PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
Departamento de Ingeniería Industrial y de Sistemas
Departamento de Ingeniería de Transporte y Logística

Curso: ICS3105
Semestre: 02-2021
Profesor: Mathias A. Klapp

Tarea 3 - Optimización Dinámica

Fecha de entrega: 7 de Noviembre 2021 : 23:59 HRS

Instrucciones: Se recomienda usar Latex. Evite respuestas largas y sólo responda lo que se pregunta. El corrector puede solicitar revisar su código fuente, pero su respuesta debe ir completamente en el reporte. Se prohíbe discutir la tarea fuera del grupo.

Problema 1: Algo de Teoría (20 puntos)

Considere un MDP de maximización con horizonte infinito descontado por $\lambda \in [0, 1)$ con espacio de estados $\mathbb{S}$ finito; espacio de decisiones $\mathbb{X}(s)$ finito para cada $s \in \mathbb{S}$; probabilidad de transición $p(s'|s, x)$; y función de valor inmediato $r(s, x)$ finito. Considere el valor óptimo $V^*(s)$ para cada $s \in \mathbb{S}$, donde $\mathbf{V}^*$ es su vector.

1. Considere política estacionaria cualquiera definida por la regla de decisión $d' : \mathbb{S} \to \mathbb{X}$, donde su matriz de transición de estados $p(j|i, d(i))$ para cada par de estados $i, j \in \mathbb{S}$ se denota como $\mathbf{P}_{d'}$ y su retorno inmediato $r(i, d(i))$ para cada estado $i \in \mathbb{S}$ se denota como $\mathbf{r}_{d'}$.

   El vector de valor esperado futuro descontado asociado a la política d' se define igual a

   $$\mathbf{V}^{d'} = \sum_{t=1}^{\infty} \lambda^{t-1} \cdot \mathbf{P}_{d'}^{t-1} \cdot \mathbf{r}_{d'}$$

   (a) Desde la definición de $\mathbf{V}^{d'}$ derive la ecuación de recursión $\mathbf{V}^{d'} = \mathbf{r}_{d'} + \lambda \cdot \mathbf{P}_{d'} \cdot \mathbf{V}^{d'}$ y luego obtenga una fórmula exacta para $\mathbf{V}^{d'}$ (5 puntos).

   Solución:

   Es simple, pues:

   $$\begin{aligned}
   \mathbf{V}^{d'} &= \sum_{t=1}^{\infty} \lambda^{t-1} \cdot \mathbf{P}_{d'}^{t-1} \cdot \mathbf{r}_{d'} && (1)\\
   &= \sum_{t=0}^{\infty} \lambda^{t} \cdot \mathbf{P}_{d'}^{t} \cdot \mathbf{r}_{d'} && (2)\\
   &= \mathbf{r}_{d'} + \lambda \cdot \mathbf{P}_{d'} \cdot \sum_{t=1}^{\infty} \lambda^{t-1} \cdot \mathbf{P}_{d'}^{t-1} \cdot \mathbf{r}_{d'} && (3)\\
   &= \mathbf{r}_{d'} + \lambda \cdot \mathbf{P}_{d'} \cdot \mathbf{V}^{d'} && (4)\\
   & && (5)
   \end{aligned}$$

   Luego, reordenando se obtiene $(\mathbf{I} - \lambda \cdot \mathbf{P}_{d'}) \cdot \mathbf{V}^{d'} = \mathbf{r}_{d'}$. Multiplicando por la matriz inversa (que siempre existe) se llega a $\mathbf{V}^{d'} = (\mathbf{I} - \lambda \mathbf{P}_{d'})^{-1} \cdot \mathbf{r}_{d'}$

   (b) Defina la función vectorial $L_{d'} : \mathbb{R}^{|\mathbb{S}|} \to \mathbb{R}^{|\mathbb{S}|}$ como:

   $$L_{d'}(\mathbf{V}) = \mathbf{r}_{d'} + \lambda \cdot \mathbf{P}_{d'} \cdot \mathbf{V}.$$

   Considere la serie $\mathbf{V}^{n+1} = L_{d'}(\mathbf{V}^n)$ para $n \geq 0$ considerando un punto de partida $\mathbf{V}^0$ cualquiera. Muestre que si para algún $n' \geq 0$ se cumple que $\mathbf{V}^{n'+1} \geq \mathbf{V}^{n'}$, entonces se cumple que $\mathbf{V}^{m+1} \geq$

$\mathbf{V}^m, \quad \forall m \geq n'$ (5 puntos).

