STAT 5000 Final

Lucas Bloomenstein Boos

December 2024

Imagine, for a moment, that you were a kid obsessed with the card game Magic The Gathering in the year 1995 when the set Ice Age was released and wanted to collect every single card in the set the old fashioned way: opening booster packs. Magic The Gathering is a collectible trading card game in which players role-play as wizards casting spells to reduce the life total of their opponent from 20 to 0 (the game is played as 1 versus 1). In order to cast spells, players draw cards from their pre-constructed deck of cards obtained from booster packs. Booster packs contain 15 unique Magic The Gathering cards with 10 common cards, 1 land card, 3 uncommon cards, and 1 rare card. The set we will be examining, Ice Age, had 383 unique cards in total; 121 commons, 20 lands, 121 uncommons, and 121 rares. What these numbers mean is that if we are lucky, we only have to open 121 booster packs to obtain every card in the set. If we're unlucky, however, we may never collect every single card in the set no matter how many booster packs we open. So, the question is simple: how many packs on average do we have to open in order to obtain every single one of the 383 unique cards belonging to the set Ice Age?

To answer this question, we will simplify it and compare the analytic solutions of these simplified examples to the mean of 100 Markov Chain processes (a sequence of possible events where the next state is dependent on the current state). By comparing these two expectations, we can show that the Markov process is a good estimation of the expectation for which we do not have an analytic solution. Each process does the same thing with different stipulations: open card packs until all 383 unique cards are collected.

The first simplification we are looking at is assuming that the 383 cards are uniformly distributed, as in it is equally likely to pull any of the 383 cards and pulling 1 card at a time instead of in packs of 15. This is a specific example of the famous Coupon Collector's problem with N = 383 (given N coupons, how many coupons do you expect to draw with replacement before having drawn each coupon at least once?). As it turns out, there is an analytic solution to this question. First, assume that we have N different types of coupons (or trading cards) and let X_i be the random variable that denotes the number of additional coupons needed to get the next type of coupon for $i = 0, 1, ..., N-1$. For instance $X_0 = 1$ because when you have no coupons the first coupon you pull is the "next type of coupon". We will continue to pull coupons until we have all N types resulting in the sum $X = \sum_{i=0}^{N-1} X_i$. We want to find the expectation of this sum is $E[X] = \sum_{i=0}^{N-1} E[X_i]$ by the fact that expectation distributes over sums. To find the general distribution of $P(X_i = k)$, the probability that getting the i+1st type of coupon takes k pulls, consider $P(X_1 = k)$. We already have 1 coupon so the probability that we pull it again is $\frac{1}{N}$ because our distribution is uniform. The probability that we do not pull the coupon we already have is $\frac{N-1}{N}$. Assuming pulls are independent (which I believe is a fair assumption) and it takes k pulls to get the second type of coupon $P(X_1 = k) = (\frac{1}{n})^{k-1}(\frac{N-1}{N})$, for $P(X_2 = k)$ we have $(\frac{2}{n})^{k-1}(\frac{N-2}{N})$ because we have two coupons and it takes us k pulls to get the third. In general we have $P(X_i = k) = (\frac{i}{n})^{k-1}(\frac{N-i}{N})$. This is a geometric distribution, with $p_i = \frac{N-i}{N}$, and with a known expectation of $\frac{1}{p}$, we know $E[X_i] = \frac{N}{N-i}$. now we know $E[X] = \sum_{i=0}^{N-1} E[X_i] = \sum_{i=0}^{N-1} \frac{N}{N-i} = N * \sum_{i=0}^{N-1} \frac{1}{N-i} = N * (\frac{1}{N} + \frac{1}{N-1} + ... + \frac{1}{1}) = N * \sum_{i=1}^{N} \frac{1}{i}$. For N = 383 we can expect to pull $383 * \sum_{i=1}^{383} \frac{1}{i} \approx 2500$ cards before collecting all 383 unique cards.

