MILLER: RESEARCH CLASS: 2019

STEVEN J. MILLER

Abstract. Below is a description of the class mechanics and possible projects for our CMU Spring 2019 class. If as you progress in your projects you find they are not as enjoyable or interesting as you expected, please let me know and we can talk about moving to something else.

Note I: Anyone who wants to work on a problem is welcome to do so. The primary goal is to get a sense of what research is like, but almost always projects lead to publications. Anyone who works on a paper is a co-author on the paper. This includes not just CMU students and faculty, but sometimes colleagues elsewhere who help. Thus the professors listed are only a subset of the professors who might be on it.

Note II: The semester will go by very fast; it is helpful to hit the ground running. Thus, if you can do some of the background reading before you arrive on campus in the spring that is terrific. This does not mean you must master every detail, but rather be aware of what is done, what techniques are useful, where to go for details, what the open questions and projects are, ... It's great to read about problems that you don't work on, as this way you get a sense of what is out there.

Contents

Date: October 19, 2018.

1. Course Mechanics

1.1. **Course Description:** One of the most exciting aspects of mathematics is how accessible many of the current research problems are to people with just a high school (or lower!) background. Are there[1] infinitely many primes differing by 2? Is every even number the sum of at most two primes? Why do so many natural and mathematical processes have a preponderance of their first digit 1 (about 30% of the time!)? Defining the Fibonaccis by 1, 2, 3, 5, 8, ..., every positive integer is uniquely the sum of non-adjacent terms; what can we say about the distribution of summands here and in other decompositions? If $f(x) = 3x + 1$ for x odd and $x/2$ for x even, if we keep iterating f will we always reach 1 (and then cycle 4-2-1)? While easy to state, these and other problems are difficult to solve and require a lot of advanced mathematics. Using these and other topics as motivation, we will explore how some of the math you have seen, as well as some natural extensions, can be applied to attack new research problems. The class will be a mix of lectures on general theory, numerical exploration, and theoretical assaults.

1.2. **Course Structure:** Students will work in small groups on research projects, which will require them to learn how to read technical papers in the field, and frequently program to gathering data to help conjecture. All groups will write a final report, hopefully leading to a publication, and give presentations at CMU and if possible conferences. Likely conferences over a two year window include the following.

- CANT (Combinatorial and Additive Number Theory): Every May in NYC (in 2019 it will be May 21-24).
- Maine-Québec Number Theory Conference: Alternates between the two locations; will be in Maine in 2019 (probably October) and Québec in 2020.
- The 34$^{\text{th}}$ Automorphic Forms Workshop: Location and date to be determined; the 33$^{\text{rd}}$ will be March 6-10 (see http://automorphicformsworkshop.org/).
- Joint Math Meetings: Every January; this year January 16-19 in Baltimore.
- 19$^{\text{th}}$ International Fibonacci Conference: I'm an organizer, will be at the University of Sarajevo in July 2020.

Class lectures will be a mix of in-person visits by Professor Miller, Skype classes, and material on-line and self-study. Some of these lectures will be on material either of general interest or of importance to all; others will be tailored to the groups. Problems and material will be chosen mostly from analysis, combinatorics, number theory and probability.

Target student: Pre-requisites: Multivariable calculus, linear algebra. Real analysis (and probability or number theory) are helpful, but not needed.

Scheduling Issues: I am on sabbatical from Williams in Spring 2019. I have two young kids and am the primary caregiver; my wife is a professor of marketing at UMass Amherst. While I won't be in residence for the semester, I'll frequently visit for 1-2 day stays (say 5-6 plane trips to campus) and communicate constantly via Skype and email as needed.

[1]When I first started guiding research this was wide open; in the past few years we have made tremendous progress, and now know that there are infinitely many differing by a known, bounded amount.

2. First Problem Topic: More Sums Than Differences

Team: TBD. Advisor Professor Miller.

There are a lot of great problems in additive number theory. These require very little background to understand. Common tools include basic number theory, combinatorics, probability and analysis. For some good surveys see articles (3), (9) and (10). One of the most important papers in the field is the one by Martin and O'Bryant (6), which shows the power of probabilistic techniques.

