Project 1: Which Urn is Which ?

Craig Brooks

Due February 6 2023

1 Introduction

There are two urns, and inside each urn are 100,000 marbles comprised of 3 colors: White, Green, and Black. In one urn, the group ratio for $White$: $Green$: $Black$ is 25 : 35 : 40 and the second is 42 : 20 : 38. We reach into one of the urns and pull a green marble on the third draw. Our task: To determine which urn most likely contains Distribution A ($Urn\ A$) and which contains B ($Urn\ B$) by doing random samples from each population by observing the median number of draws it will take to draw a green marble from one of them.

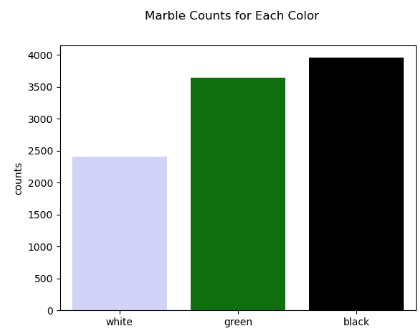
2 What We Know

If we simply want to find the distribution given an arbitrary amount of draws, our random variable will be ($X = number\ of\ successes\ drawn\ in\ n\ trials$) $\sim$ **Bernoulli**(X) in the case of 2 marble species if they are being drawn with replacement. In our case, we have 3 marble species. Thus, we can characterize our sampling with a multinomial distribution. In these situations, all draws are independent of each other, so the probability of success for each draw *does not change*.

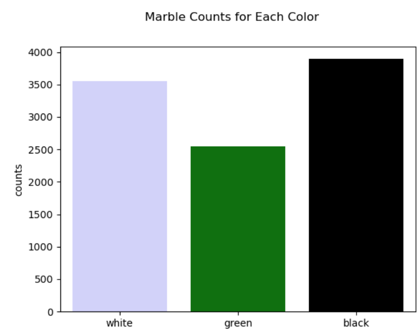
We can also determine the *number* of times we can expect to draw, on average, until we successfully draw a green marble, and this can be modeled with a geometric distribution where $X \sim \mathbf{Geo}(p)$ is the probability of success in k draws. Unlike the multinomial distribution, where we are looking at the distribution of the successes, we are counting the number of *trials* until a successful outcome.

3 Outline of the Experiment

In our experiment, we will construct our sets of marbles with a multinomial distribution. Figure 1 shows a barplot representing the two urns with the counts of marbles of each color. From the generated dataset, we will then do a large number of trials, and pick (with replacement) until we achieve a positive result for each trial, in this case picking a green marble. Next, we will calculate the



(a) Distribution A: 25/35/40



(b) Distribution B: 42/20/38

Figure 1: Two graphs representing the distribution of marbles in two urns

average and/or median number of draws it takes until drawing a green marble, then plot a histogram of the counts over n trials. From there, we will plot box-and-whisker plots for the trials, and therefore determine the interquartile range (IQR) and $\pm$ 1.5 times the median of n trials. Finally, we will calculate the log-likelihoods of drawing a green in k tries given a particular urn in n trials to a 95 % confidence interval.

In our experiment, the priors that we have are that each urn contains 100000 marbles and that one urn has distribution $A = 25 : 35 : 40$ and the other $B = 42 : 20 : 38$. We cannot see inside the urns, but we are allowed to take an arbitrary amount of draws of the marbles. In theory, we could take random sample of perhaps 50 draws, count each color and compute the fraction of White, Green, and Black marbles. The problem with this approach is that it is unclear whether one random sampling of marbles is really representative of the entire population. To overcome this, you could continue taking random samples, with replacement, as many times as desired and then find how the distribution of the marbles emerges. Unfortunately, this is a slow process, and realistically we do not have the time or resources to undertake such a process.