Solución:

Probemos primero que si se cumple $\mathbf{V} \geq \mathbf{U}$ para dos vectores cualquiera, entonces se debe cumplir que $L_{d'}(\mathbf{V}) \geq L_{d'}(\mathbf{U})$. Es cierto, pues desde $\mathbf{V} \geq \mathbf{U}$ se deriva que $\lambda \cdot \mathbf{P}_{d'} \cdot \mathbf{V} \geq \lambda \cdot \mathbf{P}_{d'} \cdot \mathbf{U}$ (son combinaciones lineales de las desigualdades anteriores ponderadas por coeficientes no negativos). Luego, se debe cumplir lo pedido: $\mathbf{r}_{d'} + \lambda \cdot \mathbf{P}_{d'} \cdot \mathbf{V} \geq \mathbf{r}_{d'} + \lambda \cdot \mathbf{P}_{d'} \cdot \mathbf{U}$.
Aplicando el resultado, si se cumple que $\mathbf{V}^{n'+1} \geq \mathbf{V}^{n'}$ entonces $\mathbf{V}^{n'+2} = L_{d'}(\mathbf{V}^{n'+1}) \geq L_{d'}(\mathbf{V}^{n'}) = \mathbf{V}^{n'+1}$. Por inducción se debe cumplir para todo $m \geq n'$.

2. Defina la función vectorial $L : \mathbb{R}^{|\mathbb{S}|} \to \mathbb{R}^{|\mathbb{S}|}$ para cada $s \in \mathbb{S}$ como:

$$L(\mathbf{V})(s) = \max_{x \in \mathbb{X}(s)} \left\{ r(s,x) + \lambda \cdot \sum_{s' \in \mathbb{S}} p(s'|s,x) V(s') \right\}.$$

(a) Muestre que si $\mathbf{V} \leq L(\mathbf{V})$, entonces $\mathbf{V}$ es una cota inferior (componente a componente) del vector de value-to-go $\mathbf{V}^*$ que resuelve $\mathbf{V}^* = L(\mathbf{V}^*)$ (4 puntos).

Solución:

Si se cumple para todo $s \in \mathbb{S}$ que:

$$V(s) \leq L(\mathbf{V})(s) = \max_{x \in \mathbb{X}(s)} \left\{ r(s,x) + \lambda \sum_{s' \in \mathbb{S}} p(s'|s,x) V(s') \right\},$$

entonces existe una regla de decisión $d'(s)$ (la regla que optimiza $L(\mathbf{V})(s)$) tal que

$$\begin{aligned} V(s) &\leq r\big(s, d'(s)\big) + \lambda \sum_{s' \in \mathbb{S}} p(s'|s, d'(s)) V(s') \quad \forall s \in \mathbb{S} \\ &\Rightarrow \mathbf{V} \leq r_{d'} + \lambda \mathbf{P}_{d'} \mathbf{V} \\ &\Rightarrow \underbrace{(I - \lambda \mathbf{P}_{d'})^{-1} \mathbf{r}_{d'}}_{\mathbf{V}_{d'}} \geq \mathbf{V} \end{aligned}$$

Por lo tanto, existe una regla de decisión d' tal que $\mathbf{V}_{d'} \geq \mathbf{V}$, lo que implica que $\mathbf{V}^* \geq \mathbf{V}_{d'} \geq \mathbf{V}$.

(b) Considere que usted está ejecutando Iteración de Valor, es decir, está computando la serie $\mathbf{V}^{n+1} = L(\mathbf{V}^n)$ desde un punto de partida $\mathbf{V}^0$ cualquiera. Muestre que si en una iteración intermedia n' se cumple que $\mathbf{V}^{n'+1} \geq \mathbf{V}^{n'}$, entonces para todo $m \geq n'$ se debe cumplir que $\mathbf{V}^{n'} \leq \mathbf{V}^m \leq \mathbf{V}^*$ (4 puntos).

Solución:

Probemos primero que si se cumple $\mathbf{V} \geq \mathbf{U}$ para dos vectores cualquiera, entonces se debe cumplir que $L(\mathbf{V}) \geq L(\mathbf{U})$.
Para probar aquello, defina la regla de decisión $d(s) := \text{argmax}_{x \in \mathbb{X}(s)} \left\{ r(s,x) + \lambda \sum_{s' \in \mathbb{S}} p(s'|s,x) U(s') \right\}$ para todo $s \in \mathbb{S}$. Luego:

$$\begin{aligned} L(\mathbf{U})(s) &= r(s, d(s)) + \lambda \sum_{s' \in \mathbb{S}} p(s'|s, d(s)) U(s') \\ &\leq r(s, d(s)) + \lambda \sum_{s' \in \mathbb{S}} p(s'|s, d(s)) V(s') \text{ (ya que} \mathbf{U} \leq \mathbf{V}) \\ &\leq \max_{x \in \mathbb{X}(s)} \{ r(s,x) + \lambda \sum_{s' \in \mathbb{S}} p(s'|s,x) V(s') \} \\ &= L(\mathbf{V})(s). \end{aligned}$$

Aplicando este resultado: si $\mathbf{V}^{n'+1} \geq \mathbf{V}^{n'}$, entonce $\mathbf{V}^{n'+2} = L(\mathbf{V}^{n'+1}) \geq \mathbf{V}^{n'+1} = L(\mathbf{V}^{n'})(s)$. Por inducción, se cumple lo solicitado.

