Chinese Restaurant Problem

Marques Chacon, Zoheb Sidiqui, Nile Wynar

June 18, 2019

1 Introduction to Topics

1.1 Pólya's Urn

Imagine that you have an urn with one black ball and one red ball. Next, you randomly sample one of the balls in the urn. If you draw a red ball, you put it back in the urn while adding another red ball. Similarly, if you draw a black ball, return it to the urn and add another black ball. The next time you draw a ball, the probabilities of drawing a red or black ball will change depending on the proportion of red and black balls currently in the urn.

Suppose we draw 5 balls from the urn – 3 of which are red. There are multiple ways we can draw this sequence:

1. RRRBB
2. RBRBR
3. BRRBR
4. ...

Let's consider 2 of these combinations: What is the probability we draw RRRBB in that order? Since we started with 1 black and 1 red ball, then the probability is $\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} \cdot \frac{1}{5} \cdot \frac{2}{6} = \frac{12}{720} = \frac{1}{60}$ What is the probability we draw RBRBR? $\frac{1}{2} \cdot \frac{1}{3} \cdot \frac{2}{4} \cdot \frac{2}{5} \cdot \frac{3}{6} = \frac{12}{720} = \frac{1}{60}$

In fact, no matter which sequence containing 3 Red that we draw, the probability of drawing any such sequence is $\frac{1}{60}$. Hence, the probability of drawing 3 Red in 5 draws is $\binom{5}{3} \cdot \frac{1}{60} = \frac{10}{60} = \frac{1}{6}$

We can generalize this for drawing any number of balls. Let R_n = number of red balls you draw out of n draws. Then $P(R_n = k) = \binom{n}{k} \cdot \frac{k!(n-k)!}{(n+1)!} = \frac{1}{n+1}$ Thus, R_n is uniform on $\{0, 1, \dots, n\}$. Taking the limit $n \to \infty$, we find that the proportion of red balls that you draw converges to the Uniform $(0, 1)$ distribution.

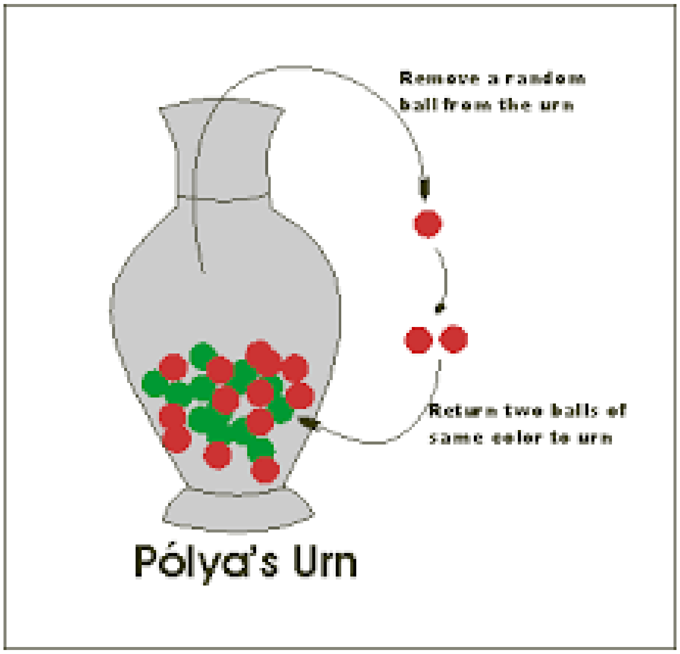

Figure 1: This is a visualization showing how the Pólya Urn process works at each step.

2 Chinese Restaurant Problem

2.1 Original

With Pólya's urn, we were limited to only a fixed amount of colors. However, we can imagine the urn to start with a rainbow-colored ball representing a color not yet in the urn. If at any given time we draw the rainbow ball, you add a color not yet represented in the urn. In this way, we can have infinite colors and represent the Chinese Restaurant Process exactly.

In the original Chinese Restaurant Problem (CRP), customers arrive at the restaurant in discrete time and no customers leave their table once they are seated. Customer n chooses to sit at a new table with $\frac{1}{n}$ probability and they choose to sit at a table with m people with $\frac{m}{n}$ probability. Each table can hold infinite customers and the restaurant can hold infinitely many tables. If we think of tables as different colors and each ball representing a person sitting at that colored table, Pólya's Urn with the addition of the rainbow ball is exactly the same process as this original CRP.

