PHSX815 Project 3: Maximum Likelihood Estimation for Bernoulli Distributions

Craig Brooks

April 2023

Revisiting Urns: N Bernoulli Trials

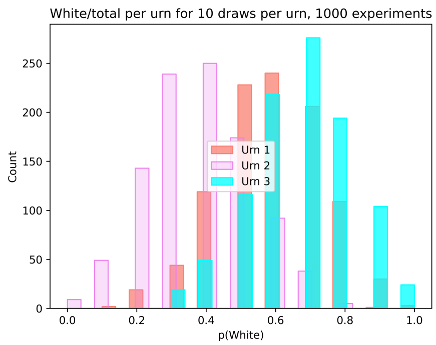
Suppose we have 3 urns, each containing White and Black marbles, and the ratios of White to Black marbles is unknown. The task here is to estimate the most likely ratio of White to Black marbles for each urn from random sampling. We will accomplish this by drawing 10, 100, then 1000 marbles from each urn. We can then calculate the White(Black) fraction to the total number of marbles drawn. We can then plot histograms of the White(Black) fractions for each urn, then calculate the maximum likelihood for each urn.

Distribution

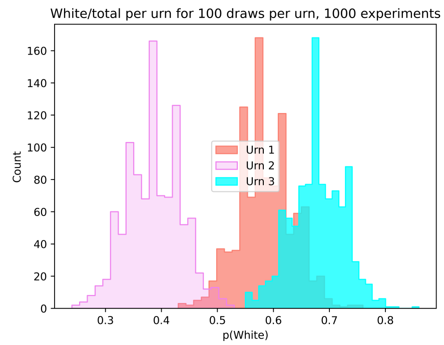
In our experiment, the parameter we are maximizing is p, which we will define as the fraction of White marbles in the urn. Since we only have two possible categories, White and Black, the most natural distribution would be to use the binomial distribution. In the binomial distribution, we would track the counts of the number of White marbles and Black marbles in n draws

However, while the binomial distribution is suitable if we wanted to know the likely count of White marbles given a fixed n, we are only concerned about the parameter p, we can instead use the Bernoulli distribution. Since the binomial distribution is simply a series of Bernoulli trials, we can simply track success as 1 (marble is White) or 0 (marble is Black) for an arbitrarily large n.

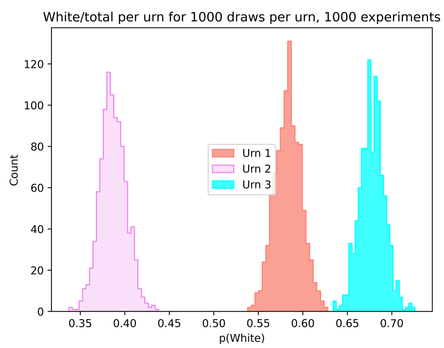
Figure 1 plots histograms for binomially distributed White and Black marbles for each urn for $n = 10, 100, and$ 1000 marbles drawn from the urn over 1000 trials. As we can see, it seems we should be able to distinguish significantly Urn 3 from Urns 1 and 2. Additionally, it is clear in all cases, the most probable outcomes for White marbles tend to cluster around $p \approx .57$ in Urn 1, $p \approx .38$ Urn 2, and $p \approx .68$ for Urn 3. Naturally, since the marbles are binomially distributed, for Black marbles $p \approx .43$ in Urn 1, $p \approx .62$ Urn 2, $p \approx .32$ Furthermore, as we increase the number of draws in each experiment, the histograms become more peaked and the FWHM becomes narrower. The analysis in the following sections will determine how good these *visual* estimates from the histograms are.