(c) Considere que usted ejecutó Iteración de Valor hasta un error de 10^{-2}, es decir, computó la serie $\mathbf{V}^{n+1} = L(\mathbf{V}^n)$ desde un punto de partida $\mathbf{V}^0 = \mathbf{0}$ hasta una iteración n tal que $||\mathbf{V}^{n+1} - \mathbf{V}^n||_\infty = 10^{-2}$ (norma infinito). Su regla de decisión es $d'(s) = \text{argmax}_{x \in \mathbb{X}(s)}\{r(s,x) + \lambda \sum_{s' \in \mathbb{S}} p(s'|s,x) V^{n+1}(s')\}$ para todo $s \in \mathbb{S}$.

Si $\lambda = 0,95$ ¿A qué garantía de optimalidad se encuentra d' de la política óptima? (2 puntos)

Solución:

Se sabe que $10^{-2} = \frac{\epsilon \cdot (1-\lambda)}{2 \cdot \lambda}$. Basta con despejar y obtener ϵ

Problema 2: Estudio óptimo (13 puntos)

Un alumno desea maximizar el uso de su tiempo y decidir una política de estudio cada semana. Suponga que el alumno puede estar en tres estados: descansado (0), estresado (1) y colapsado (2).

Si el alumno está en los estados descansado o estresado, entonces puede estudiar poco ($x = 0$) y obtener un beneficio semanal de 1 o estudiar mucho ($x = 1$) y obtener un beneficio semanal de 2. Un alumno colapsado no tiene decisiones y su beneficio semanal es 0 hasta terminar el año (a perpetuidad).

Si el alumno está descansado y estudia poco, entonces estará descansado para la próxima semana. Por el contrario, si está descansado y estudia mucho existe un 50% de probabilidades de estresarse para la próxima semana, de lo contrario termina descansado.

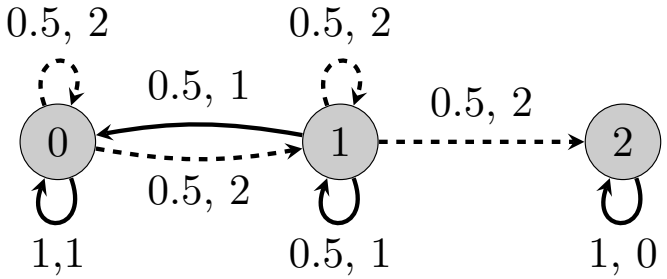
Si el alumno está estresado y estudia poco, hay un 50% de probabilidad de terminar descansado para la próxima semana, de lo contrario termina estresado. Si estudia mucho al estar estresado, tiene un 50% de probabilidad de terminar colapsado para la próxima semana, de lo contrario termina estresado.

Preguntas:

1. Modele este problema como un MDP de horizonte infinito descontado con $\lambda = 4/5$. Específicamente, defina su función de value-to-go y las ecuaciones de optimalidad de Bellman (1 punto).

   Solución:

   Los arcos continuos corresponden a la decisión de estudiar poco ($x = 0$) y los punteados a la de estudiar mucho ($x = 1$).



   La formulación del MDP resulta ser:

$$V(0) = \max\left(1 + \frac{4}{5}V(0)\ ,\ 2 + \frac{4}{5}\left(\frac{V(0)+V(1)}{2}\right)\right)$$
$$V(1) = \max\left(1 + \frac{4}{5}\left(\frac{V(0)+V(1)}{2}\right)\ ,\ 2 + \frac{4}{5}\left(\frac{V(1)+V(2)}{2}\right)\right)$$
$$V(2) = 0$$

2. Escriba las ecuaciones que determinan el valor futuro esperado en cada estado de la política "estudiar poco independiente del estado". Resuélvalas, obteniendo el valor de dicha política en cada estado (4 puntos).

