2. MARKOV ~~DECISION~~ PROCESSES

THIS IS A VERY BREIF INTRODUCTION TO MARKOV CHAINS.

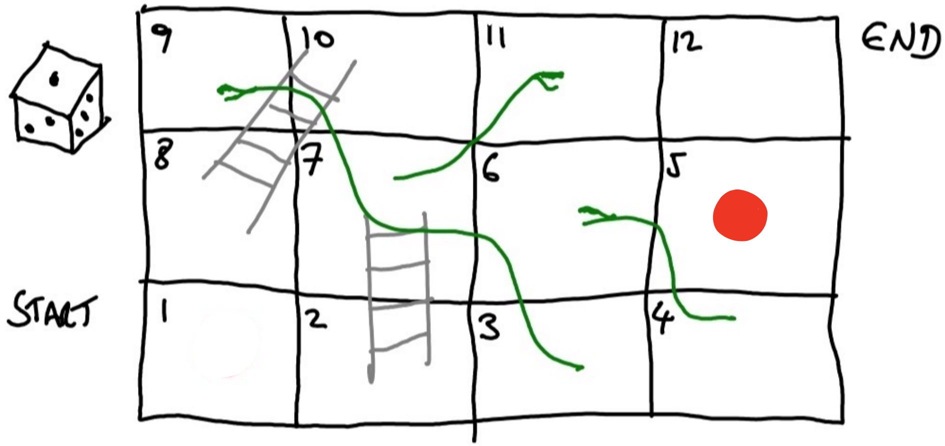
REFERENCES

J.R. Norris. *Markov Chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

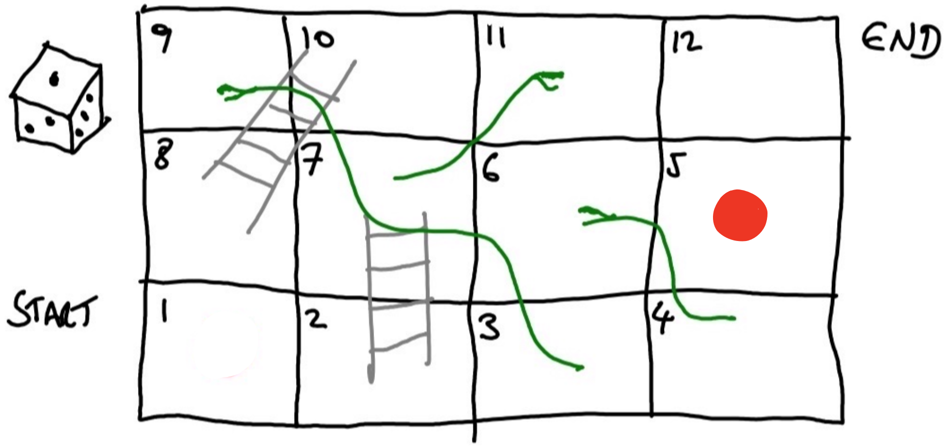
P Bremaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Texts in Applied Mathematics. Springer New York, 2013.

D A Levin, Y Peres, and E L Wilmer. *Markov Chains and Mixing Times*. American Mathematical Soc.

EXAMPLE: (SNAKES & LADDERS)



EXAMPLE: (SNAKES & LADDERS)



TWO OBSERVATIONS:

1. TO GET THE NEXT STATE NEED: CURRENT STATE & RANDOM DICE THROW

2. HOW YOU GET TO THE CURRENT STATE DOES NOT INFLUENCE THE FUTURE OF THE GAME.

FUTURE IS INDEPENDENT OF THE PAST GIVEN THE PRESENT.