(a) 10 draws

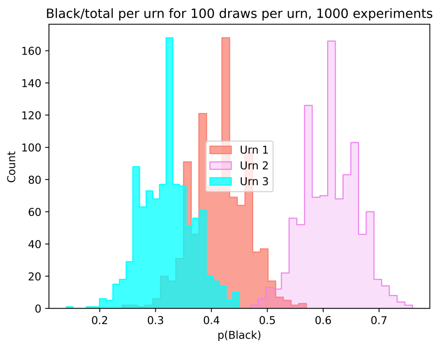


(b) 100 draws

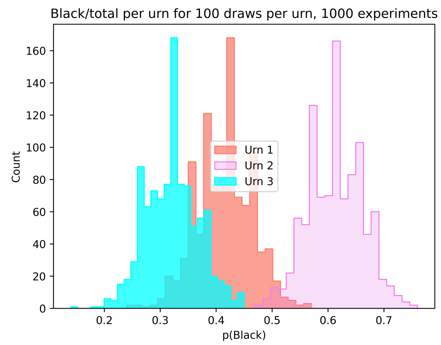


(c) 1000 draws

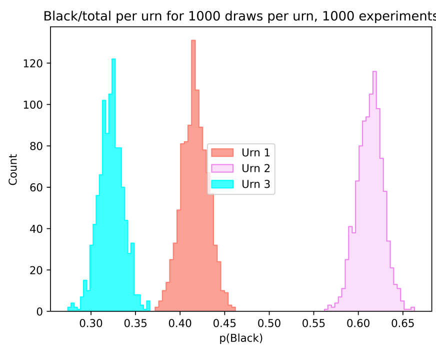
Figure 1: Histograms of the fraction of White marbles for different number of draws



(a) 10 draws



(b) 100 draws



(c) 1000 draws

Figure 2: Histograms of fractions of Black marbles for different number of draws

Algorithm Analysis

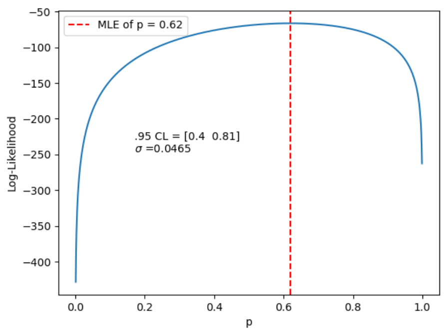
First, using the script called `urn_mle.py`, we create 3 urns having different ratios of White to Black marbles, and store them in an array called `urns`. We set the number of marbles with `N_marbles_urn = 10000` Then, we pass this into a function called `bern(p, n)` to generate `trials` number of urns (By default `trials = 10`) from each urn ratio and stores in an array called `outcomes_list`. Next, for each urn generated, `N_marbles_sample` are drawn for each trial for each urn. Then, the number of White and Black marbles are computed and stored in a file called *urn_{urn}_data_mle_{n}.csv*

Next, a script called `urn_mle_analysis.py` will read in the data that the user inputs, then calculated the maximum likelihood estimate from the table data. To do this, we will minimize the Bernoulli likelihood function given our data-sets. We then calculate the uncertainty and 95 % confidence interval for the experiments.

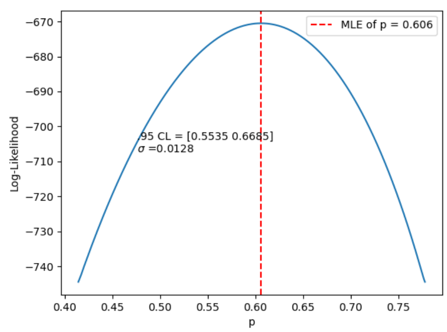
Results

Below are plots that show MLE for p (White) and for each urn for 10, 100, and 1000 draws. and for 100 draws, where the p_{true} in Urn 1 as $p = .584$, Urn 2 as $p = .387$, and Urn 3 as $p = .678$. It should be noted that for 1000 draws, the plot only shows the maximum likelihood estimate, and not the likelihood plot, as the likelihood isn't finite for that many trials. However, for the sake of demonstration and completeness, they have been included.

