FINITE STATE MARKOV CHAINS

$w_t = w_t(\omega)$ $\omega \in (\Omega, \mathcal{F}, \mathbb{P})$

$x_{t+1} = f(x_t, w_t)$ ⟶ discrete time stochastic systems

$x_0 = x$

x_t = state

w_t = exogenous input

w_t = stochastic process ⇒ x_t = stochastic process

Hp: w_t is an i.i.d sequence

x_t = random variable $\forall t$

usually, $x_t \in \mathbb{R}^n$, suppose instead that $\forall t$

$x_t \in \mathcal{X}$ a finite set w.l.o.g. $x_t \in \{1, 2, \dots, N\}$

⇒ x_t = finite state Markov chain

$x_{t+1} = f(x_t, w_t)$ induces a distribution over $\{x_t\}$ and in particular the so called transition probabilities

$$P(i \mid j) = \mathbb{P}(x_{t+1} = i \mid x_t = j) \quad \leftarrow \text{invariant}$$

transition probabilities enough to characterize the evolution of the Markov chain

thanks to

$x_{t+1} = f(x_t, w_t)$ + i.i.d ⇒ MARKOV property: "the future is independent of the past given the present"

↳ depends on x_t and w_t only

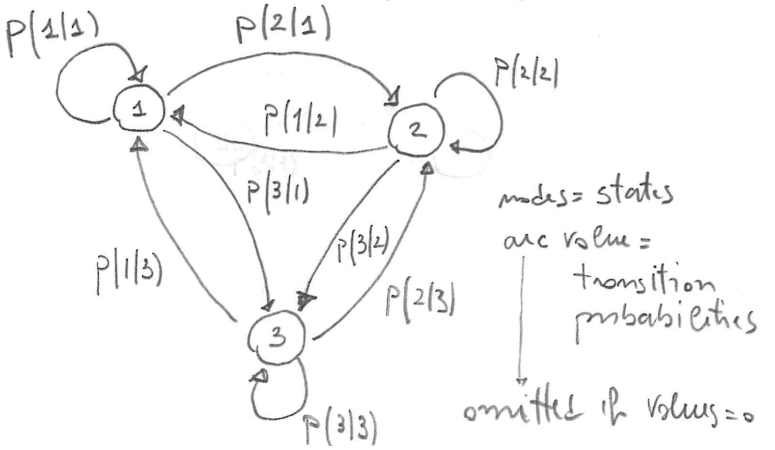
$$P(x_{t+1} = i \mid x_t = j \wedge x_{t-1} = k \wedge \dots) = P(x_{t+1} = i \mid x_t = j)$$

Clearly, $\sum_{i=1}^{N} P(i|j) = 1$ and note that $P(i|i) \neq 0$ in general

$[P(i|j)]_{N \times N}$ = transition matrix = $\mathbf{P}$ ← conveys the whole info

$[\mathbb{P}(x_t = i)]_N = \mathbf{P}^t \cdot [\mathbb{P}(x_0 = i)]$ for example

example $\mathcal{X} = \{1, 2, 3\}$

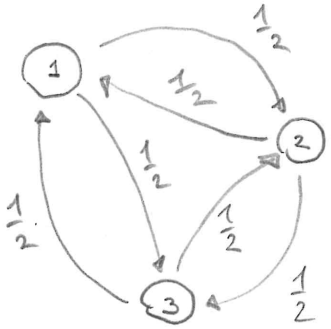


nodes = states

arc value = transition probabilities

omitted if values = 0

e.g.



≡

$$\mathbf{P} = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

$x_{t+1} = f(x_t, u_t, W_t)$ $\quad W_t$ = i.i.d. sequence

↳ control action

(finite state/input) Markov Decision Process - MDP (controlled Markov chain)

$\forall t, \quad x_t \in X = \{1, \dots, n\} = \{1_x, 2_x, \dots, n_x\}$

$u_t \in U = \{1, \dots, m\} = \{1_u, 2_u, \dots, m_u\}$

transition probabilities

$$p(i|j,u) = \mathbb{P}\{x_{t+1} = i \mid x_t = j \wedge u_t = u\}$$

$i, j \in X \quad u \in U$

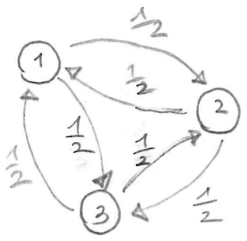
MARKOV property ↓

$$= \mathbb{P}\{x_{t+1} = i \mid x_t = j \wedge u_t = u \wedge x_{t-1} = h \wedge u_{t-1} = k \wedge \dots\}$$