ABSTRACT DEFINITION:

$x \in \mathcal{X}$ - STATES

P_{xy} - TRANSITION MATRIX $\qquad \sum_y P_{xy} = 1.$

(PROBABILITY OF GOING TO y FROM x)

λ_x - INITIAL DISTRIBUTION $\qquad \sum_x \lambda_x = 1$

(PROBABILITY OF STARTING IN $x \in \mathcal{X}$)

THE MARKOV PROPERTY

IF

$$\mathbb{P}(\underbrace{X_{t+1} = x_{t+1}}_{\text{FUTURE}} \mid \underbrace{X_t = x_t}_{\text{PRESENT}}, \underbrace{\ldots, X_0 = x_0}_{\text{PAST}}) = \mathbb{P}(\underbrace{X_{t+1} = x_{t+1}}_{\text{FUTURE}} \mid \underbrace{X_t = x_t}_{\text{PRESENT}})$$

$$= P_{x_t x_{t+1}}$$

THEN THE PROCESS $(X_t : t \in \mathbb{Z}_+)$ IS A MARKOV CHAIN.

[THE MARKOV PROPERTY SHOWS THE FUTURE IS INDEPENDENT OF THE PAST GIVEN THE PRESENT]

PROBABILITIES & EXPECTATIONS ARE MATRIX CALCULATIONS:

1) $\mathbb{P}(X_0 = x_0, X_1 = x_1, X_2 = x_2)$

[BAYES RULE]

$$= \underbrace{\mathbb{P}(X_0 = x_0)}_{\lambda_{x_0}} \underbrace{\mathbb{P}(X_1 = x_1 \mid X_0 = x_0)}_{P_{x_0 x_1}} \underbrace{\mathbb{P}(X_2 = x_2 \mid X_1 = x_1, \cancel{X_0 = x_0})}_{P_{x_1 x_2}}$$

MARKOV PROPERTY

$$= \lambda_{x_0} P_{x_0 x_1} P_{x_1 x_2}$$

← MULTIPLYING COMPONENTS

2) $$\mathbb{P}(X_2 = x_2) = \sum_{x_1, x_0} \mathbb{P}(X_2 = x_2, X_1 = x_1, X_0 = x_0)$$

$$= \sum_{x_0, x_1} \lambda_{x_0} P_{x_0 x_1} P_{x_1 x_2} = [\lambda P^2]_{x_2}$$

MULTIPLYING VECTORS

PROBABILITIES & EXPECTATIONS ARE MATRIX CALCULATIONS:

3) $$\mathbb{E}[r(X_2)] = \sum_{x_2} \mathbb{P}(X_2 = x_2)\, r(x_2)$$

$$= \sum_{x_2} [\lambda P^2]_{x_2}\, r(x_2) = \lambda P^2 r$$

NOTATION: $\mathbb{E}_x[r(X_n)] := \mathbb{E}[r(X_n) \mid X_0 = x] =: P^n r(x)$

MEANS x-th COMPONENT OF $P^n r$

4) $$\mathbb{E}_x\left[\sum_{t=0}^{\infty} \beta^t r(X_t)\right] = \sum_{t=0}^{\infty} \beta^t\, \mathbb{E}_x[r(X_t)]$$

$$= \sum_{t=0}^{\infty} \beta^t P^t r(x)$$

ALL YOU NEED IS A DICE!

