

# IFT-6521

## PROGRAMMATION DYNAMIQUE

### Chapitre 5:

### État partiellement observé

Pierre L'Ecuyer

DIRO, Université de Montréal

Février 2015

## Information partielle et conversion au modèle avec information parfaite

Dans plusieurs situations pratiques, l'état du système n'est pas observable complètement, de sorte qu'une politique admissible ne peut pas être n'importe quelle fonction de l'état. Les décisions ne doivent dépendre que de ce qui est observable.

Comment traiter cette situation? La réponse courte: simplement remplacer (redéfinir) l'état (partiellement observable) par l'information disponible, ou encore par une fonction de l'information disponible qui nous donne autant d'information utile mais sous une forme plus agrégée (une statistique exhaustive). On se ramène alors au cadre connu, mais avec un état défini différemment. L'aggrégation des états (via une statistique exhaustive ou une autre méthode d'approximation) est souvent essentielle pour limiter la dimension de l'espace d'états, pour pouvoir résoudre.

C'est essentiellement ce que raconte la chapitre 5 du livre, avec quelques détails en plus.

## Modèle.

À l'étape  $k$ , le système est dans l'état  $x_k$ , mais on ne peut observer que

$$z_k = h_k(x_k, u_{k-1}, v_k),$$

où  $z_k \in Z_k$  et  $v_k \in V_k$  est une v.a. dont la loi

$$\mathbb{P}[v_k \in \cdot \mid x_k, \dots, x_0, u_{k-1}, \dots, u_0, w_{k-1}, \dots, w_0, v_{k-1}, \dots, v_0]$$

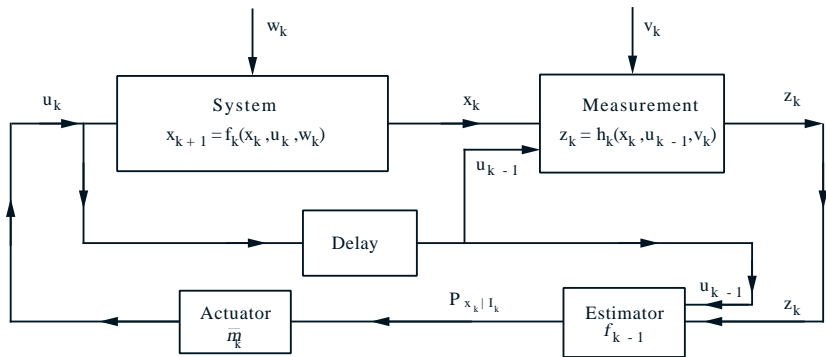
dépend de la suite des états, décisions, et aléas précédents.

L'état initial peut aussi être aléatoire, de loi  $\mathbb{P}[x_0 \in \cdot]$ .

L'information disponible à l'étape  $k$  est

$$I_k = (z_0, z_1, \dots, z_k, u_0, u_1, \dots, u_{k-1}), \quad k = 0, 1, \dots, N-1,$$

et la décision  $u_k \in U_k$  ne peut dépendre que de cette information. On suppose ici que  $U_k$  ne dépend pas de  $x_k$ . Ensuite une variable aléatoire  $w_k$  est "générée" selon une loi  $P_k(\cdot \mid x_k, u_k)$ , on doit payer un coût  $g_k(x_k, u_k, w_k)$ , et l'état à la prochaine étape est  $x_{k+1} = f_k(x_k, u_k, w_k)$ .



On cherche une **politique** admissible de la forme  $\pi = (\mu_0, \dots, \mu_{N-1})$ , où  $\mu_k(I_k) \in U_k$ , qui minimise

$$\mathbb{E} \left[ g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k) \right],$$

sous les contraintes

$$\begin{aligned} x_{k+1} &= f_k(x_k, \mu_k(I_k), w_k), \\ z_0 &= h_0(x_0, v_0), \\ z_k &= h_k(x_k, \mu_{k-1}(I_{k-1}), v_k), \quad k = 1, \dots, N-1. \end{aligned}$$

## Reformulation.

Si on remplace l'état par  $I_k$  et la fonction de coût par étape par

$$\tilde{g}_k(I_k, u_k) = \mathbb{E}[g_k(x_k, u_k, w_k) \mid I_k, u_k],$$

on se retrouve dans cadre "standard" où l'état est complètement observé.

