Un ejemplo de procesos de decisión de Markov sensibles al riesgo: Un enfoque por matrices no negativas

María Chávez, Hugo Cruz y Hortensia Reyes

M. Chávez, H. Cruz y H. Reyes
Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla. Avenida San Claudio y 18 sur
nagiroge@hotmail.com

M. Ramos., V.Aguilera., (eds.).Ciencias Naturales y Exactas, Handbook -©ECORFAN- Valle de Santiago, Guanajuato, 2014.

Abstract

In this paper we focus attention on risk-sensitive Markov decisión chains. We are interested in characterizations of policies maximizing growth rate expected utility. In contrast to the existing literature, the problem is handled using methods of nonnegative matrices.

24 Introducción

El trabajo está relacionado con Procesos de Decisión de Markov (PDMs) con espacio de estados finitos. Los PDMs son usados para modelar sistemas que son observados de forma discreta en el tiempo por un controlador en un periodo finito o infinito de tiempo, en el cual el sistema puede presentar una variación en su movimiento. Los PDMs son aplicados en áreas como economía, biología, ingeniería, etc.

Los PDMs se encuentran caracterizados por un modelo conocido como Modelo de Control de Markov (MCM), cuyas componentes permiten describir su desarrollo en el transcurso del tiempo. La evolución de un PDM está dado de acuerdo al siguiente procedimiento. Sea $x_t = x \in S$ un estado al tiempo $t = 0,1,...$ y $a_t = a \in A(x)$ la acción (control) elegida en ese tiempo, entonces el sistema transita del estado x al estado $x_{t+1} = y \in S$ con probabilidad $P(y|x,a)$, obteniendo una recompensa $r(x,a)$. Una vez que la transición al siguiente estado ha ocurrido, una nueva acción es elegida y, el proceso es repetido.

A la sucesión de controles que el proceso genera se le conocerá como política. Para evaluar la calidad de cada política se contará con un criterio de rendimiento que medirá la eficiencia de las políticas en función de los costos o recompensas que generan. En el presente trabajo se considera el criterio promedio.

Así, el Problema de Control Óptimo (PCO) consiste en encontrar una política que optimice el criterio de rendimiento. A la política que optimiza el criterio de rendimiento se le llama política óptima y, al criterio de rendimiento evaluado en tal política óptima se le conoce como la función de valores óptimos.

Una manera de resolver el PCO es mediante la técnica conocida como Programación Dinámica (PD) iniciada a mediados de los años 50's por Richard E. Bellman (véase Bellman (1950)). El principio de Programación Dinámica permite resolver problemas en los que es necesario tomar decisiones en etapas sucesivas que condicionan la evolución futura del sistema, afectando a las situaciones en las que el sistema se encontrará en el futuro (estados), y a las decisiones (acciones) que se plantearán; todo esto se lleva a cabo mediante la Ecuación de Optimalidad de Bellman (EO), la cual permite establecer una forma recursiva que permite resolver el problema a tratar.

Además se supondrá que el controlador es sensible al riesgo. Que el controlador sea sensible al riesgo significa que la ganancia obtenida por el proceso de decisión es evaluada por una función de utilidad, la cual es una función creciente y dependiente de g.

Algunos de los trabajos donde se estudia el caso de PDMs dotados con un criterio de rendimiento sensible al riesgo son: Howard and Matheson (1972) donde se estudian modelos finitos y se usa la teoría de Perron Frobenius para establecer la existencia de una solución. En Cavazos-Cadena and Fernández-Gaucherand (2002) se da una aproximación diferente usando el criterio de costo total sensible al riesgo.

En Di Masi and Stettner (1999) se usa un costo descontado. Finalmente, en Sladky (2008) se caracteriza una función de costo óptima sensible al riesgo mediante la teoría de Perron Frobenius.

Así, a diferencia de la literatura existente acerca de este tema donde es usada la ecuación de optimalidad para determinar a la solución óptima, en este trabajo dicha función será determinada mediante la teoría de matrices no negativas.

Finalmente, se presenta un problema donde se determina la función de valor óptimo mediante el uso de la teoría dada en el trabajo de Sladky.

