Card Search

Problem:
Consider a deck of m cards. Suppose we turn over the first card, record what it is, put it back in the deck and shuffle. If we repeat this process, how many times do we expect we have to do this until we see every card in the deck?

Solution:
Let N_n be a random variable denoting the number of times we need to perform this action until we are left with n cards we have yet to see in a deck of m cards. Then, the probability of turning over a card we have seen already is $(m-n)/m$, and the probability of turning over a new card is n/m. Thus, the expected value of this game can be calculated by conditioning on whether or not the next card turned over has been seen or not. This is an application of the partition theorem and yields

$$\begin{aligned}
\mathbb{E}(N_n) &= \frac{m-n}{m}\Big[1+\mathbb{E}(N_n)\Big] + \frac{n}{m}\Big[1+\mathbb{E}(N_{n-1})\Big] \\
&= 1 + \frac{m-n}{m}\mathbb{E}(N_n) + \frac{n}{m}\mathbb{E}(N_{n-1}) \\
\implies \; m\mathbb{E}(N_n) &= m + (m-n)\mathbb{E}(N_n) + n\mathbb{E}(N_{n-1}) \\
\implies \; \mathbb{E}(N_n) &= \frac{m}{n} + \mathbb{E}(N_{n-1}) \\
&= \frac{m}{n} + \frac{m}{n-1} + \mathbb{E}(N_{n-2}) \\
&= m\left[\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{2} + 1\right] + \mathbb{E}(N_0) \\
&= m\sum_{i=1}^{n}\frac{1}{i},
\end{aligned}$$

where we have used that $\mathbb{E}(N_0) = 0$, since when there are no cards left to see the game is over. Note that this equality only holds for $1 \leq n \leq 51$, since we need to turn the first card over to initiate the game. The summation we have derived is a Harmonic series, and can be approximated as

$$\sum_{i=1}^{n}\frac{1}{i} = \log n + \gamma,$$

where $\gamma \approx 0.577$ is the Euler-Mascheroni constant and is the limit of the Harmonic series and the natural logarithm of n as n tends to ininity. Therefore, when $m = 52$, the asymptotic estimate for the expected value is given by

$$\mathbb{E}(N_{52}) = 1 + \mathbb{E}(N_{51}) = 1 + 52(\gamma + \log 51) \sim 235.$$

To support this analysis, we performed Monte Carlo simulations and found the results to be consistent.