Aplicaciones de los Procesos Estocásticos

Roi Naveiro Flores

Junio 2020

Modelos Estocásticos en Biología

0.1. Introducción

En esta parte de la asignatura nos centraremos en la construcción y el análisis de modelos Markovianos en biología. En particular, construiremos modelos para analizar la evolución de una epidemia y describir algunos aspectos importantes de esta evolución. Nos centraremos en los modelos SIS y SIR. Comenzaremos describiendo estos modelos de manera determinista, basándonos en ecuaciones diferenciales ordinarias. Posteriormente formularemos estos modelos de manera estocástica, como CTMCs.

0.2. Modelo SIS

Estudiaremos el modelo SIS desde un punto de vista determinista así como desde el punto de vista estocástico. Una de las diferencias más importantes entre los modelos epidémicos deterministas y los estocásticos reside en su dinámica asintótica. Eventualmente, las soluciones estocásticas convergen a un estado libre de enfermedad mientras que las correspondientes soluciones deterministas convergen a un equilibrio endémico.

0.2.1. Modelo SIS Determinista

En el modelo SIS, los individuos de la población pueden pertenecer a dos grupos: individuos infectados por la enfermedad (I) e individuos no infectados pero suceptibles de estarlo (S). En este modelo, un individuo susceptible, tras un contacto existoso con uno infectado, se enferma, pero nunca desarrolla inmunidad. Por tanto, una vez recuperado, este individuo vuelve a pertenecer a la clase susceptible. Además, asumimos que la población bajo estudio tiene un tamaño constante N (no hay muertes, no hay nacimientos...).

Sea β la tasa de contacto por unidad de tiempo y por persona infectada. Sea γ la tasa de recuperación por unidad de tiempo, es decir la fracción de infectados que se recuperan por unidad de tiempo. Si denominamos $I(t)$ al número de

infectados a tiempo t y $S(t)$ al número de susceptibles a tiempo t entonces las ecuaciones diferenciales ordinarias que describen la dinámica de esta población son

$$\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}t} &= -\frac{\beta}{N}SI + \gamma I \\ \frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\beta}{N}SI - \gamma I \end{aligned} \tag{1}$$

El segundo término de la ecuación es fácil de comprender pues γI no es más es el número total de infectados que se recuperan por unidad de tiempo. El primer término es algo más complejo. Podemos interpretar β como el número de individuos que contagia un infectado por unidad de tiempo cuando se produce contacto con ellos. Además S/N es la probabilidad de que un individuo infectado dado, entre en contacto con uno sano (es decir, la probabilidad de que cada individuo con el cual un infectado entra en contacto, sea susceptible). Luego I individuos infectados causan $\beta IS/N$ infecciones por unidad de tiempo.

La dinámica asintótica de este modelo viene dada por el siguiente Teorema.

Teorema 1. *Sean $S(t)$ e $I(t)$ soluciones a (1) y $R_0 = \frac{\beta}{\gamma}$ el número reproductivo básico. Se tiene*

1. *Si $R_0 \leq 1$ entonces $\lim_{t\to\infty}(S(t),\ I(t)) = (N, 0)$, es decir la epidemia se extingue y no quedan enfermos al final.*

2. *Si $R_0 > 1$ entonces $\lim_{t\to\infty}(S(t),\ I(t)) = \left(\frac{N}{R_0},\ N\left[1 - \frac{1}{R_0}\right]\right)$, es decir, la epidemia alcanza un equilibrio endémico.*

0.2.2. Modelo SIS estocástico

Ahora consideramos el caso en el que $S(t)$ e $I(t)$ son variables aleatorias discretas que toman valores en el espacio de estados $\{0, 1, 2, \ldots, N\}$. Como la población es fija, tan solo necesitamos estudiar una variable aleatoria, por ejemplo $I(t)$, pues $S(t)$ está determinada una vez conocemos $I(t)$, $S(t) = N - I(t)$. Además, asumiremos tiempo continuo.

Sea $\{I(t) : t \geq 0\}$ el proceso estocástico definido por el número de infectados a cada tiempo t. Además denotemos $p_i(t)$ a la probabilidad de que $I(t) = i$ a tiempo t. Asumiremos que este proceso verifica la propiedad de Markov

$$P\{I(t_{n+1})|I(t_0)\ldots I(t_n)\} = P\{I(t_{n+1})|I(t_0)\}$$

para cualquier secuencia $0 \leq t_0 < t_1 \leq \ldots < t_{n+1}$. Además, asumiremos que las probabilidades de transición solo dependen del tiempo transcurrido y no de tiempos concretos. Con esto, $\{I(t) : t \geq 0\}$ es una CTMC. Las probabilidades infinitesimales de transición en un intervalo de tiempo Δt muy pequeño vienen dadas por

$$p_{ij}(\Delta t) = \begin{cases} \frac{\beta}{N} i(N-i)\Delta t + o(\Delta t), & \text{si } \ j = i+1 \\ \gamma i \Delta t + o(\Delta t), & \text{si } \ j = i-1 \\ 1 - \left[\frac{\beta}{N} i(N-i) + \gamma i\right] \Delta t + o(\Delta t) & \text{si } \ j = i \\ o(\Delta t) & \text{si de otro modo} \end{cases}$$

Es decir, como Δt es muy pequeño, solo existen transiciones de hasta un paso. Denotemos $\lambda_i = \frac{\beta}{N} i(N-i)$ y $\mu_i = \gamma i$. Si aplocamos la propiedad de Markov y la forma de las probabilidades de transición llegamos a

$$p_i(t+\Delta t) = p_{i-1}(t)\lambda_{i-1}\Delta t + p_{i+1}(t)\mu_{i+1}\Delta t + p_i(t)\left[1 - (\mu_i + \lambda_i)\right]\Delta t + o(\Delta t)$$

Restando $p_i(t)$, dividiendo por Δt y tomando el límite $\Delta t \to 0$ obtenemos

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = p_{i-1}\lambda_{i-1} + p_{i+1}\mu_{i+1} - p_i\left[\lambda_i + \mu_i\right] \tag{2}$$

por tanto podemos escribir

$$\frac{\mathrm{d}p}{\mathrm{d}t} = Qp$$

donde Q es el generador infinitesimal de un proceso de nacimiento y muerte con tasa de nacimiento (o infección de un individuo sano) λ_i y tasas de muerte (o recuperación de un indviduo infectado) μ_i. Vemos pues que el espacio de estados admite ser descompuesto en dos partes: el estado $\{0\}$ que es un estado absorbente ($\lambda_0 = 0$) y la clase de estados transitorios $\{1, 2, \ldots, N\}$.

Una vez formulado el modelo SIS como un proceso de nacimiento y muerte podemos estudiar una serie de propiedades de interés del mismo.

ODEs para media y momentos de orden más alto de $I(t)$

En este apartado veremos cómo obtener ecuaciones diferenciales ordinarias que permitan calcular la media y momentos de orden más alto de $I(t)$. Para ello definiremos la función generatriz de momentos

$$M(\theta, t) = \mathbb{E}\left(\exp\left[\theta I(t)\right]\right) = \sum_{i=0}^{N} p_i(t)\exp\left(i\theta\right)$$

Con esto, los momentos de la distribución de $I(t)$ pueden ser calculados utilizando

$$\frac{\partial^k M}{\partial \theta^k}\Big|_{\theta=0} = \mathbb{E}[I^k(t)]$$

Primero, derivamos la ecuación diferencial que satisface la función generatriz de momentos. Multiplicando (2) por $e^{i\theta}$ y sumando en i

$$\frac{\partial M}{\partial t} = e^{\theta}\sum_{i=1}^{N} p_{i-1}e^{(i-1)\theta}\lambda_{i-1} + e^{-\theta}\sum_{i=0}^{N-1} p_{i+1}e^{(i+1)\theta}\mu_{i+1} - \sum_{i=0}^{N} p_i e^{i\theta}[\lambda_i + \mu_i]$$

Simplificando y substituyendo las formas concretas de las tasas de nacimiento y muerte llegamos a

$$\frac{\partial M}{\partial t} = \beta(e^{\theta}-1)\sum_{i=1}^{N} i p_i e^{i\theta} + \gamma(e^{-\theta}-1)\sum_{i=1}^{N} i p_i e^{i\theta} - \frac{\beta}{N}(e^{\theta}-1)\sum_{i=1}^{N} i^2 p_i e^{i\theta}$$

Por último vemos que

$$\frac{\partial M}{\partial t} = \left[\beta(e^{\theta}-1) + \gamma(e^{-\theta}-1)\right]\frac{\partial M}{\partial \theta} - \frac{\beta}{N}(e^{\theta}-1)\frac{\partial^2 M}{\partial \theta^2}$$

Ahora ya estamos en disposición de encontrar ODE para la media. Derivando la ecuación anterior con respecto a θ y evaluando en 0 obtenemos

$$\frac{\mathrm{d}\mathbb{E}[I(t)]}{\mathrm{d}t} = [\beta-\gamma]\mathbb{E}[I(t)] - \frac{\beta}{N}\mathbb{E}[I^2(t)]$$

Como la ODE de la media depende del segundo momento, no se puede resolver directamente. Lo mismo sucede con las ODEs de momentos más altos. Lo que suele hacerse es aproximar momentos de orden alto, por momentos de orden más abajo haciendo hipótesis acerca de la forma de las distribuciones (moment closure techniques). Entonces se pueden resolver las ODEs correspondientes.

Por último, usando que $\mathbb{E}[I^2(t)] \geq \mathbb{E}^2[I(t)]$ vemos que

$$\begin{aligned}\frac{\mathrm{d}\mathbb{E}[I(t)]}{\mathrm{d}t} &\leq [\beta-\gamma]\mathbb{E}[I(t)] - \frac{\beta}{N}\mathbb{E}^2[I(t)] \\ &= \frac{\beta}{N}\left(N - \mathbb{E}[I(t)]\right)\cdot\mathbb{E}[I(t)] - \gamma\mathbb{E}[I(t)]\end{aligned}$$

La parte derecha de la ecuación es la misma que la ODE determinista para $I(t)$. Esto implica que la media de $I(t)$ es menor que la solución determinista para $I(t)$.

Longitud de un brote

Supongamos que en el instante inicial hay $I(0) = i$ individuos infectados. Definimos la variable aleatoria L_i como el tiempo que transcurre hasta la extinción de la infección, supuesto que la población contiene i individuos infectados inicialmente.

Notemos que L_i es una variable aleatoria contínua $PH(e_N(i), Q_{S_T,S_T})$, donde $e_N(i)$ es un vector de tamaño N de ceros con un uno en la entrada i-ésima y

$$Q_{S_T,S_T} = \begin{pmatrix} -(\lambda_1+\mu_1) & \lambda_1 & & & \\ \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & & \\ & & \ddots & \ddots & \\ & & \mu_{N-1} & -(\lambda_{N-1}+\mu_{N-1}) & \lambda_{N-1} \\ & & & \mu_N & -\mu_N \end{pmatrix}$$

Como consecuencia, conocemos la forma de la función distribución y la tranformada de Laplace-Stieltjes de L_i.

$$\begin{aligned} F_{L_i}(x) &= 1 - e_N(i)\exp\{Q_{S_T,S_T}x\}\,\mathbf{1}_N \\ \psi_i(s) &= e_N(i)(s\mathbf{I}_N - Q_{S_T,S_T})^{-1}q_{S_T,S_A} \\ \mathbb{E}[L_i^k] &= k!e_N(i)(-Q_{S_T,S_T})^{-1}\mathbf{1}_N \end{aligned}$$

donde $q_{S_T,S_A} = -Q_{S_T,S_T}\mathbf{1}_N$. La teoría general de distribuciones PH permite demostra que existe una representación que satisface que la absorción en el estado 0 sucede con probabilidad 1. Esto implica que $\boldsymbol{Q}_{\mathcal{S}_T\mathcal{S}_T}$ es no singular y $\left(\boldsymbol{Q}^{-1}_{\mathcal{S}_T\mathcal{S}_T}\right)_{ij}$ es el tiempo medio total que la población permanece con j individuos infectados durante un brote que comienza con i infectados.

Como evaluar las función de distribución de L_i requiere la evaluación numérica de exponenciales matriciales, damos procedimiento alternativo a esta evaluación.

Primero, derivemos una ecuación recursiva de la transformada de Laplace. Para ello usamos un argumento de primer paso: estando en el estado i podemos ir al estado $i+1$ con probabilidad $\frac{\lambda_i}{\lambda_i+\mu_i}$ o al $i-1$ con probabilidad $\frac{\mu_i}{\lambda_i+\mu_i}$. El tiempo de la transición tiene distribución exponencial de parámetros $\lambda_i+\mu_i$ pues es el mínimo de dos tiempos exponenciales. La transformada de LS de este tiempo es $\frac{\lambda_i+\mu_i}{s+\lambda_i+\mu_i}$. Usando además que la transformada de LS de sumas variables aleatorias independientes es igual al producto de transformadas, tenemos

$$\psi_i(s) = \frac{\lambda_i}{\lambda_i+\mu_i}\frac{\lambda_i+\mu_i}{s+\lambda_i+\mu_i}\psi_{i+1}(s) + \frac{\mu_i}{\lambda_i+\mu_i}\frac{\lambda_i+\mu_i}{s+\lambda_i+\mu_i}\psi_{i-1}(s) \tag{3}$$

Además $\psi(0) = 1$. Es posible obtener funciones de densida de L_i con técnicas de inversión numérica de la transformada de LS.

Para obtener los momentos se puede derivar el sistema anterior. Por ejemplo, para $k = 1$ se obtiene

$$\begin{aligned} \frac{\mathrm{d}\psi_i(s)}{\mathrm{d}s}\Big|_{s=0} &= \frac{-\lambda_i}{(\lambda_i+\mu_i)^2}\psi_{i+1}(0) + \frac{\lambda_i}{\lambda_i+\mu_i}\frac{\mathrm{d}\psi_{i+1}(s)}{\mathrm{d}s}\Big|_{s=0} \\ &+ \frac{-\mu_i}{(\lambda_i+\mu_i)^2}\psi_{i-1}(0) + \frac{\mu_i}{\lambda_i+\mu_i}\frac{\mathrm{d}\psi_{i-1}(s)}{\mathrm{d}s}\Big|_{s=0} \end{aligned}$$

Usando que $(-1)^n\frac{\mathrm{d}^n\psi_i(s)}{\mathrm{d}s^n}\Big|_{s=0} = \mathbb{E}[L_i^n]$ tenemos

$$\mathbb{E}[L_i] = \frac{1}{\lambda_i+\mu_i} + \frac{\lambda_i}{\lambda_i+\mu_i}\mathbb{E}[L_{i+1}] + \frac{\mu_i}{\lambda_i+\mu_i}\mathbb{E}[L_{i-1}]$$

Además, $\mathbb{E}[L_0] = 0$.

Tamaño final de la epidemia

En este apartado estudiaremos el comportamiento de la variable aleatoria N_i^R, definida como el número total de recuperaciones observadas hasta la extinción de la epidemia, dado que $I(0) = i$. Otra variable de interés es N_i^I que es el número total de infecciones observadas hasta la extinción, dado que $I(0) = i$. Evidentemente se cumple $N_i^R = i + N_i^I$, y por tanto basta con estudiar una de las variables. Nos centraremos en N_i^R que denominamos N_i.