Here are some possible problems. There are almost surely more, but this should be a good starting point.

- Hegarty and Miller (1) were able to show a phase transition in behavior with the cardinality of the sumset and difference set. Hogan and Miller (2) have extended this to more summands, but only for fast and critical decay; the important case of slow decay (the hardest) is still open (and resisted attack in a previous research I guided). Step 1 is to understand what happens with three summands. This might require opening up the strong concentration paper of Vu (11) as a start.

- While it is known due to (6) that a positive percentage of sets are sum-dominant in the uniform model, there is no explicit construction that gives a positive percentage. Original families had horrible densities; of the 2^n subsets of $\{0, \dots, n-1\}$ we had $Cn^d/2^{n/2}$ were sum-dominant in our explicit families. The first major breakthrough was by Miller, Orosz and Scheinerman (7) who got C/n^4 (though with trivial modifications one gets C/n^2). The world record is held by Zhao (12), who gets C/n. Can one get a positive percentage elementarily and not probabilistically? Note (8) might be worth looking at as well.

- In (5) many problems are studied. One project is to try and apply the graph perspective to other systems. Another, ***which I would love to see as I think it is one of the most important in the subject***, involves investigating the behavior of the divot as a function of the probability an element is chosen (currently being this was begun with a CMU student with me in 2016, and pushed significantly further with some of my summer students in 2018). The divot (the number of missing sums in a sumset is less for 7 missing sums than for 6 and 8, meaning we do not have a unimodal distribution) is only analyzed when each element is in A with probability $1/2$; see Figure 1. What happens if the probability equals a fixed number p? What if the probability decays like $p(N)$? What if we have more summands? One could investigate similar questions for difference sets, but this will be significantly harder as there is a lack of independence which squares the run-time. This project will require serious computational resources; we may be able to interest Kevin O'Bryant to collaborate again.

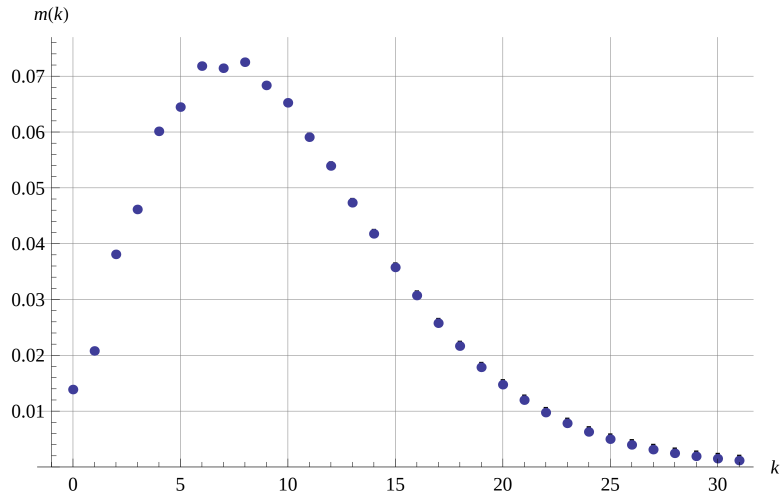

FIGURE 1. Experimental values of $m(k)$, with vertical bars depicting the values allowed by our rigorous bounds. In most cases, the allowed interval is smaller than the dot indicating the experimental value. The data comes from generating 2^{28} sets uniformly forced to contain 0 from $[0, 256)$.

- I believe in (6) and (13) there is work on the joint distribution of sumset and difference set sizes; i.e., how often $|A+A| = i$ and $|A-A| = j$. Generalize this to the fast, critical and slow decay regimes of $p(N)$ from (1). Thanks to Francesco Cellarosi for suggesting this.