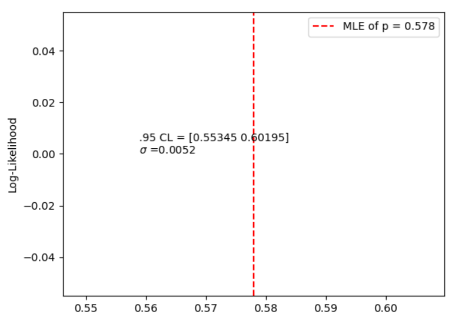
In Urn 1, the MLE was determined to be $p = .620 \pm .047$ for 10 draws, $p = .606 \pm .0128$ for 100 draws, and $p = .578 \pm .0052$ for 1000 draws.



(a) 10



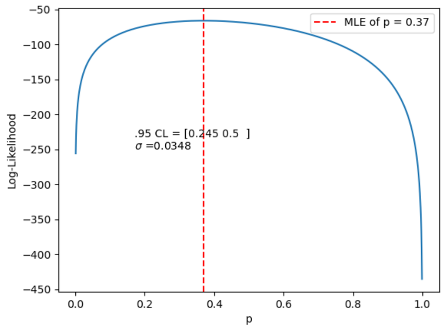
(b) 100 draws



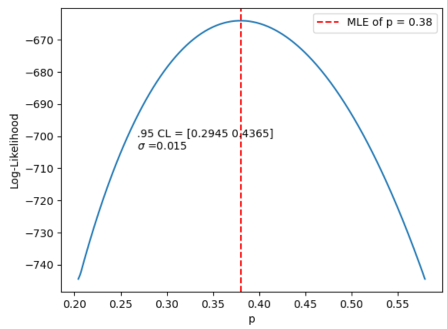
(c) 1000 draws

Figure 3: MLE for Urn 1, all trials

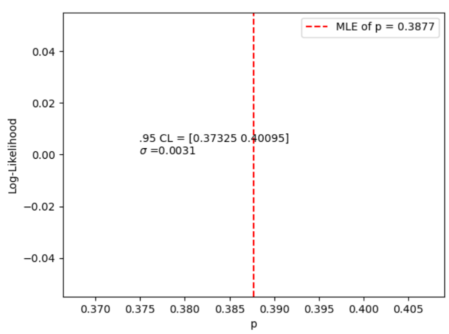
In Urn 2, the MLE was determined to be $p = .37 \pm .035$ for 10 draws, $p = .380 \pm .015$ for 100 draws, and $p = .3877 \pm .0031$ for 1000 draws.



(a) 10



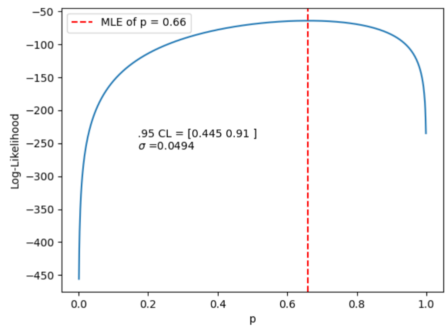
(b) 100 draws



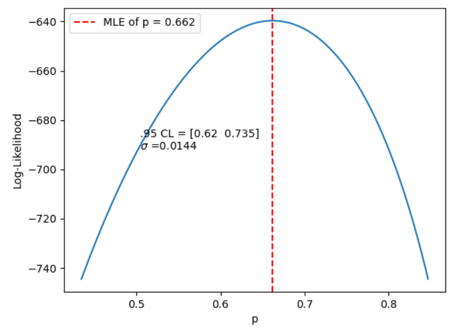
(c) 1000 draws

Figure 4: MLE for Urn 2, all trials

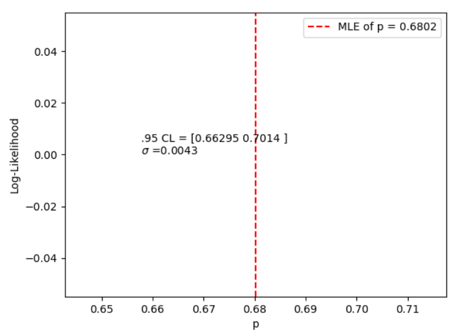
The results for Urn 3 are shown in Figure 5. The MLE was determined to be $p = .660 \pm .049$ for 10 draws , $p = .662 \pm .0144$ for 100 draws, and $p = .6807 \pm .0043$ for 1000 draws.



