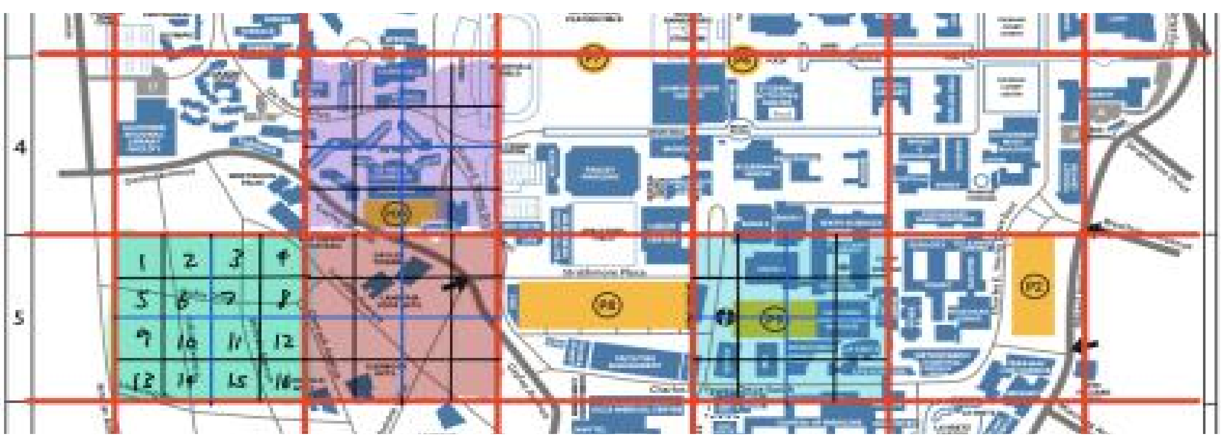
Kuan-Ting Chen
Victoria Delk
Aaron Lee
Charles Zhang

Stats 100A Homework #3

Part 1: Poisson Processes and Linear Programs for Allocating Shared Vehicles

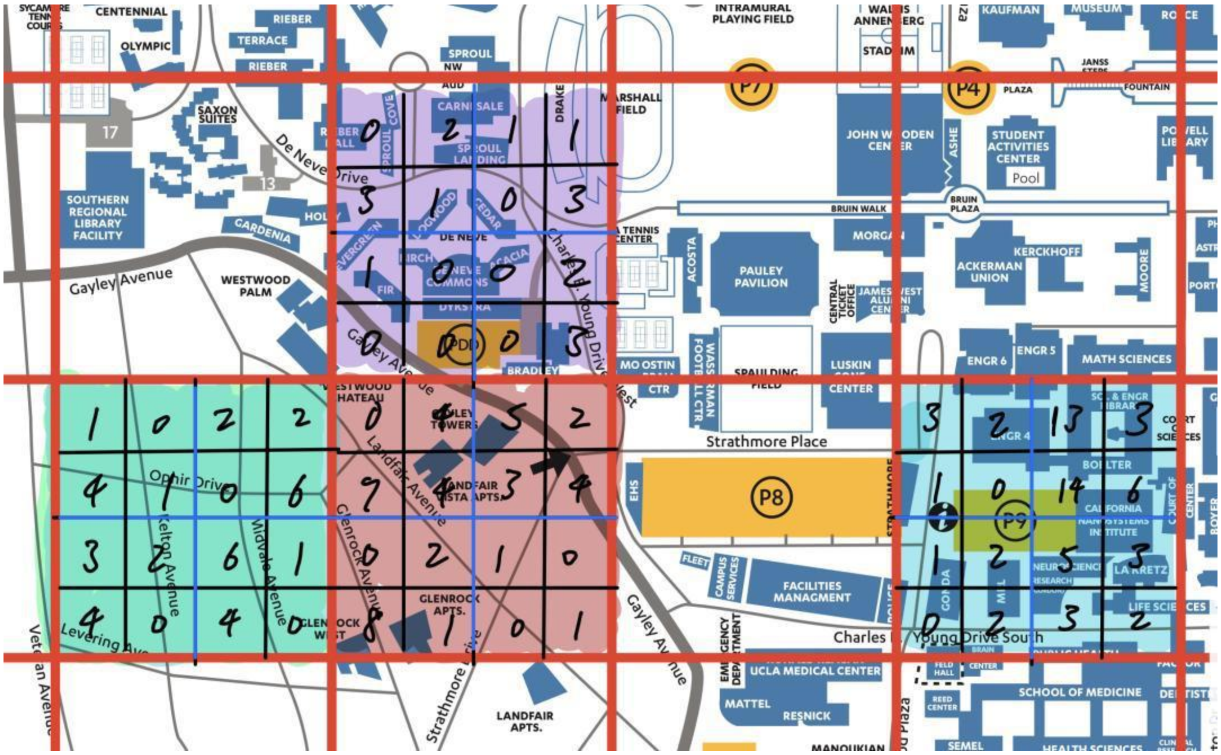
In the article about micro-mobility, Fields makes a distinction between utilization and demand, the latter being the set of rides customers would have taken if there were infinite vehicles everywhere. Companies collect scooter location data to analyze and target consumer demand using non-homogeneous Poisson point processes. Therefore, the number of consumers desiring rides within an area in a given time frame is Poisson distributed with its expected value as a function of time and area. Two properties of Poisson processes that we can capitalize on in modeling micro-mobility are Poisson subprocesses and exponential interarrival times. Considering locations A and B, the arrival rates follow a Poisson process with rates λp and $\lambda(1 - p)$. If there are k vehicles within a bounded location, then each will have $\frac{1}{k}$ the rate of the overall process. By joining these subprocesses – vehicle-specific, time-specific, and area-specific – the overall arrival rate and demand for the entire cell can be modeled.