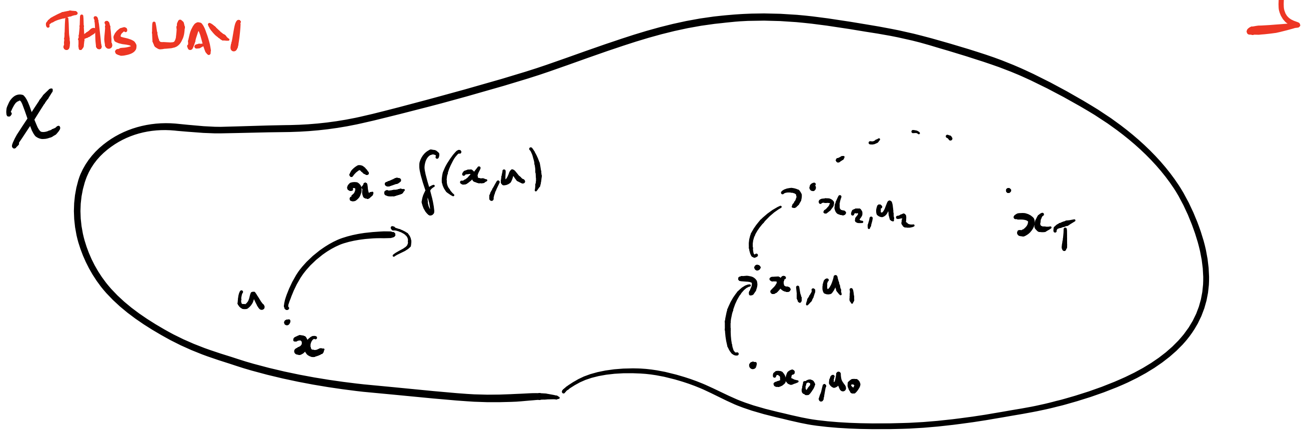
$U_0, U_1, \ldots$ – UNIFORM $[0,1]$ R.V.s.

$f(x,u)$ – A FUNCTION $f : \mathcal{X} \times [0,1] \to \mathcal{X}$

THEN $\quad X_{t+1} = f(X_t, U) \quad , \quad t = 0,1,2,\ldots$

DEFINES A MARKOV CHAIN.

[IN FACT ALL MARKOV CHAINS CAN BE CONSTRUCTED THIS WAY]



SUMMING REWARDS

IF $$R_T(x) = \mathbb{E}_x\left[\sum_{t=0}^{T} r(X_t)\right]$$

THEN $$R_T(x) = r(x) + \mathbb{E}_x\left[R_{T-1}(X_1)\right]$$

PROOF

$$R_T(x) = \mathbb{E}_x\left[r(x) + \sum_{t=1}^{T} r(X_t)\right]$$

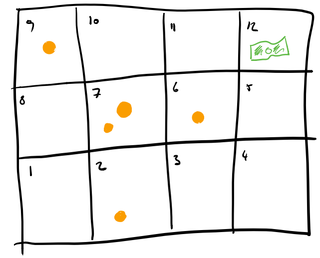
$$= r(x) + \mathbb{E}_x\Big[\underbrace{\mathbb{E}\Big[\sum_{t=1}^{T} r(X_t) \,\Big|\, X_1, \cancel{X_0 = x}\Big]}_{R_{T-1}(X_1)}\Big]$$

MARKOV

$$= r(x) + \mathbb{E}_x\left[R_{T-1}(X_1)\right]$$

$\square$

SUMMING REWARDS.



DEFINE

$$R(x) = \mathbb{E}_x \left[\sum_{t=0}^{\infty} \beta^t r(X_t) \right]$$

[CONDITIONAL ON $X_0 = x$] [$\in [0,1]$] [$r : X \to \mathbb{R}_+$]

THEN

$$R(x) = r(x) + \beta P R(x) = r(x) + \beta \, \mathbb{E}_x [R(\hat{x})] \quad (+)$$

THINK BELLMAN EQUATION

$$\underline{R} = \underline{r} + \beta P \underline{R} \Rightarrow (I - \beta P)\underline{R} = r \Rightarrow \underline{R} = (I - \beta P)^{-1} r$$

OR EQUIVALENTLY

$$R = (I - \beta P)^{-1} r \qquad (++)$$

MOREOVER, THE SOLUTION TO (+) IS UNIQUE.