L'équation de récurrence se réécrit alors comme

$$\begin{aligned} J_k(I_k) &= \text{coût espéré total optimal de l'étape } k \text{ à la fin,} \\ &\quad \text{si l'information disponible à l'étape } k \text{ est } I_k \\ &= \min_{u_k \in U_k} [\tilde{g}_k(I_k, u_k) + \mathbb{E}[J_{k+1}(I_{k+1})]] \end{aligned}$$

où  $z_{k+1} = h_{k+1}(x_{k+1}, u_k, v_{k+1})$  et  $I_{k+1} = (I_k, z_{k+1}, u_k)$ .

DPOC traite en détail le cas des systèmes linéaires à coût quadratique, puis examine plusieurs exemples.

# Statistique exhaustive

Une statistique exhaustive est une fonction  $S_k$  qui associe à chaque  $I_k$  une valeur  $S_k = S_k(I_k)$ , souvent plus compacte, telle que l'on peut réécrire

$$J_k(I_k) = \min_{u_k \in U_k} H_k(S_k(I_k), u_k)$$

pour une certaine fonction  $H_k$ . En d'autres mots, on peut écrire  $J_k$  et une politique optimale comme fonctions de  $S_k = S_k(I_k)$  au lieu de  $I_k$ .

Dans ce cas, on peut remplacer l'état  $I_k$  par  $S_k$ .

## Loi conditionnelle de l'état $x_k$ .

Dans le cas fréquent où

$$\begin{aligned} & \mathbb{P}[v_k \in \cdot \mid x_k, \dots, x_0, u_{k-1}, \dots, u_0, w_{k-1}, \dots, w_0, v_{k-1}, \dots, v_0] \\ = & \mathbb{P}[v_k \in \cdot \mid x_k, x_{k-1}, u_{k-1}, w_{k-1}], \end{aligned}$$

on peut prendre  $S_k = S_k(I_k) = \mathbb{P}[x_k \in \cdot \mid I_k]$ , la loi de probabilité de  $x_k$  conditionnelle à l'information connue  $I_k$ . On peut mettre à jour

$$S_{k+1} = \Phi_k(\mathbb{P}[x_k \in \cdot \mid I_k], u_k, z_{k+1}) = \Phi_k(S_k, u_k, z_{k+1})$$

pour une certaine fonction  $\Phi_k$ . Le coût par étape est remplacée par

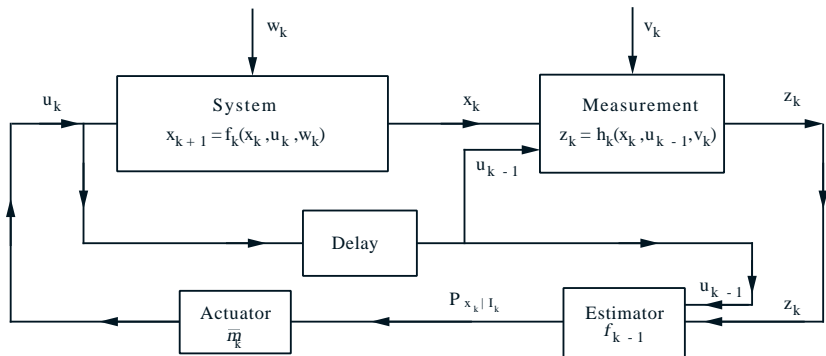
$$\tilde{g}_k(S_k, u_k) = \mathbb{E}[g_k(x_k, u_k, w_k) \mid S_k, u_k] = \int g_k(x_k, u_k, w_k) d\mathbb{P}(x_k, w_k \mid S_k, u_k).$$

et on peut alors écrire

$$J_k(S_k) = \min_{u_k \in U_k} [\tilde{g}_k(S_k, u_k) + \mathbb{E}[J_{k+1}(S_{k+1})]].$$