Now we can move on to pulling 15 cards at a time, which is known as the batch Coupon Collector's problem. We start by assuming we have n coupons and pull m at a time. X can be the random variable denoting the number of packs it takes to obtain all n coupons. For any amount of S (the set of coupons you've already collected), the probably that we don't pull anything from the set S in the first i pulls is $(\frac{\binom{n-|S|}{m}}{\binom{n}{m}})^i$. We can say this because each pull is independent. Now we consider the probability to avoid at least 1 duplicate coupon in the first i pulls using the inclusion exclusion principle is $\binom{n}{1} * (\frac{\binom{n-|S|}{m}}{\binom{n}{m}})^i - \binom{n}{2} * (\frac{\binom{n-|S|}{m}}{\binom{n}{m}})^i + \binom{n}{3} * (\frac{\binom{n-|S|}{m}}{\binom{n}{m}})^i$...... With $|S|$ duplicates the probability to avoid this is $\sum_{s=1}^{n} \binom{n}{s} * (-1)^{s+1} * (\frac{\binom{n-|S|}{m}}{\binom{n}{m}})^i$, or the probability that the number of packs you open in order to obtain all n coupons exceeds i $P(X > i) = \sum_{s=1}^{n} \binom{n}{s} * (-1)^{s+1} * (\frac{\binom{n-|S|}{m}}{\binom{n}{m}})^i$. We can now find the expectation of X using the tail sum expectation formula which says $E[X] = \sum_{i=0}^{\infty} P(X > i) = \sum_{i=0}^{\infty} \sum_{s=1}^{n} \binom{n}{s} * (-1)^{s+1} * (\frac{\binom{n-|S|}{m}}{\binom{n}{m}})^i$. Using tricks from Calculus 2 yields a final expectation formula of $E[X] = \binom{n}{m} \sum_{s=1}^{n} \frac{(-1)^{s+1}\binom{n}{s}}{\binom{n}{m} - \binom{n-s}{m}}$. When trying to compute this analytical result with n = 383 and m = 15 I was unable to get a reasonable result, most likely due to small floating point errors (the sum converges to unreasonably small or large values).

Now that we understand how the expectation is computed analytically, we can move on to estimating the expectation by using a Markov Chain. The Markov process used for the regular Coupon problem was to pull 1 card (a possible event) that is equally likely to be any of the 383 cards, check wether it is a unique card to our collection or not (the state attained in previous events), do this until all 383 unique cards were acquired, and count the number of cards we had to pull to obtain them. Because the above process doesn't yield a constant number, I repeated it 100 times, took note of the number of pulls it took each time, then took the mean of these 100 trails to estimate the expectation. Below is a histogram for the number of pulls it took to get all 383 unique cards 100 times.

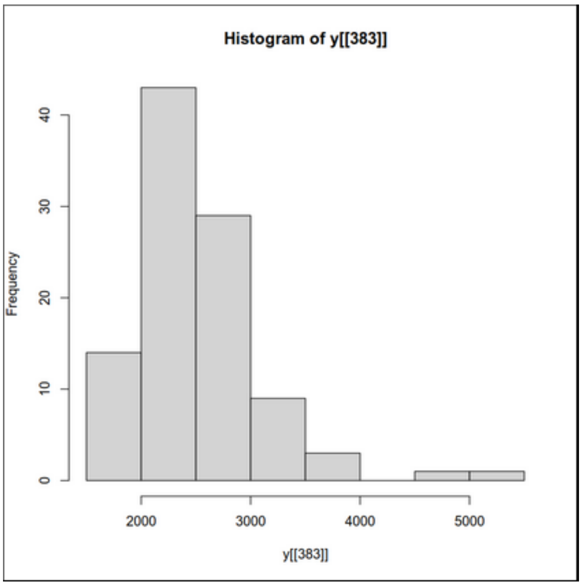

Figure 1: Coupon Problem, n = 383

The mean of these 100 samples is approximately 2522 pulls which is very close to the analytical expectation of problem with n = 383 (2500 pulls). The histogram also has a mode of about 2400 or 2300, which is close to the analytic solution. I also created box plots for n = 1 to 500, as there are rarely ever more than 500 unique cards in a set, to show how the expectation increases as n increases.

