Département de Génie Mathématique et Modélisation INSA TOULOUSE

EFITA 2001

MDPs for constrained management of renewable ressources

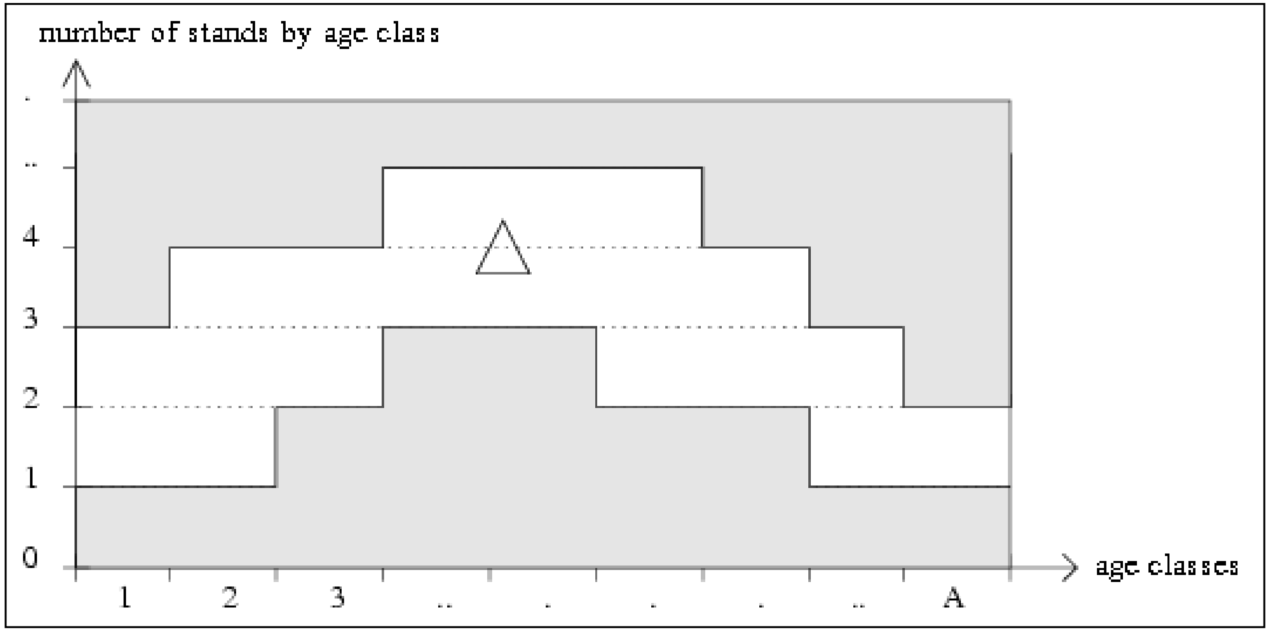
Institut National de la Recherche Agronomique

INRA BIA

SOL Fabien
INSAT GMM
5th year training period
19/06/2001

Purpose

- **Maintain bio-diversity in the forest, which goes through keeping a kind of homogeneity in the repartition of ages of the trees in the stands:**



Δ : set of satisfying age repartitions

2 approaches

- **To keep this homogeneity along the time, a first approach is to consider a classical MDP problem with a reward $r_t = 1$ if $s_t \in \Delta$ and 0 otherwise**

- **A second approach, based on solving a problem of optimization under constraints, is to use a Constrained MDP, the repartition of ages being an additional constraint to the maximisation of the timber sale revenue**

First approach: classical MDP

- **In this case, we consider at each time t a reward $r_t = 1$ if $s_t \in \Delta$ and 0 otherwise**

- **For the average reward function, our problem is then:**

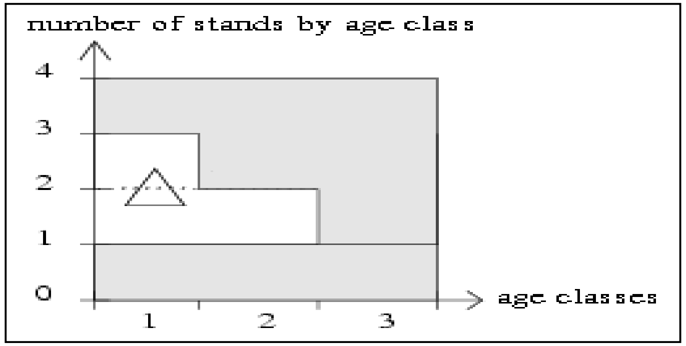
find $\pi^* \in \arg\max_{\pi} V^{\pi}$

with $V^{\pi}(s) = \begin{cases} \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} E^{\pi}[r_t \mid s_0 = s] \quad \forall s \\ V^{\pi} \quad if \quad unichain \end{cases}$