Fortunately, we can use statistics to model random sampling with replacement. Since we have 3 different species of marbles at different ratios, we can model our urn using a Multinomial distribution. Furthermore, because these marbles have different ratios, then we can distinguish between them based on the drawing marbles. However, instead of taking lots of random samples and counting each color, we are able to determine the median or mean for the number of draws it will take for us to achieve some threshold X, in our case, the number of draw it would take, on average, until a green marble is drawn. If we let $X = k$ be a random variable corresponding to the number of draws and p be the probability of getting green in a Bernoulli trial, then the probability mass function is

$$P(X = k) = (1 - p)^{k-1}\, p$$

4 Hypothesis

We want to determine the likelihood that the urn we are drawing from has distribution A given that we drew a green marble on the kth draw, (in our case, $k = 3$. We have prior knowledge about the distribution of each urn, and we believe that we are more likely to draw a green marble in fewer draws when we have Dist $A \to P(Green) = .35$. Thus, we will state the null hypothesis $\mathbf{H}_0$ and alternative $\mathbf{H}_1$ as

$$\mathbf{H}_0 :\ X \sim \mathbf{Geo}(.35)\ \to (Urn\ A) \quad , \quad \mathbf{H}_1 : X \sim \mathbf{Geo}(.25)\ \to (Urn\ B)$$

5 Algorithm Analysis

We will have two scripts Urn.py and UrnAnalysis.py. In Urn.py, we will generate random data to represent the urn containing 100,000 marbles. To do this, we use a function from *numpy.stats* called *numpy.random.multinomial* to generate population data in either default ratio or a random ratio created by passing 3 floats whose sum is equal to 1 into an array. After the urns are generated, we run random sampling trials that are stored in an array. Next, we write the data contained in the array into a file called urn.txt.

In the UrnAnalysis.py script, we read the data from the array, then for each trial, determine which draw the first green marble is picked and append each to an array called *success*. This can be accomplished simply because the color choices are appended in the order that they were drawn, so we simply can append the "index + 1" where the success occurred.

Next, we can compute the median and mean, then plot a histogram of that data. From there, we can calculate the log-likelihood ratios for each case assuming that $\mathbf{H}_0$ and then $\mathbf{H}_0$ is true.

We then calculate the likelihood for an alternative hypothesis (in this case that the urn has Multinomial distribution of different parameters. From here, we compute the log-likelihood ratio of obtaining a green marble in k draws and/or the log-likelihood ratio of obtaining a green marble in k or fewer draws.

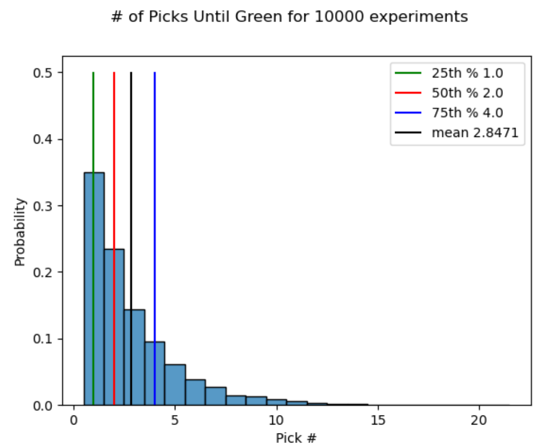
The log-likelihood for $\mathbf{H}_0$ can be found by

$$\sum_i \log\left(\frac{P(X_i|P_A)}{P(X_i|P_B)}\right)$$

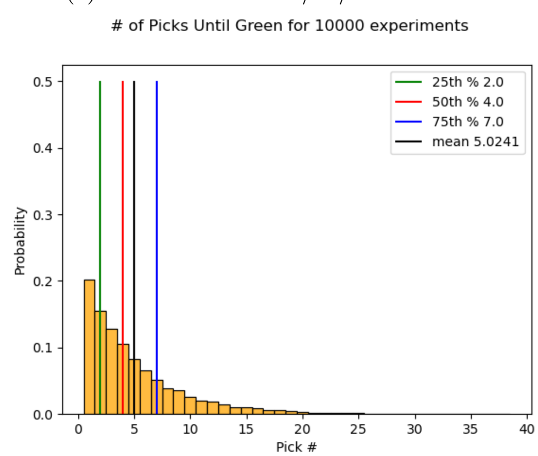
Similarly, the log-likelihood for the alternative hypothesis $\mathbf{H}_1$ is can be found by