24.1 Método

Procesos de Decisión de Markov sensibles al riesgo

Sea $\left(S,A,\{A(x)|x\in S\},R,P\right)$ un modelo de control de Markov donde S y A son conjuntos finitos, llamados el espacio de estados y acciones, respectivamente. Para cada $x\in S, A(x)\subset A$ es el conjunto de acciones admisibles al estado x; se define a la clase de parejas admisibles por $\mathbb{K}:=\{(x,a)|x\in S, a\in A(x)\}$. La siguiente componente, es la función de recompensa, $R\in\mathcal{B}(\mathbb{K})$, donde $\mathcal{B}(\mathbb{K})$ es el espacio de funciones real valuadas y $P=\left[p_{xy}(\cdot)\right]$ es la ley de transición.

El modelo anterior es interpretado como sigue: Sea $x_t=x\in S$ un estado al tiempo t=0,1,... y $a_t=a\in A(x)$ la acción (control) elegida en ese tiempo, entonces el sistema se traslada del estado x al estado $x_{t+1}=y\in S$ con probabilidad $p_{xy}(a)$ obteniendo una recompensa $R(x,a)$.

Para cada $t\in\square$, el espacio $\mathbb{H}_t$ de historias hasta el tiempo t es dado por $\mathbb{H}_0:=S$ y $\mathbb{H}_t:=\mathbb{H}_{t-1}\times\mathbb{K},\ t\geq 1$. Un elemento de $\mathbb{H}_t$ es denotado por $h_t=(x_0,a_0,\ldots,x_t)$. Una política $p=(f^0,f^1,\square)$ es una sucesión de vectores de decisión $\{f^n,n=0,1,\square\}$ donde $f^n\in\mathbb{F}$ con $\mathbb{F}=\prod_{i=1}^{N}A(i)$ para cada $n=0,1,\square$ y $f_i^n\in A(i)$ es la decisión (o acción) elegida en la n- ésima transición si la cadena se encuentra en el estado i.

La clase de todas las políticas es denotada por $\mathcal{P}$. Dada una política $\pi\in\mathcal{P}$ y un estado inicial $X_0=x\in S$, la distribución del proceso $\{(X_t,A_t)\}$ está determinada de manera única; tal distribución será denotada por P_x^p, mientras que la esperanza con respecto a tal distribución será denotada por E_x^p.

Para cada $g\in\square$ la función de utilidad correspondiente es la función $u^g:\square\to\square$ especificada como sigue: para cada $x\in\square$

$$u^g(x)=\begin{cases}sign(g)e^{gx}, & g\neq 0\\ x, & g=0.\end{cases}\qquad(24)$$

Se tiene que $u^g(x)$ es continua, estrictamente creciente y convexa (respectivamente, cóncava) para $g > 0$, el caso propenso al riesgo (respectivamente, $g < 0$, el caso averso al riesgo). En el caso de $g = 0$ la función de utilidad será neutral al riesgo.

Suponga que la cadena comienza en el estado $X_0 = i$ y se sigue la política $p = (f^n)$, entonces la utilidad esperada en las n siguientes transiciones esta dada por:

$$\bar{U}_i^p(g,0,n) := (\text{sign } g)U_i^p(g,0,n), \tag{24.1}$$

Donde

$$U_i^p(g,0,n) := E_i^p\left[\exp\left(g\sum_{k=0}^{n-1} R\left(X_k, f^k{}_{x_x}\right)\right)\right] > 0. \tag{24.2}$$

Similarmente, para $m < n$ si el estado inicial es $X_m = i$, escribimos

$$U_i^p(g,m,n) := E_i^p\left[\exp\left(g\sum_{k=m}^{n-1} R\left(X_k, f_{x_k}{}^k\right)\right)\right]. \tag{24.3}$$

además, sea $G_i^p(g) \in \mathbb{R}^+$ la *tasa de crecimiento* de $U_i^p(g,0,n)$ dada implícitamente por

$$a_1\left(G_i^p(g)\right)^n \le U_i^p(g,0,n) \le a_2\left(G_i^p(g)\right)^n \tag{24.4}$$

donde $a_2 > a_1 > 0$ son números reales.

