Master Equation & HJB Control in Stochastic Games

Snakes & Ladders with Optimal Re-roll Policy

Hooman Zare

SUT

August 1, 2025

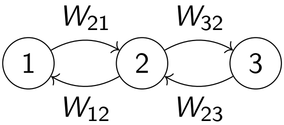
What is the Master Equation?

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = \sum_{\mathbf{x}'} \left[W(\mathbf{x}|\mathbf{x}') P(\mathbf{x}', t) - W(\mathbf{x}'|\mathbf{x}) P(\mathbf{x}, t) \right]$$

- Probability conservation equation
- Governs Markov processes
- Balance of probability flows

Discrete-State Formulation

$$\frac{d\mathbf{p}}{dt} = \mathbf{W}\mathbf{p}$$



- $\mathbf{p}$: Probability vector
- $\mathbf{W}$: Transition matrix
- $W_{ij} \geq 0$ for $i \neq j$
- $\sum_j W_{ij} = 0$

Spectral Properties

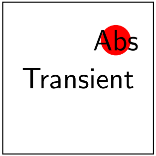
$$\mathbf{W} = \sum_k \lambda_k \mathbf{v}_k \mathbf{u}_k^T$$

- $\lambda_0 = 0$: Steady state
- $\mathrm{Re}(\lambda_k) < 0$: Relaxation modes
- $\tau_k = |\mathrm{Re}(\lambda_k)|^{-1}$: Characteristic times

$$\mathbf{p}(t) = \mathbf{p}_{\mathrm{ss}} + \sum_{k>0} c_k e^{\lambda_k t} \mathbf{v}_k$$

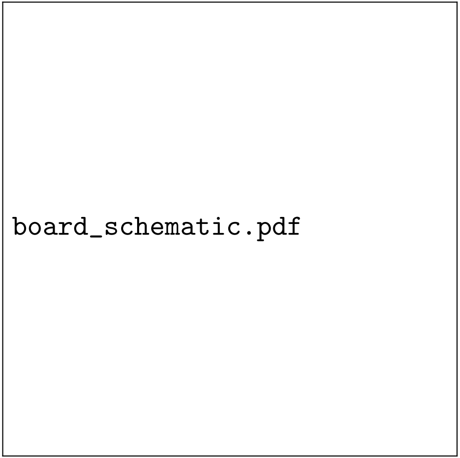
Absorption Problems

$$\frac{dp_i}{dt} = \sum_j W_{ij} p_j - \kappa_i p_i$$



- Absorbing states: $\kappa_i = 0$
- Mean absorption time: τ_i
- Fundamental matrix solution

Game as Markov Chain



- States: $s \in \{1, \dots, 100\}$
- Transitions: $s \to s + d$
- Snakes/ladders: $s \to T(s)$

Transition Matrix Construction

$$P_{ij} = \frac{1}{6} \sum_{d=1}^{6} \delta_{j, T(\min(i+d, 100))}$$

Algorithm

For each state s:

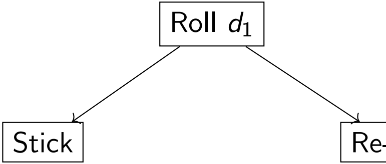
1. For $d = 1$ to 6:
2. $s' \leftarrow \min(s + d, 100)$
3. $s'' \leftarrow T(s')$
4. $P_{s,s''} \leftarrow P_{s,s''} + \frac{1}{6}$

Master Equation for Expected Moves

$$\mathbf{v} = (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{1}$$

- v_i: Expected moves from state i
- $\mathbf{Q}$: Transient submatrix
- Poisson equation solution

Extended Game: Re-roll Option



- After first roll d_1:
  - Stick: Move by d_1
  - Re-roll: Move by d_2
- Each turn $=$ 1 move

MDP Formulation

- **States**: (s, d_1) pairs
- **Actions**: $\{\text{Stick, Re-roll}\}$
- **Cost**: $c(s, a) = 1$ per turn
- **Value functions**:
  - $V_0(s)$: Pre-roll value
  - $Q(s, d_1)$: Post-roll value

HJB Equation

$$Q(s, d_1) = \min \begin{cases} 1 + V_0(T(s + d_1)) \\ 1 + \frac{1}{6} \sum_{d_2} V_0(T(s + d_2)) \end{cases}$$

$$V_0(s) = \frac{1}{6} \sum_{d_1=1}^{6} Q(s, d_1)$$

Boundary: $V_0(100) = 0$

Value Iteration

Initialize $V_0^{(0)} \leftarrow 0$ $k = 1$ to K each state s each roll d_1
$Q^{(k)}(s, d_1) \leftarrow \min(\text{stick}, \text{re-roll})$ $V_0^{(k)}(s) \leftarrow \frac{1}{6} \sum Q^{(k)}(s, d_1)$

Optimal Policy

$$\pi^*(s, d_1) = \begin{cases} \text{Stick} & \text{if } V_0(T(s+d_1)) \leq \mathbb{E}[V_0(s+d_2)] \\ \text{Re-roll} & \text{otherwise} \end{cases}$$

- Risk-averse control
- Exploitation vs exploration

Physical Interpretation

potential_landscape.pdf

- $V_0(s)$: Potential energy
- Snakes: Potential barriers
- Ladders: Potential wells
- Policy: Gradient descent

Conclusion

- Master Equation: Stochastic dynamics
- HJB: Optimal control extension
- Re-rolls: Entropy injection
- Physics of decision-making