La commande optimale se décompose alors en deux parties: (a) estimation de la loi conditionnelle de l'état; (b) choix de la décision. En pratique, de nombreuses heuristiques (sous optimales) sont basées sur des versions approximatives de ce schéma.



## Exemple: prises de décision sous un modèle Bayésien

Une loterie bien particulière vend des billets  $C$  dollars. On pense que chaque billet permet de gagner  $V$  dollars avec probabilité  $\beta > 0$  (cas A) mais il est aussi possible que la probabilité de gagner soit de zéro à tous les tirages (cas B). Notre probabilité a priori que l'on soit dans le cas A est  $p_0 > 0$ . Dès que l'on a gagné une fois, on ne peut plus jouer.

Soit  $p_k$  la probabilité que l'on soit dans le cas A après avoir acheté  $k$  billets sans gagner. Ces  $p_k$  suivent la récurrence (formule de Bayes):

$$\begin{aligned} p_k &= \mathbb{P}[\text{cas A} \mid k \text{ échecs}] = \frac{\mathbb{P}[\text{cas A et } k \text{ échecs}]}{\mathbb{P}[k \text{ échecs}]} \\ &= \frac{\mathbb{P}[k \text{ échecs} \mid \text{cas A}]p_0}{\mathbb{P}[k \text{ échecs} \mid \text{cas A}]p_0 + \mathbb{P}[k \text{ échecs} \mid \text{cas B}](1 - p_0)} \\ &= \frac{(1 - \beta)^k p_0}{(1 - \beta)^k p_0 + 1 - p_0} = \frac{(1 - \beta)p_{k-1}}{(1 - \beta)p_{k-1} + 1 - p_{k-1}}. \end{aligned}$$

On voit que  $p_k$  est décroissant en  $k$  et  $p_k \rightarrow 0$  quand  $k \rightarrow \infty$ .

La dernière égalité est facile à vérifier.

Comme **état** à l'étape  $k$ , on peut prendre  $p_k$  (statistique exhaustive), ou même simplement  $k$ , puisque l'état n'est utile que lorsqu'on n'a pas encore gagné. Si  $J_k$  est le gain espéré optimal après  $k$  échecs, alors on a

$$J_k = \max[0, p_k \beta V - C + (1 - p_k \beta) J_{k+1}]$$

pour  $k = 0, 1, 2, \dots$

Il est clair que  $J_k \leq \max(0, p_k V - C)$ , car c'est le cas où on gagne à coup sûr au prochain coup, de sorte que  $J_k = 0$  dès que  $p_k \leq C/V$ . Si

$N = \min\{k : p_k \leq C/V\}$ , alors  $J_k = 0$  pour tout  $k \geq N$ . La récurrence implique aussi que  $J_k = 0$  tant que  $p_k \beta V - C \leq 0$ , et que  $J_k > 0$  dès que  $p_k \beta V - C > 0$ .

La **politique optimale** est donc d'acheter au maximum  $k^*$  billets, où

$$k^* = \max\{k : p_k \beta V > C\}.$$

## Exemple: modèle simplifié d'entretien d'une machine

Une machine peut être dans l'état 1 (elle fonctionne correctement) ou 0 (elle est défectueuse).