$r_t = 1$ **if** $s_t \in \Delta$ **and 0 otherwise**

Example

- For example, for a forest of 4 stands, 3 age classes, 2 prevention levels, some given fire probabilities and the following Δ :



- We find an optimal deterministic markovien policy π^* with

$$V^{\pi^*}(s) = V^* = 0.8757 \qquad \forall s \quad (unichain)$$

$$\pi^* \quad defined \quad by$$

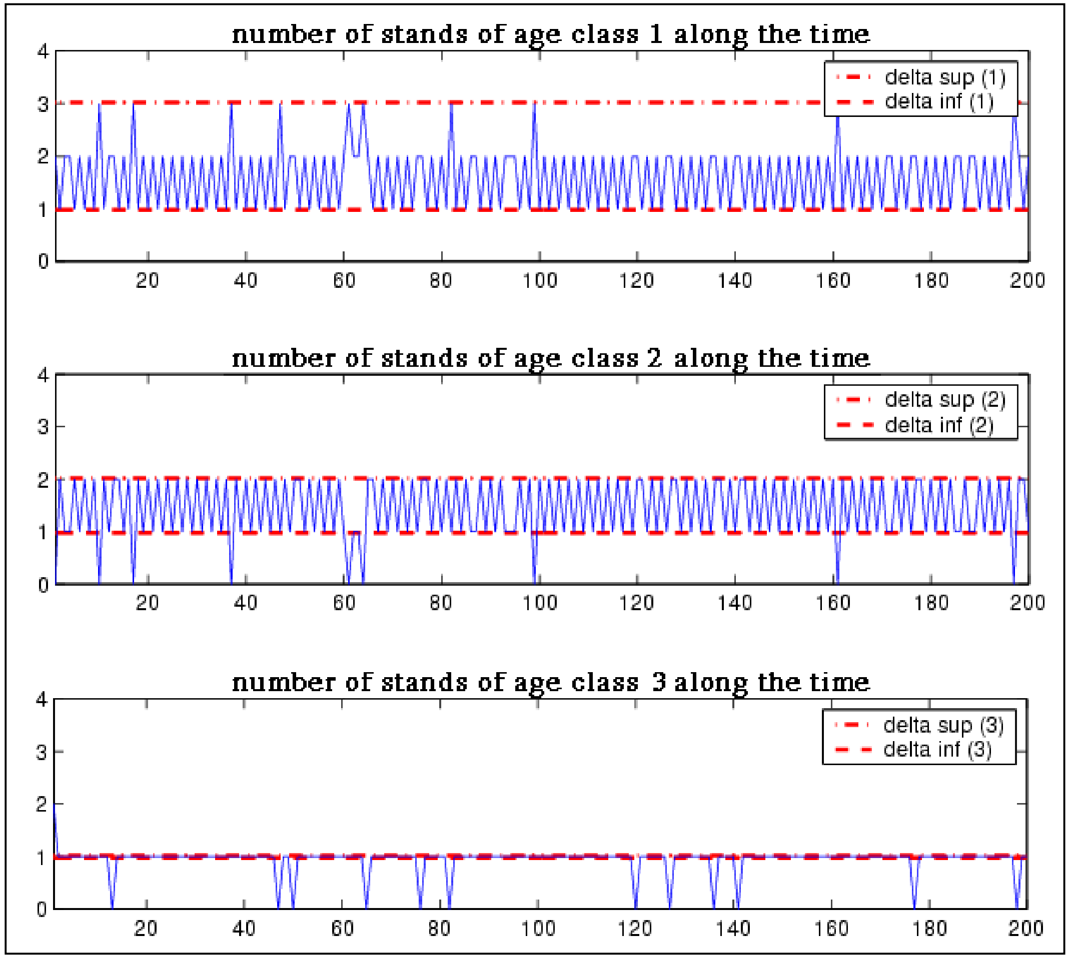
$$U^{\pi}(s) + V^{\pi} = \max_{a \in A}\left\{ r(s,a) + \sum_{s' \in S} p(s' \mid s,a) U^{\pi}(s') \right\} \forall s \in S$$

(Bellman's equation)

- So, the forest should be 87% of the time in a good repartition if we follow π^*

Example

- We can simulate the management for a given initial state to see if the optimal policy found is really effective:



- We stay in a good repartition most of the time (87% of the time)