ALSO IF $\tilde{R}(x) \geq r(x) + \beta P \tilde{R}(x)$ THEN $\tilde{R}(x) \geq R(x)$

SUMMING REWARDS. (PART 1. R SOLVES $R = r + \beta P R$)

PROOF.

$$R(x) = \mathbb{E}_x\left[\sum_{t=0}^{\infty} \beta^t r(X_t)\right]$$

$$= \mathbb{E}_x\left[r(x) + \underbrace{\mathbb{E}\left[\beta \sum_{t=1}^{\infty} \beta^{t-1} r(X_t) \,\middle|\, X_1\right]}_{= \beta R(X_1)} \right]$$

$$= r(x) + \beta\, \mathbb{E}_x R(X_1)$$

$$= r(x) + \beta P R(x) \qquad \square$$

[ROLLING UP !]

SUMMING REWARDS. (PART 1. R SOLVES $R = r + \beta P R$)

PROOF. (ALTERNATIVE)

$$R(x) = \mathbb{E}_x\left[\sum_{t=0}^{\infty} \beta^t r(X_t)\right] = \sum_{t=0}^{\infty} \beta^t P^t r(x)$$

GEOMETRIC SERIES.

i.e.

VECTOR

$$R = r + \cancel{\beta P r} + \cancel{\beta^2 P^2 r} + \cancel{\ldots}$$

$$\therefore \quad \beta P R = \cancel{\beta P r} + \cancel{\beta^2 P^2 r^2} + \cancel{\beta^3 P^3 r} + \ldots$$

$$\therefore \quad R - \beta P R = r \qquad \therefore \quad R = r - \beta R r \qquad (\dagger) \text{ AS REQUIRED}$$

OR ALTERATIVELY

$$R = (I - \beta P)^{-1} r \qquad (\ddagger) \text{ AS REQUIRED}$$

$\square$

SUMMING REWARDS. (PART 2, CONTRACTIONS)

$$R(x) \longmapsto \hat{R}(x) \doteq r(x) + \mathbb{E}_x[R(X_1)]$$

WITH RESPECT TO

IS A CONTRACTION W.R.T. UNIFORM NORM

UNIFORM NORM – $\|R\|_\infty \doteq \max_x |R(x)|$

CONTRACTION – $\|\hat{R}_1 - \hat{R}_2\|_\infty \leq \beta \|R_1 - R_2\|_\infty$

PROOF:

$$\max_x |\hat{R}_1(x) - \hat{R}_2(x)| = \max_x |\beta \mathbb{E}_x[R_1(X_1)] - \beta \mathbb{E}_x[R_2(X_1)]|$$

$$= \beta \max_x \Big|\mathbb{E}\big[\underbrace{R_1(X_1) - R_2(X_1)}_{\leq \max_y |R_1(y) - R_2(y)|}\big]\Big| \leq \beta \max_y |R_1(y) - R_2(y)|$$

$\square$

<u>SUMMING REWARDS</u>. (PART 3, $R = r + \beta PR$ IS UNIQUE)

FOR $\beta \in (0,1)$,

$$R(x) = \mathbb{E}_x\left[\sum_{t=0}^{\infty} \beta^t r(X_t)\right]$$ IS THE UNIQUE SOLUTION TO

$$R(x) = \underbrace{r(x) + PR(x)}_{\hat{R}(x)}$$

<u>PROOF</u>: TAKE ANOTHER SOLUTION $R_1(x)$, SINCE A CONTRACTION

$$\| R - R_1 \|_\infty \leq \beta \| \hat{R} - \hat{R}_1 \|_\infty = \beta \| R - R_1 \|_\infty$$

BUT $\beta < 1$ $\therefore$ ONLY SOLUTION IS $\| R - R_1 \|_\infty = 0$

$$\therefore \quad R(x) = R_1(x) \quad \forall x$$

$\square$

SUMMING REWARDS. (PART 4, MONOTONE).

