Thesis Proposal

Saleh Hindi

September 21, 2016

1 Markov Chains and Intransitive Games

I want to study Markov Chains for my capstone project because the subject combines graph theory, probability, and linear algebra so it seems like a fitting stop to my undergraduate math degree. Specifically, I want to explore the Penney Ante problem which can be stated as

Problem 1.1 (Penney Ante Problem) *Given a sequence $S_n \in S_0, s_1, s_2, \ldots, S_n$ of countable length n, what is $E(S_i)$ the expected waiting time for S_i to appear? Furthermore, what is $P(S_i)$, where $P(S_i)$ is the probability that S_i will appear first? Does $P(S_i) = P(S_j)$ imply $E(S_i) = E(S_j)$?*

This is the martingale approach to the problem discovered by the eponymous Walter Penny [**?**]. Nickerson, Gardner, and Breen describe Penney's problem as given two sequences of length $n-1$ coin flips, $S_i = HHHHH$ and $S_j = THTHHT$, which coin flip is expected to come first in a series of random coin flips? What is the expected waiting time for these two sequences? As it turns out although $E(S_i) > E(S_j)$, $P(S_i) > P(S_j)$ [**?**] [**?**] [**?**]. This counterintuitive result fascinated my attention. This problem could also be formulated in terms of a sequence of rolls of a die with similarly defined waiting times and first appearence probabilies [**?**].

From this formulation, Li and Wendell develop formulas for the stopping time of S_i, $E(S_i)$ and for the probability that S_i appears before S_j [**?**]. While Li uses a martingale approach, Wendel bases his approach in Markov chains. [**?**]

The result of the Penney Ante problem is that for the sequences $A = HHHHH$ and $B = HTHHT$, the expected waiting time (in tosses) of A is greater than the expected waiting time of B but the probability of seeing A first is also greater than the probability of seeing B first [?]. This problem has applications to biology as a DNA sequence is a sequence of nucleotides represented by letters A, T, C, and G [?]. This problem also has application in computer science as the Boyr-Moore string matching algorithm is based on this result [?]. I will attempt to study properties of this game including how $E(S_i)$ varies with n and the number of choices for each n. I will also explore if there other games with this counterintuitive result.

References

[1] Breen, Stephen, Waterman Michael S., and Zhang Ning. "Renewal Theory for Several Patterns." Journal of Applied Probability 22.1 (1985): 228-34. Web.

[2] Gardner, Martin. "Mathematical Games: On the Paradoxical Situations That Arise from Nontransitive Relations." Scientific American 10 (1974): 120-25. Print.

[3] Li, Shuo-Yen Robert. "A Martingale Approach to the Study of Occurrence of Sequence Patterns in Repeated Experiments." The Annals of Probability 8.6 (1980): 1171-176. Web.

[4] Nickerson, R. S. "Penney Ante: Counterintuitive Probabilities in Coin Tossing." The UMAP Journal 28.4 (2007): 503-32. JSTOR. Web. 8 Sept. 2016.

[5] L.J Guibas, A.M Odlyzko, String overlaps, pattern matching, and nontransitive games, Journal of Combinatorial Theory, Series A, Volume 30, Issue 2, 1981, Pages 183-208, ISSN 0097-3165, http://dx.doi.org/10.1016/0097-3165(81)90005-4. (http://www.sciencedirect.com/science/article/pii/0097316581900054)

2 Elliptic Curves

Another consideration for my thesis involves studying the Elliptic Curves, specifically the Sato Tate conjecture on elliptic curves. Let $E := y^2 = x^3 + ax + b$ for $a, b \in \mathbb{R}$ be the elliptic curve and define E_p to be $E_p := E \mod p$ for any prime p. E_p is called the elliptic curve over finite field with p elements [?]. We can define an addition operation on any points P_1, P_2 on the elliptic curve. Let $P_1 + P_2 = P_3$ where P_3 is the point found by finding P_3', the point found by intersecting the line through P_1 and P_2 and the elliptic curve. The third point P_3 is the reflection of P_3' over the x-axis. This is not the same as addition of coordinate points on the plane [?]. However, this operation turns the set of points in E into an "additive abelian group with ∞ as the identity element" [**?**].

One idea for my paper is to explore properties of subsets of the curves, like E_p over the complex numbers or rational numbers. Sutherland references some properties of E_p like the Trace of Frobenius. Let $\mathbb{F}_p$ be a finite field and let the "trace of Frobenius" be defined by $a_p := p+1-\#E_p(\mathbb{F}_p)$ [**?**]. What happens to a_p as p varies? There is also possibility of exploring geometric properties of E and E_p. Are any property of the elliptic curve invariant under isomorphism or endomorphism?

References

[1] Sutherland, Andrew. "Sato-Tate Distributions." Arizona Winter School 2016. Course notes.

[2] Clozel, L. "The Sato-Tate Conjecture." Current Developments in Mathematics 2006.1 (2006): 1-34. JSTOR. Web. 7 Sept. 2016.

[3] Baier, Stephan. "The Sato-Tate Conjecture on Average for Small Angles." Transactions of the American Mathematical Society 361.4 (2009): 1811-832. JSTOR. Web. 8 Sept. 2016.

[4] Washington, Lawrence C. Elliptic Curves Number Theory and Cryptography. Boca Raton: CRC, 2008. Print.

My Thesis

I have thought long and hard about what I want my thesis to be. It's going to be my final project at Haverford so I want to make it something I am proud of and so I can present to people outside of the college. I want to end my last year on a high note. That said, my first choice is the Penney Ante problem because I have not gotten a chance to explore very much of either combinatorics or probability at Haverford. As I said, the mix of linear algebra and probability is particularly exciting to me and I think this topic relates to my personal interests. The topics presented best match the research interests of Curtis Greene and Heidi Goodson so I prefer to work with one of these two professors. I would be elated to work with either faculty member so I will be happy with any choice the department makes.