   Solución:

   La política de decisión es $d^0 = (0, 0, 0)$. Así, el sistema queda

$$\begin{aligned} V^0(0) &= 1 + \frac{4}{5}V^0(0) \\ V^0(1) &= 1 + \frac{4}{10}(V^0(0) + V^0(1)) \\ V^0(2) &= 0 \end{aligned}$$

   Resolviendo, se obtiene que $V^0(0) = 5$, $V^0(1) = 5$ y $V^0(2) = 0$.

3. Resuelva el problema a optimalidad mediante Iteración de Política desde la política "estudiar poco independiente del estado" (4 puntos).

4. Resuelva el problema a optimalidad mediante Iteración de Valor desde value-to-go nulos. Compare esta solución con la anterior en términos de resultado y cómputo (4 puntos).

   Solución:

   Ejecutamos con V^0 (valor de $d^0 = (0, 0, 0)$).

$$\begin{aligned} d^1(0) &\in \text{argmax}\{1 + \frac{4}{5} \cdot 5 \ , \ 2 + \frac{4}{10}(5 + 5)\} \\ &\Rightarrow d^1(0) = 1 \end{aligned}$$

   Por otro lado,

$$\begin{aligned} d^1(1) &\in \text{argmax}\{1 + \frac{4}{10}(5 + 5) \ , \ 2 + \frac{4}{10}(5 + 0)\} \\ &\Rightarrow d^1(1) = 0 \end{aligned}$$

   Como $d^1 = (1, 0, 0) \neq d^0$, seguimos iterando ahora con d^1. Obtenemos,

$$\begin{aligned} V^1(0) &= 2 + \frac{4}{10}(V^1(0) + V^1(1)) \\ V^1(1) &= 1 + \frac{4}{10}(V^1(0) + V^1(1)) \\ V^1(2) &= 0 \end{aligned}$$

   Resolviendo, se obtiene que $V^1(0) = 8$, $V^1(1) = 7$ y $V^1(2) = 0$. Ejecutamos con V^1.

$$\begin{aligned} d^2(0) &\in \text{argmax}\{1 + \frac{4}{5} \cdot 8 \ , \ 2 + \frac{4}{10}(8 + 7)\} \\ &\Rightarrow d^2(0) = 1 \end{aligned}$$

   Por otro lado,

$$\begin{aligned} d^2(1) &\in \text{argmax}\{1 + \frac{4}{10}(8 + 7) \ , \ 2 + \frac{4}{10}(7 + 0)\} \\ &\Rightarrow d^2(1) = 0 \end{aligned}$$

   Dado que $d^2 = d^1$, se logra convergencia por lo que

$$d^* = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad V^* = \begin{bmatrix} 8 \\ 7 \\ 0 \end{bmatrix}$$

Problema 3: Robot (13 puntos)

Supongamos que usted debe darle instrucciones a un robot saltarín que debe subir por los peldaños de una escalera de 100 pisos. El robot saltarín comienza en el piso 1 y su meta es llegar al piso 100. En cada piso s usted debe definir la acción del robot entre cuatro disponibles definidas por un parámetro $p \in \left\{\frac{1}{10}; \frac{1}{5}; \frac{2}{5}; 1\right\}$. Si se escoge la acción p, el robot saltara un piso (terminará en el piso $s+1$) con probabilidad p, saltará 2 con probabilidad $p \times (1-p)$ y así sucesivamente (saltará q pisos con probabilidad $p \cdot (1-p)^{q-1}$). El problema es que si salta más allá del piso 100 caerá al piso 1 y comenzará desde cero. Suponga que el costo de cada movida es \$1 y que el factor de descuento es $\lambda = 0,99$.

1. Modele el problema como un MDP de horizonte infinito (6 puntos).

   Solución:

   Estados: piso en el que está el robot.

   $$s \in \{1, ..., 100\}$$

   Etapas: intento de salto.

   $$t \in \{1, ..., \infty\}$$

   Acción: selección de parámetro p

   $$p \in \{\frac{1}{10}, \frac{1}{5}, \frac{2}{5}, 1\}$$

   Costo inmediato: costo de cada movida.

   $$r_t = 1$$

   Función de transición:

   $$\mathbb{P}(s_2|p, s_1) = p \cdot (1-p)^{s_2 - s_1 - 1} \quad \forall s_1, s_2 : s_2 > s_1, s_2 \leq 100$$

   $$\mathbb{P}(s_2|p, s_1) = 0 \quad \forall s_1, s_2 : s_2 < s_1, s_2 \neq 1$$

   $$\mathbb{P}(s_2|p, s_1) = \sum_{q=101-s_1}^{\infty} p \cdot (1-p)^{q-1} = 1 - \sum_{q=1}^{101-s_1} p \cdot (1-p)^{q-1} \quad \forall s_1, s_2 : s_2 < s_1, s_2 = 1$$

   $$\mathbb{P}(s_2|p, s_1) = 0 \quad \forall s_1 = s_2$$

   Ecuaciones de Bellman:

   $$V_t(s_1) = 1 + \lambda \sum_{s_2=1}^{100} \cdot \mathbb{P}(s_2|s_1, x) \cdot V_t(s_2)$$