Part 2: Setting up the Experiment



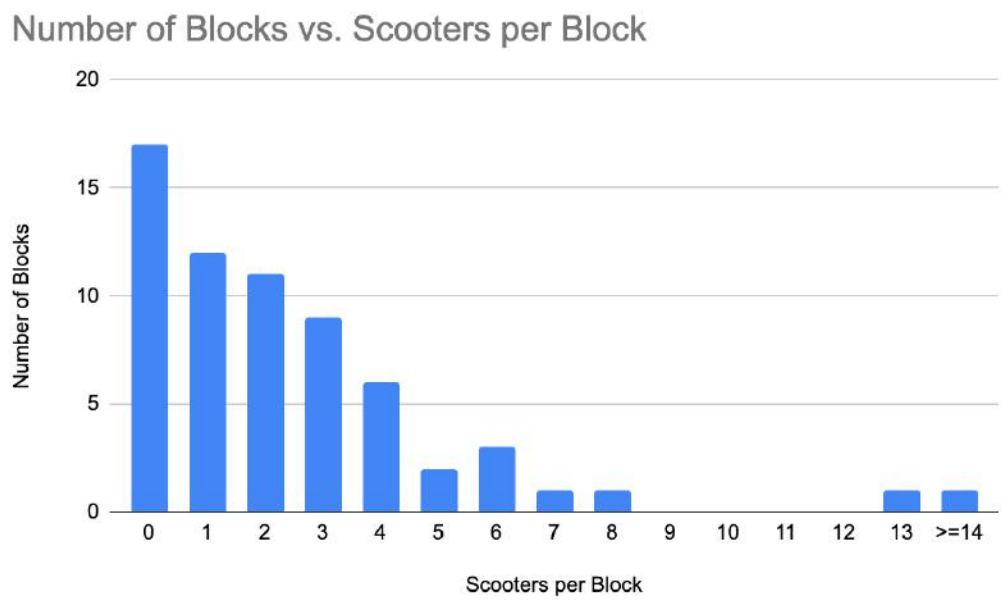
All observations were made from approximately 3 - 4 p.m. on Wednesday, May 25th. Kuan-Ting Chen covered the purple area, Victoria Delk the green area, Aaron Lee the blue area, and Charles Zhang the red area. Each area was split into a grid of 16 squares numbered from 1 at the top left to 16 at the bottom right corner, as shown in the green cell, resulting in 64 total observations.

Part 3: Determining Uniform Scatter



	Scooter Count			
Cell #	Kuan-Ting Chen	Victoria Delk	Aaron Lee	Charles Zhang
1	0	1	3	0
2	2	0	2	4
3	1	2	13	5
4	1	2	3	2
5	3	4	1	7
6	1	1	0	4
7	0	0	14	3
8	3	6	6	4
9	1	3	1	0
10	0	2	2	2

