Chaînes de Markov avec coûts associés (tous les états persistants et apériodiques)

Coût d'une transition: $A_{i,j}$ (en moyenne) pour une transition de i à j (ne dépend pas du temps ni de l'histoire)

Coût moyen d'une transition: de l'état i:

$$a_i = \sum_{j=1}^{N} P_{i,j} A_{i,j}$$

Tendance vers la moyenne:

Le coût moyen par transition tend vers g où g est le coût moyen d'une transition dans la distribution limite Π

On peut facilement calculer g:

$$\sum_{i=1}^{N} \Pi_i a_i = g$$

Différences entre les états:
pour tout i on sait que la limite de
$cout\ des\ t\ premieres\ transitions/t$ est g mais tous les états ne sont pas égaux; pour un état i on a que la limite de
$cout\ des\ t\ premieres\ transitions - tg$ existe, disons δ_i;
comment calculer les δ_i pour savoir quels sont les bons états?

$$\delta_i \ = \ a_i + \sum_{j=1}^{n} P_{i,j}\delta_j - g$$

mais cela ne suffit pas pour déterminer les δ;

on ajoute

$$\sum_{i=1}^{N} \Pi_i \delta_i = 0$$

Décisions qui peuvent changer les probabilités de transition

La possibilité de choix: supposons que dans l'état i on peut choisir entre plusieurs *décisions* et que pour la décision d, la probabilité d'aller en état j est $P^d_{i,j}$ avec un coût $a^d_{i,j}$

on peut calculer l'espérance du coût d'une transition:

$$a^d_i = \sum_{j=1}^{N} P^d_{i,j} a^d_{i,j}$$

mais minimiser ca (la politique gloutonne) n'est pas forcément la bonne politique

Le nombre de politiques possibles peut être trop grand pour permettre de calculer le g de chacune
et peut-être ce n'est pas la politique possédant le plus petit g qu'il convient de choisir...

Actualisation des coûts

Un compromis entre la politique gloutonne et minimiser g

Actualisation: Un euro aujourd'hui est plus utile qu'un euro l'année prochaine;
Un coût d'un euro aujourd'hui est plus grave qu'un pareil coût l'année prochaine;

On veut calculer les coûts sur la supposition qu'un euro au temps $t+1$ vaut ρ euros au temps t; $0 < \rho < 1$.

Rapport avec le taux d'intérêt: si le taux réel d'intérêt (payé ou reçu) est τ, on a
$\rho = 1/(1+\tau)$.

Calcul des coûts

Les équations pour les coûts totaux: On a toujours un coût immédiat (moyen) en l'état i de a_i et une probabilité $P_{i,j}$ de transition de l'état i à j:

$$v_i = a_i + \rho \sum_{j=1}^{N} P_{i,j} v_j$$

Système non singulier: on peut calculer les v_i de ces équations parce que $\rho < 1$.

Le coût dans la distribution limite: Si le coût moyen d'une transition dans la distribution limite (sans actualisation) est g, le coût (total) si on commence dans cette distribution est $g(1 + \rho + \rho^2 + \rho^3 \cdots) = g/(1 - \rho)$. Donc $g\rho^t/(1-\rho)$ est une approximation au coût à partir de temps t (t grand et la chaîne régulière).

Optimisation des coûts

Selon notre point de vue nous voulons choisir une politique de façon à minimiser

- soit g
- soit v_i pour l'état initial i.

Existence d'algorithmes assez efficaces qui convergent sur une politique optimale.

Dans les deux cas l'algorithme consiste à commencer avec une politique arbitraire et de l'améliorer tant que l'amélioration est non nulle:
principe d'amélioration: calculer les valeurs rélatives des états (différences entre les états de l'espérance de la somme de tous les coûts de transition jusqu'à l'infini, actualisée ou non; c'est-à-dire les δ_i ou les v_i);
décider une nouvelle politique basée sur le coût de la première transition plus la valeur de l'etat après cette transition (actualisée le cas échéant.)

