

Optimal Play for a Non-Ergodic Card Game

Mees van Dartel, Lorenzo Gregoris

September 28, 2025

1 Problem formulation

Consider the following card game ¹. A decision maker faces a deck $\mathcal{D} = \{x_1, x_2, \dots, x_n\}$ containing n cards, labeled $i \in \{1, 2, \dots, n\}$. The value of card $x_i = i$. At every time step, the decision maker draws from the deck with uniform probability without replacement. Denote by D_t the set of remaining cards in the deck at time t . The probability of drawing a card x_i , conditional on the remaining cards D_t , is given by the following PMF:

$$\mathbb{P}(X = x_i \mid D_t) = \begin{cases} \frac{1}{|D_t|} & \text{for } x_i \in D_t \\ 0 & \text{else} \end{cases}. \quad (1)$$

Let x_t be the card drawn at time t . The decision maker sequentially draws cards, and faces the following choice. She can *pick* the drawn card x_t , and receive reward $r_t = x_t$. However, she must thereafter draw and discard x_t cards from the deck, such that:

$$D_{t+1} \mid \text{Pick} = D_t \setminus \{y_1, y_2, \dots, y_{x_t}, x_t\}, \quad (2)$$

where y_j is drawn uniformly at random without replacement $y_j \sim U(D_t \setminus \{y_1, \dots, y_{j-1}\})$. Alternatively, she can choose to *skip* the card, whereafter she receives reward $r_t = 0$, and can draw a new card: We then have:

$$D_{t+1} \mid \text{Skip} = D_t \setminus \{x_t\}. \quad (3)$$

The decision maker's information set at time t is given by $\Omega_t = \{\mathcal{D}, D_t\}$, such that she can observe which cards are left in the deck. At $t = 0$, we have $D_0 = \mathcal{D}$. The decision maker's objective is to maximize her total expected sum of rewards:

$$R = \mathbb{E} \left[\sum_{t=0}^{\infty} r_t \right]. \quad (4)$$

Whenever a time step τ occurs where there are no cards remaining, such that $D_\tau = \emptyset$, all subsequent rewards are 0, i.e., $r_t = 0$ for all $t \geq \tau$. D_τ is an absorbing state. D_τ eventually occurs with $\mathbb{P} = 1$, at the latest at $t = n$, if the decision maker never picks any card. This is clearly suboptimal, as her total sum of rewards will be $R = 0$. She can easily improve upon this with the following policy:

$$\pi(x_t, D_t) = \begin{cases} \text{Pick} & \text{if } x_t = n \\ \text{Skip} & \text{if } x_t \neq n \end{cases}, \quad (5)$$

¹Of course single player games are not games in the formal sense.

such that her total sum of rewards is $R = n$.

THIS IS A NON-HOMOGENOUS, NON-ERGODIC MARKOV CHAIN? WHAT IS THE STRUCTURE, WHAT DOES THIS SAY ABOUT THE OPTIMAL POLICY?

2 Optimal play

Consider the following policy:

$$\pi^{\max}(x_t, D_t) = \begin{cases} \text{Pick} & \text{if } x_t = \max\{D_t\} \\ \text{Skip} & \text{if } x_t \neq \max\{D_t\} \end{cases}, \quad (6)$$

The policy π^s simply searches for the largest card remaining in the deck and ends the game. Once this policy chooses *pick*, we enter the absorbing state D_τ , ending the game. For any state D_t , the value of this policy is given by:

$$V^{\pi^s}(x_t, D_t) = \sum_t^\infty r_t = \max\{D_t\}, \quad (7)$$

We can think of an example of a deck D^s for which the following policy is optimal, namely a deck for which, for every $x_i \in D^s, x_i > |D^s|$. Picking any card from D^s will generate the absorbing state D_τ , ending the game, so clearly it is optimal to search for the largest card in the deck and end the game.

Our proposition for the optimal policy is as follows:

Proposition 2.1. *The optimal policy for the card game is given by:*

$$\pi^m(x_t, D_t) = \arg \max_{a \in \{\text{Pick}, \text{Skip}\}} \left\{ r(x_t, a) + \mathbb{E} \left[V^{\pi^s}(x_{t+1}, D_{t+1}) \right] \right\}. \quad (8)$$

Proof. The Bellman equation for the card game is given by:

$$V(x_t, D_t) = \max_{a \in \{\text{Pick}, \text{Skip}\}} \left\{ r(x_t, a) + \mathbb{E} \left[V(x_{t+1}, D_{t+1}) \mid a \right] \right\}. \quad (9)$$