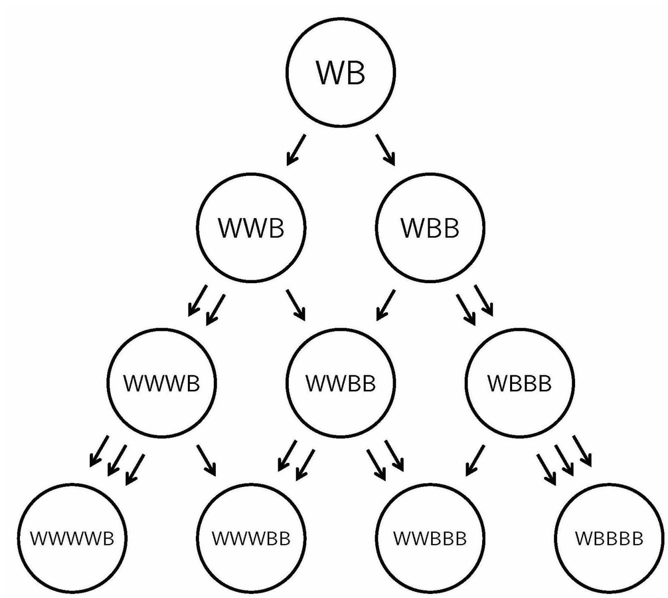

Figure 2: This tree represents an urn that starts with a white ball (W) and a black ball (B). It shows the possibilities of adding a W or B based off the current state of the urn.

2.2 (α, θ) Generalization

In this version of the CRP instead of a $\frac{1}{n}$ probability of customer n sitting at a new table there is a α probability of sitting at a new table to the right of any occupied table. Also, for a table with m customers occupying it, there is a $\frac{m+\alpha}{n+\theta}$ likely hood of the n^{th} customer sitting there. Additionally there is a θ probability of sitting at a new table to the left of all of the existing tables. Figure 3, is a visualization of this process with two tables, the first with 3 customers and the second with 2 customers, represented by blue ovals. For our project, we focused on a non-zero α and zero θ, but in the future we will implement a non-zero θ.

2.3 Ordered

The ordered CRP is a process we used to set up the initial tables in our simulation. This involves keeping track of the ordering of the tables and being able to add tables to the right of existing tables as in the (α, θ) generalized CRP. No customers may leave the table once they sit and so this process stops when the restaurant reaches a maximum occupancy limit. This process might stop with many small tables and a few larger tables and in that case the small tables would be ignored and not used in the initial set up of the simulation.

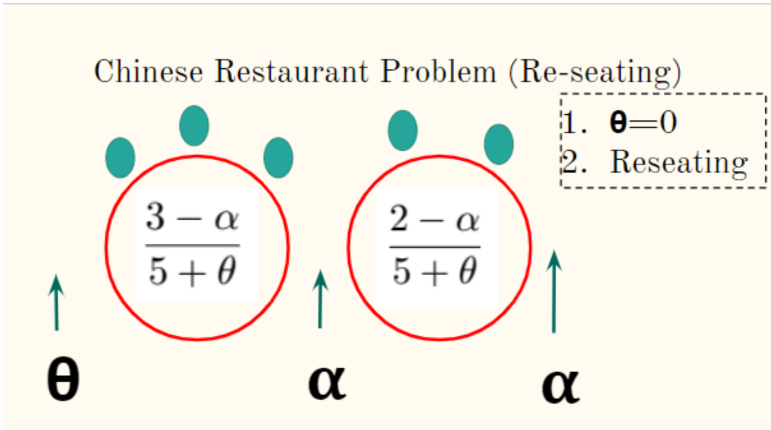

Figure 3: This is a visualization of the (α, θ) generalized CRP with two tables and 5 customers in the restaurant.

2.4 Re-seating

In the re-seating CRP, customers have the option of leaving the restaurant after they have sat down at a table. So as customers enter and exit the restaurant, tables will gain and lose customers throughout time. When a table loses all of its customers, we say that the table has "died." This is the problem that we simulated.

3 Scaffold Spindle Skewer

3.1 Simulation

3.1.1 Spindles

Tables in a CRP simulation needs to encode the number of people sitting at the table at any point in time. This information is illustrated in a spindle. Each spindle is a polygon that represents one table. The width of the spindle corresponds to the number of customers sitting at it. Therefore, the height of a spindle is the amount of time that the table has customers sitting at it. Each spindle is a different color which represents different tables.

3.1.2 Scaffolding

It is advantageous to visually represent the relationship between tables. When a table is created to the right of an existing table, it is considered the child of the table to its left. In this simulation, this parent-child relationship is represented by lines of a constant slope connecting the tops and bottoms of spindles. In addition the scaffolding's slope is a way to visualize the table losing its' customers

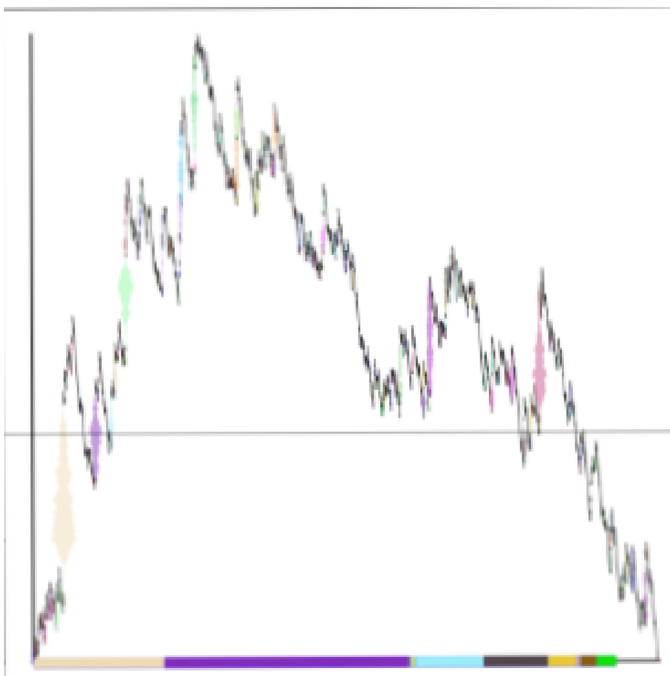

Figure 4: This is an image of a skewer through a Scaffolding-Spindle-Skewer image along with the interval partitioned into the different tables indicated by different colors.