2. Encuentre una política óptima que minimice el costo total esperado de subir al piso 100 (7 puntos).
   Solución:

   La política óptima corresponde a:

   $$x(s) = 0.1 \quad s \in \{1, ..., 79\}$$
   $$x(s) = 0.2 \quad s \in \{80, ..., 89\}$$
   $$x(s) = 0.4 \quad s \in \{90, ..., 96\}$$
   $$x(s) = 1 \quad s \in \{97, ..., 100\}$$

   El valor de comenzar en el piso 1 es 16,02.

Problema 4: Inventario Estocástico (14 puntos)

Considere un problema de inventario estocástico como el desarrollado en clase, donde un tomador de decisiones (*i.e.*, usted) debe ordenar productos al comienzo de cada semana (después de observar inventario) a lo largo de un horizonte de ejecución infinito. Suponga una tasa de descuento semanal λ. Específicamente, suponga que inicia en la semana 1 con $y_1 = 30$ unidades almacenadas y que su bodega posee capacidad para $Q = 200$ unidades. La dinámica de eventos en cada semana $t \in \{1, \ldots, \infty\}$ es la siguiente:

1. Usted observa el inventario del producto en bodega $y_t \in \{0, 1, 2, \ldots, Q\}$.
2. Luego, decide ordenar $x_t \in \{0, 1, \ldots, Q - y_t\}$ unidades de producto a un costo de ordenamiento $\$1600 + 20 \cdot x_t$. El inventario ordenado es repuesto instantáneamente hasta $y'_t = y_t + x_t$.
3. Observa la demanda D_t de la semana t, donde D_t es una variable aleatoria independiente (i.i.d. para cada semana). El cliente recibe todo el producto demandado disponible (es decir, $\min(D_t, y'_t)$) y el sobrante se almacena para la siguiente semana. El costo por unidad-semana en inventario es $h = \$4$ (a pagar si un producto pasa en inventario de una semana a la otra) y el costo por unidad demandada no satisfecha es $q = \$75$. Asuma que la demanda semanal D_t es discreta y distribuida uniforme en $\{30, \ldots, 60\}$.

Preguntas:

- Modele el problema que planifica una política de decisión que minimiza el costo total esperado como un Proceso de Decisión Markoviana (MDP) de horizonte infinito (2 puntos).

  Solución:

  Modelaremos el problema. Lo que queremos hacer es minimizar el costo total esperado a través de un MDP de horizonte infinito. La formulación MDP se expone a continuación.

  Decisión ($x(s)$): cantidad de producto a pedir si tengo inventario s.

  $$\mathbb{X}_t(s) = \{x_t \in \mathbb{Z}^+ : x_T \leq Q - s\} \quad t = 1, ..., \infty$$

  Con Q la capacidad máxima de la bodega. En este caso en particular $Q = 200$.

  Estados (s): nivel de inventario al inicio de la etapa t.

  $$\mathbb{S}_t = \{0, ..., Q\} \quad t = 1, ..., \infty$$

  Función de transición: la probabilidad de que dada una decisión x_t en el estado s_t llegue al estado s_{t+1}.

  - Si $s_{t+1} = 0 \rightarrow \mathbb{P}(s_{t+1}|x_t, s_t) = \mathbb{P}(D_t \geq x_t + s_t)$
  - Si $0 < s_{t+1} \leq x_t + s_t \rightarrow \mathbb{P}(s_{t+1}|x_t, s_t) = \mathbb{P}(D_t = s_{t+1} - (x_t + s_t))$
  - Si $s_{t+1} > x_t + s_t \rightarrow \mathbb{P}(s_{t+1}|x_t, s_t) = 0$

  Costo inmediato: costo en la etapa t de tomar la decisión x_t en el estado s_t.

  $$r_t(x_t, s_t) = 1600 \cdot \mathbb{I}_{x_t > 0} + 20 \cdot x_t + 75 \cdot \mathbb{E}_{D_t}[(D_t - (x_t + s_t))^+] + 4 \cdot \mathbb{E}_{D_t}[(x_t + s_t - D_t)^+]$$

  Donde:

  - $\mathbb{E}_{D_t}[(D_t - (x_t + s_t))^+] = \sum_{d_t = \min\{\mu_t - \delta_t, x_t + s_t + 1\}}^{\mu_t + \delta_t} (d_t - (x_t + s_t)) \cdot \dfrac{1}{2 \cdot \delta_t + 1}$
  - $\mathbb{E}_{D_t}[(x_t + s_t - D_t)^+] = \sum_{d_t = \mu_t - \delta_t}^{\max\{\mu_t - \delta_t, x_t + s_t - 1\}} (x_t + s_t - d_t) \cdot \dfrac{1}{2 \cdot \delta_t + 1}$

–

El valor de $d_t = x_t + s_t$ queda implícito en las relaciones debido a que no aporta en ninguno de los dos costos, pero es parte de las probabilidades totales.