Second approach: Constrained MDP

- **In this case, we consider at each time t a reward r_t of management of the forest (timber sales)**
- **We take into account the homogeneity of the forest as a constraint**
- **Our problem is now (also for the average reward function):**

find $\pi^* \in \arg\max_{\pi} V^{\pi}$ **subject to** $D^{\pi} \leq C$

with

$$V^{\pi}(s) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} E^{\pi}[r_t \mid s_0 = s] = V^{\pi} \quad \forall s \quad (unichain)$$

$$r_t \quad (timber \quad sales)$$

$$D^{\pi}(s) = \lim_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} E^{\pi}[d_t \mid s_0 = s] = D^{\pi} \quad \forall s \quad (unichain)$$

$d_t = 0$ if $s_t \in \Delta$ and 1 otherwise

C Constraint value

Theory

- **Transformation in a Linear Program**

- **Approach based on occupation measures:**

$$\mu : \pi \to \mu(\pi)$$

$$V^{\pi} = \sum_{s \in S} \sum_{a \in A} \mu(s,a) r(s,a)$$

$$D^{\pi} = \sum_{s \in S} \sum_{a \in A} \mu(s,a) d(s,a)$$

NB: The optimal policy is here markovien stochastic

Linear Program

find $\pi^* \in \arg\max_{\pi} V^{\pi}$ **subject to** $D^{\pi} \leq C$

⇕

find $$\max_{\mu} \sum_{s \in S} \sum_{a \in A} \mu(s,a) r(s,a)$$

subject to $$\sum_{s \in S} \sum_{a \in A} \mu(s,a) = 1, \qquad \forall s,a \quad \mu(s,a) \geq 0$$

$$\sum_{a \in A} \mu(s',a) = \sum_{s \in S} \sum_{a \in A} \mu(s,a) p(s' | s,a), \quad \forall s' \in S$$

$$\sum_{s \in S} \sum_{a \in A} \mu(s,a) d(s,a) \leq C$$

our optimal policy π^* **then chooses action a at state s with probability:** $$\frac{\mu(s,a)}{\sum_{a' \in A} \mu(s,a')}$$

Example

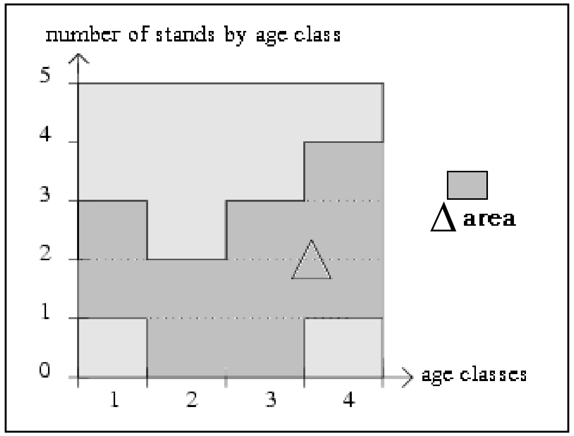
- In fact, the nature of the optimal policy depends on the value of C and Δ

- The optimal policy can be - deterministic
  - stochastic

  moreover, there can be no feasible policy for the problem

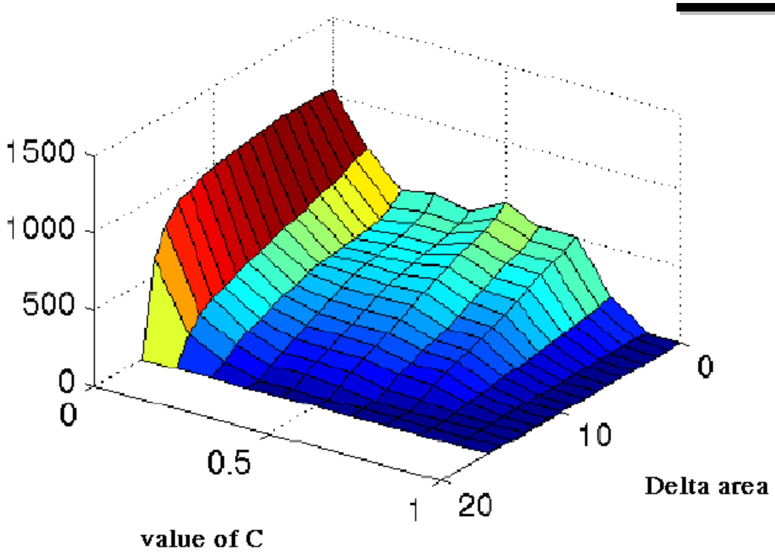
- For example, for a forest of 4 stands, 4 age classes, 2 prevention levels, some given rewards, some given fire probabilities, we can analyse the results for different values of C and Δ