À chaque période  $k$ , l'état  $x_k \in \{0, 1\}$  n'est pas observé, mais on inspecte la machine et on observe

$z_k = G$  (semble ok) ou  $z_k = B$  (semble défectueuse).

$$\mathbb{P}[z_k = G \mid x_k = 1] = 1 - \mathbb{P}[z_k = B \mid x_k = 1] = 3/4,$$

$$\mathbb{P}[z_k = B \mid x_k = 0] = 1 - \mathbb{P}[z_k = G \mid x_k = 0] = 3/4,$$

$$\mathbb{P}[x_k = 1 \mid x_{k-1} = 1] = 1 - \mathbb{P}[x_k = 0 \mid x_{k-1} = 1] = 2/3,$$

$$\mathbb{P}[x_k = 0 \mid x_{k-1} = 0] = 1.$$

Après l'inspection, on peut arrêter la machine pour connaître son état véritable (action  $u_k = S$ ) ou encore continuer (action  $u_k = C$ ). Si on arrête et trouve la machine défectueuse, on la répare (remet dans l'état 1).

Coûts par étape:  $g(1, C) = 0$ ,  $g(0, C) = 2$ ,  $g(1, S) = g(0, S) = 1$ .

Vecteur d'information:  $I_k = (z_0, z_1, \dots, z_k, u_0, \dots, u_{k-1})$ .

$J_k(I_k)$  = coût espéré de l'étape  $k$  à la fin. On a  $J_N(I_k) = 0$  et

$$\begin{aligned} J_k(I_k) &= \min [\mathbb{P}(x_k = 1 \mid I_k)g(1, C) + \mathbb{P}(x_k = 0 \mid I_k)g(0, C) \\ &\quad + \mathbb{E}[J_{k+1}(I_k, C, z_{k+1}) \mid I_k, C], \\ &\quad \mathbb{P}(x_k = 1 \mid I_k)g(1, S) + \mathbb{P}(x_k = 0 \mid I_k)g(0, S) \\ &\quad + \mathbb{E}[J_{k+1}(I_k, S, z_{k+1}) \mid I_k, S]] \\ &= \min [2\mathbb{P}(x_k = 0 \mid I_k) + \mathbb{E}[J_{k+1}(I_k, C, z_{k+1}) \mid I_k, C], \\ &\quad 1 + \mathbb{E}[J_{k+1}(I_k, S, z_{k+1}) \mid I_k, S]]. \end{aligned}$$

Statistique exhaustive:  $p_k = \mathbb{P}(x_k = 1 \mid I_k)$ .

$$\begin{aligned} J_k(p_k) &= \min [2(1 - p_k) + \mathbb{E}[J_{k+1}(p_{k+1}) \mid p_k, C], \\ &\quad 1 + \mathbb{E}[J_{k+1}(p_{k+1}) \mid p_k, S]]. \end{aligned}$$

On peut calculer  $p_{k+1}$  à partir de  $p_k, u_k, z_{k+1}$ :

$$\begin{aligned}
 p_{k+1} &= \mathbb{P}[x_{k+1} = 1 \mid I_k, u_k, z_{k+1}] \\
 &= \frac{\mathbb{P}[z_{k+1}, x_{k+1} = 1 \mid I_k, u_k]}{\mathbb{P}[z_{k+1} \mid I_k, u_k]} \\
 &= \frac{\mathbb{P}[z_{k+1} \mid x_{k+1} = 1] \cdot \mathbb{P}[x_{k+1} = 1 \mid I_k, u_k]}{\mathbb{P}[z_{k+1} \mid I_k, u_k]}.
 \end{aligned}$$

Par exemple, si  $u_k = S$  et  $z_k = G$ :  $\mathbb{P}[z_{k+1} = G \mid x_{k+1} = 1] = 3/4$  et  $\mathbb{P}[x_{k+1} = 1 \mid I_k, u_k = S] = 2/3$ .

Si  $u_k = C$  et  $z_k = G$ :  $\mathbb{P}[z_{k+1} = G \mid x_{k+1} = 1] = 3/4$  et  $\mathbb{P}[x_{k+1} = 1 \mid I_k, u_k = C] = (2/3)p_k$ .