Definimos su función generatriz de probabilidad como $\Phi_i(z) = \sum_{k=i}^{\infty} z^k P(N_i = k)$ (obsérvese que el número de recuperaciones al menos es i y puede llegar hasta infinito). Un argumento de primer paso conduce a las igualdades $\Phi_0(z) = 1$ y

$$\Phi_i(z) = \frac{z\mu_i}{\lambda_i + \mu_i}\Phi_{i-1}(z) + \frac{\lambda_i}{\lambda_i + \mu_i}\Phi_{i+1}(z) \tag{4}$$

Esto sale del siguiente argumento de primer paso, que puede ser utilizado para calcular las probabilidades $P(N_i = k)$. Llamemos a estas probabilidades x_i^k. Está claro que

$$\begin{aligned} x_i^0 = &\quad \delta_{0,i} \quad i \in \{0, 1, \dots, N\} \\ x_0^k = &\quad \delta_{0,k} \quad k \in \mathbb{N}_0 \end{aligned}$$

Además, $x_i^k = 0$ para $k < i$, al menos tiene que haber i recuperaciones. Además

$$x_i^k = \frac{\mu_i}{\lambda_i + \mu_i}x_{i-1}^{k-1} + \frac{\lambda_i}{\lambda_i + \mu_i}x_{i+1}^k \qquad i \in \{1, \dots, \text{mín}\{k, N\}\}$$

Utilizando esta expresión es sencillo obtener las ecuaciones recursivas de la función generatriz (4), para $i \in \{1, \dots, N\}$ y $k \in \{i, i+1, \dots\}$. Además, estas ecuaciones permiten el cálculo de las probabilidades x_i^k. Otra alternativa sería resolver el sistema (4) y utilizar métodos numéricos de inversión de funciones generatrices.

Por último, derivando sobre (4) con respecto a z y evaluando en $z = 1$, podemos obtener ecuaciones para los momentos factoriales $M_i^k = \mathbb{E}[N_i(N_i - 1)\dots(N_i - k + 1)]$. En concreto, el valor esperado de N_i es igual a $M_i^1 := M_i$

$$M_i = \frac{\mu_i}{\lambda_i + \mu_i} + \frac{\mu_i}{\lambda_i + \mu_i}M_{i-1} + \frac{\lambda_i}{\lambda_i + \mu_i}M_{i+1} \tag{5}$$

Combinando esto con que $M_i^0 = P(N_i < \infty) = \Phi_i(1) = 1$ y que $M_0^k = 0$ para $k \geq 1$ es posible resolver este sistema.

0.3. Modelo SIR

Estudiaremos el modelo SIR desde un punto de vista determinista así como desde el punto de vista estocástico.

0.3.1. Modelo SIR determinista

En el modelo SIR, los individuos de la población pueden pertenecer a tres grupos: individuos infectados por la enfermedad (I) individuos no infefctados pero suceptibles de estarlo (S) e individuous recuperados inmunes (R). En este modelo, un individuo susceptible, tras un contacto existoso con uno infectado, se enferma. Una vez recuperado de la enfermedad, pasa a la clase recuperado, se vuelve inmune y por tanto no puede contraer de nuevo la enfermedad. Además, asumimos que la población bajo estudio tiene un tamaño constante N (no hay muertes, no hay nacimientos, no hay migraciones, ...). Por tanto, en todo instante de tiempo $N = S(t) + I(t) + R(t)$.

Sea β la tasa de contacto por unidad de tiempo y por persona infectada. Sea γ la tasa de recuperación por unidad de tiempo, es decir la fracción de infectados que se recuperan por unidad de tiempo. Si denominamos $I(t)$ al número de infectados a tiempo t, $S(t)$ al número de susceptibles a tiempo t y $R(t)$ al número de inmunes a tiempo t, entonces las ecuaciones diferenciales ordinarias que describen la dinámica de esta población son

$$\begin{aligned} \frac{\mathrm{d}S}{\mathrm{d}t} &= -\frac{\beta}{N}SI \\ \frac{\mathrm{d}I}{\mathrm{d}t} &= \frac{\beta}{N}SI - \gamma I \\ \frac{\mathrm{d}R}{\mathrm{d}t} &= \gamma I \end{aligned} \tag{6}$$

En este caso, la dinámica asintótica del modelo viene dada por el siguiente Teorema.

Teorema 2. *Sean $S(t)$, $I(t)$ y $R(t)$ soluciones a (6) y $R_0 = \frac{\beta}{\gamma}$ el número reproductivo básico. Se tiene*

1. *Si $R_0 \leq 1$ entonces $\lim_{t\to\infty} I(t) = 0$, es decir la epidemia se extingue y no quedan enfermos al final. Toda la población pertence a la clase recuperada al final en este caso.*
2. *Si $R_0 > 1$ entonces $\lim_{t\to\infty}(S(t),\ I(t),\ R(t)) = \left(\frac{N}{R_0},\ 0,\ N\left[1 - \frac{1}{R_0}\right]\right)$.*
3. *Si $R_0\frac{S(0)}{N} > 1$, entonces hay un crecimiento inicial en el número de infectados $I(t)$ (epidemia). Si por el contrario $R_0\frac{S(0)}{N} \leq 1$, entonces $I(t)$ decrece de manera monotona a 0.*

0.3.2. Modelo SIR estocástico

Ahora consideremos el caso en el que $S(t)$, $I(t)$ y $R(t)$ son variables aleatorias. Podemos formular la versión estocástica del modelo SIR como un CTMC bidimensional $\mathcal{X} = \{I(t), S(t) : t \geq 0\}$. La probabilidad conjunto ahora es

$$p_{(i,j)}(t) = P(I(t) = i, S(t) = j)$$

De nuevo, este proceso bivariante tiene la propiedad de Markov y es homogéneo en el tiempo. Las probabilidades infinitesimales de transición en un intervalo de tiempo Δt muy pequeño vienen dadas por

$$p_{(i,j),(i+l,j+k)}(\Delta t) = \begin{cases} \frac{\beta}{N} ij\Delta t + o(\Delta t), & \text{si } (l,k) = (1,-1) \\ \gamma i \Delta t + o(\Delta t), & \text{si } (l,k) = (-1,0) \\ 1 - \left[\frac{\beta}{N} ij + \gamma i\right] \Delta t + o(\Delta t) & \text{si } (l,k) = (0,0) \\ o(\Delta t) & \text{si de otro modo} \end{cases}$$

Es decir, como Δt es muy pequeño, solo existen transiciones de hasta un paso. De manera análoga a como hicimos anteriormente, podemos llegar a las ecuaciones de Kolmogorov que rigen este proceso.

$$\frac{\mathrm{d}p_{(i,j)}}{\mathrm{d}t} = p_{(i-1,j+1)} \frac{\beta}{N}(i-1)(j+1) + p_{(i+1,j)}\gamma(i+1) - p_{(i,j)} \left[\frac{\beta}{N} ij + \gamma i\right]$$

Por comodidad denotaremos $\lambda_{i,j} = \frac{\beta}{N} ij$ y $\mu_i = \gamma i$. Por tanto, la cadena efectúa las transiciones:

- Infección de individuo susceptible: $(i,j) \to (i+1, j-1)$ con tasa $\lambda_{i,j}$.
- Recuperación de individuo infectado: $(i,j) \to (i-1,j)$ con tasa μ_i.

Asumamos que la población en el estado inicial contiene a $I(0) = m$ infectados y $S(0) = n$ susceptibles. El espacio de estados del modelo SIR en este caso es

$$\mathcal{S} = \{(i,j) : 0 \le j \le n, 0 \le i \le m + n - j\}$$

siendo $\mathcal{S}_A = \{(0,j) : 0 \le j \le n\}$ el conjunto de estados absorbentes y $\mathcal{S}_T = \mathcal{S} - \mathcal{S}_A$ una clase de estados transitorios. La clase $\mathcal{S}_T$ contiene $J(m,n) = \frac{1}{2}(2m+n)(n+1)$ estados pues en total hay

$$\sum_{j=0}^{n} \sum_{i=0}^{m+n-j} 1 = \sum_{j=0}^{n} (m+n-j+1) = (m+n+1)(n+1) - \frac{n(n+1)}{2} = \frac{1}{2}(n+1)[2m+n+2]$$

y por tanto

$$J(m,n) = \frac{1}{2}(n+1)[2m+n+2] - (n+1) = \frac{1}{2}(2m+n)(n+1)$$

Longitud del brote

En este caso estudiamos el comportamiento de la variable aleatoria $L_{i,j}$, que es el tiempo hasta el final de la epidemia dado que el estado actual es (i,j). Las probabilidades $u_{i,j} = P(L_{i,j} < \infty)$ vienen dadas por $u_{i,j} = 1$ para cada

estado de la clase $\mathcal{S}_T$ pues la clase de estados transitorias es finita. Además L_{ij} es el tiempo hasta la absorción en algúno estado de la calse $\mathcal{S}_A$ de con matriz de tasas asociada a los estados transitorios $Q_{\mathcal{S}_T,\mathcal{S}_T}$ y por tanto

$$L_{i.j} \sim \text{PH}(\bar{e}_{J(m,n)}(i,j), Q_{\mathcal{S}_T,\mathcal{S}_T})$$

donde $\bar{e}_{J(m,n)}(i,j)$ es un vector de ceros con $J(m,n)$ elementos, con un solo 1 en la posición asociada al estado (i,j). Como consecuencia, conocemos la forma de la función distribución y la tranformada de Laplace-Stieltjes de $L_{[}i,j]$.

$$\begin{aligned} F_{L_{i,j}}(x) &= 1 - \bar{e}_{J(m,n)}(i,j) \exp\{Q_{S_T,S_T} x\} \mathbf{1}_{J(m,n)} \\ \zeta_{i,j}(s) &= \bar{e}_{J(m,n)}(i,j)(s\mathbf{I}_{J(m,n)} - Q_{S_T,S_T})^{-1} q_{S_T,S_A} \\ \mathbb{E}[L_i^k] &= k!\bar{e}_{J(m,n)}(i,j)(-Q_{S_T,S_T})^{-1}\mathbf{1}_{J(m,n)} \end{aligned}$$

donde $q_{S_T,S_A} = -Q_{S_T,S_T}\mathbf{1}_{J(m,n)}$. Como evaluar las función de distribución de $L_{i,j}$ requiere la evaluación numérica de exponenciales matriciales y esto es costoso, damos procedimiento alternativo a esta evaluación.

Utilizando un argumento de primer paso, vemos que la transformada de Laplace-Stieltjes verifica

$$\zeta_{i,j}(s) = \frac{\mu_i}{s + \lambda_{i,j} + \mu_i}\zeta_{i-1,j}(s) + \frac{\lambda_{i,j}}{s + \lambda_{i,j} + \mu_i}\zeta_{i+1,j-1}(s)$$

Como $\lambda_{i,0} = 0$ y $\zeta_{0,j} = 1$ entonces

$$\zeta_{i,0}(s) = \prod_{k=1}^{i} \frac{\mu_k}{s + \mu_k}$$

Entonces, es posible entonces computar la transformada, y usando argumentos de inversión numérica de la transformada LS derivar la función distribución y densidad de $L_{i,j}$. Un fenómeno curioso que se observa es que si $\beta < \gamma$ la distribución tiene una única moda en el cero. Si $\beta > \gamma$ esta distribución es bimodal, lo que quiere decir que hay epidemias más largas.

Para obtener los momentos de la distribución de $L_{i,j}$ se puede derivar el sistema anterior y usar que $(-1)^n \frac{\mathrm{d}^n \zeta_{i,j}(s)}{\mathrm{d}s^n}\Big|_{s=0} = \mathbb{E}[L_{i,j}^n] := m_{i,j}(n)$. Por ejemplo, tomando $m_{i,j}(0) = 1$ si $j \in \{0,\ldots,n\}$ e $i \in \{0,\ldots,m+n-j\}$ y $m_{0,j}(n) = 0$ si $j \in \{0,\ldots,n\}$, el primer momento de $L_{i,0}$ satisface

$$m_{i,0} = \frac{1}{\mu_i} + m_{i-1,0} = \sum_{i=1}^{l} \frac{1}{\mu_l}$$

y para $j \in \{1,\ldots,n\}$ e $i \in \{1,\ldots,m+n-j\}$

$$m_{i,j} = \frac{\mu_i}{\lambda_{i,j} + \mu_i} m_{i-1,j} + \frac{\lambda_{i,j}}{\lambda_{i,j} + \mu_i} m_{i+1,j-1} + \frac{1}{\lambda_{i,j} + \mu_i}$$

Tamaño final de la epidemia

En el modelo SIR la epidemia termina eventualmente. Una cantidad de interés es el tamaño final de la epidemia. Este puede determinarse a partir de la función de masa de $P(S(L_{i,j}) = j - j')$ del número de supervivientes para $j' \in \{0, \ldots, j\}$.

Notemos que el número de saltos del proceso $\mathcal{X}$ sigue una PH discreta con matriz de probabilidades de transición similar a la $Q_{\mathcal{S}_T,\mathcal{S}_T}$ pero dividiendo por los tiempos de permanencia y poniendo la diagonal a 0. (Ver Allen pag 113).

El evento $\{S(L_{i,j}) = j - j'\}$ equivale al hecho de que, en la DTMC encajada, el número de saltos hasta la absorción sea $2j' + i$ (es decir, j' nuevas infecciones e $i + j'$ recuperaciones). Como para una PH discreta se tiene $P(X = k) = \tau T^{k-1} t$, tenemos

$$P(S(L_{i,j}) = j - j') = \bar{e}_{J(m,n)}(i,j) P_{\mathcal{S}_T,\mathcal{S}_T}^{i+2j'-1} p_A$$

donde el vector columna p_A contiene las probabilidades de transición en una etapa desde los estados transitorios y los absorbentes.