$$\sum_i \log\left(\frac{P(X_i|P_B)}{P(X_i|P_A)}\right)$$

6 Analysis



(a) Distribution A: 25/35/40



(b) Distribution B: 42/20/38

Figure 2: Histograms for the two possible ratios in the urns

Figure 2 are the histograms for distributions A and B. For distribution A, on the first draw, the probability of drawing *Green* is .35 and for distribution B, the probability is .2. On the second draw, the probability drops for both, but if we look at the patterns, A is always more probable to successfully draw green on the k^{th} draw. Additionally, the box-and-whisker plot in Figure 3 shows the median draws and the IQR under both hypotheses. If we examine Urn 1, the median is 2.0 with an IQR of $[1, 4]$, while Urn be has a median of 4 and an IQR of $[2, 7]$

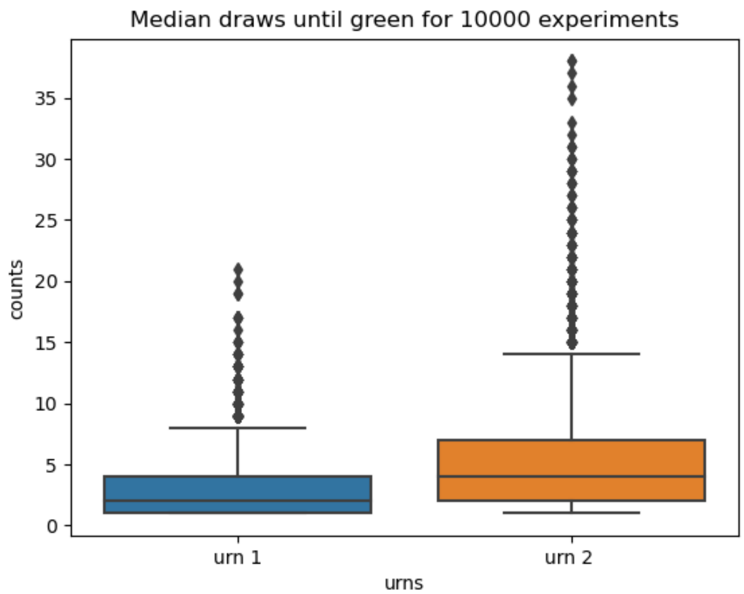


Figure 3: Box and whisker plot of H_0 and H_1

Figure 4 shows the Probability vs Log-Likelihood plot for $\mathbf{H}_0$ in blue and $\mathbf{H}_1$ in orange. From this plot, it is slightly more likely to draw a green in ≤ 3 tries under $\mathbf{H}_0$ than it is for $\mathbf{H}_1$. However, as we move away from $LLR = 0$, the likelihood becomes larger $\mathbf{H}_0$ is true as the number of draws it takes to successfully draw a green becomes small. On the contrary, if it takes more draws it takes to draw a green marble, then $\mathbf{H}_1$ becomes more likely.

In our scenario, we are able to draw a green marble on the third draw. Therefore, it is more likely that we are drawing from Urn 1 with Distribution A.