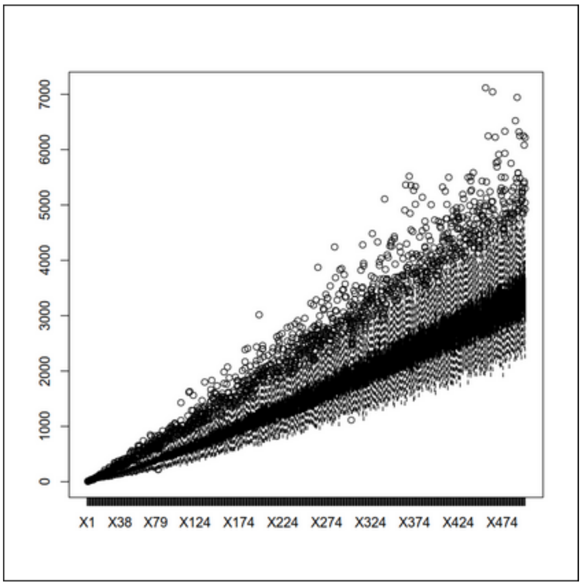

Figure 2: Coupon Problem n = 1 to 500

For the batch coupon problem I assumed a uniform distribution and drew in batches of 15. The process here was much like the one above: pull 15 cards (a possible event) that are equally likely to be any of the 383 cards (although the 15 cards are all different from one another), check each of the 15 cards to determine wether it is a unique card for our collected, do this until all 383 unique cards were acquired, and count the number groups of 15 cards we pull in order to collect all 383 unique cards. This also doesn't yield a constant number so I repeated it 100 times, took note of the number of packs it took each time, then took the mean of these 100 trails to estimate the expectation. Below is a histogram for the number of packs it took to get all 383 unique cards 100 times.

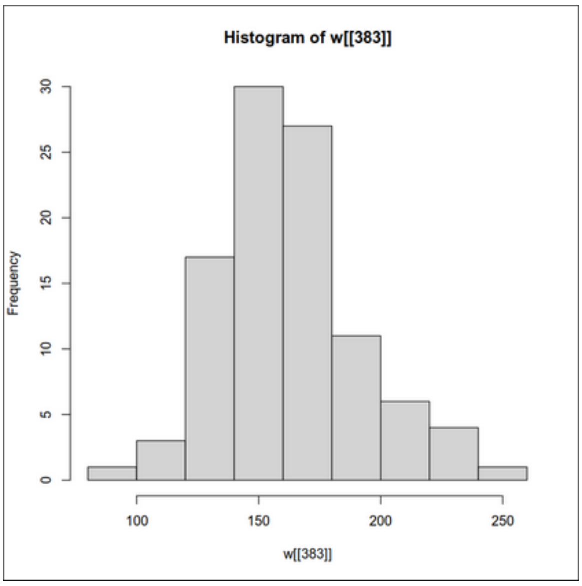

Figure 3: Batch Coupon Problem, n = 383, m = 15

The mean of these samples was 164 packs of cards which I cannot compare to the analytic solution as it seams to have floating point errors, but the Markov Estimation can be compared for smaller n and m. The analytical solution for n = 10 and m = 3 is about 9 packs, and looking at the histogram of the 100 sampled Markov chains for n = 10 and m = 3 yields a similar result with a mean of 9.17, which confirms that sampling in this way comes close to the true expectation.

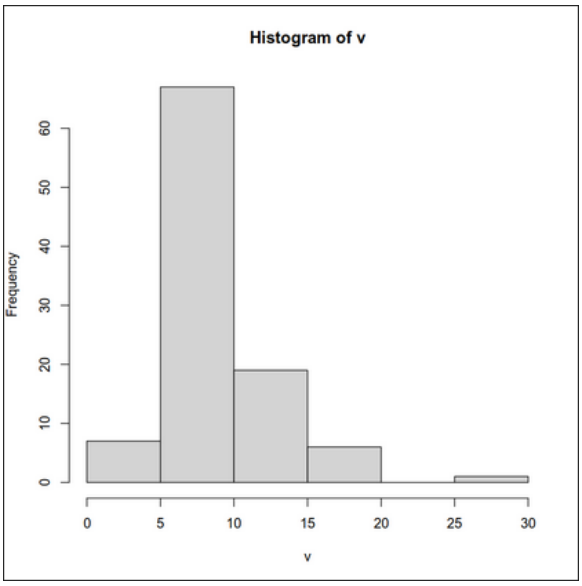

Figure 4: Batch Coupon Problem, n = 10, m = 3

I also created a box plot of the batch Coupon problem for n = 15 to 500 taking batches of m = 15 each time to show how the expectation increases as n increase with a constant number of batches.