- Consider the MSTD problem on more general lattices, such as $\mathbb{Z}^+ \times \mathbb{Z}^+$. A lot of the arguments for $\mathbb{Z}^+$ involve the fact that these sets have a fringe, which is an easy boundary to analyze. If you recall, the Fundamental Theorem of Calculus (Stokes Theorem) is far more interesting in dimension 2 and higher as the only closed curve on the line is a point. Another paper that might be interesting is *Lattice polytopes with distinct pair-sums* by Choi, Lam and Reznick (I have a copy). Thus, let's look at subsets of $\mathbb{Z}^d$, intersected with different regions (say spheres, boxes); maybe by a change of variables any lattice can be converted to the standard one in $\mathbb{Z}^d$ at a cost of changing the intersecting region. These sets have different fringe structures. How does the shape of the fringe affect the answer? We can play with the relative sizes of the length and width of a box in two dimensions, for example. Progress on this has been made recently by other students of mine.

- New problem from CANT 2013 discussions: In determining if A is sum-dominant or difference-dominant, it doesn't matter how much larger one is than the other. Try and find a natural weighting on the sets, try to take into account by how much one beats the other.
- New problem from CANT 2013 discussions: Is there a set A such that $|A + A| > |A - A|$ *and* $|A \cdot A| > |A/A|$? If yes, can you find an explicit, infinite family? What is the density of such sets?

- New problem from CANT 2013 discussions: The 84% lower bound from work of Mizan Khan and his colleagues on modular hyperbolas is almost surely not the true answer. What do numerical investigations suggest? What is the correct limiting behavior?

Action Plan.

- Everyone should read some of the classic papers in the subject. This does not mean you must understand everything in them, but you should have a good sense of the questions and techniques. Please read (1), (3) (the more technical paper, with details, is (4)), and (6). What you read next depends on what you are interested in studying.

- Read the problem descriptions in this section and the papers listed in the bullet point above, and then email me which sub-projects you are interested in. Once you choose a sub-project, read the referenced papers (I can share the preprints of work with my 2018 students).

Bibliography for this section.

(1) P. V. Hegarty and S. J. Miller, *When almost all sets are difference dominated*, Random Structures and Algorithms **35** (2009), no. 1, 118–136.
http://arxiv.org/abs/0707.3417

(2) G. Hogan and S. J. Miller, *When Generalized Sumsets are Difference Dominated*, preprint.
http://arxiv.org/abs/1301.5703

(3) G. Iyer, O. Lazarev, S. J. Miller and L. Zhang, *Finding and Counting MSTD sets*, to appear in the conference proceedings of the 2011 Combinatorial and Additive Number Theory Conference.
http://arxiv.org/abs/1107.2719

(4) G. Iyer, O. Lazarev, S. J. Miller and L. Zhang, *Generalized More Sums Than Differences Sets*, Journal of Number Theory **132** (2012), no. 5, 1054–1073.
http://arxiv.org/abs/1108.4500

(5) O. Lazarev, S. J. Miller and K. O'Bryant, *Distribution of Missing Sums in Sumsets*, to appear in Experimental Mathematics.
http://arxiv.org/abs/1109.4700

(6) G. Martin and K. O'Bryant, *Many sets have more sums than differences*, Proceedings of CRM-Clay Conference on Additive Combinatorics, Montréal 2006.
http://arxiv.org/abs/math/0608131

(7) S. J. Miller, B. Orosz and D. Scheinerman, *Explicit constructions of infinite families of MSTD sets*, Journal of Number Theory **130** (2010), 1221–1233.
http://arxiv.org/abs/0809.4621

(8) S. J. Miller, S. Pegado and S. Robinson, *Explicit Constructions of Large Families of Generalized More Sums Than Differences Sets*, Integers **12** (2012), #A30.
http://arxiv.org/abs/1303.0605

(9) M. B. Nathanson, *Problems in additive number theory, 1*, Proceedings of CRM-Clay Conference on Additive Combinatorics, Montréal 2006.
http://arxiv.org/abs/math/0604340

(10) M. B. Nathanson, *Sets with more sums than differences*, Integers : Electronic Journal of Combinatorial Number Theory **7** (2007), Paper A5 (24pp).
http://arxiv.org/abs/math/0608148

(11) V. H. Vu, *Concentration of non-Lipschitz functions and Applications*, Random Structures and Algorithms **20** (2002), no. 3, 262-316.

