CMUSR

1:00 - 1:10

Charles Grant Beck, Stanley Jian, Catherine Lyu, Harrison Wang
Faculty Mentor: Milind Hegde

Asymptotic Behavior of Last Passage Percolation on Complete Graphs

Abstract. In this project, we examined a complete graph where each edge was assigned a random number using the same distribution, and we considered the set of all paths between two vertices on the random graph. For each path, we add the weights of each edge, giving us a total. Using various bounding methods (both lower and upper bounds), we could examine the behavior of this weight for different distributions. For the upper bound we consider the largest n edges in the graph, and use these edges as an upper bound. One method we used for our lower bound was to simply consider the largest edge in the graph. This method works well for heavier tailed distributions. Another method we used was to construct a greedy algorithm (an algorithm which generates a path of large weight), and then describe how such a greedy algorithm would grow. Finally, another method used Random Graph theory to prove that if we consider the $Cn\log(n)$ largest edges in a graph, then there will exist a path of roughly length n. Using these methods, we found that for the exponential and Rayleigh distributions, W_n grew like $n\log(n)$ and $n\sqrt{\log(n)}$, respectively. For the power law we found that W_n grew like $n^{2/\alpha}$, $n^2\log(n)$, and $n^{1+1/\alpha}$ for $\alpha < 1$, $\alpha = 1$, and $\alpha > 1$, respectively.

CSUREMM

1:15 - 1:25

Brianna Han, Carlos Mendonca, Andrada Nicolae, Jane Zhang
Graduate Mentor: Enlin Qian

Using Stochastic Processes to Characterize the Metastatic Progression of Breast Cancer

Abstract. Due to the biochemical, physiological, and genetic complexity of cancer metastasis, a deterministic model which accurately predicts metastatic progression of breast cancer is not currently feasible. Therefore, we utilized a probabilistic approach, modeling the presence of cancer at sites within the body as states in a Markov chain. We used a transition matrix to represent the probabilities of metastasis from one site to another. The algorithm used to obtain such a transition matrix is inspired by an iterative adjustment method used in the analysis of lung cancer metastasis. Additionally, we sought to improve current models by further developing Monte Carlo simulations to more adequately account for a patient's metastatic history. After running 1000 simulations, we recorded the average time that each site was first reached. We were able to deduce other clinically relevant information from our simulations, such as the most likely metastatic pathways given metastases at specific sites.

1:30 - 1:40

Samuel Farrell, San Winter, Robert Young
Graduate Mentor: Enlin Qian

Determining an Optimal Route for a Futuristic Probe to Visit Earth-like Exoplanets

Abstract. As the number of discovered exoplanets increases, many dream of finding life on another planet. Although it is known many specific conditions are necessary for the development of life, many of these conditions are difficult to measure from Earth. This has led to questions about how to identify habitable exoplanets using data available from remote sensing techniques. In a 2011 paper, Schulze-Makuch, Dirk, et

al. defined the Earth Similarity Index (ESI) as a method for determining exoplanet habitability. The ESI metric is calculated using values that are easily attainable via remote sensing, like planetary radius and mass, to determine an exoplanet's percent similarity to Earth on a scale from 0 to 1. However, this metric fails to incorporate several factors that are harder to attain, such as the chemical composition of an exoplanet's atmosphere. The aim of this research was to determine a route to visit high-ESI planets in order to collect such data. The first method creates a route to all exoplanets with an ESI higher than Mars (ESI = 0.70). In the result for the first method, the program finds a route that has an overall distance of 1971.36 light-years. However, this high travel distance led to a second approach, which chooses the visited planets via a profit function. The profit function is based on a theoretical "reward" reaped on each exoplanet (based on ESI) and the "cost," which is akin to distance. This approach resulted in a 60.06 light-year route that visits a total of 7 exoplanets in 2 different star systems (Trappist-1 and GJ 273).

1:45 - 1:55

Jack Henry Cleeve, Ekene Ezeunala, Gilead Turok
Graduate Mentor: Joe Suk

Extending the Sparse Projections Fine-mapping Model by Learning Genomic Annotation Weights

Abstract. The problem of variable selection in genetic fine-mapping is that of identifying the genetic variants that causally affect some given trait. The sparse projections (SparsePro) model accomplishes a solution to the fine-mapping problem of genotype-phenotype causality by representing regression coefficients as probability distributions in a linear regression framework. These distributions are approximated with variational inference and their parameters are learned with the coordinate-based optimization algorithm CAVI. We modify SparsePro to instead fit these variational parameters with the gradient-based optimization algorithm Adam, resulting in an alternative approximation to the true posterior distribution. We evaluate our model using simulated data obtained from GWAS.

