Devoir 4
MTH 6415

Olivier Sirois - 1626107
Corey Ducharme - 1626614

2018-03-28

4.22 (3.22 ed. 4)

On peut voir ce problème comme un problème de temps d'arrêt optimal. Ce problème est similaire à celui du voleur qui sait compter. Définisson donc le problème de cette manière.
Soit x_k notre variable d'état représentant l'argent cumulé jusqu'à l'état k. Le but de l'extorqueur est de maximiser son profit espéré sur les N étapes. Soit,

$$max(E[x_N])$$

Les décisions possibles à chaque étape k pour notre extorqueur est de continuer son extorsion en demandant un montant d'argent u_k ou de s'arrêter et recevoir un montant R. De plus, s'il demande de l'argent, il peut soit la recevoir avec un probabilité $(1 - u_k)$ ou il peut se faire raporter à la police avec une probabilité u_k. Nous exprimons, ces différents états de la décisions comme ceci :

$$U = [0, 1] \cup \Delta$$

où Δ représente l'état où l'extorqueur ne peu plus faire de l'extorsion soit parce qu'il à décider d'arrêter ou parce qu'il s'est fait reporter à la police.
Nous écrivons maintenant les équations de Bellmann du problème.

$$J_N(x_N) = x_N$$
$$J_k(x_k) = \max_{u \in U(x_k)} (x_k + R, (1 - u_k)E[J_{k+1}(x_k + u_k)] + u_k x_k)$$

Nous remarquons que notre problème peut être vue comme un problème d'un coup à l'avance. L'état du un coup à l'avance peut donc être écrit comme :

$$\begin{aligned} T_{N-1} &= \{x \mid x + R \geq (1 - u)(x + u) + ux\} \\ &= \{x \mid R \geq (1 - u)(x + u) + ux - x\} \\ &= \{x \mid R \geq u(1 - u)\} \end{aligned}$$

Nous déterminons la valeur maximale de $u(1 - u)$ dans cette situation comme étant $u = 0.5$. Ainsi nous obtenons

$$T_{N-1} = \left\{x \mid R \geq \frac{1}{4}\right\}$$

Donc, nous pouvons conclure que la décision optimale de l'extorqueur est indépendante du montant x_k. Sa décision optimale est se retirer si le montant R est supérieur à $\frac{1}{4}$ ou de continuer à extorquer de l'argent jusqu'à ce qu'il se fasse reporter à la police.

0.1 4.23 (3.23 ed. 4)

Nous remarquons que ce problème a une politique optimale en boucle ouverte. En effet, le problème est stochastique, mais l'information obtenue au cours des premières étapes n'est pas utile pour améliorer les décisions futures. Ainsi, nous pouvons chercher un ordonnacement optimale pour ce problème en utilisant l'argument d'échange des voisins.
Soit L un ordonnacement optimales des N décisions de Theseus et L' une permutation de cette ordonnacement.

$$L = (i_0, \ldots, i_{k-1}, i, j, i_{k+2}, \ldots, i_{N-1})$$
$$L' = (i_0, \ldots, i_{k-1}, j, i, i_{k+2}, \ldots, i_{N-1})$$

Soit $J(x)$, la probabilité que Theseus s'échappe suivant un ordonnacement x. Dans ce problème, nous remarquons qu'à chaque tour la probabilité de réussite est p_k, la probabilité d'échec est de q_k. Donc, la probabilité de pouvoir se réesseyer au tour $k+1$ est de $(1 - p_k - q_k)$. Ainsi, nous pouvons écrire les équations pour $J(L)$ et $J(L')$.

$$\begin{aligned} J(L) = &\geq p_j + (1 - p_j - q_j)p_i \\ -q_i p_j &\geq -q_j p_i \\ q_j p_i &\geq q_i p_j \\ \frac{p_i}{q_i} &\geq \frac{p_j}{q_j} \end{aligned}$$

Donc, la politique optimale de Theseus est de choisir les passages selon l'ordre décroissant des ratios p_i/q_i.

0.2 p.450 # 16

a)

Pour ce problème, on nous demande de prouver que la politique du problème est sous la forme :

- si $i > i*$, on répare
- sinon, on continue.

Il n'y a pas beaucoup d'information donnés dans l'énoncé.. seulement que:

- $g(1) \leq g(2) \leq g(3) \leq g(4)...g(n)$

- deux seule transition sont possible: $i' = 0, i' = i + 1$ avec $0 < p < 1$

Évidemment, on peut déduire certaines charactéristique de nos fonctions J. Premièrement, le problème est un problème de 'discout' ou $\alpha < 1$. Cela implique, selon les propositions 5.4.1 du livre de Bertsekas que la fonctions J_k aura la forme suivante:

$$J_{k+1} = min_u[g(i,u) + \alpha \sum_{j=1}^{n} p_{ij}(u)J_k(j)$$

Par contre, nous avons déjà établie que les seules transitions possible était qu'on reste au même endroit ou qu'on incrémente la valeur de i. On peut traduire cela:

$$J_{k+1} = min_u[g(i,u) + \alpha(p * J_{k+1} + (1-p) * J_k)]$$

On peut facilement voir que la fonction J_i est strictement croissante selon i. Supposons alors que notre première étape (J_0) est égale à 0. Étant donné que la fonction est strictement croissante, il arrivera un point $i*$ ou les valeurs :

$$J_0 + R < g(i*, u) + \alpha(p * J(i + 1 + (1-p) * J_k$$

C'est à ce point i ou la décision sera de réparer la machine et de mettre i à 0. Évidemment, étant donné que la fonction J est strictement croissante, la décision restera la même jusqu'à temps qu'on répare la machine.

b)

Pour cette question, on peut utiliser l'algorithm d'itération de politique pour trouver une politique optimale pour le problème. Comme mentionner dans l'énoncé et dans a), la politique prendre forme:

- si $i > i*$, on répare,

- sinon on continue.

Évidemment, notre algorithme va évaluer la performance de cette politique selon la seuil. Cela peut être fait analytiquement et en simulation dépendamment des valeurs de g() et de p. Notre algorithme va essayer toutes les valeurs de seuils possible entre $\{0, 1, 2...n\}$ et va conserver la politique ayant la meilleur performance.