Además, si $g \neq 0$ para la certeza equivalente asociado, digamos $Z_i^p(g,0,n)$, definida implícitamente por $u^g\left(Z_i^p(g,0,n)\right) := E_i^p\left\{u^g\left(\sum_{k=0}^{n-1} R(X_k, A_k)\right)\right\}$, y para su valor asintótico, digamos $J_i^p(g,0)$, tenemos

$$\begin{aligned} Z_i^p(g,0,n) &= \frac{1}{g}\ln\left\{E_i^p\left[\exp\left(g\sum_{k=0}^{n-1} R(X_k, A_k)\right)\right]\right\} \\ J_i^p(g,0) &= \limsup_{n\to\infty} \frac{1}{n} Z_i^p(g,0,n). \end{aligned} \tag{24.5}$$

El símbolo ***I*** representará a la matriz identidad y ***e*** al vector (columna) unidad. Además, para cualquier $f \in \mathbb{F}$, sea

$$Q^{(g)}(f) = \left[q_{ij}^{(g)}(f_i)\right] \tag{24.6}$$

una matriz $N \times N$ no negativa con elementos

$$q_{ij}^{(g)}(f_i) := p_{ij}(f_i)\exp\left(gR_{ij}(f_i)\right). \tag{24.7}$$

Así, se tiene que

$$U_i^p(g,0,n) = \sum_{j \in S} q_{ij}^{(g)}(f_i^0) \cdot U_j^p(g,1,n) \tag{24.8}$$

con $U_i^p(g,n,n) = 1,$ o en notación vector

$$U^p(g,0,n) = Q^{(f)}(f^0) \cdot U^p(g,1,n) \tag{24.9}$$

con $U^p(g,n,n) = e.$ (24.10)

En particular, la política $\hat{p}^{(n)} = (\hat{f}^{(k,n)})$ mazimizando a $U^p(g,0,n),$ es decir
$U^{\hat{p}}(g,0,n) = \max_p U^p(g,0,n)$ debe cumplir la siguiente ecuación de Programación Dinámica

$$\begin{aligned} U^{\hat{\pi}}(\gamma,k,n) &= \max_{f \in \mathbb{F}} \{Q^{(\gamma)}(f) U^{\pi}(\gamma,k+1,n)\} \\ &=: Q^{(\gamma)}(\hat{f}^{(k,n)}) U^{\pi}(\gamma,k+1,n) \end{aligned} \tag{24.11}$$

para $k = 0,1,\ldots,n-1$ y

$$\begin{aligned} U^{\hat{\pi}}(\gamma,n-1,n) &= \max_{f \in \mathbb{F}} \{Q^{(\gamma)}(f) \cdot e\} \\ &=: Q^{(\gamma)}(\hat{f}^{(n-1,n)}) \cdot e. \end{aligned} \tag{24.12}$$

Ya que $Q^{(g)}(f)$ es una matriz no negativa, por el teorema de Perron-Frobenius el radio espectral $r^{(g)}(f)$ de $Q^{(g)}(f)$ es igual al máximo valor propio positivo de $Q^{(g)}(f)$ y los correspondientes vectores propios izquierdo (fila) y derecho (columna), digamos $y^{(g)}(f), x^{(g)}(f),$ (llamados vectores propios de Perron) pueden ser elegidos no negativos.

Además, bajo la condición de que $x^{(g)}(f) > 0$ para cada $f \in \mathbb{F}$, se tiene que existe un vector de decisión $\hat{f} \in \mathbb{F}$ tal que $\hat{r}^{(g)} \circ r^{(g)}(\hat{f})$ es el máximo vector propio posible de $Q^{(g)}(f)$ sobre toda $f \in \mathbb{F}$ y

$$\begin{aligned} &Q^{(\gamma)}(f) x^{(\gamma)}(f) \le \max_{f \in \mathbb{F}} \{Q^{(\gamma)}(f) x^{(\gamma)}(\hat{f})\} \\ &= Q^{(\gamma)}(\hat{f}) x^{(\gamma)}(\hat{f}) = \rho^{(\gamma)}(\hat{f}) x^{(\gamma)}(\hat{f}). \end{aligned} \tag{24.13}$$

Así, obtenemos lo siguiente

Suposición 1. Existen $\hat{r}^{(g)} \circ r^{(g)}(\hat{f})$ y $\hat{x}^{(g)} \circ x^{(g)}(\hat{f}) > 0$ (único salvo la adición de una constante) tales que (para un valor dado del coeficiente de aversión al riesgo

$$g)\ \hat{\rho}^{(\gamma)} \hat{x}^{(\gamma)} = \max_{f \in \mathbb{F}} \{Q^{(\gamma)}(f) \cdot \hat{x}^{(\gamma)}\} = Q^{(\gamma)}(\hat{f}) \cdot \hat{x}^{(\gamma)}. \tag{24.14}$$

Suposición 2. Para cada $f \in \mathbb{F}$ la matriz de transición $P(f)$ tiene una única clase recurrente.