(12) Y. Zhao, *Constructing MSTD Sets Using Bidirectional Ballot Sequences*, Journal of Number Theory **130** (2010), no. 5, 1212–1220.
http://arxiv.org/abs/0908.4442

(13) Y. Zhao, *Sets Characterized by Missing Sums and Differences*, Journal of Number Theory **131** (2011), 2107–2134.
http://arxiv.org/abs/0911.2292

3. Second Problem Topic: Benford's Law

Team: TBD. Advisor Professor Miller.

There is an enormous wealth of problems on Benford's law of digit bias. Needed background includes Fourier analysis and probability (both are a must!); some basic combinatorics and number theory are useful for some problems. If you are interested in these problems, email me and I will pass along some introductory notes and chapters of a book I edited. I have copies of most of the papers below, so if you want any just ask. Good surveys (in addition to my notes which I can send) are (2), (3), (4), (12), (16). For applications see (13) and (14) (though Knuth (8) talks about its importance in computer science).

Here are some possible problems. There are more, but this is a good starting point.

- My main interest this year is in extending (1). What I like about this project is that it involves dependent random variables. There are a lot of ways to go. We can look at partition functions from number theory, amalgamating as well as fragmenting, using different densities for the cuts.

- A standard problem is to find interesting systems that satisfy either Benford's law or its cousins (there are other digit biases; these often arise from power law behavior, and include Benford's law as a special case). Potentially interesting cases include roots of polynomials (which might be known), zeros of the Riemann zeta function, special values of special functions, ... These projects would require a serious literature search; I have some results from students working with me at Princeton N years ago. There are so many fun things to play with. For example, choose a real number x in $[0, 1]$ at random and look at its continued fraction expansion $[a_1(x), a_2(x), a_3(x), \dots]$. Look at the sequence $\{b_n(x)\}_{n=1}^{\infty}$, with $b_n(x) = a_1(x)a_2(x)\cdots a_n(x)$. What can we say about the leading digits of this sequence (for almost all x).

- We have bounds on convergence to Benford of systems such as $x_n = \alpha^n$ where $\log_{10}\alpha$ is irrational. These bounds involve the irrationality exponent, but may not be sharp. It might be interesting to numerically investigate convergence issues from numbers with similar irrationality exponents. A good start is the discussions related to the $3x + 1$ Problem in (7).

- I'm not sure if the Benfordness of sequences such as $\{p^m q^n\}_{m,n=0}^{\infty}$ is known; if not this could be a good case to study (of course, more generally look at $\{p_1^{m_1} \cdots p_k^{m_k}\}_{m_1,\dots,m_k=0}^{\infty}$.

- Read (6) and think about generalizations (this is about chains of random variables); can we do anything with chains involving two free parameters?

- Read (11) on order statistics and find an interesting generalization.

- Explore the connections between Benford's law and copulas. While I think this is an interesting project, previous students have had trouble saying something meaningful.

- In 2018 some students and I explored non-linear recurrences; that work is in progress and has several natural possibilities. I am also involved in some projects related to fraud detection; I cannot publicly post such links, but there are several items worth studying.

Action Plan.

- Email me for some introductory chapters on Benford's law from a book I edited. If you are not familiar with Fourier analysis, remedy that! It's one of my favorite methods for attacking such problems. If you need notes on this let me know.

- Read some of the papers I've authored, so you have a sense of the problems I like and the methods I use. Good ones are (1), (6), (7), (10), (11).

- Read the problem descriptions in this section and the papers listed in the bullet point above, and then email me which sub-projects you are interested in. Once you choose a sub-project, read the referenced papers.

Bibliography for this section.

(1) T. Becker, D. Burt, T. C. Corcoran, A. Greaves-Tunnell, J. R. Iafrate, J. Jing, S. J. Miller, J. D. Porfilio, R. Ronan, J. Samranvedhya, F. W. Strauch and B. Talbut, *Benford's Law and Continuous Dependent Random Variables*, Annals of Physics **388** (2018), 350–381.
http://arxiv.org/abs/1309.5603

(2) F. Benford, *The law of anomalous numbers*, Proceedings of the American Philosophical Society **78** (1938), 551–572.