FOR $\beta \in [0,1]$, IF $R_1(x) \geq r(x) + \beta P R_1(x)$

THEN
$$R_1(x) \geq R(x) = \mathbb{E}_x\left[\sum_{t=0}^{\infty} \beta^t r(X_t)\right]$$

PROOF:

$$R_1(x) \geq r(x) + \beta\, \mathbb{E}_x\left[R_1(X_1)\right]$$
$$\geq r(x) + \beta\, \mathbb{E}_x\left[r(X_1) + \beta\, \mathbb{E}\left[R_1(X_2) \mid X_1\right]\right]$$
$$= \mathbb{E}\left[r(X_0) + \beta r(X_1) + \beta^2 R(X_2)\right]$$
$$\vdots$$
$$= \mathbb{E}\left[r(X_0) + \beta r(X_1) + \beta^2 r(X_2) + \dots + \beta^{T-1} r(X_{T-1})\right] + \beta^T \underbrace{\mathbb{E}\left[R_1(X_T)\right]}_{\geq 0}$$
$$\xrightarrow{T \to \infty}$$
$$\geq \mathbb{E}_x\left[\sum_{t=0}^{\infty} \beta^t r(X_t)\right] = R(x) \quad \leftarrow \text{[ROLLING OUT]} \quad \square$$

MARKOV CHAINS & MARTINGALES:

[MARTINGALES – SEE APPENDIX TO NOTES
OR, BETTER STILL, SEE WILLIAMS "PROBABILITY WITH MARTINGALES" CUP.]

GIVEN $R(x)$ & $r(x)$ BOUNDED, LET

$$M_t = R(X_0) - \beta^t R(X_t) - \sum_{s=0}^{t-1} \beta^t r(X_t), \quad t \in \mathbb{Z}_+$$

THEN

$$R(x) = \mathbb{E}_x\left[\sum_{t=0}^{\infty} \beta^t r(X_t)\right] \quad \text{IFF} \quad M_t \text{ IS A MARTINGALE.}$$

PROOF:

$\Rightarrow$) WE KNOW $\underbrace{R(x) = r(x) + \beta \mathbb{E}_x[R(\hat{x})]}_{\text{PREVIOUS SLIDE}}$.

$$\mathbb{E}[M_{t+1} - M_t \mid X_1, \dots, X_t] = \mathbb{E}\left[\beta^t R(X_t) - \beta^{t+1} R(X_{t+1}) - \beta^t r(X_t) \mid X_t\right]$$

$$= \beta^t \left[R(X_t) - \beta \mathbb{E}_{X_t}[R(X_{t+1})] - r(X_t)\right] = 0$$

SO M_t IS A MARTINGALE

PROOF (CONTINUED).

$(\Leftarrow)$

$$0 = \mathbb{E}_x[M_t] = R(x) - \underbrace{\beta^t \mathbb{E}[R(X_t)]}_{\to 0} - \underbrace{\mathbb{E}\left[\sum_{t=0}^{T-1} \beta^t r(X_t)\right]}_{\to \mathbb{E}\left[\sum_{t=0}^{\infty} \beta^t r(X_t)\right]}$$

So $R(x) = \mathbb{E}\left[\sum_t \beta^t r(X_t)\right]$

$\square$

SUMMARY:

- MARKOV PROPERTY: CAN FORGET THE PAST

$$\mathbb{P}(X_{t+1}=x_{t+1} \mid X_t = x_t, \underbrace{X_{t-1}=x_{t-1}, \dots, X_0 = x_0}_{\text{PAST}}) = \mathbb{P}(X_{t+1}=x_{t+1} \mid X_t = x_t)$$

- PROBABILITIES & EXPECTATIONS ARE VECTOR MATRIX CALCULATIONS

$$\mathbb{E}_x[r(X_n)] = P^n r(x)$$

INDEPENDENT R.V.

- SIMULATION IS DICE THROWING: $X_{n+1} = f(X_n, U_n)$

- SUMMING REWARDS: $R(x) = \mathbb{E}_x\left[\sum_{t=0}^{\infty} \beta^t r(X_t)\right] \iff R(x) = \hat{R}(x) := r(x) + \mathbb{E}_x[R(\hat{x})]$

β - CONTRACTION