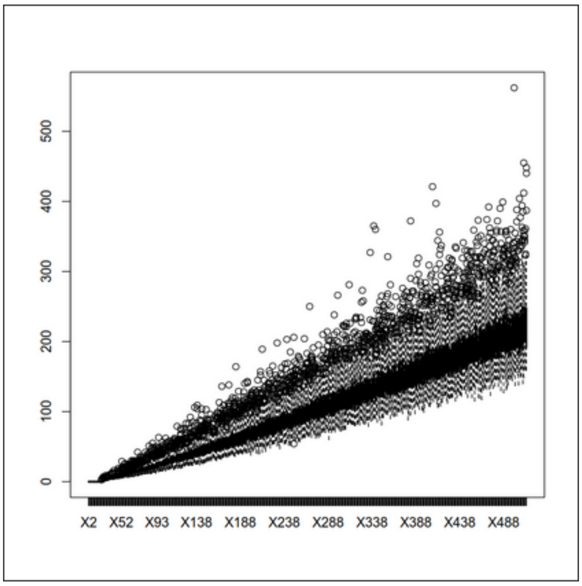

Figure 5: Batch Coupon Problem n = 15 to 500, m = 15

Now that we have seen that this Markov process is a good estimator of the expectation in simple cases, we can use it to find the expectation in a case where the analytic solution is unknown. In this case, we are trying to find the expected number of card packs needed to be opened to obtain all 383 unique cards from the set Ice Age. Ice Age has 121 commons, 121 uncommons, 121 rares, and 20 lands all normally distrusted among themselves (in other words we assume you are equally likely to get any of the 121 rares for your 1 rare per pack). Booster packs contain 15 cards 10 of which are common, 3 are uncommon, 1 is a land, and 1 is a rare, and each booster pack contains no repeat cards (we assume it is imposable to get two of the same card in 1 booster pack). The process here was to create a booster pack with the above distribution of cards (the possible event), see if any of the cards where new to our collection (the state), repeat the process until all 383 cards where obtained, then finally count the number of packs we opened to get there. Once again, I preformed this process 100 times because the number of packs needed to obtain all 383 cards is not constant. I created a histogram for the number of packs opened to obtain all 383 cards with the process described above.

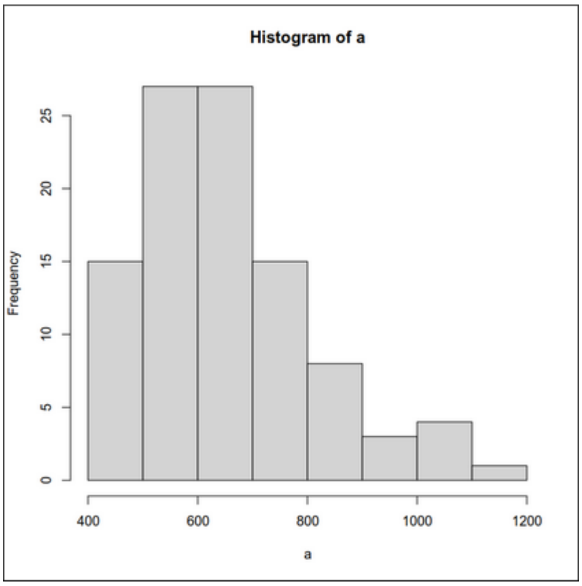

Figure 6: Enter Caption

The mean of these 100 samples was approximately 651, and given that our other means were very close to the analytic expectations, it is safe to say that this is very close to the true number of packs one should expect to open before obtaining every card in the set Ice Age.

This modeling process shows that it takes a great deal of packs to be opened in order for every card in a set of Magic The Gathering cards to be obtainable to use for play (In other words, if you were a Magic obsessed kid in 1995 with the above goal, you'd better be rich). In the future I would like to extend this research by changing some of my assumptions, like the assumption that a certain rarity of a card is equally likely as cards of the same rarity, in order to better predict the expected number of packs that need to be opened to collect a full set of cards. I would also like to preform the expected value of cards in a booster pack or in a box of booster packs based on the occurrences of certain cards and their distribution throughout the set (perhaps something like: how many of x expensive card occurs in a set number of booster packs?). I think my current question is important because while an individual does not have to open all these packs of cards, the community does as a whole in order for everyone to be able to use new cards that are released each set. Wizards of the Coast (who create and own the property of Magic), could use this expectation to predict the number of packs they will sell (this requires tweaking as people often need multiple copies of a card).

Stadje, W. (1990). The Collector's Problem with Group Drawings. *Advances in Applied Probability*, 22(4), 868-873

Wikimedia Foundation. (2024, December 3). Coupon Collector's problem. Wikipedia. link

Wikimedia Foundation. (2024b, December 10). Markov chain. Wikipedia. link