11	0	6	5	1
12	2	1	3	0
13	0	4	0	8
14	0	0	2	1
15	0	4	3	0
16	3	0	2	1

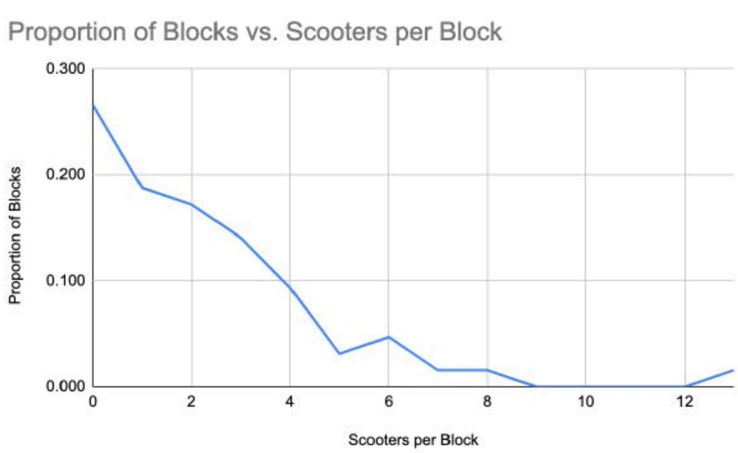


Based on the recorded data, we shouldn't proceed using a Poisson process. The data above is clearly not uniformly distributed, with multiple blocks having significantly more scooters than others. As a result, it is unlikely that a Poisson model would be appropriate for this dataset.

Part 4: Fitting a Poisson Model to the Data

# Scooters	Observed Tally	Total Skooters	Empirical Probability	Theoretical Probability	Expected	(O-E)^2	[(O-E)^2] / E
0	17	0	0.266	0.089	5.68	128.135	22.558
1	12	12	0.188	0.215	13.76	3.087	0.224
2	11	22	0.172	0.260	16.66	32.023	1.922
3	9	27	0.141	0.210	13.45	19.790	1.472
4	6	24	0.094	0.127	8.14	4.591	0.564
5	2	10	0.031	0.062	3.94	3.780	0.958
6	3	18	0.047	0.025	1.59	1.982	1.245
7	1	7	0.016	0.009	0.55	0.202	0.366
8	1	8	0.016	0.003	0.17	0.694	4.164
9	0	0	0.000	0.001	0.04	0.002	0.045
10	0	0	0.000	0.000	0.01	0.000	0.011
11	0	0	0.000	0.000	0.00	0.000	0.002
12	0	0	0.000	0.000	0.00	0.000	0.000
13	1	13	0.016	0.000	0.00	1.000	11113.650
>=14	1	14	0.016	0.000	0.00	1.000	53998.144
Totals:	64	155	1.000	1.000	64.00	196.287	65145.327
Expected Scooters / Block:							
2.422							

The Chi-square statistic is the sum of $\frac{(O-E)^2}{E}$, which equals 65145.327 for our group's scooter data. From the Chi-square distribution applet, we calculate the probability of a Chi-square with 14 degrees of freedom to be greater than 65145.327 as essentially 0. In other words, since the maximum number of scooters observed in a single cell is 14, $P($"Chi-square with 14 degrees of freedom" $> 65145.327) \approx 0$. Because the P-square statistic of 0 is not greater than 0.05, we conclude that a Poisson model with the parameter $\lambda = 2.422$ not a good fit for the scooter distribution data. The parameter λ, 2.422, comes from the expected number of scooters per block, calculated by dividing the sum of total scooters by the number of blocks observed between members.



Part 5: Conclusions of the Experiment

In conducting this experiment, we found the number of scooters that were available for people to ride, or in other words, the supply of scooters available to students. This would logically mean that the demand of scooters, would be the number of scooters in use.

From the article, the most significant point of contention between the author's words and our results lay in the use of the Poisson distribution. The author mentions that they use the Poisson distribution as the basis for their calculations and models. However, we found that the Poisson distribution was actually not a good fit for the distribution of scooters. The second major point of comparison between our results and the article is that the article mostly focuses on calculating demand with a Poisson model, not supply as we did. Perhaps in our case, the demand for scooters in UCLA could fit the Poisson model. Regardless, being able to choose the best model, and knowing what you are trying to model are both very important factors in building an accurate and representative model, especially in the business world where decisions that can cost lots of money are made based on these types of results.

One pro was that we all coordinated to count the scooters at the same time. This allowed us to remove any biases such as rush hour when everyone is running to class or at nighttime when all the scooters are taken away to be charged. Another pro was that we were able to split our coverage area over a rather large part of campus, and as we know from the Law of Large Numbers, the more area and scooters we consider, the closer we will arrive at the true probability.