Es fácil ver que

$$\mathbb{E}[S(L_{i,j})] = j + 1 - \bar{e}_{J(m,n)}(i,j) \left[\boldsymbol{I}_{J(m,n)} - P_{\mathcal{S}_T,\mathcal{S}_T}^2 \right]^{-1} \mathbf{1}_{J(m,n)}$$

donde $\mathbf{1}_{J(m,n)}$ es un vector de 1's de longitud $J(m,n)$. Para ello observamos que

$$\begin{aligned}
\mathbb{E}[S(L_{i,j})] &= \sum_{j'=0}^{j} (j - j') \bar{e}_{J(m,n)}(i,j) P_{\mathcal{S}_T,\mathcal{S}_T}^{i+2j'-1} p_A \\
&= j \sum_{j'=0}^{j} \bar{e}_{J(m,n)}(i,j) P_{\mathcal{S}_T,\mathcal{S}_T}^{i+2j'-1} p_A - \sum_{j'=0}^{j} j' \bar{e}_{J(m,n)}(i,j) P_{\mathcal{S}_T,\mathcal{S}_T}^{i+2j'-1} p_A \\
&= j \sum_{j'=0}^{j} \bar{e}_{J(m,n)}(i,j) P_{\mathcal{S}_T,\mathcal{S}_T}^{i+2j'-1} p_A - \sum_{k=1}^{j+1} (k-1) \bar{e}_{J(m,n)}(i,j) P_{\mathcal{S}_T,\mathcal{S}_T}^{i+2(k-1)-1} p_A \\
&= (j+1) - \sum_{k=1}^{j+1} k \bar{e}_{J(m,n)}(i,j) P_{\mathcal{S}_T,\mathcal{S}_T}^{2(k-1)} q_A
\end{aligned}$$

donde $q_A = P_{\mathcal{S}_T,\mathcal{S}_T}^{i-1} p_A$. Este segundo término, corresponde al primer momento de una variable aleatoria PH discreta con representación $(\bar{e}_{J(m,n)}(i,j), P_{\mathcal{S}_T,\mathcal{S}_T}^2)$, ya que q_A puede ser interpretado como el vector de probailidades de estados transitorios a estados absorbentes de ese proceso (verifica $q_A + P_{\mathcal{S}_T,\mathcal{S}_T}^2 \mathbf{1}_{J(m,n)} = \mathbf{1}_{J(m,n)}$). Por tanto

$$\sum_{k=1}^{j+1} k \bar{e}_{J(m,n)}(i,j) P_{\mathcal{S}_T,\mathcal{S}_T}^{2(k-1)} q_A = \bar{e}_{J(m,n)}(i,j) \left[\boldsymbol{I}_{J(m,n)} - P_{\mathcal{S}_T,\mathcal{S}_T}^2 \right]^{-1} \mathbf{1}_{J(m,n)}$$

obteniéndose la expresión de partida.

Diseño Óptimo

0.4. Introducción

Hasta ahora se han estudiado maneras de cuantificar el rendimiento de sistemas estocásticos, desarrollando métodos que permiten predecir el futuro de los mismos, e.g. distribuciones límite, distribuciones de ocupación, etc. En la práctica, no basta solo com predercir el futuro de los sistemas estocásticos, necesitamos además poder controlarlo para que su rendimiento cumpla una serie de objetivos dados. Este control de los sistemas puede dividirse en dos grupos: control *estático* y control *dinámico*.

En problemas de control estático, los parámetros del sistema se fijan en un valor tal que el sistema alcance el comportamiento deseado. Una vez fijos, su valor no cambia, sea cual sea el comportamiento observado del sistema. Los parámetros bajo control en este caso se denominan *variables de diseño* y el problema de encontrar las mejores variables de diseño es un problema de *diseño óptimo*.

En problemas de control dinámico, los parámetros del sistema van cambiando dinámicamente en respuesta a la evolución observada del sistema. En este caso, los parámetros bajo control se denominan *variables de control o de decisión* y el problema de encontrar las mejores variables de control es un problema de *control óptimo* o *decisión óptima*.

En este capítulo estudiaremos una serie de ejemplos de problemas de diseño óptimo. El método básico para calcular el valor óptimo de las variables de diseño contiene dos fases:

1. Analizar los modelos descriptivos del sistema para derivar una relación funcional entre el rendimiento del mismo y los valores de los parámetros.

2. Usar técnicas númericas o analíticas para calcular el valor de los parámetros que optimiza el rendimiento.

0.5. Cantidad óptima de pedido

Una tienda almacena un producto de temporada. La número total K de unidades de este producto se pide de una sola vez. La demanda del producto es desconocida, pero se conoce su distribución de probabilidad. El coste de

obtención de cada unidad es c y el precio de venta es p. Si la demanda excede K entonces la ganancia es $(p-c)K$. Si la demanda es menor que K, las unidades sobrantes se revenden a precio $s < c$ y por tanto la ganancia es $(p-c)D + (s-c)(K-D)$. Por tanto, podemos escribir la ganancia como $(p-c)\,\text{mín}(D,K) + (s-c)\,\text{máx}(K-D,0)$. Por tanto, la ganancia esperada al fijar el pedido en K unidades

$$\begin{aligned} G(K) &= \mathbb{E}\left((p-c)\,\text{mín}(D,K) + (s-c)\,\text{máx}(K-D,0)\right) = \\ &= (p-c)\sum_{d=0}^{K-1} P(D>d) + (s-c)\sum_{d=0}^{K-1} P(D\leq d) \end{aligned}$$

(Ver apuntes Artalejo para demostración detallada de las esperanzas del mínimo y el máximo).

Para encontrar el valor de K que maximiza la ganancia esperada, estudiamos el comportamiento de $G(K)$

$$G(K+1) - G(K) = (s-p)P(D\leq K) + (p-c)$$

Por tanto se tiene que

$$\begin{aligned} P(D\leq K) &< \frac{p-c}{p-s} \Rightarrow G(K+1) > G(K) \\ P(D\leq K) &\geq \frac{p-c}{p-s} \Rightarrow G(K+1) \leq G(K) \end{aligned}$$

De aquí es inmediato ver que existe un K^* tal que si $K < K^*$ entonces $G(K+1) > G(K)$ y si $K \geq K^*$ entonces $G(K+1) \leq G(K)$. Esto es así pues $\frac{p-c}{p-s} \in (0,1)$ y $P(D \leq d)$ crece de forma monótona desde 0 hasta 1. Luego $G(K)$ tiene un máximo en $K = K^*$.

Si $p = 29{,}99$, $c = 15$, $s = 9{,}99$ y la demanda sigue una distribución de Poisson de media 300, entonces $\frac{p-c}{p-s} = 0{,}7495$. Además

$$P(D\leq 311) = 0{,}7484 \qquad P(D\leq 312) = 0{,}7662$$

Luego $K^* = 312$.

0.6. El vendedor de periódicos

Un vendedor de periódicos compra unidades a un distribuidor a c céntimos la copia, para después venderlos a p céntimos por copia. Cada día compra un número fijo K e intenta venderlos antes de las cinco de la tarde. Si vende todas

las copias antes de esta hora, se retira. Sino, recicla las copias no vendidas, sin sacar beneficio alguno. Por experiencia, el vendedor conoce la función de distribución de la demanda de periódicos $P(D)$. Calcúlese el valor esperado de la ganancia como función de K, $G(K)$ y el número de unidades que maximiza su beneficio esperado.

Podemos escribir la ganancia como $(p-c)\,\text{mín}(D,K) - c\,\text{máx}(K-D,0)$. Por tanto, la ganancia esperada al fijar el pedido en K unidades

$$\begin{aligned} G(K) &= \mathbb{E}\left((p-c)\,\text{mín}(D,K) + -c\,\text{máx}(K-D,0)\right) = \\ &= (p-c)\sum_{d=0}^{K-1} P(D>d) - c\sum_{d=0}^{K-1} P(D\leq d) \end{aligned}$$

Para encontrar el valor de K que maximiza la ganancia esperada, estudiamos el comportamiento de $G(K)$

$$G(K+1) - G(K) = -p\cdot P(D\leq K) + (p-c)$$

Por tanto se tiene que

$$\begin{aligned} P(D\leq K) < \frac{p-c}{p} &\Rightarrow G(K+1) > G(K) \\ P(D\leq K) \geq \frac{p-c}{p} &\Rightarrow G(K+1) \leq G(K) \end{aligned}$$

Como $0 < \frac{p-c}{p} < 1$, y $P(D\leq K)$ crece de 0 a 1 con K, existirá un K^* tal que si $K < K^*$ entonces $P(D\leq K) < \frac{p-c}{p}$ y si $K \geq K^*$ entonces $P(D\leq K) \geq \frac{p-c}{p}$. O lo que es lo mismo

$$\begin{aligned} G(K+1) > G(K) \quad &\text{si} \quad K < K^* \\ G(K+1) \leq G(K) \quad &\text{si} \quad K \geq K^* \end{aligned}$$

Luego el coste se maximiza en $K = K^*$.

Si $C = 40$, $P = 50$ y la distribución de la demanda es binomial con parámetros $n = 200$ y $p = 0{,}8$, ¿qué número de unidades maximiza el beneficio esperado? En este caso, $\frac{p-c}{p} = 0{,}2$. Además, se tiene que $P(D \leq 155) = 0{,}211$ y $P(D \leq 154) = 0{,}165$, luego $K^* = 155$.

De aquí es inmediato ver que existe un K^* tal que si $K < K^*$ entonces $G(K+1) > G(K)$ y si $K \geq K^*$ entonces $G(K+1) \leq G(K)$. Luego $G(K)$ tiene un máximo en $K = K^*$.

0.7. Número óptimo de cajeros

Los clientes que llegan a un banco forman una única cola para ser atendidos por los cajeros disponibles. Asumamos que el director del banco decide usar s cajeros. En el banco pueden estar como máximo K personas al mismo tiempo, pues no hay espacio para más. Ningún cliente puede entrar en el banco cuando este está lleno. Asumamos que el coste de los cajeros es C_t euros por hora y el coste de espera se cuantifica como C_w euros por cliente por hora. El objetivos es determinar el número de cajeros s que minimice el coste por tiempo a largo plazo.

Asumimos que los clientes llegan al banco siguiendo un proceso de Poisson de parámetro λ y requieren un tiempo de servicio iid con distribución exponencial de parámetro μ. Sea $X(t)$ el número de clientes que están en el banco a tiempo t, (clientes haciendo cola y clientes en caja). Entonces $\{X(t), t \geq 0\}$ es el proceso longitud de cola de una cola $M/M/s/K$, con espacio de estados $\{0, 1, 2 \ldots, K\}$. Hemos de escoger el número de cajeros s que minimice el coste esperado por hora en el largo plazo. El coste por hora cuando hay i clientes en el banco y s cajeros operativos, viene dado por

$$c(i, s) = iC_w + sC_t$$

Sea $g(i, T, s)$ el coste total esperado a tiempo T cuando $X(0) = i$ y existen s cajeros operativas, i.e.

$$g(i, T, s) = \mathbb{E}\left(\int_0^T c(X(t), s)dt \;\Big|\; X(0) = i\right)$$

Veamos que $g(i, T, s) = \sum_{j=0}^{K} m_{i,j}(T)c(j, s)$ donde $m_{i,j}(T)$ son los elementos de la matriz de ocupación, es decir; $m_{i,j}(T)$ es el tiempo esperado que el proceso pasa en el estado j cuando empieza en i. Primero veamos que $m_{i,j}(T) = \int_0^T p_{i,j}(t)dt$, donde $p_{i,j}(t) = P(X(t) = j|X(0) = i)$. Definamos la siguiente variable auxiliar

$$Y_j(t) = \begin{cases} 1, & \text{si } \; X(t) = j \\ 0, & \text{si } \; X(t) \neq j \end{cases}$$

Entonces, que

$$\begin{aligned} m_{i,j}(T) &= \mathbb{E}\left(\int_0^T Y_j(t)dt\big|X(0) = i\right) = \int_0^T \mathbb{E}\left(Y_j(t)\big|X(0) = i\right) dt \\ &= \int_0^T P(Y_j(t) = 1|X(0) = i)dt = \int_0^T P(X(t) = j|X(0) = i) \\ &= \int_0^T p_{i,j}(t)dt \end{aligned}$$

Con esto

$$\begin{aligned} g(i,T,s) &= \mathbb{E}\left(\int_0^T c(X(t),s)dt \;\middle|\; X(0)=i\right) \\ &= \int_0^T \mathbb{E}\left(c(X(t),s)|\ X(0)=i\right)dt \\ &= \int_0^T \sum_{j=0}^{K} c(j,s)P(X(t)=j|X(0)=i)dt \\ &= \sum_{j=0}^{K} c(j,s)\int_0^T P(X(t)=j|X(0)=i)dt = \sum_{j=0}^{K} c(j,s)m_{i,j}(T) \end{aligned}$$

Se define coste esperado por hora en el largo plazo como

$$g(s) = \lim_{T\to\infty} \frac{g(i,T,s)}{T}$$

y es fácil ver que verifica $g(s) = \sum_{j=0}^{K} c(j,s)p_j$.

$$\begin{aligned} g(s) &= \lim_{T\to\infty} \frac{g(i,T,s)}{T} \\ &= \lim_{T\to\infty} \frac{\sum_{j=0}^{K} c(j,s)m_{i,j}(T)}{T} \\ &= \sum_{j=0}^{K} c(j,s) \lim_{T\to\infty} \frac{m_{i,j}(T)}{T} = \\ &= \sum_{j=0}^{K} c(j,s)p_j \end{aligned}$$

donde $p_j = \lim_{t\to\infty} P(X(t)=j)$ es la distribución límite. Aquí estamos utilizando que

$$\lim_{T\to\infty} \frac{m_{i,j}(T)}{T} = p_j$$

Esto es fácil de ver intuitivamente pues $\lim_{T\to\infty} \frac{m_{i,j}(T)}{T}$ es la fracción de tiempo que el proceso pasa, en media, en el estado j en el largo plazo, cuando $X(0)=i$. Esta cantidad es la componente j-ésima de la distribución límite. No obstante, la demostración formar cae fuera del alcance del curso.

Para nuestro problema, necesitamos por tanto calcular la distribución límite del proceso $X(t)$. $X(t)$ puede ser visto como un proceso de nacimiento muerte con tasa de nacimiento $\lambda_i = \lambda$ para $i = 0,1,\ldots,K-1$ y tasas de muerte $\mu_i = \min(s,i)\mu$ para $i = 0,1,\ldots,K$. La distribución límite de un proceso de nacimiento muerte viene dada por

$$p_i = \frac{\rho_i}{\sum_{j=0}^{K} \rho_j}$$

donde $\rho_i = \frac{\prod_{j=0}^{i-1}\lambda_j}{\prod_{j=1}^{i}\mu_i}$. (La derivación es trivial, tan solo se requiere escribir las ecuaciones de balance $p_j r_j = \sum_{i=1}^{N} p_i r_{ij}$, despejar de la primera, e ir sumando y despejando consecutivamente). Para nuestro caso tenemos

$$\rho_i = \begin{cases} \frac{1}{i!}\left(\frac{\lambda}{\mu}\right)^i, & \text{si } \ 0 \le i \le s-1 \\ \frac{s^s}{s!}\rho^i, & \text{si } \ s \le i \le K \end{cases}$$

además

$$\begin{aligned} \sum_{j=0}^{K}\rho_j &= \sum_{j=0}^{s-1}\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j + \sum_{j=s}^{K}\frac{s^s}{s!}\rho^j = \sum_{j=0}^{s-1}\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j + \frac{s^s}{s!}\rho^s\sum_{j=s}^{K}\rho^{j-s} \\ &= \sum_{j=0}^{s-1}\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j + \frac{s^s}{s!}\rho^s\sum_{j=0}^{K-s}\rho^j = \sum_{j=0}^{s-1}\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j + \frac{s^s}{s!}\rho^s\frac{1-\rho^{K-s+1}}{1-\rho} \end{aligned}$$

donde $\rho = \lambda/(\mu s)$. Por tanto, la distribución límite es $p_i = p_0\rho_i$, donde

$$p_0 = \left[\sum_{j=0}^{s-1}\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j + \frac{s^s}{s!}\rho^s\sum_{j=0}^{K-s}\rho^j\right]^{-1} = \left[\sum_{j=0}^{s-1}\frac{1}{j!}\left(\frac{\lambda}{\mu}\right)^j + \frac{(\lambda/\mu)^s}{s!}\frac{1-\rho^{K-s+1}}{1-\rho}\right]^{-1}$$

Por tanto podemos computar el coste por minuto a largo como

$$g(s) = \sum_{j=0}^{K} c(j,s)p_j = sC_t + C_w\sum_{j=0}^{K} jp_j$$

Que puede ser optimizado numéricamente. Por ejemplo, supongamos que los cajeros cuestan 15 euros por hora y el coste de espera se estima en 10 euros por cliente y por hora. La tasa de llegada es de 10 clientes por hora y el tiempo medio de servicio son 10 minutos. La capacidad del banco es de 15 clientes. Estos datos se traducen en $K = 15$, $C_w = 10$, $\lambda = 10$ y $\mu = 6$. En este caso se puede ver que el coste se minmiza cuando $s = 3$. Ahora bien, estimar el coste de espera es difícil puesto que requiere evaluar aspectos no cuantitativos. Por tanto, es importante hacer un análisis de sensibilidad de este parámetro. Los cálculos numéricos muestran que un solo cajero es óptimo cuando $0 < C_w \le 1{,}66$, dos cajeros cuando $1{,}67 < C_w \le 6{,}14$, tres cuando $6{,}15 < C_w \le 50{,}1$, cuatro cuando $50{,}11 < C_w \le 258{,}53$, etc.