BREAK

2:05 - 2:15

Daniela Perez, Krystal Yimeng Sun, Shivani Tripathi
Graduate Mentor: Xiaoxiao Sun

The Impact of Climate Change Induced Natural Disasters on Education and Child Marriage Rates

Abstract. This research models the relationship between the frequency and intensity of climate change induced natural disasters and education enrollment rates and child marriage rates on a global scale. Previous studies show that exposure to floods was correlated with a decrease in the average number of completed grades of children in Ethiopia, Vietnam, and India. Our generalized linear mixed model aims to quantify the effects of all climatological, meteorological, and hydrological natural disasters on children in as many countries as possible, and takes confounding variables including life expectancy and government expenditure on education into account. Our current results have demonstrated a statistically significant negative correlation between natural disaster frequency and gross primary enrollment rates. A similar negative correlation was found between frequency and net primary enrollment rates when a 7 year enrollment rate lag was taken into account, illustrating the long-lasting effect of natural disasters. The results of this study could be used to advise countries' post-disaster responses pertaining to education and child marriage, while also raising awareness about the multi-dimensional consequences of climate change on our most vulnerable population and the planet's future - children.

2:20 - 2:30

Makoto Powers, Ethan Turok, Haoran (Dragon) Yuan
Graduate Mentor: Joe Suk

Data-Driven Newsvendor

Abstract. We study the classical newsvendor problem in which the decision maker does not have access to the underlying demand distribution. Sample Average Approximation (SAA) is often applied to this problem because it has strong performance guarantees (BM22). However, SAA assumes past demand is independently and identically distributed (i.i.d). We extend SAA into the non-i.i.d setting with a remarkably simple approach: decomposing demand into a trend, seasonality, and residual. This allows us to perform SAA directly on the trend (which resembles a less noisy demand sequence) or directly on the residual (which is i.i.d), both of which are fundamentally easier problems. Indeed, numerical experiments confirm that both SAA decomposition techniques have a lower cost than SAA on synthetic and real world data.

2:35 - 2:45

Wynna Huang, Sadie Klaus, Lillian Rountree
Graduate Mentor: Han Yong Wunrow

Tuberculosis transmission in United States correctional facilities pre- and post-2020

Abstract. The main objective of this study is to model the transmission of tuberculosis among incarcerated populations in the United States to explore potential impacts of SARS-CoV-2 pandemic regulations on TB's disease burden. This was done using the Wells-Riley model, originally developed for tuberculosis transmission and used in previous studies for prison settings. Our model follows a previous study about the spread of tuberculosis within Brazilian prisons, but expands the existing Wells-Riley model for application on a state-wide, rather than facility-wide, scale. Using data on the number of confirmed TB cases per year alongside regulations on the dimensions of cells and prison common areas, we created functions pertaining to the number of infectors as well as the ventilation rate within various facilities to estimate key parameters for the overall model. Our analysis showed that the onset of the SARS-CoV-2 pandemic did not lead to a significant change in TB transmission rates, and that the probability of transmission does not appear correlated to how overcrowded a facility is.

2:50 - 3:00

Melody Harwood, Yoojin Lee, Ashley Ren
Graduate Mentor: Xiaoxiao Sun

Modeling the Dynamics of Language Shift in Bilingual Communities

Abstract. This project aims to model language shift in order to understand which conditions contribute to the stable survival of both languages in a bilingual community. In past literature, research has generally moved from focus on differential models which represent entire population proportions to agent-based models which capture stochastic individual language choices. Even with inclusion of individual choices, past models have made simplifying assumptions that may misrepresent the dynamic behavior of bilinguals. Sociological research shows that bilinguals are not often at first-language proficiency levels in both languages at once. This indicates that bilinguals must be represented variably, with speakers who tend to speak one language more frequently than the other. In this paper we create a model which features various levels of bilingualism and other modifications to represent how language communities interact to improve upon previous models. Our model simulates language shift by initializing a population with language use data. When compared to empirical language shift trends, our model is able to follow patterns observed in real life. Improving the realism of our model allows us to analyze the impact of specific variables, like language prestige, population proportions, and population connectivity on language shift. Our current results show that isolation between

minority and majority language communities is positively correlated to the survival of a group of monolingual minority language speakers.