Objetivo: minimizar el costo esperado desde la etapa 0.

$$\max\{V^*(s)\} = \max_{s\in S}\left\{\mathbb{E}\left(\lim_{n\to\infty}\sum_{t=0}^{n}\lambda^{t-1}r_t(S_t, d_t^{\pi}(S_t))|S_0 = 30\right)\right\}$$

- Obtenga una política estacionaria d que minimiza el costo total descontado esperado del sistema ejecutando iteración de valor para $\lambda \in \{90\%, 95\%, 99\%\}$ con 10^{-4} unidades de precisión absoluta, *i.e.*, $\max_{s\in\mathbb{S}}(V^d(s) - V^*(s)) < 10^{-4}$. Verifique (en cada caso) que efectivamente la política óptima obtenida posee una estructura (s, S) para cada semana (3 puntos).

  Solución:

  La política óptima para cada λ es:

  $\lambda = \mathbf{0.90}$:

  $\pi = \{0:148, 1:147, 2:146, 3:145, 4:144, 5:143, 6:142, 7:141, 8:140, 9:139, 10:138, 11:137, 12:136, 13:135, 14:134, 15:133, 16:132, 17:131, 18:130, 19:129, 20:128, 21:127, 22:126, 23:125, 24:124, 25:123, 26:122, 27:121, 28:0, 29:0, 30:0, 31:0, 32:0, 33:0, 34:0, 35:0, 36:0, 37:0, 38:0, 39:0, 40:0, 41:0, 42:0, 43:0, 44:0, 45:0, 46:0, 47:0, 48:0, 49:0, 50:0, 51:0, 52:0, 53:0, 54:0, 55:0, 56:0, 57:0, 58:0, 59:0, 60:0, 61:0, 62:0, 63:0, 64:0, 65:0, 66:0, 67:0, 68:0, 69:0, 70:0, 71:0, 72:0, 73:0, 74:0, 75:0, 76:0, 77:0, 78:0, 79:0, 80:0, 81:0, 82:0, 83:0, 84:0, 85:0, 86:0, 87:0, 88:0, 89:0, 90:0, 91:0, 92:0, 93:0, 94:0, 95:0, 96:0, 97:0, 98:0, 99:0, 100:0, 101:0, 102:0, 103:0, 104:0, 105:0, 106:0, 107:0, 108:0, 109:0, 110:0, 111:0, 112:0, 113:0, 114:0, 115:0, 116:0, 117:0, 118:0, 119:0, 120:0, 121:0, 122:0, 123:0, 124:0, 125:0, 126:0, 127:0, 128:0, 129:0, 130:0, 131:0, 132:0, 133:0, 134:0, 135:0, 136:0, 137:0, 138:0, 139:0, 140:0, 141:0, 142:0, 143:0, 144:0, 145:0, 146:0, 147:0, 148:0, 149:0, 150:0, 151:0, 152:0, 153:0, 154:0, 155:0, 156:0, 157:0, 158:0, 159:0, 160:0, 161:0, 162:0, 163:0, 164:0, 165:0, 166:0, 167:0, 168:0, 169:0, 170:0, 171:0, 172:0, 173:0, 174:0, 175:0, 176:0, 177:0, 178:0, 179:0, 180:0, 181:0, 182:0, 183:0, 184:0, 185:0, 186:0, 187:0, 188:0, 189:0, 190:0, 191:0, 192:0, 193:0, 194:0, 195:0, 196:0, 197:0, 198:0, 199:0, 200:0\}$

  $\lambda = \mathbf{0.95}$:

  $\pi = \{0:186, 1:185, 2:184, 3:183, 4:182, 5:181, 6:180, 7:179, 8:178, 9:177, 10:176, 11:175, 12:174, 13:173, 14:172, 15:171, 16:170, 17:169, 18:168, 19:167, 20:166, 21:165, 22:164, 23:163, 24:162, 25:161, 26:160, 27:159, 28:158, 29:157, 30:0, 31:0, 32:0, 33:0, 34:0, 35:0, 36:0, 37:0, 38:0, 39:0, 40:0, 41:0, 42:0, 43:0, 44:0, 45:0, 46:0, 47:0, 48:0, 49:0, 50:0, 51:0, 52:0, 53:0, 54:0, 55:0, 56:0, 57:0, 58:0, 59:0, 60:0, 61:0, 62:0, 63:0, 64:0, 65:0, 66:0, 67:0, 68:0, 69:0, 70:0, 71:0, 72:0, 73:0, 74:0, 75:0, 76:0, 77:0, 78:0, 79:0, 80:0, 81:0, 82:0, 83:0, 84:0, 85:0, 86:0, 87:0, 88:0, 89:0, 90:0, 91:0, 92:0, 93:0, 94:0, 95:0, 96:0, 97:0, 98:0, 99:0, 100:0, 101:0, 102:0, 103:0, 104:0, 105:0, 106:0, 107:0, 108:0, 109:0, 110:0, 111:0, 112:0, 113:0, 114:0, 115:0, 116:0, 117:0, 118:0, 119:0, 120:0, 121:0, 122:0, 123:0, 124:0, 125:0, 126:0, 127:0, 128:0, 129:0, 130:0, 131:0, 132:0, 133:0, 134:0, 135:0, 136:0, 137:0, 138:0, 139:0, 140:0, 141:0, 142:0, 143:0, 144:0, 145:0, 146:0, 147:0, 148:0, 149:0, 150:0, 151:0, 152:0, 153:0, 154:0, 155:0, 156:0, 157:0, 158:0, 159:0, 160:0, 161:0, 162:0, 163:0, 164:0, 165:0, 166:0, 167:0, 168:0, 169:0, 170:0, 171:0, 172:0, 173:0, 174:0, 175:0, 176:0, 177:0, 178:0, 179:0, 180:0, 181:0, 182:0, 183:0, 184:0, 185:0, 186:0, 187:0, 188:0, 189:0, 190:0, 191:0, 192:0, 193:0, 194:0, 195:0, 196:0, 197:0, 198:0, 199:0, 200:0\}$

  $\lambda = \mathbf{0.99}$

  $\pi = \{0:195, 1:194, 2:193, 3:192, 4:191, 5:190, 6:189, 7:188, 8:187, 9:186, 10:185, 11:184, 12:183, 13:182, 14:181, 15:180, 16:179, 17:178, 18:177, 19:176, 20:175, 21:174, 22:173, 23:172, 24:171, 25:170, 26:169, 27:168, 28:167, 29:166, 30:165, 31:164, 32:0, 33:0, 34:0, 35:0, 36:0, 37:0, 38:0, 39:0, 40:0, 41:0, 42:0, 43:0, 44:0, 45:0, 46:0, 47:0, 48:0, 49:0, 50:0, 51:0, 52:0, 53:0, 54:0, 55:0, 56:0, 57:0, 58:0, 59:0, 60:0, 61:0, 62:0, 63:0, 64:0, 65:0, 66:0, 67:0, 68:0, 69:0, 70:0, 71:0, 72:0, 73:0, 74:0, 75:0, 76:0, 77:0, 78:0, 79:0, 80:0, 81:0, 82:0, 83:0, 84:0, 85:0, 86:0, 87:0, 88:$

$0,89:0,90:0,91:0,92:0,93:0,94:0,95:0,96:0,97:0,98:0,99:0,100:0,101:0,102:0,103:0,104:0,105:0,106:0,107:0,108:0,109:0,110:0,111:0,112:0,113:0,114:0,115:0,116:0,117:0,118:0,119:0,120:0,121:0,122:0,123:0,124:0,125:0,126:0,127:0,128:0,129:0,130:0,131:0,132:0,133:0,134:0,135:0,136:0,137:0,138:0,139:0,140:0,141:0,142:0,143:0,144:0,145:0,146:0,147:0,148:0,149:0,150:0,151:0,152:0,153:0,154:0,155:0,156:0,157:0,158:0,159:0,160:0,161:0,162:0,163:0,164:0,165:0,166:0,167:0,168:0,169:0,170:0,171:0,172:0,173:0,174:0,175:0,176:0,177:0,178:0,179:0,180:0,181:0,182:0,183:0,184:0,185:0,186:0,187:0,188:0,189:0,190:0,191:0,192:0,193:0,194:0,195:0,196:0,197:0,198:0,199:0,200:0\}$

- Optimice el problema de nuevo, pero ejecutando iteración de política para cada tasa de descuento (comience desde la política $(0,0)$). Explote en su cómputo que la política óptima posee una estructura (s,S).

  $\lambda = \mathbf{0.90}$:

  $(s,S) = (27,148)$.