⇝ m transition matrices $P_u = [p(i|j,u)]_{n \times n}$

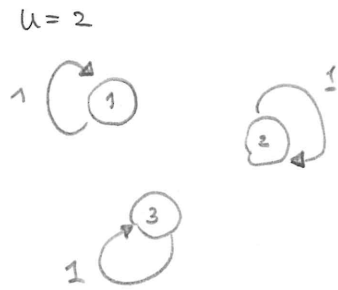
e.g. $X = \{1,2,3\}$ $\quad U = \{1,2\}$

$u = 1$



$$P_1 = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

$u = 2$



$$P_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

u_t to be chosen so as to impose a desired behavior ⇝ optimal control problem

every transition from x_t to x_{t+1} generates a (stage) cost

$g(x_t, u_t, W_t)$ ⇝ e.g. $g(x_{t+1}) = g(f(x_t, u_t, W_t))$ but here more general to represent various situations

use the degree of freedom so as to minimize a "total" (average) cost ↳ to be precisely defined later

Example inventory control

x_t = site of inventory at the end of day t (n° of items)

u_t = no. of items ordered at the end of day t

W_t = demand on day $t+1$

$x_t \in \{0, 1, \dots, M\}$ M = max no. of items that can be stored

$u_t \in \{0, 1, \dots, M\}$ → to fill up inventory when $x_t = 0$

$w_t \in \mathbb{Z}$

$$x_{t+1} = \min\left(\max(x_t + u_t, M) - w_t, 0\right)$$

$$g(x_t, u_t, w_t) = K \cdot \mathbb{1}_{u_t > 0} + c \cdot u_t + h \cdot x_t - p \cdot \min\left(\max(x_t + u_t, M), w_t\right)$$

- K: fixed cost for placing an order
- $c \cdot u_t$: transportation costs
- $h \cdot x_t$: inventory maintainance costs
- p: price
- $\min(\max(x_t + u_t, M), w_t)$: actually sold items
- $p \cdot \min(\dots)$: income

risk measure = expected value

expected stage cost : $\mathbb{E}\left[g(x_t, u_t, w_t)\right]$

we consider the evolution of the MDP up to time T

▸ finite horizon T ← terminal cost

$$J\left(x_0, \{u_t\}_{t=0}^{T-1}\right) = \sum_{t=0}^{T-1} \mathbb{E}\left[g(x_t, u_t, w_t)\right] + \mathbb{E}\left[g_T(x_T)\right]$$

t-varying

OBS : for the finite horizon problem we may admit $f = f_t$ $g = g_t$ results are unchanged

⇝ to obtain good solutions, u_t must be stochastic and adapted to x_t

Hp: x_t is observable and u_t is chosen as a function of x_t by means of a policy $\pi = \{\mu_0, \mu_1, \dots, \mu_{T-1}\}$ where each $\mu_t : X \to U$. ← motivated by the Markov property

In other words, $u_t = \mu_t(x_t)$ (state feedback) so that

$$\begin{cases} x_{t+1} = f(x_t, \mu_t(x_t), w_t) \\ x_0 = x \end{cases}$$

$x \in X$ initial condition

and $J\left(x, \{\mu_t(x_t)\}_{t=0}^{T-1}\right) =$

$$= \sum_{t=0}^{T-1} \mathbb{E}\left[g(x_t, \mu_t(x_t), w_t)\right] + \mathbb{E}\left[g_T(x_T)\right]$$

Goal: $\min_\pi V^\pi(x) = V^\pi(x)$ ← value function associated to policy π

well posed $\forall x \in X$ (finite ns. sh cases)

optimal value function $V^*(x) = \min_\pi V^\pi(x)$

↓

as $x_0 = x$ deterministic, only $\mu_0(x)$ counts, while, as for $\mu_1, \dots, \mu_{T-1}$ potentially different x could lead to different π

⇒ there exists a unique policy $\pi^* = \{\mu_0^*, \mu_1^*, \dots, \mu_{T-1}^*\}$ such that

