Chaînes de Markov avec coûts associés
et actualisation des coûts

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

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

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: cette fois on peut calculer les v_i de ces équations sans ajouter une $(n+1)$-ième.

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 après temps t (t grand et la chaîne régulière).

Calcul d'une politique optimale avec actualisation

On considère (enfin) le cas plus réaliste où on peut choisir entre plusieurs décisions en chaque état et on veut minimiser le coût total avec actualisation.
Encore, c'est évident que la décision optimale en chaque état ne dépend ni de l'histoire ni du temps.

Le même algorithme! On peut utiliser le même 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 $\hat{d}$ qui minimise $a_i^{\hat{d}} + \rho \sum_{j=1}^{N} P_{i,j}^{\hat{d}} v_j$; répéter tant qu'il y a une réduction; (minimiser le coût qu'on aurait si on choisissait une nouvelle politique pour la première étape et gardait l'ancienne politique après)

Cette fois c'est clair pourquoi l'algorithme réduit les coûts à chaque étape:

D'abord avec les décisions d on a:

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

On choisit $\widehat{d}$ tel que

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

(et on continue si on a $<$ pour au moins un i)

Les vrais nouveaux coûts sont donnés par:

$$\widehat{v}_i^{\widehat{d}} = a_i^{\widehat{d}} + \rho \sum_{j=1}^{N} P_{i,j}^{\widehat{d}} \widehat{v}_j^{\widehat{d}}$$

Donc:

$$v_i^d - \widehat{v}_i^{\widehat{d}} \geq \rho \sum_{j=1}^{N} P_{i,j}^{\widehat{d}} (v_j^d - \widehat{v}_j^{\widehat{d}})$$

(et on a $>$ pour au moins un i)

Considérons le i qui donne le minimum de $v_i^d - \widehat{v}_i^{\widehat{d}}$: on a

$$v_i^d - \widehat{v}_i^{\widehat{d}} \geq \rho \sum_{j=1}^{N} P_{i,j}^{\widehat{d}}(v_i^d - \widehat{v}_i^{\widehat{d}})$$

$$v_i^d - \widehat{v}_i^{\widehat{d}} \geq \rho(v_i^d - \widehat{v}_i^{\widehat{d}})$$

$$v_i^d - \widehat{v}_i^{\widehat{d}} \geq 0$$

Et cette dernière est vraie aussi pour tout i. Finalement on ne peut pas avoir $= 0$ pour tout i; et donc on a une amélioration pour tout i...

Dépendance de ρ

Comportement quand $\rho \to 1$: sans grande surprise, pour ρ suffisamment proche de 1, la politique optimale est la même que dans le calcul sans actualisation.

Peut-être plusieurs politiques optimales, chacune optimale pour plusieurs intervalles de ρ: Pour chaque politique, le coût total à partir de chaque état est approximé par un polynôme en ρ. Entre les valeurs de ρ où deux de ces polynômes sont égaux, c'est la même politique qui est optimale. Puisque deux polynômes peuvent se croiser plusieurs fois, une politique pourrait être optimale en plusieurs intervalles séparés de ρ!

Un exemple: assurance d'un véhicule

Le coût de l'assurance d'un véhicule est 1000, 2000, 4000 ou 6000.
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 réparation est 5000

Devrait-on déclarer un accident mineur ou payer la réparation?
($\rho = 0.9$)

Calcul de la politique optimale

Je commence avec la politique gloutonne

$$\begin{aligned}
v_1 &= 1000 + 0.9(.8v_1 + .2v_2) \\
v_2 &= 2000 + 0.9(.8v_1 + .2v_3) \\
v_3 &= 4000 + 0.9(.8v_2 + .2v_4) \\
v_4 &= 6000 + 0.9(.8v_3 + .2v_4)
\end{aligned}$$

et je calcule $\mathbf{v} = [13147, 14895, 19053, 24046]$ Donc une meilleur politique pour les niveau 2000 et 4000 est de payer les réparations (par exemple pour v_2, $2000 + 0.9(.8v_1 + .2v_3) > 2750 + 0.9(.95v_1 + .05v_3)$):

$$\begin{aligned}
v_1 &= 1000 + 0.9(.8v_1 + .2v_2) \\
v_2 &= 2750 + 0.9(.95v_1 + .05v_3) \\
v_3 &= 4750 + 0.9(.95v_2 + .05v_4) \\
v_4 &= 6000 + 0.9(.8v_3 + .2v_4)
\end{aligned}$$

et $\mathbf{v} = [13037, 14724, 18395, 23469]$

Maintenant on voit qu'une meilleur politique à v_2 est de déclarer l'accident (parce que maintenant
$2000 + 0.9(.8v_1 + .2v_3) < 2750 + 0.9(.95v_1 + .05v_3))$

$$\begin{aligned}
v_1 &= 1000 + 0.9(.8v_1 + .2v_2) \\
v_2 &= 2000 + 0.9(.8v_1 + .2v_3) \\
v_3 &= 4750 + 0.9(.95v_2 + .05v_4) \\
v_4 &= 6000 + 0.9(.8v_3 + .2v_4)
\end{aligned}$$

et $\mathbf{v} = [12991, 14654, 18332, 23414]$
la solution optimale!