Teorema 1. Si la Suposición 1 es válida, entonces para un g dado existen números $a_2^{(g)} > a_1^{(g)} > 0$ tales que

$$a_1^{(g)} x^{(g)}(\hat{f}) \le \left(\hat{r}^{(g)}\right)^{-n} \sum_{k=0}^{n-1} Q^{(g)}(f^k) x^{(g)} \le a_2^{(g)} x^{(g)}(\hat{f}), \tag{24.15}$$

para cualquier política $p = (f^k)$ maximizando el crecimiento de $U^p(g,n)$ para $n = 0,1,\ldots$. Además, la desigualdad anterior es también válida para la política estacionaria $\hat{\pi} \sim (\hat{f})$.

Demostración. Si la Suposición 1 es válida (con $x^{(g)}(f)$ no necesariamente estrictamente positivo para cada $f \in \mathbb{F}$), podemos elegir $x^{(g)}(\hat{f}) > 0$ tal que $x^{(g)}(\hat{f}) \ge e$ o $x^{(g)}(\hat{f}) \le e$. Se tiene que iterando (1) podemos concluir que para $x^{(g)}(\hat{f}) \ge e$ y cualquier política $p = (f^k)$,

$$\begin{aligned} \sum_{k=0}^{n-1} Q^{(g)}(f^k) e &\le \sum_{k=0}^{n-1} Q^{(g)}(f^k) x^{(g)}(\hat{f}) \\ &\le \left(Q^{(g)}(\hat{f})\right)^n x^{(g)}(\hat{f}) \\ &= \left(\hat{r}^{(g)}\right)^n x^{(g)}(\hat{f}) \end{aligned} \tag{24.16}$$

y por lo tanto el comportamiento asintótico de $U^p(g,n)$ depende fuertemente de $r^{(g)}(\hat{f})$ o $\hat{r}^{(g)}$, y los elementos de $\sum_{k=0}^{n-1} Q^{(g)}(f^k) x^{(g)}(\hat{f})$ deben ser acotados superiormente por $\left(\hat{r}^{(g)}\right)^n x^{(g)}(\hat{f})$.

Similarmente, si elegimos $x^{(g)}(\hat{f}) \le e$ de (0) y (1), se tiene que para cualquier política $\hat{p}^{(n)} = (\hat{f}^{(k,n)})$ maximizando a $U^p(g,n)$:

$$\begin{aligned} \sum_{k=0}^{n-1} Q^{(g)}(\hat{f}^{(k,n)}) e &\ge \sum_{k=0}^{n-1} Q^{(g)}(\hat{f}) e \\ &\ge \left(Q^{(g)}(\hat{f})\right)^n x^{(g)}(\hat{f}) \\ &= \left(\hat{r}^{(g)}\right)^n x^{(g)}(\hat{f}) \end{aligned} \tag{24.17}$$

Por lo tanto, el crecimiento de $U^p(g,n)$ si la política que maximiza a $U^p(g,n)$ es seguida, está acotado por $\left(\hat{r}^{(g)}\right)^n x^{(g)}(\hat{f})$.

Así, de (2) y (3) se tiene la conclusión deseada.

Corolario 1. Bajo la Suposición 1, si la política $p=\left(f^{n}\right)$ que minimiza $U^{p}\left(g,n\right)$ es seguida, la tasa de crecimiento de cada elemento de $U^{p}\left(g,n\right)=\sum_{k=0}^{n-1}Q^{(g)}\left(f^{k}\right)\times e$ es igual a $\hat{r}^{(g)}$. Denotamos los elementos de $x^{(g)}\left(\hat{f}\right)>0$ por $x_{j}^{(g)}\left(f\right)>0$ para $j=1,\ldots,N$ y los elementos de la matriz $N\acute{}\ N$, $Q^{(g)}\left(f\right)$ por $q_{ij}^{(g)}\left(f_{i}\right)$, para $g\left(f\right):=g^{-1}\ln\left(r^{(g)}\left(f\right)\right)$, $w_{j}\left(f\right):=g^{-1}\ln\left(x_{j}^{(g)}\left(f\right)\right)$ con $j=1,\ldots,N$, entonces (1) y la ecuación dada en la Suposición 1 pueden ser escritas como el siguiente conjunto de ecuaciones (no-lineales)

