=2$	$e_1=3$	$e_1=4$	$e_1=5$	V_1^*	e_1^*
1	1.98	-2.524	1.87	4.598	2.514	-2	4,598	3

Para el primer período el inventario inicial es conocido y vale 1. La mejor acción a seguir es ordenar 3 unidades y el beneficio acumulado que se obtiene para los tres meses es de $4.598.

2.3.2 Problema de Asignación

La formulación del modelo de programación dinámica de este capítulo se enmarcó en un contexto de problemas de decisión secuencial a través del tiempo. En general, la programación dinámica puede ser usada para resolver una gama mucho más amplia de problemas que no necesariamente incluyen decisiones a través del tiempo. Con el objeto de introducir otras aplicaciones de la programación dinámica formularemos y resolveremos un problema de asignación mediante esta técnica.

Consideremos el problema de encontrar las cantidades no negativas $(u_0, ..., u_{N-1})$ de N productos a comprar de modo de satisfacer la restricción

$$\sum_{k=0}^{N-1} \lambda_k \cdot u_k = A$$

al menor costo

$$\sum_{k=0}^{N-1} g_k(u_k)$$

donde $A, \lambda_1, ..., \lambda_{N-1}$ son constantes positivas conocidas y $g_0, ..., g_{N-1}$ son funciones reales a valores positivos conocidas.

Definamos como variables de estado $\{x_k\}$ que representan el aporte agregado de los k primeros productos a la restricción $\sum_{k=0}^{N-1} \lambda_k \cdot u_k = A$, es decir,

$$x_{k+1} = x_k + \lambda_k \cdot u_k \quad k = 0, 1, ..., N-1$$

De esta forma el problema a resolver consiste en

$$\begin{gathered} \min \sum_{k=0}^{N-1} g_k(u_k) \ (1) \\ \text{s.a} \\ x_{k+1} = x_k + \lambda_k \cdot u_k \quad k = 0, 1, .., N-1 \ (2) \\ x_0 = 0 \ (3) \\ x_N = A \ (4) \end{gathered}$$

Claramente el problema anterior es susceptible de ser resulto usando las técnicas de programación dinámica vistas en este capítulo. Tiene una función de utilidad aditiva (1), las variables de estado están ligadas mediante una relación recursiva (2) y se tienen condiciones de borde inicial (3) y final (4).

2.4 Ejercicios

1. Considere el problema de asignación anterior. Suponga que la canasta de productos está constituida por cuatro bienes B, C, D y E. Se sabe que $\lambda_B = 1$, $\lambda_C = 2$, $\lambda_D = 4$, $\lambda_E = 7$, $A = 20$ y $g_B(x) = 3x$, $g_C(x) = 6x$, $g_D(x) = x \cdot e^x$ y $g_E(x) = x^2$. Determine los niveles de productos x_B, x_C, x_D, x_E que minimizan el costo de la compra utilizando una formulación de programación dinámica.