

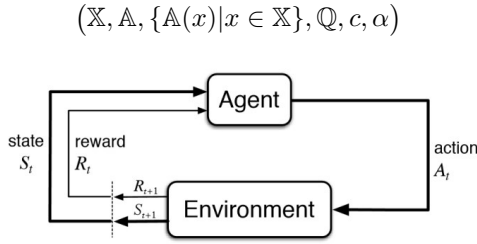
Aplicación de Procesos de Decisión de Markov: Observabilidad Completa y Parcial Inventarios

RUY LÓPEZ & HUGO CRUZ

Objetivo: Encontrar decisiones (acciones) óptimas que minimicen (o maximicen) un costo (o recompensa) esperado dado un estado inicial y donde participen elementos aleatorios en la dinámica del sistema.

1 Introducción

Un modelo de control de Markov



con criterio de rendimiento a optimizar:

$$J(\pi, x) := \mathbb{E}_x^\pi \left[\sum_{t=0}^{\infty} \alpha^t c(x_t, a_t) \right], \quad \forall x \in \mathbb{X}, \pi \in \Pi,$$

donde $\alpha \in (0, 1)$ es un factor de descuento.

Una política π^* satisfaciendo

$$J(\pi^*, x) = \inf_{\Pi} J(\pi, x) =: J^*(x), \quad \forall x \in \mathbb{X},$$

se dice ser *política óptima α -descontada*, y J^* es llamada *función de valor (función objetivo o función de utilidad óptima) α -descontada*.

2 Problema de Inventario

Sistema de inventario-producción con variable estado $x_t \in \mathbb{X} = \mathbb{R}$ (nivel de stock en el inicio de período) con evolución:

$$x_{t+1} = x_t + a_t - \xi_t$$

$a_t \equiv$ cantidad ordenada (o producida) $\mathbb{A} = \mathbb{A}(x) = [0, \infty)$.
 $\xi_t \equiv$ demanda, perturbación o v.a. exógena (i.i.d.). no-negativas e independientes de x_0 , con distribución $\nu(s)$.

El costo es:

$$c(x_t, a_t, \xi_t) = b a_t + h \max(0, x_{t+1}) + p \max(0, -x_{t+1})$$

donde $b :=$ costo de producción unitario, $h :=$ costo por almacenaje unitario y $p :=$ costo unitario por demanda no satisfecha. b, h y $p > 0$. Se asume que $p > b$.

Las EPD para el problema de inventario-producción son:

$$J_N(x) := 0,$$

$$J_t(x) := \min_{a \in \mathbb{A}(x)} \mathbb{E} \{ c(x, a, \xi) + \alpha J_{t+1}(x_{t+1}) \}.$$

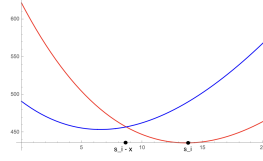


Fig. 1: Rendimientos

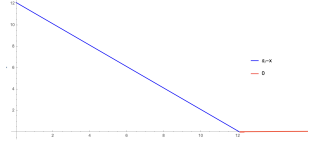


Fig. 2: Políticas

En general, la solución en cada etapa es: (Fig. 2)

$$f_t(x) = \begin{cases} 0 & \text{si } x \geq s_t \\ s_t - x & \text{si } x < s_t \end{cases}$$

$$J_t(x) = \begin{cases} L(x) + \alpha \mathbb{E}\{J_{t+1}(x - \xi_t)\} & \text{si } x \geq s_t \\ b(s_t - x) + L(s_t) + \alpha \mathbb{E}\{J_{t+1}(s_t - \xi_t)\} & \text{si } x < s_t \end{cases}$$

(Fig. 1), con

$$L(y) := h \int_{-\infty}^y (y - s) d\nu(s) + p \int_y^{\infty} (s - y) d\nu(s)$$

i	s _i	a* [x]	a* [x ₀]	y*	G _i [y*]	J _i [x ₀]	α ⁱ
0	S ₀	A ₀	a ₀	y ₀	G ₀ [y ₀]	0	
1	12.2621	(12.2621 - x)	7.26209	12.2621	448.518	436.018	
2	12.1453	(12.1453 - x)	7.14528	12.1453	468.102	455.602	
3	12.138	(12.138 - x)	7.13803	12.138	469.002	456.502	
4	12.1377	(12.1377 - x)	7.13771	12.1377	469.043	456.543	
5	12.1377	(12.1377 - x)	7.13769	12.1377	469.045	456.545	
6	12.1377	(12.1377 - x)	7.13769	12.1377	469.045	456.545	
7	12.1377	(12.1377 - x)	7.13769	12.1377	469.045	456.545	
8	12.1377	(12.1377 - x)	7.13769	12.1377	469.045	456.545	
9	12.1377	(12.1377 - x)	7.13769	12.1377	469.045	456.545	
10	12.1377	(12.1377 - x)	7.13769	12.1377	469.045	456.545	

Tabla. Política óptima y rendimientos óptimos

3 PDM's Parcialmente Observables

Se puede transformar el problema *Parcialmente Observable* a uno *Completamente Observable* considerando al proceso de estados $\{\nu_t\}$ una distribución de la variable parcialmente observada (ventas) que evoluciona así:

$$\nu_0 = \mathcal{H}_0(p, y_0), \quad y,$$

$$\nu_{t+1} = \mathcal{H}(\nu_t, a_t, y_{t+1}), \quad \text{donde } t = 0, 1, \dots \quad \forall t \geq 1,$$

$\{a_t\}$ y $\{y_t\}$ son las sucesiones de control y observación.

$$\nu_n(d) = \mathbb{P}\{d_n = d \mid \mathcal{Z}_{n-1}\}$$

$$\nu_{n+1}(d) = \mathbb{I}_{z_n < x_n + a_n} p(d \mid z_n) + \mathbb{I}_{z_n \geq x_n + a_n} \frac{\int_{x_n + a_n}^{\infty} \nu_n(\xi) p(d \mid \xi) d\xi}{\int_{x_n + a_n}^{\infty} \nu_n(\xi) d\xi}$$

Referencias:

Hernández-Lerma. *Discrete-time Markov control processes*. Springer, (1991).
Bensoussan. *Inventory problems with partially observed demands and lost sales*. Springer, (2008).