0.8. Arrendamiento óptimo de líneas de teléfono

Una pequeña compañía de teléfono arrienda líneas de otra más grande, y carga a sus clientes por llamar en ellas. Si todas las líneas están cubierta, un

potencial cliente se pierde, si no, el cliente entra en una línea vacía de manera automática. Cada cliente aceptado paga c céntimos por minuto. El coste del alquiler de cada línea es de D euros por año. La única variable que controla la compañía es el número de líneas que alquila. ¿Cuál es el valor que maximiza la ganancia esperada de la compañía?

Supongamos que las llamadas llegan de acuerdo con un proceso de Poisson de tasa λ y que cada llamada aceptada dura un tiempo con distribución exponencial de parámetro μ. Las duraciones de las llamadas son independientes entre sí. Sea $X(t)$ el número de llamadas en progreso a tiempo t. Vemos que $\{X(t), t \geq 0\}$ es el proceso longitud de cola de una cola $M/M/K/K$ con tasa de llegada λ y tasa de tiempo de servicio μ. Hemos de escoger K que maximice el beneficio neto esperado por minuto en el largo plazo. Es fácil ver que cuando hay i llamadas en el sistema, el beneficio neto por minuto es $r(i) = ic - Kd$, donde $d = 100D/(3652460)$ es el coste de arrendamiento por línea por minuto.

Necesitamos buscar el valor de K que maximice el beneficio neto esperado por minuto en el largo plazo, $g(K) = \sum_{i=0}^{K} r(i)p_i$. Por tanto, necesitamos calcular la distribución límite de $X(t)$.

$X(t)$ puede ser visto como un proceso de nacimiento muerte con tasa de nacimiento $\lambda_i = \lambda$ para $i = 0, 1, \ldots, K-1$ y tasas de muerte $\mu_i = i\mu$ para $i = 0, 1, \ldots, K$. La distribución límite de este proceso es

$$p_i(K) = \frac{\rho_i}{\sum_{j=0}^{K} \rho_j}$$

donde $\rho_i = \frac{(\lambda/\mu)^i}{i!}$. Sustituyendo y simplificando se tiene

$$g(K) = c\rho \left(1 - \frac{\frac{\rho^K}{K!}}{\sum_{j=0}^{K} \frac{\rho^j}{j!}} \right) - Kd$$

donde $\rho = (\lambda/\mu)$. Esta expresión puede ser optmizada numéricamente.

0.9. Sustitución óptima I

Supongamos que para cierta persona es muy inconveniente que su coche le deje tirado porque la batería se estropee. Esta persona decide cambiar la batería del coche a los T años, si esta no se ha estropeado antes, en cuyo caso la cambiaría en cuanto se estropee. ¿Cuál es el valor de T que minimiza el coste de mantenimiento por año en el largo plazo?

Supongamos que la vida de las distintas baterías son variables aleatorias iid con cdf $G(t)$. Cada nueva batería cuesta C euros. Si la batería falla antes del tiempo T entonces existen D euros adicionales de coste, por molestias. El objetivo es calcular el coste por año en el largo plazo, $g(T)$ como función de T.

Sea L_n la vida de la enésima batería y $T_n = \text{mín}(L_n, T)$ el enésimo tiempo entre sustituciones. Un proceso de renovación es aquel proceso de conteo en el

que los tiempos entre eventos son variables aleatorias iid. El proceso $\{N(t), t \geq 0\}$, donde $N(t)$ es el número de baterías reemplazadas durante los primeros t años, es evidentemente un proceso de renovación, ya que L_n son variables iid y por tanto T_n también lo son. Es fácil probar que la tasa de renovación a largo plazo, definida como

$$\lim_{t\to\infty} \frac{N(t)}{t}$$

verifica

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mathbb{E}[T_1]}$$

Este resultado será útil posteriormente.

Sea un sistema que incurre en costes y sea $C(t)$ el coste neto incurrido en $[0, t]$. Dividamos el tiempo en ciclos, siendo T_n la longitud del ciclo n. Además sea $S_0 = 0$ y $S_n = \sum_{i=1}^{n} T_n$. Con lo que el enésimo intervalo es $(S_{n-1}, S_n]$, y el coste neto incurrido en el enésimo intervalo es $C_n = C(S_n) - C(S_{n-1})$. El proceso estocástico $\{C(t), t \geq 0\}$ se denomina proceso cumulativo si $\{(T_n, C_n), n \geq 1\}$ es una secuencia de variables aleatorias bivariantes iid, o lo que es lo mismo: los tiempos T_n son iid y los costes incurridos en estos tiempos son iid. Sea, en nuestro problema, C_n el coste incurrido en el intervalo anterior a la enésima reparación. Se tiene que

$$C_n = \begin{cases} C + D, & \text{si } L_n < T \\ C, & \text{si } L_n \geq T \end{cases}$$

Sea $C(t)$ el coste neto total hasta tiempo t. Es claro que $\{C(t), t \geq 0\}$, es un proceso cumulativo. El coste por año a largo plazo viene definido por

$$\lim_{t\to\infty} \frac{C(t)}{t}$$

en nuestro caso tenemos

$$\begin{aligned} \lim_{t\to\infty} \frac{C(t)}{t} &= \lim_{t\to\infty} \frac{\sum_{n=1}^{N(t)} C_n}{t} = \lim_{t\to\infty} \frac{\sum_{n=1}^{N(t)} C_n}{N(t)} \frac{N(t)}{t} = \\ &= \lim_{t\to\infty} \frac{\sum_{n=1}^{N(t)} C_n}{N(t)} \lim_{t\to\infty} \frac{N(t)}{t} = \lim_{k\to\infty} \frac{\sum_{n=1}^{k} C_n}{k} \frac{1}{\mathbb{E}[T_1]} = \\ &= \mathbb{E}[C_1] \frac{1}{\mathbb{E}[T_1]} \end{aligned}$$

Además

$$\begin{aligned} \mathbb{E}[C_1] &= \mathbb{E}[C_1 | L_1 > T] P(L_1 > T) + \mathbb{E}[C_1 | L_1 \leq T] P(L_1 \leq T) \\ &= C P(L_1 > T) + (C + D) P(L_1 \leq T) = \\ &= C + D G(T) \end{aligned}$$

y

$$\mathbb{E}[T_1] = \mathbb{E}[\text{mín}(L_1, T)] = \int_0^T (1 - G(t))dt$$

Con lo que

$$g(T) = \frac{C + DG(T)}{\int_0^T (1 - G(t))dt}$$

que puede ser optimizado para conseguir el valor óptimo de T.

Supóngase que la vida media de las baterías siguen una distribución de Erlang con media 4 años y varianza 8. El coste de sustitución es de 25 euros y el coste de fallo 200.

Si las vidas de las baterías siguen una distribución $\text{Erl}(k, \lambda)$, como la media es k/λ y la varianza k/λ^2, entonces $\lambda = 0{,}5$ y $k = 2{,}0$. Con esto

$$G(T) = 1 - e^{-T/2}[1 + T/2]$$

y

$$\int_0^T [1 - G(T)] = 4 - (4 + T)e^{-T/2}$$

Además, $C = 75$, $D = 25$. Por tanto

$$g(T) = \frac{75 + 27(1 - e^{-T/2}[1 + T/2])}{4 - (4 + T)e^{-T/2}}$$

Evaluando esta función vemos que el coste por unidad de tiempo disminuye de infinito para $T = 0$ a 42.56 a $T = 1{,}41$ y después aumenta a 56,25. Por tanto la edad óptima para sustituir la batería es 1.48.

0.10. Sustitución óptima II

Considérese un modelo de remplazamiento de un máquina como el que sigue. Supóngase que la vida de las máquinas (en años) son variables aleatorias iid con función de distribución $G(t)$ y media τ. Una máquina nueva cuesta C euros. Las máquinas se reemplazan cuando fallan o cuando alcanzan la edad T. Una máquina que funciona, puede revenderse a precio $Re^{-t/\tau}$. Las máquinas que han fallado no dan beneficio alguno. Calculese el coste por año a largo plazo y escríbase un programa de ordenador que calcule la edad óptima de reemplazamiento si las vidas de las máquinas tienen una distribución $U(a, b)$.

En este caso se tiene que

$$C_n = \begin{cases} C, & \text{si } L_n \leq T \\ C - R \cdot e^{-\frac{T}{\tau}}, & \text{si } L_n > T \end{cases}$$

Si $C(t)$ es el coste neto a tiempo t entonces $\{C(t), t \geq 0\}$ es un proceso cumulativo y de nuevo tenemos que

$$\lim_{t\to\infty} \frac{C(t)}{t} = \mathbb{E}[C_1]\frac{1}{\mathbb{E}[T_1]}$$

Tenemos que

$$\mathbb{E}[T_1] = \mathbb{E}[\text{mín}(L_1, T)] = \int_0^T (1 - G(t))dt$$

y además

$$\begin{aligned}
\mathbb{E}[C_1] &= \mathbb{E}[C_1|L_1 \leq T]P(L_1 \leq T) + \mathbb{E}[C_1|L_1 > T]P(L_1 > T) \\
&= CG(T) + (1 - G(T))\left[C - Re^{-\frac{T}{\tau}}\right] \\
&= C - Re^{-\frac{T}{\tau}}\left[1 - G(T)\right]
\end{aligned}$$

Con esto

$$g(T) = \frac{C - Re^{-\frac{T}{\tau}}\left[1 - G(T)\right]}{\int_0^T (1 - G(t))dt}$$

que puede ser optimizada numéricamente.

Estudiemos el caso en el que la vida media de las máquinas sigue una distribución $U(a, b)$. En este caso tenemos que

$$G(t) = \begin{cases} 0, & \text{si } \ t < a \\ \frac{t-a}{b-a}, & \text{si } \ t \in [a, b) \\ 1, & \text{si } \ t \geq b \end{cases}$$

Además

$$\int_0^T (1 - G(t))dt = \begin{cases} T, & \text{si } \ T < a \\ a + \frac{(a-T)(a-2b+T)}{2(b-a)}, & \text{si } \ T \in [a, b) \\ \frac{a+b}{2}, & \text{si } \ T \geq b \end{cases}$$

Con esto

$$g(T) = \begin{cases} \frac{1}{T}\left[C - Re^{-T/\tau}\right], & \text{si } \ T < a \\ \frac{1}{a + \frac{(a-T)(a-2b+T)}{2(b-a)}}\left\{C - Re^{-T/\tau}\frac{b-T}{b-a}\right\}, & \text{si } \ T \in [a, b) \\ \frac{2C}{a+b}, & \text{si } \ T \geq b \end{cases}$$

donde $\tau = \frac{a+b}{2}$.

Control Óptimo

0.11. Introducción

El problema de control óptimo, en esencia, es aquel de cambiar dinámicamente los parámetros de un sistema en respuesta a su evolución para así optimizar su rendimiento. En el contexto del control óptimo, se suele utilizar las expresión *escoger una acción* para referirse a *cambiar los parámetros*. La regla que determina qué acción coger en respuesta a la evolución del sistema se denomina *política*. El objetivo es por tanto, encontrar la política óptima. Estudiaremos cómo elegir políticas óptimas en un tipo de particular de proceso denominado Procesos de Decisión de Markov (MDP).

En primer lugar, consideramos procesos que son observados en una serie discreta de tiempos $t = 0, 1, 2, \ldots$. El conjunto de posibles estados es contable y será etiquetado con enteros no negativos. Después de observar el estado del proceso, ha de tomarse una acción de un conjunto A de posibles acciones. Si el proceso se encuentra en el estado i a tiempo t y se escoge la acción a, ocurren dos cosas

- Incurrimos en un coste $C(i, a)$.
- El próximo estado del sistema es escogido de acuerdo con las probabilidades de transicion $P_{i,j}(a)$.

Si X_t es el estado del sistema a tiempo t y a_t es la acción tomada a tiempo t, entonces la última hipótesis es equivalente a

$$P\{X_{t+1} = j | X_0, a_0, X_1, a_1, \ldots, X_t = i, a_t = a\} = P_{i,j}(a)$$

Por tanto, ambos los costes y las probabilidades de transición son funciones únicamente del último estado del sistema y la acción tomada en el mismo. Además, asumiremos que los costes están acotados $|C(i, a)| < M$ para todo i, a.

Para escoger acciones, debemos seguir una política. Si no restringimos la clase de políticas, la política puede, por ejemplo, depender de toda la historia pasada o incluso estar aleatorizada. Una subclase importante de políticas es la clase de políticas estacionarias, aquellas no aleatorizadas en las que la acción escogida a tiempo t solo depende del estado del proceso a tiempo t. Es decir, una política estacionaria es una función que mapea estados en acciones. Es fácil ver

que bajo una política estacionaria, la secuencia de estados $\{X_t, t = 0, 1, 2, \dots\}$ forma una cadena de markov con probabilidades de transición $P_{ij} = P_{ij}[f(i)]$. Es por esta razón que este proceso se denomina Proceso de Decisión de Markov (MDP).

En este capítulo intentaremos desarrollar teoría que nos permita encontrar políticas óptimas en algún sentido. Y para esto, lo primero que necesitamos es definir un criterio de optimalidad. Estudiaremos tres criterios de optimalidad: Coste descontado esperado, coste total esperado y coste medio esperado por unidad de tiempo en el largo plazo.

0.12. Coste descontado esperado

Este criterio asume un factor de descuento $\alpha \in (0, 1)$ e intenta minimizar el coste descontado esperado.

Para cualquier política π, definamos el valor del estado i como

$$V_\pi(i) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} \alpha^t C(X_t, a_t) \Big| X_0 = i \right] \tag{7}$$

es decir, $V_\pi(i)$ representa el valor esperado del coste total descontado incurrido cuando se sigue la política π y el estado inicial es i. (Nótese que (7) está bien definido pues los costes están acotados y $\alpha < 1$.)

