Optimal Play for a Non-Ergodic Card Game

Mees van Dartel, Lorenzo Gregoris

September 17, 2025

1 Problem formulation

Consider the following card game [1]. A decision maker faces a deck $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ containing n cards, labeled $i \in \{1, 2, \ldots, n\}$. 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:

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

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:

$$D_{t+1} \mid \text{Pick} = D_t \setminus \{y_1, y_2, \ldots, y_{x_t}, x_t\}, \tag{2}$$

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

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

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:

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

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:

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

[1] Of 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-ERGORIC MARKOV CHAIN? WHAT IS THE STRUCTURE, WHAT DOES THIS SAY ABOUT THE OPTIMAL POLICY?.

2 Optimal play

Consider the following policy:

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

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:

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

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:*

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

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

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

It follows that if we can show that

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

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

$$\pi^m(x_t, D_t) = \pi^*(x_t, D_t). \tag{11}$$

We can rewrite the RHS as:

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

Where:

$$\mathbb{P}\Big(\max\{D_{t+1}\} = \max\{D_t\} \Big) = \frac{|D_t|-1}{|D_t|} \cdot \frac{|D_t|-2}{|D_t-1} \cdot \ \dots \ \cdot \frac{|D_t|-x_t}{|D_t-x_t+1} \cdot \max\{D_t\}, \tag{13}$$

$$\mathbb{P}\Big(\max\{D_{t+1}\} = \max\big\{ D_t \setminus \max\{D_t\} \big\} \Big) \tag{14}$$

$$\mathbb{E}\big[\max\{D_{t+1}\} \big] = \tag{15}$$

□