Project 2 Peer Review Input: The Polya Urn

Craig Brooks

March 2023

Introduction

In this project, we are returning to urn modelling for Project 1. To recap, in Project 1, we constructed two different urns, both of which had marbles of 3 different colors with a multinomial distribution. In one of the urns, the probability of picking green in $< k$ tries was higher. While we knew the probabilities prior to drawing, we did not know which urn contained which probability.

To determine the likelihood for which urn was which, we drew from an urn until we successfully drew a green with replacement, and based on the kth draw, determine the likelihood that we drew from the higher/lower probability urn.

In the Polya Urns, we are still drawing from urns, but what is not known are the probabilities from which we are drawing from. Instead, we know some other priors α_k, which will produce an ensemble of urns of 3 colors with multinomial distributions within the urn, but across the entire ensemble, each individual color may follow an entirely different distribution.

The Polya Urn

The Polya urn is a way to model sampling from 'real world' objects like particles. Unlike in simpler models, where you may draw marbles from an urn with or without replacement, when a draw is made from the Polya urn, one of the same color is added back into the urn. Therefore, the more often a color is drawn, the more likely one is to draw in future draws.

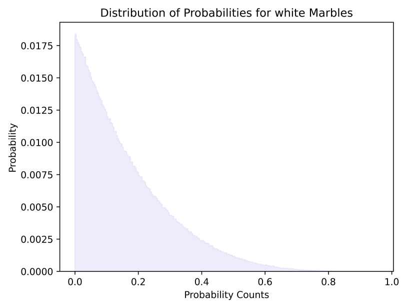
Probability Distribution

While within urns, the marbles are multinomially distributed, the conjugate prior follows a Dirichlet-multinomial distribution. The key difference between a 'normal' multinomial distribution and the Dirichlet version is that the *probabilities* for each of the marble colors follow a distribution based on a parameter α.

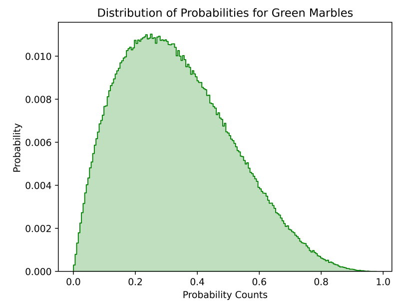
The marginal distribution for probability vector $\mathbf{p}$ is.

$$Pr(\mathbf{x}|n,\alpha) = \int_p Multi(\mathbf{x}|n,\mathbf{p})Dir(\mathbf{p}|\alpha)d\mathbf{p}$$

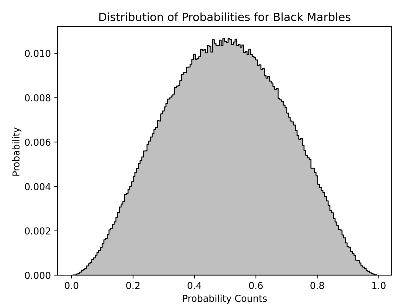
where $\mathbf{x}$ is a vector of category counts, $\mathbf{p}$ is a vector containing the probability for each category, n is the number of trials, and α are the 'priors' that we know, which are essentially the group ratios for the means of each color probability.



(a)



(b)



(c)

Figure 1: Species distributions for white, green, and black marbles from Dirichlet multinomial distribution

Experiment

What we will do for our experiment is to choose some $\alpha_k{}'s$ which correspond to the dispersion of the colors within the ensemble. The $\alpha_k{}'s$ can also be seen as the group ratios of the means for each category $\mathbf{X}$ we are considering. We sample from the DM distribution for n trials.

We will then sample from two generated urns and draw from each urn 20 times for M trials. We will then try to determine the likelihood of drawing each urn $\alpha_k{}'s$ We can accomplish this by calculating the likelihood of each species probability for each urn, and compare them.

An experiment such as this can be a way to model how populations can evolve due to natural selection. we can compare subgroups within a population and determine how representative they are of the overall population.