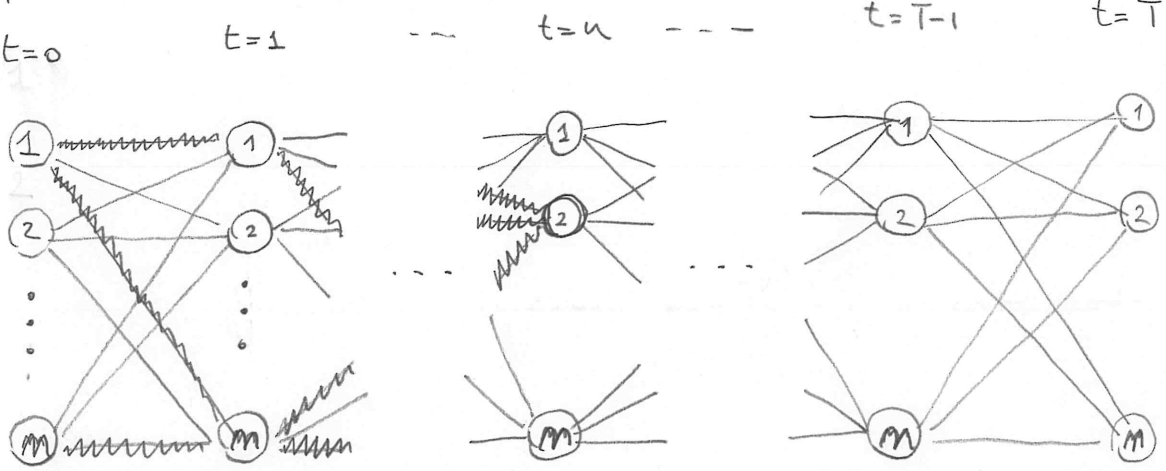
$V^{\pi^*}(x) = V^*(x) \quad \forall x \in X$ ↳ simultaneously optimal for all initial conditions

⇒ consequence of the Bellman optimality principle : given any $x \in X$, let $\pi^* = \{\mu_0^*, \mu_1^*, \dots, \mu_{T-1}^*\}$ optimal for problem $\min_\pi V^\pi(x)$ and assume that using π^*, $x_k = x'$ arises. Consider the sub-problem where

$x_k = x'$ and we wish to minimize the "cost-to-go" from $t=k$ to

$t = T$: $V_k^{\pi_k}(x') = \sum_{t=k}^{T-1} \mathbb{E}\left[g(x_t, \mu_t(x_t), w_t)\right] + \mathbb{E}\left[g_T(x_T)\right]$ where $x_k = x'$

and $\pi_k = \{\mu_k, \dots, \mu_{T-1}\}$. Then $\pi_k^* = \{\mu_k^*, \dots, \mu_{T-1}^*\}$ is optimal for this sub-problem



once at $t=k$ I'm arrived at $x_k = 2$, the future evolution doesn't depend on the past (Markov), if I'm not optimal we can switch to an optimal policy and simultaneously improve all the situations

⇒ Dynamic Programming algorithm : break down multistage optimization into single stage optimizations; generate backwards μ_{T-1}^*, solving the cost-to-go from T-1 to T subproblem for all $x \in X$, $x_{T-1} = x$, then μ_{T-2}^*, solving the cost-to-go from T-2 to T subproblem where μ_{T-1}^* is already implemented for all $x \in X$ and $x_{T-2} = x$, and so forth and so on till μ_0^*

init : $V_T^*(x) = g_T(x) \quad \forall x \in X$

for $t = T-1, \dots, 0$ compute $\forall x \in X$

$$V_t^*(x) = \min_{u \in U} \mathbb{E}\left[g(x,u,w_t) + V_{t+1}^*\left(f(x,u,w_t)\right)\right] \qquad (\star)$$

$$\mu_t^*(x) = \operatorname*{argmin}_{u \in U} \mathbb{E}\left[g(x,u,w_t) + V_{t+1}^*\left(f(x,u,w_t)\right)\right]$$

end

then $\pi^* = \{\mu_0^*, \dots, \mu_{T-1}^*\}$ is optimal for $\min_{\pi} V^{\pi}(x) \ \forall x \in X$ and