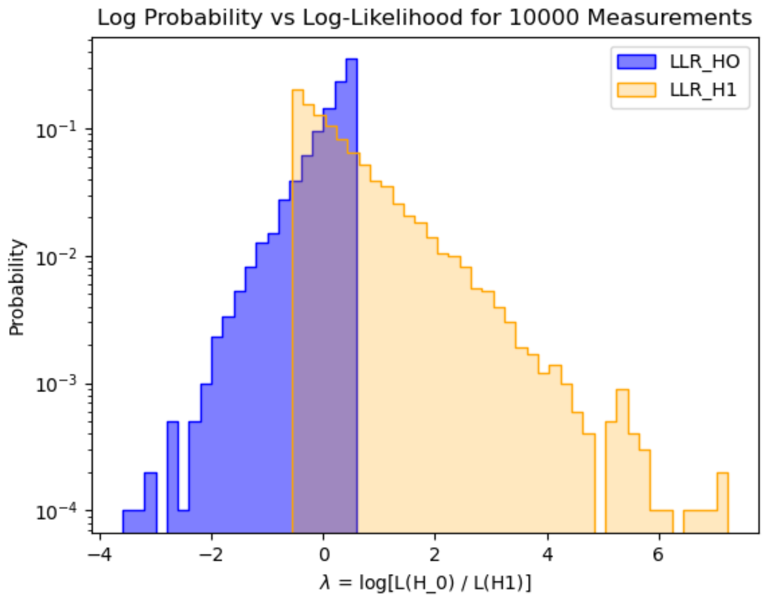


Figure 4: Log-likelihood plots of H_0 and H_1

If we have Urn A, then the probability of drawing a green in 3 or fewer trials is

$$P(X < 3) = .35 + (.65) \cdot .35 + (.65)^2 \cdot .35 = .73$$

On the other hand, if we are drawing from Urn B, this becomes

$$P(X < 3) = .2 + (.8) \cdot .2 + (.8)^2 \cdot .2 = .488$$

As we can see, it is significantly more *probable* to draw a green marble under $\mathbf{H}_0$ than $\mathbf{H}_1$.

For our experiment, we will model this for $1 \leq k \leq 3$. We would also like to calculate the log-likelihood of drawing a green marble in k or fewer trials given *Distribution A* and *Distribution B*. For $k = 3$, the log-likelihood of $\mathbf{H}_0$ is

$$\log\left(\frac{.35(.65)^2)}{.2(.8)^2)}\right) = .14$$

Therefore, we cannot reject the null hypothesis $\mathbf{H}_0$. Alternatively, the log-likelihood of $\mathbf{H}_1$ is true is -0.14, so we do have justification to reject the alternative hypothesis, since the null hypothesis is more *likely*.

Significance and power of the test

Under $\mathbf{H}_0$, and $\alpha = .05$ the significance of the test is

$$(1-p)^{k-1} \rightarrow (.65)^2 = .42 > .05$$

Therefore we fail to reject the null hypothesis $\mathbf{H}_0$

To calculate the critical value λ under the assumption that $p = .35$

$$\begin{aligned} X \sim Geo(.35), \qquad , \qquad P(X \geq k) = (1-p)^{k-1} < .05 \\ \rightarrow (x-1)\log(.65) < \log(.05) \\ \rightarrow k-1 > \frac{\log(.05)}{\log(.65)} \\ \rightarrow k > 8 \end{aligned}$$

This means that if we were to draw a green on pick ≥ 8 with a significance of $\leq .05$, we would have evidence to reject the null hypothesis. In terms of the log-likelihood $\log \frac{L(\mathbf{H}_0)}{L(\mathbf{H}_1)} \leq -1.4534$. The size of the test, or the probability of making a Type 1 Error, is the probability of rejecting $\mathbf{H}_0$ when $\mathbf{H}_0$ is true. This can be determined from the actual significance level in the rejection region $k \geq 8$.

$$\begin{aligned} P(Reject\ \mathbf{H}_0 \mid \mathbf{H}_0\ True) = (1-p)^k \\ \rightarrow (.65)^8 = .0490 \end{aligned}$$

The power of the test, which is the probability of rejecting $\mathbf{H}_0$ when $\mathbf{H}_0$ is false is

$$\begin{aligned} P(Reject\ \mathbf{H}_0 \mid \mathbf{H}_0\ False) = P(k \geq 8 | p = .2) \\ \rightarrow (.8)^7 = .2097 \end{aligned}$$

7 Summary

This test failed to reject the null hypothesis. We were able to determine that if the number of draws until a successful green marble is less than 3, it is more likely being drawn from an urn with $X \sim \mathbf{Geo}(.35)$ than it is from one with $X \sim \mathbf{Geo}(.2)$ with a log-likelihood of .14 with a 95% confidence interval. If the number of marbles drawn until successfully drawing a green ≥ 8, then we have sufficient evidence to reject the null hypothesis $\mathbf{H}_0$ and therefore accept the alternative hypothesis $\mathbf{H}_1$.