- Δ area:

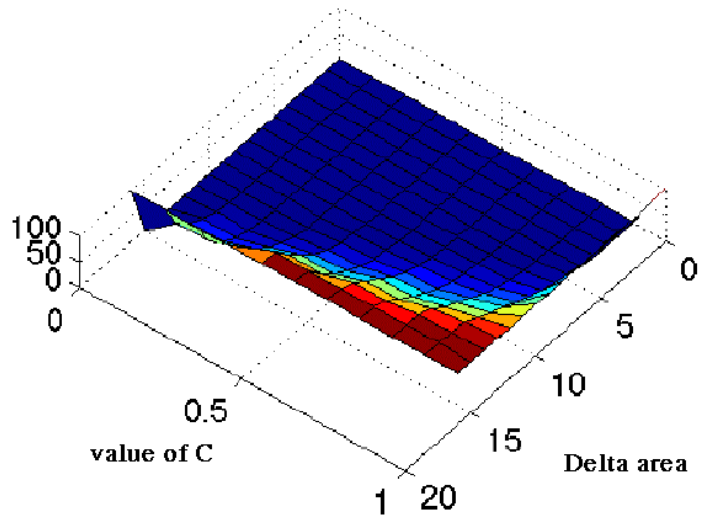


Example

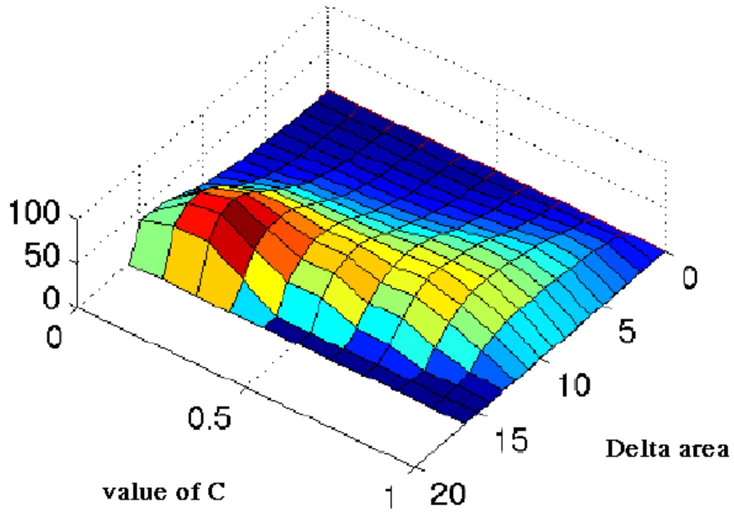
TOTAL NUMBER OF RANDOMIZATIONS



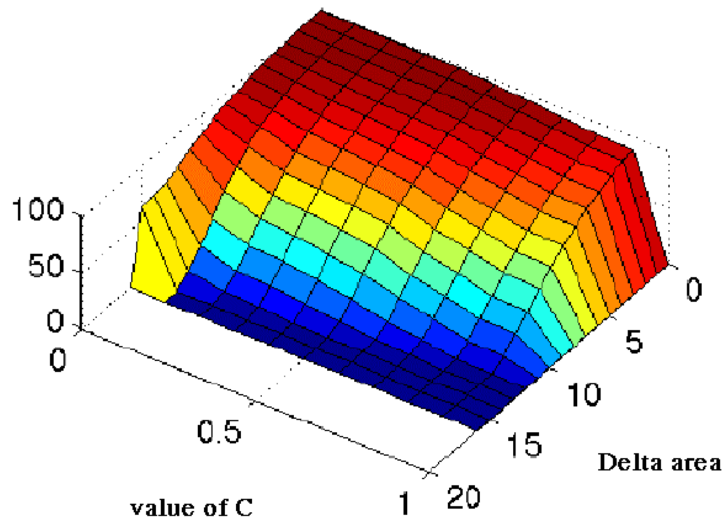
% OF DETERMINISTIC OPTIMAL POLICIES



% OF STOCHASTIC OPTIMAL POLICIES



% OF NON FEASIBILITY



Conclusion

MDP and CMDP: 2 approaches for constrained problems

- **MDP: “r=constraint (1 if s satisfying, 0 else)”**

  - easy to implement (known techniques)
  - does not take objective function optimization into account

- **CMDP: “r=timber sales & d=constraint”**

  - allows optimization while constraint satisfaction
  - may give stochastic optimal policies