before it was able to give birth to a new table. If the scaffolding attached to the top of a spindle reaches the bottom of the spindle before a new child was produced, the spindle did not have any children. A simple picture of this is in Figure 5.

3.1.3 Skewers

In the scaffolding spindle picture, the y-axis shows how table sizes change over time. In this way, drawing a horizontal line so that it slices through spindles at a certain point in time, the state of the restaurant can be reproduced. Since the width of each table represents the number of customers, we can visualize the entire restaurant with an interval partitioned into different colors representing different tables. Each color covers more or less of the interval depending on the size of that table. Figure 4 is an example of this.

3.2 Depoissonization

In order to animate our simulation, we need to sample multiple skewers over time, which is the y-axis in the Scaffolding-Spindle-Skewer picture. We need

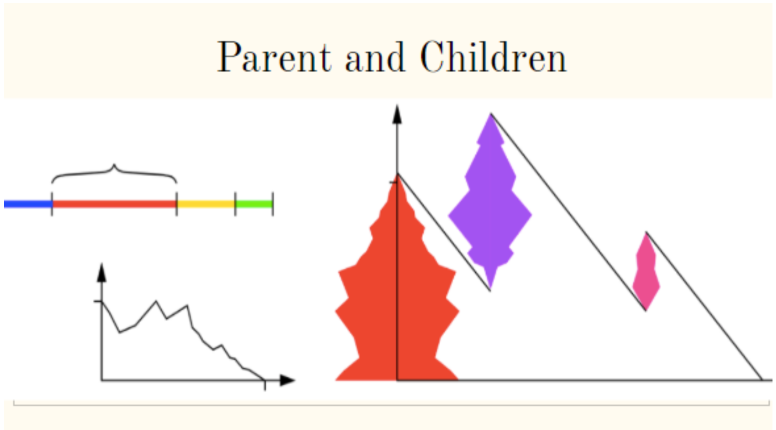

Figure 5: This is a visualization of one parent and two children in a scaffolding spindle and skewer picture. To the left, there is a line graph for the number of people sitting at the red table over time which is encoded entirely by the red spindle in the scaffolding spindle skewer picture.

to decide the spacing between the skewers we sample from our picture. In the CRP, customers enter the restaurant in discrete time. However, in our simulation customers enter the restaurant in continuous time. To account for this we need a time change from continuous to discrete time. In the end, this translates to

$$y_{i+1} = yi + \beta_i * du$$

where y_i is the height at step i, β_i is the number of customers sitting in the restaurant at step i, and du is a "small enough" number.

3.3 Biological Story

Another way to visualize the CRP overtime is through particles in space. In this visualization, instead of tables being represented by spindles of changing widths, they are represented by circles of changing diameters. Because of this, the only way to represent the full restaurant over time is with an animation where the circles change diameters. Additionally, the circles move in the two dimensional space with random Brownian Motion. When a parent table gives birth, its child starts at the same coordinate as its parent, however it starts moving independently from its parent immediately.

In this visualization there is additional information about the gene space of tables. If we think about parent tables as species and children as genetic mutations of their parent species then the two dimensional space of these particle tables in space can represent the changes in the tables genetics over time.

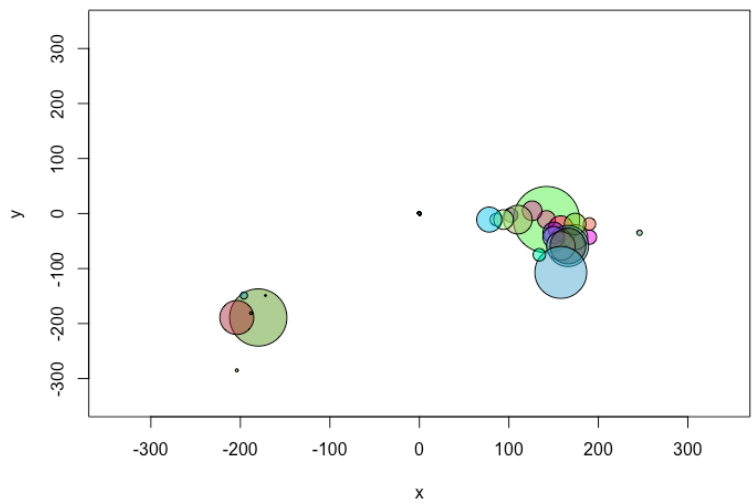

Figure 6: This is a snapshot of the particle picture.