Calcul d'une politique optimale avec actualisation

On peut utiliser l'algorithme d'amélioration d'une politique initiale: calculer le coût total (avec actualisation) v_j à partir de chaque état j; choisir à chaque état i la décision $*d$ qui minimise $a_i^{*d} + \rho \sum_{j=1}^{N} P_{i,j}^{*d} v_j$; répéter tant qu'il y a une réduction;
(Ca équivaut à minimiser le coût qu'on aurait si on choisissait une nouvelle politique pour la première étape et gardait l'ancienne politique après, ce qui explique pourquoi il y a amélioration)

Terminaison de l'algorithme

- Réduction des coûts à chaque itération
- Un nombre fini de politiques possibles
- Donc, terminaison garantie
- En principe pourrait être très lente (nombre de politiques très élevé)
- Mais "en pratique" l'algorithme trouve la politique optimale après un nombre raisonnable d'itérations

Le cas sans actualisation

Ce cas est plus compliqué parce que les équations:

$$v_i \;=\; a_i + \rho \sum_{j=1}^{N} P_{i,j} v_j$$

ne sont pas utilisables si $\rho = 1$; (soit aucune solution, soit plusieurs)

Rappel : on a $\delta_i = a_i - g + \sum_{j=1}^{N} P_{i,j}\delta_j$

On essaie d'optimiser la politique en utilisant la même idée que dans le cas avec actualisation:

- Choisir une politique initiale
- Calculer Π, g et les δ_i
- Utiliser ces δ_i pour chercher une meilleure décision en chaque état

- Itérer tant qu'il y a amélioration

Un exemple: assurance d'un véhicule

Le coût de l'assurance d'un véhicule est 100, 200, 400 ou 600.
Après un an sans aucun accident déclaré, le prix est réduit au niveau inférieur
Après un an avec un accident déclaré, le prix est augmenté au niveau supérieur
La probabilité d'un accident grave est 5% par an
La probabilité d'un accident mineur est 15% par an et le coût de sa réparation est 500

Devrait-on déclarer un accident mineur ou payer la réparation?
(nous prenons $\rho = 0.9$ et supposons les coûts actualisés au début de l'année)

Calcul de la politique optimale

Je commence avec la politique de déclarer tout incident

$$\begin{aligned} v_1 &= 100 + 0.9(.8v_1 + .2v_2) \\ v_2 &= 200 + 0.9(.8v_1 + .2v_3) \\ v_3 &= 400 + 0.9(.8v_2 + .2v_4) \\ v_4 &= 600 + 0.9(.8v_3 + .2v_4) \end{aligned}$$

et je calcule $\mathbf{v} = [1315, 1489, 1905, 2405]$
Donc une meilleur politique pour les niveau 200 et 400 est de payer les réparations (par exemple pour v_2, $200 + 0.9(.8v_1 + .2v_3) > 275 + 0.9(.95v_1 + .05v_3)$):

$$\begin{aligned} v_1 &= 100 + 0.9(.8v_1 + .2v_2) \\ v_2 &= 275 + 0.9(.95v_1 + .05v_3) \\ v_3 &= 475 + 0.9(.95v_2 + .05v_4) \\ v_4 &= 600 + 0.9(.8v_3 + .2v_4) \end{aligned}$$

et $\mathbf{v} = [1304, 1472, 1839, 2347]$

Maintenant on voit qu'une meilleur politique à v_2 est de déclarer l'accident
(parce que maintenant
$200 + 0.9(.8v_1 + .2v_3)$
$< \quad 275 + 0.9(.95v_1 + .05v_3))$

Donc, une troisième politique:

$$\begin{aligned} v_1 &= 100 + 0.9(.8v_1 + .2v_2) \\ v_2 &= 200 + 0.9(.8v_1 + .2v_3) \\ v_3 &= 475 + 0.9(.95v_2 + .05v_4) \\ v_4 &= 600 + 0.9(.8v_3 + .2v_4) \end{aligned}$$

et $\mathbf{v} = [1299, 1465, 1833, 2341]$

la solution optimale!
Le modèle n'est pas très réaliste (tous les accidents mineurs n'ont pas le même coût) mais ce calcul peut aider à prendre de bonnes décisions, par exemple en l'état 2 déclarer un sinistre si le coût est supérieur à $\rho \times (1833 - 1299)$.