The largest con for this project is the potential that we could have accidentally double counted scooters. For example, despite the team counting the scooters at the same time, we did have to each cover a rather large area, so it is possible that a scooter in one person's area could have been counted and then taken to another area and then counted again without us knowing. This would bias our results and make it so that the distribution of scooters is not independent in each area. Idealistically, the best way to improve this would be to have no one use the scooters and then count them but since that is impossible, one possible solution is that we could take pictures of the area so that we could cover more land faster and then count the scooters later. In effect, this would make it so that we spend less time counting scooters on site so that the chances of scooters being double counted would be minimized.

Since the λ in a Poisson distribution is both the expected value and variance, we can think of our variables from that perspective. Some examples of variables that we could incorporate into our calculation of λ could be the battery percentage of the scooters. This would determine how many are left out and how many are taken away for charging, which would affect the overall supply of scooters. Another variable could be the number of broken scooters. This too would affect the supply of scooters as broken ones would be taken away for repair and therefore not usable. One more variable to consider could be the number of scooters in use. Although we are considering only supply for this experiment, the rules of economics still play a part, and as demand grows, so does the supply of scooters.

Part 6: Brain PMFs

The goal of the experiment in the paper was to determine if tonal inputs could show how the brain builds global probability distributions and use that global representation to process new information. The way a brain is theorized to produce a probability model, and then apply it to resolve new queries, is similar to how the companies that provide micro-mobility services would like to use probability models based on historical data to predict future trends. Both experiments act to predict relationships between past and future behaviors using probability models (normal and bimodal models in the brain study and Poisson models in the scooter study). While our experiment attempts to provide information that a company could then use to improve efficiency, the brain study focuses on how individual brains use this process in relation to auditory sensation.

Both the brain experiment and the scooter experiment used a university setting, with the brain experiment pulling subjects from a population of university students, and the scooter experiment pulling data from a university location. In both experiments, data was first collected with the intention of later fitting that data to a specific model (normal/bimodal in the brain experiment and Poisson in the scooter experiment). After data collection, the data was then fit to the specified models to conclude whether or not the models accurately fit the gathered data. In the case of the brain study, successful fitting determined that brains are able to build probability distributions and use them to process new information. In the case of the scooter study, successful fitting would determine the validity of using the properties of Poisson processes to infer latent demand in the micro-mobility industry.

Victoria Delk Example:

An example where my brain used probability distributions is determining when to drive based on traffic patterns. Certain time frames like rush hour or holidays create a high probability of running into traffic. Thus, I have to reference the time when predicting how long it will take for me to drive from point A to point B. If I have to drive across LA, I can use a model built off of my past driving experiences to know that driving at rush hour will take significantly longer than driving at midnight.

Charles Zhang Example:

I think that an example of when my brain would use a probability model to conclude something would be determining what to wear based on the weather. Past experience tells me that the temperature varies depending on the season. Therefore, my internal model tells me that it is much more likely to be warmer in the summer and colder in the winter. This is a model based on my past experiences of living in the area, and can be used every day to predict what the weather will be like in a given part of the year.

Kuan-Ting Chen Example:

I noticed that I would use models in my head when deciding whether I have enough time to play a match of a video game. For example, I know that the average match takes about 40 minutes, but that sometimes the opponent could surrender within 15 minutes or that we might have multiple overtimes, causing the game to last well over 2 hours. I soon realized that this was representative of a normal distribution where the average match length is in the middle and the extremes are less likely. Since I knew the average match length and the amount of time I could spare to play, I based my decision on whether or not I had time to play off of this model. If I only had 15 minutes, it was unlikely we would finish fast enough, but if I had an hour, then I would likely choose to play since the game would likely be over by then.

Aaron Lee Example:

I use a probability distribution to help me decide the optimal timing to leave my room in order to get to class on time. Many factors affect the time it takes to walk, including my walking speed, number of people I run into along the way, and the distance between my room and the class. For example, walking speed and distance can be combined into a Gaussian, or normal distribution, where it takes me an average of 10 minutes per mile with a standard deviation of 4 minutes per mile. Running into acquaintances along the way adds about 5 minutes per person met, and the number of people I run into can be modeled with a similar model that we tested the scooter distribution data against: a Poisson model. Here, I estimate that the λ parameter of expected number of people met along the way would be equal to 2. My brain then combines these two distributions to form a more accurate model in order to predict the amount of time it will take me to get to class with a given distance. As I keep walking to class and collecting time data, my brain fine-tunes the existing model using backpropagation by adjusting parameters, such as the mean and standard deviation of walking time per mile as well as the λ parameter for average number of people met along a trip, which helps me better predict walking time in the future.