$V_0^*(x) = V^*(x)$. $\quad = T-1, \dots, 0$

⇝ FORMAL PROOF: it is enough to show that $\forall k$ and $\forall x \in X$

$$V_k^*(x) = \min_{\pi_k} V_k^{\pi_k}(x) = \min_{\pi_k} \sum_{t=k}^{T-1} \mathbb{E}\Big[g\big(\underbrace{x_t}_{x_k = x}, \mu_t(x_t), w_t\big)\Big] + \mathbb{E}\big[g_T(x_T)\big]$$

as defined in (★)

and that $\pi_k^* = \{\mu_k^*, \dots, \mu_{T-1}^*\}$ as defined in (★) is optimal for $\min_{\pi_k} V_k^{\pi_k}(x)$. This can be done by induction

$$\min_{\pi_{T-1}} V_{T-1}^{\pi_{T-1}}(x) = \min_{\mu_{T-1}} \mathbb{E}\Big[g\big(x, \mu_{T-1}(x), w_{T-1}\big] + \mathbb{E}\Big[g_T\big(f(x, \mu_{T-1}(x), w_{T-1})\big)\Big] =$$

$$= \min_{\mu_{T-1}} \mathbb{E}\Big[g\big(x, \mu_{T-1}(x), w_{T-1}\big) + V_T^*\big(f(x, \mu_{T-1}(x), w_{T-1})\big)\Big] =$$

$$= \min_{u \in U} \mathbb{E}\Big[g(x, u, w_{T-1}) + V_T^*\big(f(x, \mu_{T-1}(x), w_{T-1})\big)\Big]$$

For $x \in X$, we have n distinct opt. problems

Clearly, $\mu_{T-1}^*(x) = \underset{u \in U}{\operatorname{argmin}} \dots$ defines an optimal policy π_{T-1}^*

Suppose, property is true for $t = k+1$

$$\min_{\pi_k} V_k^{\pi_k}(x) = \min_{\mu_k, \pi_{k+1}} \sum_{t=k}^{T-1} \mathbb{E}\Big[g\big(\underbrace{x_t}_{x_k = x}, \mu_t(x_t), w_t\big)\Big] + \mathbb{E}\big[g_T(x_T)\big] =$$

$$= \min_{\mu_k, \pi_{k+1}} \mathbb{E}\Big[g\big(x, \mu_k(x), w_k\big)\Big] + \sum_{t=k+1}^{T-1} \mathbb{E}\Big[g\big(\underbrace{x_t}_{x_{k+1} = f(x, \mu_k(x), w_k)}, \mu_t(x_t), w_t\big) + \mathbb{E}\big[g_T(x_T)\big] =$$

$$= \min_{\mu_k, \pi_{k+1}} \mathbb{E}\Big[g\big(x, \mu_k(x), w_k\big)\Big] + \mathbb{E}\Big[V_{k+1}^{\pi_{k+1}}\big(f(x, \mu_k(x), w_k)\big)\Big]$$

$$= \min_{\mu_k, \pi_{k+1}} \mathbb{E}\Big[g\big(x, \mu_k(x), w_k\big) + V_{k+1}^{\pi_{k+1}}\big(f(x, \mu_k(x), w_k)\big)\Big]$$

$$\geq \min_{\mu_k} \mathbb{E}\Big[g\big(x, \mu_k(x), w_k\big) + \min_{\pi_{k+1}} V_{k+1}^{\pi+1}\big(f(x, \mu_k(x), w_k)\big)\Big]$$

$$= \min_{\mu_k} \mathbb{E}\Big[g\big(x, \mu_k(x), w_k\big) + V_{k+1}^*\big(f(x, \mu_k(x), w_k)\big)\Big] \rightarrow n \text{ distinct problems}$$

$$= \min_{u \in U} \mathbb{E}\Big[g(x, u, w_k) + V_{k+1}^*\big(f(x, \mu_k(x), w_k)\big)\Big] \rightarrow \text{argmin defines a feedback } \mu_k^*$$

$$= \min_{u \in \mathcal{U}} \mathbb{E}\left[g(x, u, w_n) + V_{n+1}^{\pi^*_{n+1}}\left(f(x, u, w_n)\right)\right] =$$

$$= \mathbb{E}\left[g(x, \mu_n^*(x), w_n) + V_{n+1}^{\pi^*_{n+1}}\left(f(x, \mu_n^*(x), w_n)\right)\right] = V_n^{\pi_n^*}(x)$$

altogether gives a policy π_n^*

$$\min_{\pi_n} V_n^{\pi_n}(x) \geq V_n^*(x) = V_n^{\pi_n^*}(x)$$

must be an equality