Sea

$$V_\alpha(i) = \inf_\pi V_\pi(i),$$

entonces una política π^* es α-óptima si para todo $i \geq 0$

$$V_{\pi^*} = V_\alpha(i)$$

es decir, si el valor esperado de su coste descontado con factor descuento α es mínimo para todo estado inicial.

Teorema 3. *$V_\alpha(i)$ verifica la siguiente ecuación*

$$V_\alpha(i) = \min_a \left\{ C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) \right\} \tag{8}$$

Demostración. Sea π una política arbitraria y supongamos que π escoge la acción a a tiempo 0 con probabilidad P_a. Entonces

$$V_\pi(i) = \sum_{a \in A} P_a \left[C(i, a) + \sum_{j=0}^{\infty} P_{ij}(a) W_\pi(j) \right]$$

donde $W_\pi(j)$ representa el coste descontado esperado incurrido desde el tiempo 1 en adelante, dado que se utiliza la política π y que el estado a tiempo 1 es j. Es claro que $W_\pi(j) \geq \alpha V_\alpha(j)$ y por tanto

$$\begin{aligned} V_\pi(i) &\geq \sum_{a\in A} P_a \left[C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) \right] \\ &\geq \sum_{a\in A} P_a \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) \right\} \\ &\geq \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) \right\} \end{aligned}$$

y como π es arbitraria se tiene que

$$V_\alpha(i) \geq \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) \right\} \tag{9}$$

Ahora, sea a_0 tal que

$$C(i,a_0) + \alpha \sum_{j=0}^{\infty} P_{ij}(a_0) V_\alpha(j) = \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) \right\}$$

y sea una política π que escoge a_0 a tiempo 0, y si el siguiente estado es j, entonces vea el proceso como si se originase en el estado j y siguiese la política π_j, que verifica $V_{\pi_j} \leq V_\alpha(j) + \epsilon$. Entonces

$$\begin{aligned} V_\pi(i) &= C(i,a_0) + \alpha \sum_{j=0}^{\infty} P_{ij}(a_0) V_{\pi_j}(j) \\ &\leq C(i,a_0) + \alpha \sum_{j=0}^{\infty} P_{ij}(a_0) V_\alpha(j) + \alpha\epsilon \end{aligned}$$

y como $V_\alpha(i) \leq V_\pi(i)$ se tiene

$$V_\alpha(i) \leq C(i,a_0) + \alpha \sum_{j=0}^{\infty} P_{ij}(a_0) V_\alpha(j) + \alpha\epsilon$$

con esto

$$V_\alpha(i) \leq \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) \right\} + \alpha\epsilon \tag{10}$$

Como ϵ es arbitrario (8) se sigue de (9) y (10). □

Ahora, consideremos una política estacionaria f que, cuando el proceso se encuentra en el estado i, escoge la acción $f(i)$. Sea $B(I)$ en conjunto de todas las funciones reales acotadas en el espacio de estados. Nótese que $V_\pi \in B(I)$. Para cualquier política estacionaria, definimos el mapeo f

$$T_f : B(I) \to B(I)$$

de la siguiente manera con esto

$$(T_f u)(i) = C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)]u(j)$$

Es fácil comprobar que $T_f u$ está acotado y por tanto pertenece a $B(I)$.
Introducimos la siguiente notación $T_f^1 = T_f$ y para $n > 1$ $T_f^n = T_f(T_f^{n-1})$.

Definición 1. *Para cualesquiera dos funciones $u, v \in B(I)$, diremos que $u \leq v$ si $u(i) \leq v(i)$ para todo i. Haremos lo mismo para $u = v$. Además, para $u, u_n \in B(I)$ diremos que $u_n \to u$ si $u_n(i) \to u(i)$ uniformemente para todo i.*

Lema 4. *Para $u, v \in B(I)$ y una política estacionaria f:*

1. $u \leq v \Rightarrow T_f u \leq T_f v$
2. $T_f V_f = V_f$
3. $T_f^n u \to V_f$ *para todo* $u \in B(I)$

Demostración. Las partes 1 y 2 son triviales. Para probar 3, primero notemos que

$$\begin{aligned}
(T_f^2 u)(i) &= \\
&= C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)](T_f u)(j) \\
&= C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)] \left[C[j, f(j)] + \alpha \sum_{k=0}^{\infty} P_{jk}[f(j)]u(k) \right] \\
&= C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)]C[j, f(j)] + \alpha^2 \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} P_{ij}[f(i)]P_{jk}[f(j)]u(k)
\end{aligned}$$

O en otras palabras, T_f^2 representa el coste esperado total si usamos la política f pero terminamos tras dos períodos e incurrimos un coste final $\alpha^2 u$. Por inducción, T_f^n representa el coste esperado percibido si usamos f durante n pasos e incurrimos un coste final $\alpha^n u$. Como $\alpha < 1$ y u está acotado, se cumple 3. □

Teorema 5. *Sea f_α una política estacionaria que, cuando el proceso está en el estado i, selecciona la acción (o una de las acciones) que minimiza la parte de la derecha de* (8)*, es decir*

$$C(i, f_\alpha(i)) + \alpha \sum_{j=0}^{\infty} P_{ij}(f_\alpha(i)) V_\alpha(j) = \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) \right\}$$

Entonces la política f_α es α-óptima.

Demostración. Aplicando T_{f_α} a V_α es inmediato ver que $(T_{f_\alpha} V_\alpha)(i) = V_\alpha(i)$ y por definición $T_{f_\alpha} V_\alpha = V_\alpha$ y por inducción $T^n_{f_\alpha} V_\alpha = V_\alpha$ para todo n. El resultado se sigue del apartado 3 del lema anterior. □

Por tanto, una política α-óptima existe, puede ser estacionaria, y está determinada por la ecuación (8). Por tanto, si podemos determinar el coste esperado óptimo V_α, entonces la política estacionaria que en el estado i selecciona la acción que minimiza la parte derecha de (8) es α-óptima.

Consideremos ahora la siguiente situación. Supongamos que hemos evaluado el coste esperado de una política estacionaria f, y construimos la política f^* que cuando el proceso está en el estado i, selecciona la acción que minimiza $C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_f(i)$. Entonces tenemos el siguiente corolario.

Corolario 6. *Para la anterior política f^* se tiene*

$$V_{f^*}(i) \leq V_f(i) \qquad \forall i \geq 0$$

Demostración. Es claro que

$$\begin{aligned} (T_{f^*} V_f)(i) &= C[i, f^*(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f^*(i)] V_f(j) \\ &\leq C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)] V_f(j) = V_f(i) \end{aligned}$$

Luego $T_{f^*} V_f \leq V_f$. Aplicando T_{f^*} a ambos lados de esta desigualdad y usando que T_{f^*} es monótona (Lemma 4 parte 1) se tiene

$$T^2_{f^*} V_f \leq T_{f^*} V_f \leq V_f$$

y por inducción $T^n_{f^*} V_f \leq V_f$. En el límite $n \to \infty$, usando Lemma 4 parte 3, se llega al resultado deseado. □

Esta técnica de comenzar con una política inicial e ir mejorándola usando el resultado anterior, se conoce como *policy improvement algorithm.*

La próxima cuestión que surge es, naturalmente, cómo determinar V_α. Antes de responder, necesitaremos algunos resultados preliminares.

Definición 2. *Un mapeo $T : B(I) \to B(I)$ se dice contractivo si*

$$\|Tu - Tv\| \leq \beta \|u - v\|$$

donde $\|u\| = \sup_{i \geq 0} |u(i)|$, $u, v \in B(I)$ y $\beta < 1$.

Teorema 7. *Teorema del punto fijo para mapeos contractivos. Si T es un mapeo contractivo entonces existe una única función $g \in B(I)$ tal que $Tg = g$. Además, para cualquier $u \in B(I)$,*

$$T^n u \to g \quad \text{cuando} \quad n \to \infty$$

Este teorema no lo demostraremos.
Para poder aplicarlo, definamos el mapeo T_α como

$$(T_\alpha u)(i) = \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) u(j) \right\} \tag{11}$$

Notemos que $T_\alpha V_\alpha = V_\alpha$. Por tanto, si probamos que este mapeo es contractivo, entonces habremos probado que V_α es la única solución a (8) y esta puede ser obtenida (en el límite) aplicando sucesivamente T_α a cualquier función inicial $u \in B(I)$. Este método se denomina *aproximaciones sucesivas*.

Teorema 8. *El mapeo T_α definido en* (11) *es contractivo.*

Demostración. Para todo $u, v \in B(I)$,

$$\begin{aligned} (T_\alpha u)(i) - (T_\alpha v)(i) &= \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) u(j) \right\} \\ &- C(i,a') - \alpha \sum_{j=0}^{\infty} P_{ij}(a') v(j) \end{aligned}$$

donde a' es tal que

$$C(i,a') + \alpha \sum_{j=0}^{\infty} P_{ij}(a') v(j) = \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) v(j) \right\}$$

por tanto

$$\begin{aligned} (T_\alpha u)(i) - (T_\alpha v)(i) &\leq \alpha \sum_{j=0}^{\infty} P_{ij}(a') u(j) - \alpha \sum_{j=0}^{\infty} P_{ij}(a') v(j) \\ &\leq \alpha \sum_{j=0}^{\infty} P_{ij}(a') \sup_j [u(j) - v(j)] \leq \alpha \|u - v\| \end{aligned}$$

Entonces tenemos

$$\sup_i \{(T_\alpha u)(i) - (T_\alpha v)(i)\} \leq \alpha \|u - v\|$$

Intercambiando los papeles de u y v llegaríamos a

$$\sup_i \{(T_\alpha v)(i) - (T_\alpha u)(i)\} \leq \alpha \|u - v\|$$

y por tanto

$$\sup_i \{|(T_\alpha u)(i) - (T_\alpha v)(i)|\} \leq \alpha \|u - v\|$$

Lo cual prueba que T_α es contractivo. □

Como consecuencia directa se tiene

Corolario 9. *V_α es la única solución a*

$$V_\alpha(i) = \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) \right\}$$

y para cualquier $u \in B(I)$, $T_\alpha^n u \to V_\alpha$ a medida que $n \to \infty$.

Nótese que este corolario nos permite demostrar que en la técnica de mejora de política, la nueva política o bien es estrictamente mejor que la anterior o ambas son óptimas. Esto se sigue de que si $V_{f^*} = V_f$, entonces

$$\begin{aligned} (T_{f^*} V_{f^*})(i) = (T_{f^*} V_f)(i) &= C[i, f^*(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f^*(i)] V_f(j) \\ &= \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_f(j) \right\} \end{aligned}$$

Usando el lemma 4 parte 2, tenemos que $T_{f^*} V_{f^*} = V_{f^*} = V_f$, luego V_f satisface la ecuación de optimalidad.

Es fácil probar que T_f es un mapeo contractivo y por tanto, como $T_f V_f = V_f$, V_f es la única solución a

$$V_f(i) = C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)] V_f(j)$$

0.13. Coste total esperado

En esta sección, asumimos que todos los costes son no negativos, $C(i,a) \geq 0$ para todo i, a. No asumiremos un factor de descuento y tampoco requeriremos que los costes esté acotados.

Para cualquier política π, sea

$$V_\pi(i) = \mathbb{E}_\pi \left[\sum_{t=0}^{\infty} C(X_t, a_t) \Big| X_0 = i \right]$$

y sea

$$V(i) = \inf_\pi V_\pi(i)$$

Es posible que $V(i)$ sea infinito. Este modelo es solamente de interés si la naturaleza del problema es tal que $V(i) < \infty$ para al menos algún valor de i. De otra manera, todas las políticas serían óptimas.

Una política π^* se denomina óptima si

$$V_{\pi^*}(i) = V(i)$$

De manera análoga a como hicimos en el Teorema 3 es posible demostrar

Teorema 10. *Se verifica*

$$V(i) = \min_a \left\{ C(i,a) + \sum_{j=0}^{\infty} P_{ij}(a) V(j) \right\} \tag{12}$$

Denotemos $N(I)$ el conjunto de todas las funciones no negativas definidas en el espacio de estados, y para cualquier política estacionaria f definimos el mapeo

$$T_f \ : \ N(I) \to N(I)$$

como

$$(T_f u)(i) = C[i, f(i)] + \sum_{j=0}^{\infty} P_{ij}[f(i)] u(j)$$

Entonces $T_f u$ es el coste esperado si usamos la política f pero terminamos tras un período e incurrimos un coste final $u(j)$ cuando el estado final es j. De manera análogo a como hicimos en el Lemma 4

Lema 11. *Para $u, v \in N(I)$ y una política estacionaria f:*

1. $u \leq v \Rightarrow T_f u \leq T_f v$
2. $T_f V_f = V_f$
3. $(T_f^n 0)(i) \to V_f(i)$ *cuando* $n \to \infty$ *para cada* i.

Nótese que ahora el tercer punto únicamente cierto para la función 0 *y la convergencia es puntual en lugar de uniforme. esto es así pues no asumimos un factor de descuento. Si existe un factor de descuento* α*, el coste final es* $\alpha^n u$*, que va a cero uniformemente para* $u \in B(i)$*. Sin descuento, la única forma de asgurar que el coste final vaya a cero es que sea cero.*

Teorema 12. *Sea* f_1 *una política estacionaria que, cuando el proceso se encuentra en el estado* i*, selecciona la acción que minimiza el lado derecho de* (12)*. Entonces* $V_{f_1}(i) = V(i)$ *para todo* $i \geq 0$ *y por tanto* f_1 *es óptima.*

Demostración. Si aplicamos T_{f_1} a V, obtenemos

$$\begin{aligned}(T_{f_1}V)(i) &= C[i, f_1(i)] + \sum_{j=0}^{\infty} P_{ij}[f_1(i)]V(j) \\ &= \min_a \left\{ C(i,a) + \sum_{j=0}^{\infty} P_{ij}(a)V(j) \right\} \\ &= V(i)\end{aligned}$$

Y por tanto $T_{f_1}V = V$. Ahora, $C(i,a) \geq 0$, que implica $V \geq 0$, y por tanto porla monotonicidad de T_{f_1}, se tiene que $T_{f1}0 \leq T_{f_1}V = V$. Aplicando sucesivamente T_{f_1} vemos que $T_{f1}^n 0 \leq V$. En el límite $n \to \infty$ vemos que $V_{f_1} \leq V$, y como $V \leq V_{f_1}$ por definición, vemos que f_1 es óptima. □

Por tanto, una política óptima existe y está determinada por la ecuación (12). No obstante, en este caso no tenemos ningún mapeo contractivo y por tanto no podemos probar que V sea la única solución a (12). Además, el método de aproximaciones sucesivas no está disponible.

0.14. Coste medio esperado por unidad de tiempo en el largo plazo

En esta sección, supondremos de nuevo que los costes están acotados, y trataremos de minimizar el coste medio esperado por unidad de tiempo en el largo plazo.

Para cualquier política π, definimos

$$\phi_\pi(i) = \lim_{n\to\infty} \mathbb{E}_\pi \frac{[\sum_{t=0}^n C(X_t, a_t)|X_0 = i]}{n+1}$$

donde si el límite no existe, acordaremos usar el lim sum. Por tanto, $\phi_\pi(i)$ representa el coste medio esperado por unidad de tiempo en el largo plazo, cuando se emplea la política π y el estado inicial es i. Diremos que la política π^* es óptima bajo el criterio del coste medio si se cumple

$$\phi_{\pi^*}(i) = inf_\pi \phi_\pi(i) \qquad \forall i$$

La primera cuestión que debemos considerar es si existe una política óptima. Responderemos esta cuestión utilizando un contraejemplo.