It follows that if we can show that

$$\arg \max_{a \in \{\text{Pick}, \text{Skip}\}} \left\{ r(x_t, a) + \mathbb{E} \left[V(x_{t+1}, D_{t+1}) \right] \right\} = \arg \max_{a \in \{\text{Pick}, \text{Skip}\}} \left\{ r(x_t, a) + \mathbb{E} \left[V^{\pi^{\max}}(x_{t+1}, D_{t+1}) \right] \right\}, \quad (10)$$

then we can be sure that our policy is the optimal policy:

$$\pi^m(x_t, D_t) = \pi^*(x_t, D_t). \quad (11)$$

We can rewrite the RHS as:

$$r(x_t, a) + \mathbb{E} \left[V^{\pi^{\max}}(x_{t+1}, D_{t+1}) \right] = r(x_t, a) + \mathbb{E} \left[\max\{D_{t+1}\} \right]. \quad (12)$$

We can order the elements of D_t as $m_1 > m_2 > \dots > m_{|D_t|}$, where $\max\{D_t\} = m_1$. Note that the total number of ways we can discard x_t cards from a deck of size $|D_t|$ is given by

$$\binom{|D_t| - 1}{x_t}. \quad (13)$$

We are interested in the number of outcomes for which $\max\{D_t\} = m_j$. These are all events where all cards $m_i > m_j$ are discarded, and m_j is not discarded. The number of events such that these conditions hold is given by

$$\binom{|D_t| - 1 - j}{x_t - (j - 1)}, \quad (14)$$

since we can randomize over $x_t - (j - 1)$ cards (the cards smaller than m_j), and can draw them from $|D_t| - j$ cards. It follows that:

$$\mathbb{P}(\max\{D_{t+1}\} = m_j | \text{Pick } x_t) = \mathbb{P}_j = \frac{\binom{|D_t| - 1}{x_t}}{\binom{|D_t| - j - 1}{x_t - (j - 1)}}. \quad (15)$$

We can then expand the expected maximum after picking x_t as:

$$\mathbb{E}[\max\{D_{t+1}\} | \text{Pick } x_t] = \sum_{j=1}^{x_t} \mathbb{P}_j \cdot m_j = \sum_{j=1}^{x_t} \frac{\binom{|D_t| - 1}{x_t}}{\binom{|D_t| - j - 1}{x_t - (j - 1)}} m_j. \quad (16)$$

It follows that the decision maker should select action *Pick* iff:

$$x_t + \sum_{j=1}^{x_t} \frac{\binom{|D_t| - 1}{x_t}}{\binom{|D_t| - j - 1}{x_t - (j - 1)}} m_j > m_1. \quad (17)$$

WORK IN PROGRESS |

We need to show that

$$x_t + \mathbb{E}[V(x_{t+1}, D_{t+1}) | \text{Pick}] \geq \mathbb{E}[V(x_{t+1}, D_{t+1}) | \text{Skip}] \implies x_t + \mathbb{E}[\max\{D_{t+1}\} | \text{Pick}] \geq \max\{D_t\} \quad (18)$$

and

$$x_t + \mathbb{E}[V(x_{t+1}, D_{t+1}) | \text{Pick}] \leq \mathbb{E}[V(x_{t+1}, D_{t+1}) | \text{Skip}] \implies x_t + \mathbb{E}[\max\{D_{t+1}\} | \text{Pick}] \leq \max\{D_t\}. \quad (19)$$

Lemma 2.2. *Under the optimal policy, any game ending card should be the maximum card left in the deck.*

Consider the first proposition, where the optimal policy would prescribe to Pick. We then have:

$$x_t + \mathbb{E}[V(x_{t+1}, D_{t+1}) | \text{Pick}] \geq \mathbb{E}[V(x_{t+1}, D_{t+1}) | \text{Skip}] \quad (20)$$

We know that, by the principle of optimality, the following two inequalities must hold:

$$\mathbb{E}[\max\{D_{t+1}\}|\text{Pick}] \leq \mathbb{E}[V(x_{t+1}, D_{t+1})|\text{Pick}], \quad (21)$$

$$\max\{D_t\} \leq \mathbb{E}[V(x_{t+1}, D_{t+1})|\text{Skip}], \quad (22)$$

such that we have

$$x_t + \mathbb{E}[V(x_{t+1}, D_{t+1})|\text{Pick}] \geq \mathbb{E}[V(x_{t+1}, D_{t+1})|\text{Skip}] \geq \max\{D_t\}. \quad (23)$$

In the case where V_π^m and V_π^* are the same, they trivially prescribe the same actions. We are therefore only interested in the case where inequalities are strict:

$$x_t + \mathbb{E}[V(x_{t+1}, D_{t+1})|\text{Pick}] \geq \mathbb{E}[V(x_{t+1}, D_{t+1})|\text{Skip}] > \max\{D_t\}. \quad (24)$$

UNFINISHED.

□