(a) 10



(b) 100 draws

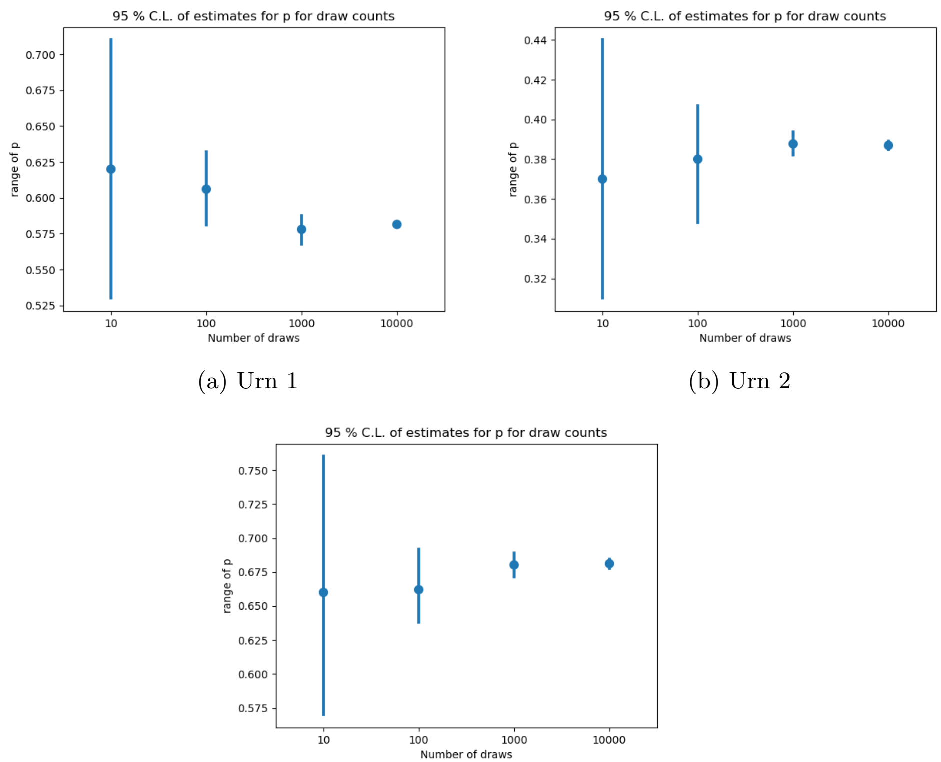


(c) 1000 draws

Figure 5: MLE for Urn 2, all trials

Uncertainties

Figure 6 shows the 95% confidence interval across all trials for each urn. It can clearly be seen that as the number of draws increases by a factor of 10, the uncertainty in our measurement decreases. In particular, it decreases by a factor of $1/\sqrt{n}$ For example, in Urn 1, the 95% confidence interval in 10 samples is $[.530, .775]$ for 10 draws. However, if we draw 100 marbles, the 95% confidence interval is $[.590, .640]$. It should be noted that $n = 10000$ was included in these plots to illustrate the size of the uncertainties, but that the differences in measurement for $n = 10000$ were within the uncertainty for $n = 1000$



(a) Urn 1

(b) Urn 2

(c) Urn 3

Figure 6: 95 % confidence interval for each urn, all trials

Conclusion

We have demonstrated that the MLE for each urn reaches a better approximation as the number of draws increases. Using $n = 1000$ for our best estimate of the maximum likelihood, the MLE_{best} for Urn 1 is $p = .578 \pm .0052$, for Urn 2 $p = .3877 \pm .0031$, and for Urn 3 $p = .6807 \pm .0043$.