$$e^{g\left(g\left(\hat{f}\right)+w_{i}\left(\hat{f}\right)\right)}=\max_{a\in A(i)}\left\{\sum_{j\in S}p_{ij}\left(a\right)e^{g\left(r_{ij}(a)+w_{j}\left(\hat{f}\right)\right)}\right\} \tag{24.18}$$

para $i=1,2,\ldots,N$, la cual es llamada la ecuación de optimalidad de g- recompensa promedio.
En forma multiplicativa la ecuación anterior se escribe como

$$r^{(g)}\left(\hat{f}\right)x_{i}^{(g)}\left(\hat{f}\right)=\max_{a\in A(i)}\left\{\sum_{j\in S}p_{ij}\left(a\right)e^{gr_{ij}(a)}x_{j}^{(g)}\left(\hat{f}\right)\right\} \tag{24.19}$$

para $i=1,2,\ldots,N.$
Observe que la solución a (4), respectivamente (5), es decir $g\left(\hat{f}\right)$, $w_{i}\left(\hat{f}\right)$, respectivamente $r^{(g)}\left(\hat{f}\right)$, $x_{i}^{(g)}\left(\hat{f}\right)$ es única salvo la adición de constantes (añadidos a $w_{i}\left(\hat{f}\right)$, respectivamente constantes multiplicativas aplicadas a $x_{i}^{(g)}\left(\hat{f}\right)$).

Usando lo anterior, el Corolario 1 puede ser reformulado como

Teorema 2. Si la Suposición 1 es válida, entonces para cualquier política $p=\left(f^{n}\right)$ el valor asintótico medio óptimo $J_{i}^{p}\left(g,0\right)$ es acotado inferiormente por $g\left(\hat{f}\right):=g^{-1}\ln\left(r^{(g)}\left(\hat{f}\right)\right)$. Además, si la política estacionaria $\hat{\pi}\sim\left(\hat{f}\right)$ lleva al valor mínimo de $J_{i}^{p}\left(g,0\right)$ el cual es independiente del estado inicial $i\in S$ y es igual a $g\left(\hat{f}\right):=g^{-1}\ln\left(r^{(g)}\left(\hat{f}\right)\right)$.

24.2 Resultado

A continuación se presenta un ejemplo donde se muestra la teoría anteriormente descrita. En tal ejemplo se considera un modelo no controlado.

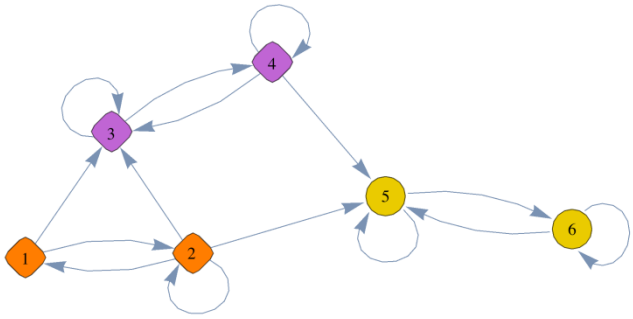
Considere un sistema no controlado, donde se cuenta con 6 estados y cuyas matrices de transición, P y de recompensa R, están dadas por

$$P=\begin{pmatrix} 0 & \frac{3}{5} & \frac{2}{5} & 0 & 0 & 0 \\ \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & 0 & \frac{1}{3} & 0 \\ 0 & 0 & \frac{3}{7} & \frac{4}{7} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{24.20}$$

y,

$$R=\begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}. \tag{24.21}$$

La siguiente gráfica reperesenta la clasificación de los estados de la cadena



Se puede observar que la cadena cuenta con una sola clase recurrente, $P_{(R)}$ dada por los estados {5,6} y dos clases de estados recurrentes, $P_{(T_1)}$ y $P_{(T_2)}$, dadas por los estados {1,2} y {3,4} respectivamente. La matriz $Q^{(g)}$ es dada a continuación:

$$Q^{(g)}=\begin{pmatrix} 0 & \frac{3}{5}e^{g} & \frac{2}{5} & 0 & 0 & 0 \\ \frac{1}{6}e^{g} & \frac{1}{3}e^{g} & \frac{1}{6} & 0 & \frac{1}{3}e^{g} & 0 \\ 0 & 0 & \frac{3}{7} & \frac{4}{7} & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2}e^{g} \\ 0 & 0 & 0 & 0 & \frac{1}{2}e^{g} & \frac{1}{2} \end{pmatrix} \tag{24.22}$$

y, se tiene que

$$Q_{T_1}^{(g)} = \begin{pmatrix} 0 & \frac{3}{5}e^{g} \\ \frac{1}{6}e^{g} & \frac{1}{3}e^{g} \end{pmatrix} \tag{24.23}$$

$$Q_{T_2}^{(g)} = \begin{pmatrix} \frac{3}{7} & \frac{4}{7} \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix} \tag{24.24}$$

y, finalmente

$$Q_{R}^{(g)} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2}e^{g} \\ \frac{1}{2}e^{g} & \frac{1}{2} \end{pmatrix}. \tag{24.25}$$