Contraejemplo 1: Sea el proceso de decisión de Markov sobre el espacio de estados $\{1, 1', 2, 2', 3, 3', \dots\}$, con dos posibles acciones y probabilidades de transición

$$\begin{aligned} P_{ii+1}(1) &= P_{i,i'}(2) = 1 \\ P_{i'i'}(1) &= P_{i'i'}(2) = 1 \end{aligned}$$

Los costes dependen únicamente del estado y vienen dados

$$\begin{aligned} C(i,\cdot) &= 1 \\ C(i',\cdot) &= 1/i \end{aligned}$$

En otras palabras, cuando el proceso se encuentra en el estado 1, se puede pagar una unidad e ir al estado $i+1$ o ir al estado i' y como resultado pagar $1/i$ todos los días a partir de ese.

Supongamos que $X_0 = 1$, y sea π una política cualquiera. Entonces existen dos casos posibles.

Caso 1: Con probabilidad 1, π siempre escoge la acción 1, en cuyo caso se tiene $\phi_\pi(1) = 1$

Caso 2: Con probabilidad positiva, π escoge la acción 2 en algún momento. En este caso se tiene que para algún n existe una probabilidad $P_n > 0$ de que π escoga la acción 2 en el estado n. Si denotamos $\bar{n}$ al n más pequeño con esto propiedad, se sigue que

$$\phi_\pi(1) \geq \frac{P_{\bar{n}}}{\bar{n}} > 0$$

Entonces, en cualquier caso $\phi_\pi(1) > 0$. No obstante, escogiendo la acción 1 suficiente tiempo y después escogiendo la acción 2, podemos hacer nuestro coste tan cercano a 0 como queramos. Por tanto, no existe una política óptima.

La segunda cuestión que cabe plantearse es si debemos restringir nuestra atención a políticas estacionarias. Por ejemplo, en el caso visto, aunque no existe una política óptima, se tiene que para cualquier política existe una política estacionaria que lo hace al menos tan bien como esa. Respondemos esta cuestión con el siguiente contraejemplo.

Contraejemplo 2:

Sea el proceso de decisión de Markov con espacio de estados 1,2,3,... y con dos acciones. Los costes y probailidades de transición son

$$\begin{aligned} P_{ii+1}(1) &= 1 = P_{ii+1}(2) \\ C(i,1) &= 1 \\ C(i,2) &= 1/i \end{aligned}$$

En otras palabras, cuando el proceso se haya en el estado i puede pagarse una unidad e ir al estado $i+1$ o pagar $1/i$ y quedarse en el estado i. Supongamos que $X_0 = 1$ y sea π cualquier política estacionaria. Se pueden dar dos casos.

Caso 1: π siempre escoge la acción 2, en cuyo caso se tiene que $\phi_\pi(1) = 1$.

Caso 2: π escoge la acción 2 por primera vez en el estado n. En este caso, el proceso irá del estado 1 al 2 al 3 hasta el n. No obstante, cuando el proceso se encuentra en n, se escoge la acción 2 y por tanto, el proceso nunca deja este estado. Por tanto $\phi_\pi(1) = 1/n > 0$.

Por tanto, para cualquier política estacionaria, $\phi_\pi(1) > 0$. Ahora sea π^* una política no estacionaria que, cuando el proceso entra en el estado i, escoge la acción 2 i veces consecutivas, y después escoge la acción 1. Como el estado inicial es $X_0 = 1$, se sigue que los costes sucesivos incurridos bajo esta política son

$$1, 1, 1/2, 1/2, 1, 1/3, 1/3, 1/3, 1, 1/4, 1/4, 1/4, 1/4, 1, 1/5 \ldots$$

El coste medio de esta secuencia es 0 y por tanto $\phi_{\pi^*}(1) = 0$. Por tanto, la política no estacionaria es mejor que cualquier política estacionaria.

Por último, veamos que en la definición de política estacionaria, no hemos permitido la posibilidad de aleatorizar. Definamos un política estacionaria aleatoria como aquella en la que las acciones escogidas, aunque solo dependen del estado actual, pueden ser aleatorias. ¿Podemos restingir nuestra atención a esta clase de políticas? Reconsideremos el contraejemplo 2. Supongamos que π es una política estacionaria aleatoria, que cuando el proceso se encuentra en el estado i, selecciona la acción 2 con porbabilida $i/i+1$ y la acción 1 con probabilidad $1/i+1$. Entonces, si empleamos esta polítia, cuando el proceso entra por primera vez en el estado i el número esperado de veces que la acción 2 es escogida consecutivamente es i. Entonces, parece razonable que esta política actuará de manera similar a π^*. De hecho puede probarse que su coste esperado es 0.

Desafortunadamanete, en general no podemos restringir nuestra atención a políticas estacionarias aleatorias. Existen ejemplos en los que una política óptima no estacionaria es mejor que cualquier política estacionaria aleatoria. A continuación, intentaremos determinar las condicionas bajo las cuales existen políticas estacionarias óptimas.

Empezaremos con un teorema.

Teorema 13. *Si existe una función acotada $h(i)$, $i = 0, 1, 2, \ldots$ y una constante g tal que*

$$g + h(i) = \min_a \left\{ C(i,a) + \sum_{j=0}^{\infty} P_{ij}(a) h(j) \right\} \tag{13}$$

entonces existe una política estacionaria π^* *tal que, para todo* i

$$g = \phi_{\pi^*}(i) = \min_\pi \phi_\pi(i)$$

y π^* *es cualquier política que para cada* i*, prescribe una acción que minimiza la parte derecha de* (13)*.*

Demostración. Sea $H_t = (X_0, a_0, \ldots, X_t, a_t)$ la historia del proceso hasta tiempo t. Para cualquier política π, usando la linealidad de la esperanza y la ley de esperanza total se tiene

$$\mathbb{E}_\pi \left\{ \sum_{t=1}^{n} [h(X_t) - \mathbb{E}_\pi(h(X_t)|H_{t-1})] \right\} = 0$$

Pero,

$$\begin{aligned}
\mathbb{E}_\pi(h(X_t)|H_{t-1}) &= \sum_{j=0}^{\infty} h(j) P_{X_{t-1}j}(a_{t-1}) \\
&= C(X_{t-1}, a_{t-1}) + \sum_{j=0}^{\infty} h(j) P_{X_{t-1}j}(a_{t-1}) - C(X_{t-1}, a_{t-1}) \\
&\geq \min_a \left\{ C(X_{t-1}, a) + \sum_{j=0}^{\infty} h(j) P_{X_{t-1}j}(a) \right\} - C(X_{t-1}, a_{t-1}) \\
&= g + h(X_{t-1}) - C(X_{t-1}, a_{t-1})
\end{aligned}$$

donde se tiene igualdad para la política π^* pues por definición π^* escoge la acción que minimiza. Por tanto

$$0 \leq \mathbb{E}_\pi \left\{ \sum_{t=1}^{n} [h(X_t) - g - h(X_{t-1}) + C(X_{t-1}, a_{t-1})] \right\}$$

o lo que es lo mismo

$$g \leq \mathbb{E}_\pi \frac{h(X_n)}{n} - \mathbb{E}_\pi \frac{h(X_0)}{n} + \mathbb{E}_\pi \frac{\sum_{t=1}^{n} C(X_{t-1}, a_{t-1})}{n}$$

Con la igualdad dándose para π^*. Tomando el límite $n \to \infty$ y usando el hecho de que h está acotada, tenemos que

$$g \leq \phi_\pi(X_0)$$

donde la igualdad se da para π y para todos los posibles valores de X_0. □

Por tanto, si se satisfacen las condiciones del Teorema 13, entonces existe una política estacionaria óptima, y puede ser caracterizada mediante la ecuación funcional (13). No obstante, el Teorema 13 no es para nada intuitivo. Y tampoco está claro cuando se satisfacen las condiciones que impone. Tratemos de arrojar algo de luz sobre el asunto.

Para el criterio del coste descontado, los estados futuros se descuentan con tasa α, mientras que para el criterio del coste medio por unidad de tiempo, todos los períodos reciben el mismo peso. Por tanto, parece razonable que bajo ciertas condiciones, el caso del coste medio por unidad de tiempo sea en algún sentido el límite del caso descontado cuando $\alpha \to 1$. Hemos visto que la función de coste $V_\alpha(i)$ α-óptima satisface

$$V_\alpha(i) = \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) V_\alpha(j) \right\}$$

y la política α-óptima selecciona las acciones que minimizan la parte derecha. Una forma posible de obtener una política óptima bajo el criterio del coste esperado, podría ser escoger la acción que minimice el límite de la expresión anterior cuando $\alpha \to 1$. No obstante, este límite no existe necesariamente y frecuentemente es infinito para todas las acciones.

Consideremos otro enfoque al problema. Fijemos un estado, el estado 0 por ejemplo, y definamos $h_\alpha(i) = V_\alpha(i) - V_\alpha(0)$, el α-coste del estado i relativo al estado 0. Se tiene que

$$(1-\alpha)V_\alpha(0) + h_\alpha(i) = \min \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) h_\alpha(j) \right\} \tag{14}$$

Además, la política que selecciona la acción que minimiza la parte derecha de esta ecuación es una política α-óptima. Ahora, si para alguna secuencia $\alpha_n \to 1$, $h_{\alpha_n}(j)$ converge a una función $h(j)$ y $(1-\alpha_n)V_{\alpha_n}(0)$ converge a una constante g entonces

$$g + h(i) = \min \left\{ C(i,a) + \sum_{j=0}^{\infty} P_{ij}(a) h(j) \right\}$$

donde hemos asumido que cambiar el sumatorio y el límite está justificado. Además, parece razonable que la política que toma la acción que minimiza la parte derecha de esta ecuación sea óptima bajo el criterio del coste medio y por tanto, el Teorema 13 sea cierto.

Ahora probamos formalmente esto.

Teorema 14. *Si existe un $N < \infty$ tal que para todo α y todo i ($h_\alpha(i)$ está acotado uniformemente)*

$$|V_\alpha(i) - V_\alpha(0)| < N$$

Entonces

1. *Existe una función acotada $h(i)$ y una constante g que satisfacen* (13).
2. *Para alguna secuencia $\alpha_n \to 1$, $h(i) = \lim_{n\to\infty} V_{\alpha_n}(i) - V_{\alpha_n}(0)$.*
3. $\lim_{\alpha\to 1}(1-\alpha)V_\alpha(0) = g$.

Demostración. Por hipótesis, $h_\alpha(i) = V_\alpha(i) - V_\alpha(0)$ está uniformemente acotada. Por tanto, dado que el espacio de estados es contable, podemos por medio del método de diagonalización de Cauchy, construír para cada i, una secuencia $\alpha_n \to 1$ tal que $\lim_{n\to\infty} h_{\alpha_n}(i) \equiv h(i)$ exista. Además, como los costes están acotados tenemos que $(1-\alpha_n)V_{\alpha_n}(0)$ está acotado y por tanto, existe una subsecuencia $\{\alpha'_n\}$ de $\{\alpha_n\}$ tal que $\lim_{n\to\infty}(1-\alpha'_n)V_{\alpha'_n}(0) \equiv g$ existe. (Pues toda secuencia acotada tiene una subsecuencia acotada convergente, Bolzano-Weierstass). Ahora, de (14) se sigue que (usamos implícitamente que una subsecuencia de una secuencia convergente converge al mismo límite).

$$(1-\alpha'_n)V_\alpha(0) + h_{\alpha'_n}(i) = \text{mín}\left\{C(i,a) + \alpha'_n \sum_{j=0}^{\infty} P_{ij}(a)h'_{\alpha n}(j)\right\}$$

Por tanto, el resultado 1. se sigue tomando el límite $n \to \infty$, y notando que el hecho que $h_{\alpha'_n}(i)$ esté acotada implica, por el teorema de convergencia acotada de Lebesgue que (versión discreta del teorema de convergencia dominada)

$$\sum_{j=0}^{\infty} P_{ij}(a)h_{'_n}(j) \to \sum_{j=0}^{\infty} P_{ij}(a)h(j)$$

Por último, como $(1-\alpha_n)V_{\alpha_n}(0)$ es una secuencia acotada, para una secuencia $\alpha_n \to 1$, existe una subsecuencia α'_n convergente. Y como el límite de esta subsecuencia es el mismo que el de la secuencia original, tenemos que $\lim_{n\to\infty} \alpha'_n = 1$. Entonces

$$\lim_{n\to\infty}(1-\alpha'_n)V_{\alpha'_n}(0) = \lim_{\alpha\to 1}(1-\alpha_n)V_\alpha(0) \equiv g$$

□

El siguiente teorema da una condición suficiente para que $V_\alpha(i) - V_\alpha(0)$ esté uniformemente acotado.

Teorema 15. *Si para algún estado, digamos el estado 0, existe una constante $N < \infty$ tal que para todo α y todo i*

$$M_{i0}(f_\alpha) < N$$

entonces $V_\alpha(i) - V_\alpha(0)$ está uniformemente acotado, donde $M_{i0}(f_\alpha)$ es el tiempo medio de recurrencia empleado en ir del estado i al estado 0 cuando se utiliza la política α-óptima f_α.

Demostración. Notemos primero que, sin perdida de generalidad, podemos suponer que los costes son estrictamente positivos. Esto es así pues los costes están acotados, y añadir una constante suficientemente grande a todos los costes $C(i,a)$ afectará a todas las reglas por igual. Sea

$$T = \text{mín}\{t : X_t = 0\}$$

Entonces, podemos escribir

$$V_\alpha(i) = \mathbb{E}_{f_\alpha} \sum_{n=0}^{T-1} C(X_n, a_n)\alpha^n + \mathbb{E}_{f_\alpha} \sum_{n=T}^{\infty} C(X_n, a_n)\alpha^n$$

donde condicionamos impícitamente las esperanzas a $X_0 = i$. Si la cota de los costes es M tenemos

$$V_\alpha(i) \leq M\mathbb{E}_{f_\alpha} \sum_{n=0}^{T-1} 1 + \mathbb{E}_{f_\alpha}\alpha^T \sum_{k=0}^{\infty} C(X_{k+T}, a_{k+T})\alpha^k \leq MN + V_\alpha(0)$$

Donde hemos usado que

$$\mathbb{E}_{f_\alpha}\left[\alpha^T \cdot \sum_{k=0}^{\infty} \alpha^k C(X_{k+T}, a_{k+T})\right] = \mathbb{E}_{f_\alpha}\left[\alpha^T\right] \cdot \mathbb{E}_{f_\alpha}\left[\sum_{k=0}^{\infty} \alpha^k C(X_{k+T}, a_{k+T})\right]$$

pues los factores son variables aleatorias independientes. Para obtener la igualdad en el otro sentido, notemos que

$$V_\alpha(i) \geq V_\alpha(0)\mathbb{E}_{f_\alpha}[\alpha^T]$$

O lo que es lo mismo

$$V_\alpha(0) \leq V_\alpha(i) + (1 - \mathbb{E}_{f_\alpha}[\alpha^T])V_\alpha(0)$$

Ahora bien, $V_\alpha(0) \leq M\sum_{n=0}^{\infty}\alpha^n = M/(1-\alpha)$. Además, usando la desigualdad de Jensen, que dice que si $g(X)$ es convexa y X una variable aleatoria entonces $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$; como α^T es convexa, $\mathbb{E}\alpha^T \geq \alpha^{\mathbb{E}T} \geq \alpha^N$. Por tanto

$$V_\alpha(0) \leq V_\alpha(i) + (1-\alpha^N)\frac{M}{1-\alpha} < V_\alpha(i) + MN$$

Donde hemos usado implícitamente que $(1-\alpha^N)/(1-\alpha) < N$, para $N > 1$. □

Como corolario tenemos

Corolario 16. *Si el espacio de estados es finito y existe un estado, digamos el 0, que es accesible desde cualquier otro para cualquier política α-óptima, entonces $V_\alpha(i) - V_\alpha(0)$ está uniformemente acotado y por tanto existe una política optima.*

Demostración. Como una cadena de Markov finita no puede tener estados recurrentes nulos, entonces, dada la política óptima f, $M_{i0}(f) < \infty$ para todo i. □

0.15. Espacio de estados finito. Estrategia computacional

En esta sección suponemos que el espacio de estados y el espacio de acciones son finitos. Denotaremos los estados como $0, 1, 2, \ldots, m$.