  $\lambda = \mathbf{0.95}$:

  $(s,S) = (29,186)$.

  $\lambda = \mathbf{0.99}$:

  $(s,S) = (31,195)$.

- Optimice el problema otra vez, pero ejecutando el método de programación lineal (dual o primal, usted elige cual). Confirme el resultado anterior y compare la eficiencia de cómputo de los tres métodos aplicados a este problema (3 puntos).

  Solución:

  La eficiencia se compara a continuación (importan los ordenes de magnitud).

Método	**Lambda**	**Tiempo de ejecución [s]**
Iteración de valor	0.90	186.86
Iteración de valor	0.95	389.47
Iteración de valor	0.99	2104.12
Iteración de política	0.90	0.26
Iteración de política	0.95	0.25
Iteración de política	0.99	0.29
LP	0.90	1.43
LP	0.95	2.30
LP	0.99	4.58

- Ahora supongan que cada vez que ordena productos existe un *Lead Time* L_t, es decir, un tiempo de entrega de la orden que tarda una semana con probabilidad 50%, dos semanas con probabilidad 35% y tres semanas con probabilidad 15%. Modele el problema de inventario estocástico de horizonte infinito con esa nueva consideración y optimice el problema que minimiza el costo total esperado descontado con el método que usted prefiera. Note que en cada período puede pedir, incluso si el pedido anterior no ha llegado (3 puntos).

Solución:

Una posible formulación para el MDP se expone a continuación:
Etapas: ∞ etapas, una por cada semana.

Decisión: cantidad de producto a ordenar (pedir) en la etapa t x_t.

$$\mathbb{X}_t(s_t) = \{x_t \in \mathbb{Z}^+ : x_t \leq Q - s_t[0]\} \quad \forall t \in \{1, ..., \infty\}$$

.
Con Q la capacidad máxima de la bodega. En este caso en particular $Q = 200$.
Estados: nivel de inventario al inicio de la etapa t, inventario en tránsito que llegará en un día y cantidad de inventario que llegará en 2 días. No se considera el inventario que llegará en 3 días, puesto que al momento de hacer una transición restan como máximo 2 días.

$$\mathbb{S}_t = \{(I_0, I_1, I_2) : \forall (I_0, I_1, I_2) \in \{0, ..., Q\}^3, \forall t \in \{1, ..., 52\}\}$$

Función de Transición: La función de transición depende del estado y la decisión.

- Con *leadtime* 1:

  $\mathbb{P}((I_1 + I_0 + x_t - d, I_2, 0)|x_t, (I_0, I_1, I_2)) = 0,5 \cdot \mathbb{P}(I_0 - D_t = d)$ con $d \geq 0$
  $\mathbb{P}((I_1 + x_t, I_2, 0)|x_t, (I_0, I_1, I_2)) = 0,5 \cdot \mathbb{P}(I_0 - D_t \leq 0)$
- Con *leadtime* 2:

  $\mathbb{P}((I_1 + I_0 - d, I_2 + x_t, 0)|x_t, (I_0, I_1, I_2)) = 0,35 \cdot \mathbb{P}(I_0 - D_t = d)$ con $d \geq 0$
  $\mathbb{P}((I_1, I_2 + x_t, 0)|x_t, (I_0, I_1, I_2)) = 0,35 \cdot \mathbb{P}(I_0 - D_t \leq 0)$

- Con *leadtime* 3:

  $\mathbb{P}((I_1 + I_0 - d, I_2, 0 + x_t)|x_t, (I_0, I_1, I_2)) = 0,15 \cdot \mathbb{P}(I_0 - D_t = d)$ con $d \geq 0$
  $\mathbb{P}((I_1, I_2, 0 + x_t)|x_t, (I_0, I_1, I_2)) = 0,15 \cdot \mathbb{P}(I_0 - D_t \leq 0)$

Costo Inmediato: costo en la etapa t de tomar la decisión x_t en el estado s_t.

$$r_t(x_t, s_t) = 1600 \cdot \mathbb{I}\{x_t > 0\} + 20 \cdot x_t + 75 \cdot \mathbb{E}D_t[(D_t - s_t[0])^+] + 4 \cdot \mathbb{E}_{D_t}[(s_t[0] - D_t)^+]$$

Objetivo: minimizar el costo esperado desde la etapa 0, donde $s_1 = (30, 0, 0)$.

$$\max\{V^*(s)\} = \max_{s \in S} \left\{ \mathbb{E} \left(\lim_{n \to \infty} \sum_{t=0}^{n} \lambda^{t-1} r_t(S_t, d_t^\pi(S_t)) | S_1 = s \right) \right\}$$