A continuación, se dan los radios espectrales de las matrices anteriores. Para la clase recurrente dada por los estados 5 y 6, se tiene que $r_R^{(g)} = 0.5 + 0.5e^{g}$; el radio espectral para la clase transitoria formada por los estados 1 y 2 es igual $r_{T_1}^{(g)} = 0.52e^{g}$. Finalmente, el radio espectral de la clase transitoria dada por los estados 3 y 4 es $r_{T_2}^{(g)} = 0.843$. Luego $r_{T_1}^{(g)} = r_{T_2}^{(g)}$ para $g = 0.476$; $r_{T_1}^{(g)} = r_R^{(g)}$ para $g = 3.03$; $r_{T_2}^{(g)} = r_R^{(g)}$ para $g = -0.374$. Luego,

Si $g < -0.374$ entonces $r_{T_2}^{(g)} > r_R^{(g)} > r_{T_1}^{(g)}$ y $r^{(g)} = r_{T_2}^{(g)} = 0.8439$.

Si $g \in (-0.374, 0.476)$ entonces $r_R^{(g)} > r_{T_2}^{(g)} > r_{T_1}^{(g)}$ y $r^{(g)} = r_R^{(g)} = 0.5 + 0.5e^{g}$.

Si $g \in (0.476, 3.03)$ entonces $r_R^{(g)} > r_{T_1}^{(g)} > r_{T_2}^{(g)}$ y $r^{(g)} = r_R^{(g)} = 0.5 + 0.5e^{g}$.

S i $g > 3.03$ entonces $r_{T_1}^{(g)} > r_R^{(g)} > r_{T_2}^{(g)}$ y $r^{(g)} = r_{T_1}^{(g)} = 0.52e^{g}$.

Por lo tanto, la tasa de crecimiento y la función de valor óptimo están dadas de la siguiente manera:

Si $g < -0.37$ entonces la tasa de crecimiento $G_i^g = 0.5 + 0.5e^{g}$ y la función de valor óptimo es $J_i^g = g^{-1}\ln\left(0.5 + 0.5e^{g}\right)$ sin importar el estado inicial.

Si $g \in (-0.37, 0.47)$ entonces la tasa de crecimiento $G_i^g = 0.5 + 0.5e^{g}$ y la función de valor óptimo es $J_i^g = g^{-1}\ln\left(0.5 + 0.5e^{g}\right)$ sin importar el estado inicial.

Si $g \in (0.47, 3.03)$ entonces la tasa de crecimiento $G_i^g = 0.5 + 0.5e^{g}$ y la función de valor óptimo es $J_i^g = g^{-1}\ln\left(0.5 + 0.5e^{g}\right)$ sin importar el estado inicial.

Si $g > 3.03$ entonces $G_i^g = 0.52e^{g}$ cuando i=1,2 y $J_i^g = g^{-1}\ln\left(0.52e^{g}\right)$ mientras que $G_i^g = 0.5 + 0.5e^{g}$ si i=3,4,5,6 y $J_i^g = g^{-1}\ln\left(0.5 + 0.5e^{g}\right)$.

A continuación se presentan los vectores propios de Perron,

Si $g < -0.374$ entonces

$$x^{(g)} = \left(\frac{-4(682.6852 - 67.4110e^{g})}{588e^{2g} + 1654.1e^{g} - 4187.9}, \frac{9.6301\left(56e^{g} + 118.1507\right)}{588e^{2g} + 1654.1e^{g} - 4187.9}, 1.3757, 1, 0, 0 \right)^{T}$$ el cual no es estrictamente positivo.