Coste descontado

Primero consideremos la técnica *policy improvement* para el caso con descuento. Para cualquier política estacionaria f, hemos visto que V_f es la única solución a

$$V_f(i) = C[i, f(i)] + \alpha \sum_{j=0}^{\infty} P_{ij}[f(i)]V_f(j)$$

para $i = 0, 1, 2, \ldots, m$. Esto es así porque T_f es un mapeo contractivo y $T_f V_f = V_f$. Tenemos pues $m+1$ ecuaciones con $m+1$ incógnitas, y por tanto $V_f(i)$ puede ser determinado utilizando métodos estándar. Una vez calculado, podemos mejorar la política f escogiendo nuestras acciones de tal manera que minimicen

$$C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a)V_f(j) \tag{15}$$

Acordemos sólo cambiar nuestra acción presente $f(i)$ si la nueva acción lleva a una mejora estricta (15). Entonces, del Corolario 9 se sigue

1. Si la política mejorada es igual a la original, entonces la original f es α-óptima.
2. Si la política mejorada no es igual a la original f, entonces la mejorada es estríctamente mejor que f para al menos un estado incial.

Una vez determinemos la política mejorada, su función de coste asociada puede ser calculada y esta política puede de nuevo mejorarse. Como hay un número finito de políticas estacionarias, esta técnica llevará eventualmente a la política α-óptima.

Otra estrategia computacional para el problema descontado se sigue del siguiente lema.

Lema 17. *Sea T_α definido como*

$$(T_\alpha u)(i) = \min_a \left\{ C(i, a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a)u(j) \right\}$$

entonces, para cualquier función u se tiene

$$T_\alpha u \geq u \Rightarrow V_\alpha \geq u$$

Demostración. Si $T_\alpha u \geq u$ entonces, por la monotonicidad de T_α se sigue que $T_\alpha^n u \geq u$ y el resultado se obtiene en el límite $n \to \infty$. □

Como $T_\alpha V_\alpha = V_\alpha$, se sigue que V_α puede ser obtenido buscando la máxima función u de aquellas que verifican $T_\alpha u \geq u$. No obstante, maximizando $u(i)$ para todo i también maximizamos $\sum_{i=0}^m u(i)$, y el problema se reduce a

$$\begin{aligned} &\text{Maximizar} \quad \sum_{i=0}^{m} u(i) \\ &\text{sujeto a} \quad \min_a \left\{ C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) u(j) \right\} \geq u(i) \qquad i = 0, 1, \ldots, m \end{aligned}$$

o lo que es lo mismo

$$\begin{aligned} &\text{Maximizar} \quad \sum_{i=0}^{m} u(i) \\ &\text{sujeto a} \quad C(i,a) + \alpha \sum_{j=0}^{\infty} P_{ij}(a) u(j) \geq u(i) \qquad i = 0, 1, \ldots, m \end{aligned}$$

Este último problema es un programa lineal y puede ser resuelto usando técnicas estándar de programación lineal.

Coste medio

Ahora consideremos el caso del coste medio por unidad de tiempo en el largo plazo. Asumimos por sencillez, que todas las políticas estacionarias dan lugar a una cadena de Markov irreducible.

En lugar de considerar únicamente políticas estacionarias, consideremos también políticas aleatorizadas, y para una política dada de esta clase, denotemos P_i^a la probabilidad de tomar la acción a cuando el estado es i. Denominemos a esta política π. Se tiene el siguiente resultado

Teorema 18. *Bajo la política π, $\{X_n; n \geq 0\}$ es una DTMC con matriz de probabilidades de transición $[P_{ij}^\pi]$ dada por*

$$P_{ij}^\pi = P(X_{n+1} = j | X_n = i) = \sum_a P_i^a P_{ij}[a]$$

Demostración.

$$P(X_{n+1} = j | X_n = i) = \sum_a P(X_{n+1} = j | X_n = i, a) P(a | X_n = i) = \sum_a P_i^a P_{ij}[a]$$

□

El objetivo ahora es calcular el coste esperado por unidad de tiempo a largo plazo, definido como

$$\phi_\pi(i) = \lim_{n\to\infty} \mathbb{E}_\pi \frac{[\sum_{t=0}^{n} C(X_t, a_t)|X_0 = i]}{n+1}$$

Para ello, hacemos uso del siguiente resultado

Teorema 19. *Sea $\{X_n; n \geq 0\}$ una DTMC irreducible con distribución límite z_i, $i = 0, 1, \ldots, m$. Entonces*

$$\phi_\pi = \sum_i z_i \sum_a P_i^a C(i, a)$$

Demostración. Cuando la DTMC se encuentra en el estado i, la acción a es escogida con probabilidad P_i^a y eso resulta en un coste incurrido $C(i, a)$. Por tanto, el coste esperado incurrido en cada visita al estado i es $C(i) = \sum_a P_i^a C(i, a)$. Como vimos en la Sección 0.7, el coste medio a largo plazo por unidad de tiempo es

$$\phi_\pi = \sum_j z_j C(j)$$

Lo que prueba el teorema. □

Por tanto, para determinar la política óptima, habría que resolver el siguiente problema de optimización

Minimizar	ϕ_π
sujeto a	π es una política que escoge a en el estado i con probabilidad P_i^a

primero definimos $z_i^a = z_i P_i^a$. z_i^a representa la proporción de tiempo a largo plazo durante la cual el sistema está en el estado i y se elige la acción a. A partir de z_i^a se puede recuperar z_i y P_i^a utilizando

$$\begin{aligned} \sum_a z_i^a &= z_i \\ P_i^a &= \frac{z_i^a}{z_i} = \frac{z_i^a}{\sum_a z_i^a} \end{aligned}$$

Además, la función objetivo puede ser expresada en términos de z_i^a como

$$\phi_\pi = \sum_i \sum_a z_i^a C(i, a)$$

De este modo, la función objetivo es una función lineal de las variables z_i^a. También, sabemos que z_i verifica

$$z_j = \sum_i z_i P_{ij}^{\pi} \qquad \sum_j z_j = 1$$

lo que en términos de z_i^a se escribe como

$$\begin{aligned} z_j &= \sum_i z_i P_{ij}^{\pi} = \sum_i z_i \sum_a P_{ij}(a) P_i^a = \sum_i \sum_a z_i^a P_{ij}(a) \\ \sum_j \sum_a z_j^a &= 1 \end{aligned}$$

Por tanto el problema de optimizacion queda reducido a

$$\begin{aligned} \text{Minimizar} \quad & \sum_i \sum_a z_i^a C(i,a) \\ \text{sujeto a} \quad & \sum_a z_j^a = \sum_i \sum_a z_i^a P_{ij}(a) \qquad \forall j \\ & \sum_j \sum_a z_j^a = 1 \qquad \forall j \\ & z_j^a \geq 0 \qquad \forall j, a \end{aligned}$$

Este problema de programación lineal puede ser resuelto utilizando un paquete numérico. Sean $\widehat{z}_i^a$ los valorpes óptimos obtenidos. Veamos cómo obtener la política óptima. Definamos $T = \{i | \sum_a \widehat{z}_i^a > 0\}$. Existen dos posibles casos

Caso 1. T es todo el espacio de estados. Entonces es claro que

$$P_i^a = \frac{\widehat{z}_i^a}{\sum_b \widehat{z}_i^b}$$

La teoría de programación lineal muestra que para un estado dado i existe una única acción $f(i)$ para la cual $\widehat{z}_i^a > 0$. Siendo esta cantidad 0 para el resto de acciones. Por tanto, vemos que la política resultante es estacionaria, tal y cómo adelantaba el Corolario 16.

Caso 2. T no es todo el espacio de estados. Entonces $\sum_a \widehat{z}_i^a = 0$ para $i \notin T$. Para los $i \in T$ la política óptima se calcula como en el caso 1. Y para los $i \notin T$ el análisis es más complejo, pues la política óptima no está bien definida.

0.16. Ejercicios

Ejercicio 1. Considérese una máquina que puede estar en cualquiera de los estados $0, 1, 2, \dots$. Supongamos que al comienzo de cada día, se observa el estado de la máquina y se toma una decisión acerca de si reemplazar la máquina o no. Si la máquina se reemplaza, la nueva máquina comienza en el estado 0.

El coste de reemplazar una máquina es R y el coste de mantenimiento incurrido cada día es $C(i)$ cuando la máquina se encuentra en el estado i. Además, P_{ij} representa la probabilidad de que la máquina, estando en el estado i al comienzo de un día, esté en el estado j al comienzo del siguiente.

Este es un modelo de decisión de Markov con dos posibles acciones, acción 1: sustituir la máquina; acción 2: no hacerlo. Los costes y las probabilidades de transición a un paso vienen dados por:

$$\begin{aligned} C(i,1) = R + C(0) \quad C(i,2) = C(i) \quad i \geq 0 \\ P_{ij}(1) = P_{0j} \quad P_{ij}(2) = P_{ij} \quad i \geq 0 \end{aligned}$$

Además, imponemos las siguientes hipótesis

1. $\{C(i), i \geq 0\}$ está acotada y es creciente en i.
2. $\sum_{j=k}^{\infty} P_{ij}$ es una función creciente de i, para cada $k \geq 0$.

Es decir, 1. dice que el coste de mantenimiento es una función creciente del estado; y 2. dice que la probabilidad de transición a cualquier bloque $\{k, k+1, \dots\}$ es una función creciente del estado actual.

Para determinar las estructura de la política óptima, necesitaremos los siguientes lemas.

Lema 20. *La hipótesis 2 implica que para cualquier función creciente $h(i)$, la función $\sum_{j=0}^{\infty} P_{ij} h(j)$ también es creciente en i.*

Demostración. Definiendo $h(-1) := 0$ tenemos

$$\begin{aligned} \sum_{j=0}^{\infty} P_{ij} h(j) &= \sum_{j=0}^{\infty} P_{ij} \left[\sum_{k=0}^{j} h(j) - h(j-1) \right] \\ &= \sum_{k=0}^{\infty} [h(j) - h(j-1)] \sum_{j=k}^{\infty} P_{ij} \\ &= h(0) \sum_{j=0}^{\infty} P_{ij} + \sum_{k=1}^{\infty} [h(j) - h(j-1)] \sum_{j=k}^{\infty} P_{ij} \\ &\geq h(0) \sum_{j=0}^{\infty} P_{mj} + \sum_{k=1}^{\infty} [h(j) - h(j-1)] \sum_{j=k}^{\infty} P_{mj} = \sum_{j=0}^{\infty} P_{mj} h(j) \end{aligned}$$

donde implícitamente hacemos uso de que $h(j) - h(j-1) \geq 0$. □

Lema 21. *Bajo las hipótesis 1. y 2., $V_\alpha(i)$ es creciente en i.*

Demostración. Definamos $V_\alpha(i,n) = (T_\alpha^n 0)(i)$. En nuestro caso tenemos

$$V_\alpha(i,1) = \text{mín}\left\{R + C(0);\ C(i)\right\}$$

y para $n > 1$, usando que $V_\alpha(i,n) = [T_\alpha V_\alpha(i,n-1)](i)$

$$V_\alpha(i,n) = \text{mín}\left\{ R + C(0) + \alpha \sum_{j=0}^{\infty} P_{0j} V_\alpha(j,n-1);\ C(i) + \alpha \sum_{j=0}^{\infty} P_{ij} V_\alpha(j,n-1) \right\}$$

De la hipótesis 1., se sigue que $V_\alpha(i,1)$ es creciente en i, y si asumimos que $V_\alpha(i,n-1)$ es creciente en i, entonces $V_\alpha(i,n)$ también es creciente en i, por el lema anterior. Por inducción $V_\alpha(i,n)$ es creciente en i para todo n y por tanto $V_\alpha(i) = \lim_{n\to\infty} V_\alpha(i,n)$ también es creciente en i. □

La estructura de la polítca óptima viene dada por el siguiente teorema.

Teorema 22. *Bajo las hipótesis 1. y 2., existe un entero $i^* \leq \infty$, tal que una política α-óptima sustituye la máquina para todo $i > i^*$ y no la sustituye para $i \leq i^*$.*

Demostración. Por el Teorema 3, tenemos que

$$V_\alpha(i) = \text{mín}\left\{ R + C(0) + \alpha \sum_{j=0}^{\infty} P_{0j} V_\alpha(j);\ C(i) + \alpha \sum_{j=0}^{\infty} P_{ij} V_\alpha(j) \right\}$$

Sea

$$i^* = \text{máx}\left\{ i :\ C(i) + \alpha \sum_{j=0}^{\infty} P_{ij} V_\alpha(j) \leq R + C(0) + \alpha \sum_{j=0}^{\infty} P_{0j} V_\alpha(j) \right\}$$

Por los lemas anteriores, tenemos que $C(i) + \alpha \sum_{j=0}^{\infty} P_{ij} V_\alpha(j)$ es creciente en i y por tanto

$$V_\alpha(i) = \begin{cases} C(i) + \alpha \sum_{j=0}^{\infty} P_{ij} V_\alpha(j), & \text{si } \ i \leq i^* \\ R + C(0) + \alpha \sum_{j=0}^{\infty} P_{0j} V_\alpha(j), & \text{si } \ i > i^* \end{cases}$$

El resultado se sigue del Teorema 5. □

Ejercicio 2. Consideremos una persona que quiere vender su casa. Al principio de cada día alguien le hace una oferta, y la persona tiene que decidir inmediatamente si aceptarla o no. Una vez rechazada, la oferta se pierde. Las

sucesivas ofertas son independientes y toman un valor i con probabilidad P_i para $i = 0, 1, 2, \ldots$. Además, sea C el coste incurrido cada día que la casa no se vende. Los costes futuros son descontados con un tasa α.