Si $g \in (-0.37, 0.47)$ entonces

$$x^{(g)} = \left(\frac{4\left(3e^{3g} - 2e^{2g} + 21e^{4g} + 12e^{g}\right)}{\left(20e^{g} - e^{2g} + 15\right)\left(7e^{2g} + e^{g} - 4\right)}, \frac{2e^{g}\left(40e^{2g} + 35e^{3g} + 3e^{g} - 10\right)}{\left(20e^{g} - e^{2g} + 15\right)\left(7e^{2g} + e^{g} - 4\right)}, \frac{4e^{g}}{\left(7e^{2g} + e^{g} - 4\right)}, \frac{e^{g}\left(7e^{g} + 1\right)}{\left(14e^{2g} + 2e^{g} - 8\right)}, 1, 1 \right)^{T} \quad (24.26)$$

el cual es estrictamente positivo.

Si $g \in (0.476, 3.03)$ entonces

$$x^{(g)} = \left(\frac{4\left(3e^{3g} - 2e^{2g} + 21e^{4g} + 12e^{g}\right)}{\left(20e^{g} - e^{2g} + 15\right)\left(7e^{2g} + e^{g} - 4\right)}, \frac{2e^{g}\left(40e^{2g} + 35e^{3g} + 3e^{g} - 10\right)}{\left(20e^{g} - e^{2g} + 15\right)\left(7e^{2g} + e^{g} - 4\right)}, \frac{4e^{g}}{\left(7e^{2g} + e^{g} - 4\right)}, \frac{e^{g}\left(7e^{g} + 1\right)}{\left(14e^{2g} + 2e^{g} - 8\right)}, 1, 1 \right)^{T} \quad (24.27)$$

el cual es estrictamente positivo.

Si $g > 3.03$ entonces $x^{(g)} = (1.1448, 1, 0, 0, 0, 0)^{T}$ el cual es estrictamente positivo.
El siguiente es el algoritmo propuesto para determinar a la política óptima:

1. Seleccionar una matriz $Q^{(g)}\left(f^{(0)}\right)$ con $f^{(0)} \in \mathbb{F}$ tal que la suma por filas sea mínima, es decir, se tiene que $Q^{(g)}\left(f^{(0)}\right)xe \leq Q^{(g)}(f)xe$ para cualquier $f \in \mathbb{F}$.

2. Para la matriz $Q^{(g)}\left(f^{(k)}\right)$ con $f^{(k)} \in \mathbb{F}$, $k = 0,1,\ldots$ calculamos su radio espectral $r\left(f^{(k)}\right)$ junto con su vector propio $x\left(f^{(k)}\right)$.

3. Se construye (si es posible) la matriz $Q^{(g)}\left(f^{(k+1)}\right)$ con $f^{(k+1)} \in \mathbb{F}$, tal que

$$Q^{(g)}\left(f^{(k+1)}\right)xx\left(f^{(k)}\right) < r\left(f^{(k)}\right)xx\left(f^{(k)}\right) = Q^{(g)}\left(f^{(k)}\right)xx\left(f^{(k)}\right). \quad (24.28)$$

4. Si tal matriz $Q^{(g)}\left(f^{(k+1)}\right)$ existe, entonces hacemos $Q^{(g)}\left(f^{(k+1)}\right) = Q^{(g)}\left(f^{(k)}\right)$ y repetimos el paso 2, de otra manera $\hat{Q}^{(g)} := Q^{(g)}\left(f^{(k)}\right)$, $\hat{f} := f^{(k)}$ y terminamos.

24.3 Conclusiones

Se estudio la teoría de procesos de decisión de Markov sensibles al riesgo y se caracterizó a la solución óptima por medio de matrices no negativas y sus valores propios, se da un ejemplo donde se muestra la aplicación de la teoría desarrollada.

Se presenta un algoritmo de iteración para determinar la política óptima, sin embargo el ejemplo presentado es no controlado. Así un posible trabajo futuro sería dar un ejemplo controlado.

24.4 Referencias

Bellman R. (1950). Dynamic Programming. Dover.

Cavazos-Cadena R. & Montes-de-Oca R. (2003). The value iteration algorithm in risk-sensitive average Markov decision chains with finite state space. *Math. Oper. Res.,* 57, 752-756.

Gantmakher, F. R. (1959). The Theory of Matrices. Chelsea, London.

Howard R. A. & Matheson J. (1972). Risk-sensite Markov decision processes. Manag. Sci. 23, 356-369.

Sladky, K. (2008). Growth rates and average optimality in risk-sensitive Markov decisión chains. Kybernetika, 44, 205–226.