Si dejamos que el estado a tiempo t sea la oferta a tiempo t, entonces esto es un proceso de decisión de Markov con dos posibles acciones. (Si la oferta se acepta, suponemos que el proceso va al estado ∞, del cual no puede regresar y en el cual los costes futuros son 0).

Si el estado es i, y la oferta se acepta, entonces el coste a un paso es $-i$ y si esta se rechaza, el coste a un paso es C. Por tanto

$$V_\alpha(i) = \text{mín} \left\{ -i;\ C + \alpha \sum_{j=0}^{N} P_j V_\alpha(j) \right\}$$

si definimos i^* como

$$i^* = \text{mín} \left\{ i : -i < C + \alpha \sum_{j=0}^{N} P_j V_\alpha(j) \right\}$$

Entonces la política α-óptima acepta cualquier oferta mayor o igual que i^* y rechaza las menores. Por tanto la estructura de la política óptima queda determinada.

Ejercicio 3. Problemas de parada óptima. Considérese un proceso con estados $0, 1, 2, \ldots$, que cuando se encuentra en el estado i, podemos parar (acción 1) y recibir una compensación $R(i)$, o bien (acción 2) pagar un coste $C(i)$ y hacer una transición al siguiente estado gobernada por las probabilidades P_{ij}.

Si decimos que el proceso va al estado ∞ cuando se toma la decisión de parar, entonces este es un proceso de decisión de Markov con dos posibles acciones, donde se tiene

$$\begin{aligned} C(i,1) &= -R(i) \\ C(i,2) &= C(i) \\ C(\infty,\cdot) &= 0 \\ P_{i\infty}(1) &= 1 \\ P_{ij}(2) &= P_{ij} \\ P_{\infty\infty}(\cdot) &= 1 \end{aligned}$$

Haremos las siguiente hipótesis

1. $\text{ínf}_{i \geq 0}\, C(i) > 0$
2. $\sup_i R(i) < \infty$

No podemos aplicar inmediatamente los resultados de la Sección 0.13 a este proceso, pues no se da el caso de que todos los costes sean no negativos. No obstante, podemos transformar este proceso en otro equivalente para el cual podemos usar los resultados de la Sección 0.13. Sea $R = \sup_i R(i)$, y consideremos el proceso alternativo que, en el estado i, nos permite o parar y y pagar un coste $R - R(i)$, o bien seguir pagando un coste de $C(i)$ y haciendo una transición al siguiente estado gobernada por P_{ij}.

Para cualquier política π, sea $V_\pi(\ )$ el coste total esperado con respecto al proceso original cuando se usa π y sea $V'_\pi(\ )$ el coste total esperado con respecto al proceso equivalene definido. Es inmediato ver que para cualquier política que toma la acción de parar en un tiempo esperado finito, se verifica $V'_\pi(\ ) = V_\pi(\ ) + R$. Además, estas son las únicas políticas que nos interesan, ya que por la hipótesis 1, cualquier política que no para en un tiempo esperado finito cumple $V_\pi(\ ) = V'_\pi(\ ) = \infty$. Por tanto, cualquier política óptima para el proceso original, también será óptima para el proceso equivalente y vice versa.

El proceso equivalente sí es un proceso de decisión de Markov con costes no negativos y por tanto podemos aplicar los resultados de la Sección 0.13. Por tanto, podemos asegurar que una política óptima existe y que la función óptima de coste $V'(i)$ satisface

$$V'(i) = \text{mín}\left\{ R - R(i);\ C(i) + \sum_{j=0}^{\infty} P_{ij}(a)V'(j) \right\}$$

Y la política que toma la acción que minimiza es una política óptima. En términos de la función de coste del proceso original tenemos

$$V(i) = \text{mín}\left\{ -R(i);\ C(i) + \sum_{j=0}^{\infty} P_{ij}(a)V(j) \right\}$$

y la política que en el estado i toma la acción que minimiza la parte derecha de esta ecuación, es una política óptima.

Sea $V_0(i) = -R(i)$ y para $n > 0$

$$V_n(i) = \text{mín}\left\{ -R(i);\ C(i) + \sum_{j=0}^{\infty} P_{ij}(a)V_{n-1}(j) \right\}$$

o en otras palabras, $V_n(i)$ es el coste esperado mínimo si empezamos en el estado i, y se nos permite avanzar como máximo n etapas antes de parar. De esta interpretación se sigue que $V_n(i) \geq V_{n+1}(i) \geq V(i)$. Por tanto $\lim_{n\to\infty} V_n(i) \geq V(i)$. Diremos que el proceso es estable si se verifica

$$\lim_{n\to\infty} V_n(i) = V(i)$$

El siguiente teorema muestra que las hipótesis 1 y 2 aseguran la estabilidad y también da cotas en cómo de rápido converge $V_n(i)$ a $V(i)$.

Teorema 23. *Sean $R = \sup_i R(i)$ y $C = \inf_i C(i)$. Asumiendo las condiciones 1 y 2 se tiene*

$$V_n(i) - V(i) \leq \frac{(2R - C)[R - R(i)]}{(n+1)C}$$

Demostración. Sea f una política óptima y T el tiempo aleatorio en el cual f toma la acción de parar. Además, sea f_n la política que escoge las mismas acciones que f en tiempos $0, 1, \ldots, n-1$, pero que para a tiempo n (si no lo ha hecho antes). Se tiene que

$$V(i) = V_f(i) = \mathbb{E}_f[X|T \leq n]P(T \leq n) + \mathbb{E}_f[X|T > n]P(T > n)$$

y

$$V_n(i) \leq V_{f_n}(i) = \mathbb{E}_f[X|T \leq n]P(T \leq n) + \mathbb{E}_{f_n}[X|T > n]P(T > n)$$

donde X denota el coste total incurrido y todo está condicionado a $X_0 = i$. Con esto

$$V_n(i) - V(i) \leq \{\mathbb{E}_{f_n}[X|T > n] - \mathbb{E}_f[X|T \leq n]\} P(T > n) \leq (R - C)P(T > n)$$

Para obtener una cota en $P(T > n)$ observemos que

$$-R(i) \geq V(i) \geq -RP(T \leq n) + (-R + (n+1)C)P(T > n) = -R + (n+1)CP(T > n)$$

con lo que

$$P(T > n) \leq \frac{R - R(i)}{(n+1)C}$$

conduciendo al resultado deseado. □

Ahora sea

$$\begin{aligned} B &= \left\{ i : -R(i) \leq C(i) - \sum_{j=0}^{\infty} P_{ij} R(j) \right\} \\ &= \left\{ i : R(i) \geq \sum_{j=0}^{\infty} P_{ij} R(j) - C(i) \right\} \end{aligned}$$

Es decir, B es el conjunto de estados para los cuales parar es al menos tan bueno como continuar un período más y parar.

Teorema 24. *Si el proceso es estable y si $P_{ij} = 0$ para $i \in B$ y $j \notin B$, entonces la política óptima para en i si y solo si $i \in B$.*

Demostración. Veamos que $V_n(i) = -R(i)$ para todo $i \in B$ y para todo n. Para $n = 0$ esto es trivial. Supongamos que se cumple para $n-1$. Entonces, para $i \in B$

$$\begin{aligned} V_n(i) &= \text{mín}\left\{-R(i);\ C(i) + \sum_{j=0}^{\infty} P_{ij}(a)V_{n-1}(j)\right\} \\ &= \text{mín}\left\{-R(i);\ C(i) + \sum_{j\in B} P_{ij}(a)V_{n-1}(j)\right\} \\ &= \text{mín}\left\{-R(i);\ C(i) - \sum_{j\in B} P_{ij}(a)R(j)\right\} = -R(i) \end{aligned}$$

donde el último paso se sigue del hecho de que $i \in B$. Por inducción, $V_n(i) = -R(i)$ para todo $i \in B$ y para todo n. Además en el límite, haciendo uso de la hipótesis de estabilidad, $V(i) = -R(i)$ para $i \in B$.

Ahora, si $i \notin B$, la política que cotinúa un paso y después para, tiene un coste esperado de

$$C(i) - \sum_{j=0}^{\infty} P_{ij}R(j)$$

que es estrictamente menor que $-R(i)$ pues $i \notin B$. Entonces, si $i \notin B$ $V(i) < -R(i)$, pues existe al menos una política mejor. Por tanto

$$V(i)\begin{cases} = -R(i), & \text{si } \ i \in B \\ < -R(i), & \text{si } \ i \notin B \end{cases}$$

□

Si las condiciones de este teorema se cumplen, se dicen que estamos en el caso monótono. En este caso, la política definida es óptima.

Ejericio 4. Imaginemos que alguien quiere vender su casa y recibe una oferta al comienzo de cada día. Las sucesivas ofertas son independientes entre sí y una oferta es de valor j con probabilidad P_j. Si una oferta no se acepta inmediatamente puede ser aceptada en cualquier tiempo futuro. Por cada día que la casa permanece sin venderse, se incurre un coste C.

Este es un problema de parada óptima, donde la acción de parar se corresponde con aceptar una oferta, y la de seguir con no aceptar. Si dejamos que el estado a tiempo t sea la mayor oferta recibida hasta el tiempo t se tiene

$$P_{ij} = \begin{cases} 0, & \text{si } \ j < i \\ \sum_{k=0}^{i} P_k, & \text{si } \ j = i \\ P_j, & \text{si } \ j > i \end{cases}$$

y por tanto

$$\begin{aligned} B &= \left\{ i : -i \leq C - \sum_{j=0}^{N} P_{ij} R(j) \right\} \\ &= \left\{ i : -i \leq C - i \sum_{k=0}^{i} P_k - \sum_{j=i+1}^{N} j P_j \right\} \\ &= \left\{ i : C \geq \sum_{j=i+1}^{N} j P_j - i \sum_{k=i+1}^{N} P_k \right\} \\ &= \left\{ i : C \geq \sum_{j=i+1}^{N} (j-i) P_j \right\} \end{aligned}$$

Como lo parte derecha de esta expresión es decreciente en i, se sigue que

$$B = \{i^*, i^* + 1, \ldots, N\}$$

donde

$$i^* = \text{mín} \left\{ i : C \geq \sum_{j=i+1}^{N} (j-i) P_j \right\}$$

Por tanto, $P_{ij} = 0$ para $i \in B$ y $j \notin B$ (pues $P_{ij} = 0$ si $j < i$ y todas las ofertas mayores a una $i \in B$ están también en B). Por tanto, la política que acepta la primera oferta que al menos vale i^* es óptima.

Ejericio 5. Análisis secuencial. Sea $Y_1, Y_2, \ldots$ una secuencia de variables aleatorias independientes e idénticamente distribuidas. Supongamos que conocemos que su distribución de probabilidad es, o bien f_0 o bien f_1, y estamos intentando decidir entre una de estas dos.

A tiempo t. tras observar $Y_1, Y_2, \ldots, Y_t$, podemos elegir entre parar y escoger o entre f_1 o f_0, o pagar un coste C para observar Y_{t+1}. Si paramos hacemos una elección acertada, recibiremos un coste nulo, si la decisión es errónea, recibimos un coste $L > 0$.

Suponemos que se nos da una probabilidad inicial p_0 de que la densidad real sea f_0, y diremos que el estado a tiempo t es p si p es la probabilidad a posteriori de que la distribución real sea f_0 a tiempo t.

Esto es un proceso de decisión de Markov con tres acciones, costes no negativos y un espacio de estados no numerable. Si el estado es p, y decidimos parar y escoger f_0 entonces el coste esperado es $(1-p)L$. Si escogemos $f1$, el coste esperado es pL. Si tomamos otra observación, entonces el valor observado será x con probabilidad $pf_0(x)+(1-p)f_1(x)$. Si el valor observado es x, entonces el siguiente estado será

$$X_{t+1} = \frac{pf_0(x)}{pf_0(x)+(1-p)f_1(x)}$$

Por tanto, la función óptima verifica

$$\begin{aligned} V(p) &= \min_{\pi\in\Delta}\left\{(1-p)L,\ pL,\ C+\int_{-\infty}^{\infty} V\left(\frac{pf_0(x)}{pf_0(x)+(1-p)f_1(x)}\right)\right. \\ &\times \left.[pf_0(x)+(1-p)f_1(x)]dx\right\} \end{aligned} \tag{16}$$

Es conveniento definir una subclase de la clase de todas las políticas posibles. Sea Δ la clase de políticas cuya acción a tiempo t es función únicamente de $Y_1, Y_2, \ldots, Y_t$ y no del estado inicial p_0. Como la distribución a posteriori de f_0 a tiempo t es solo función de p_0 y de $Y_1, Y_2, \ldots, Y_t$, se sigue que dado p_0, la política óptima puede ser descrita solamente en términos de $Y_1, Y_2, \ldots, Y_t$. Entonces se tiene que

$$V(p) = \min_{\pi\in\Delta} V_\pi(p)$$

Con esto queremos decir que para cada p existe una política $\pi_p \in \Delta$ tal que $V(p) = V_{\pi_p}(p)$, pero no decimos que exista $\pi \in Delta$ que sea óptima para cualquier estado inicial p. Usando esta última ecuación, podemos demostrar el siguiente Lema.

Lema 25. *$V(p)$ es una función cóncava de p.*

Demostración. Para $\lambda \in (0,1)$, tenemos que

$$V[\lambda p_1 + (1-\lambda)p_2] = \min_{\pi\in\Delta} V_\pi[[\lambda p_1 + (1-\lambda)p_2]$$

Como todas las políticas en Δ son independientes de la probabilidad incial, se sigue que para $\pi \in \Delta$

$$V_\pi[\lambda p_1 + (1-\lambda)p_2] = \lambda V_\pi(p_1) + (1-\lambda)V_\pi(p_2)$$

Por tanto

$$\begin{aligned} V[\lambda p_1 + (1-\lambda)p_2] &= \min_{\pi\in\Delta}\{\lambda V_\pi(p_1) + (1-\lambda)V_\pi(p_2)\} \\ &\geq \min_{\pi\in\Delta}\lambda V_\pi(p_1) + \min_{\pi\in\Delta}(1-\lambda)V_\pi(p_2) \\ &= \lambda V(p_1) + (1-\lambda)V(p_2) \end{aligned}$$

□

La estructura de la política óptima viene dada por el siguiente teorema.

Teorema 26. *Existen números p^*, p^{**} con $p* < p^{**}$, tales que cuando el estado es p, la política óptima para y escoge f_0 si $p > p^{**}$, para y escoge f_1 si $p < p^*$ y continua en el resto de casos.*

Demostración. Supongamos que p_1 y p_2 son tales que $V(p_i) = (1-p_i)L$. Entonces para cualesquiera $p = \lambda p_1 + (1-\lambda)p_2$ con $\lambda \in [0,1]$, el lemma anterior nos garantiza que

$$V(p) \geq (1-p)L$$

No obstante (16) nos asgura que $V(p) \leq (1-p)L$. Entonces $V(p) = (1-p)L$. Por tanto $\{p : V(p) = (1-p)L\}$ es un intervalo. Además, contiene al punto $p = 1$, pues en ese caso $V(1) = 0$. De esto se sigue que la política óptima es parar y escoger f_0 cuando el estado sea mayor que un cierto valor p^{**}. Un razonamiento similar se puede hacer para $\{p : V(p) = pL\}$. □