(3) T. Hill, *The first-digit phenomenon*, American Scientist **86** (1996), 358–363.

(4) T. Hill, *A statistical derivation of the significant-digit law*, Statistical Science **10** (1996), 354–363.

(5) W. Hurlimann, *Benford's Law from 1881 to 2006: a bibliography*,
http://arxiv.org/abs/math/0607168

(6) D. Jang, J. U. Kang, A. Kruckman, J. Kudo and S. J. Miller, *Chains of distributions, hierarchical Bayesian models and Benford's Law*, Journal of Algebra, Number Theory: Advances and Applications, volume 1, number 1 (March 2009), 37–60.
http://arxiv.org/abs/0805.4226

(7) A. Kontorovich and S. J. Miller, *Benford's Law, Values of L-functions and the* $3x+1$ *Problem*, Acta Arith. **120** (2005), 269–297.
http://arxiv.org/abs/math/0412003

(8) D. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, Addison-Wesley, third edition, 1997.

(9) J. Lagarias and K. Soundararajan, *Benford's Law for the* $3x+1$ *Function*, J. London Math. Soc. (2) **74** (2006), no. 2, 289–303.
http://arxiv.org/pdf/math/0608208.pdf

(10) S. J. Miller and M. Nigrini, *The Modulo* 1 *Central Limit Theorem and Benford's Law for Products*, International Journal of Algebra **2** (2008), no. 3, 119–130.
http://arxiv.org/pdf/math/0607686.pdf

(11) S. J. Miller and M. Nigrini, *Order statistics and Benford's law*, International Journal of Mathematics and Mathematical Sciences, Volume 2008 (2008), Article ID 382948, 19 pages. doi:10.1155/2008/382948.
http://arxiv.org/abs/math/0601344

(12) S. Newcomb, *Note on the frequency of use of the different digits in natural numbers*, Amer. J. Math. **4** (1881), 39-40.

(13) M. Nigrini, *Digital Analysis and the Reduction of Auditor Litigation Risk*. Pages 69–81 in *Proceedings of the 1996 Deloitte & Touche / University of Kansas Symposium on Auditing Problems*, ed. M. Ettredge, University of Kansas, Lawrence, KS, 1996.

(14) M. Nigrini, *The Use of Benford's Law as an Aid in Analytical Procedures*, Auditing: A Journal of Practice & Theory, **16** (1997), no. 2, 52–67.

(15) M. Nigrini and S. J. Miller, *Data diagnostics using second order tests of Benford's Law*, Auditing: A Journal of Practice and Theory **28** (2009), no. 2, 305–324. doi: 10.2308/aud.2009.28.2.305

(16) R. A. Raimi, *The first digit problem*, Amer. Math. Monthly **83** (1976), no. 7, 521–538.

(17) Wikipedia, *Copulas*.
http://en.wikipedia.org/wiki/Copula_(probability_theory)

4. Third Problem Topic: Elementary Number Theory

Team: TBD. Advisors Professors Miller and Thomas Wright (Wofford).

Here are some possible problems. There are almost surely more, but this should be a good starting point.

- Carmichael numbers play an important role in cryptography. In *There are infinitely many Carmichael numbers* by W. R. Alford, A. Granville, and C. Pomerance, they proved there are infinitely many Carmichael numbers (good title). Tom Wright recently proved there are infinitely many in arithmetic progressions, but did not optimize the constants. Optimize!

- Is it possible to walk from a prime to infinity always adding a bounded number of digits at each stage? If not, is it possible to find arbitrarily long sequences of primes? What if we can only add one digit at a time? What if we can add digits at the front or end (or even between existing digits)? I have some partial results with colleagues.

These are just some quick notes to record conversations Li and I had at the Sarnak conference a few years back. Useful reference here:

http://en.wikipedia.org/wiki/Cunningham_chain

Definition 4.1 (Digit function)**.** *Define the digit function $\mathcal{D}(x)$ to be the number of decimal digits of x; thus $\mathcal{D}(x) = \text{ceil}(\log_{10} x)$. Thus $\mathcal{D}(1701) = 4$.*

Definition 4.2. *We say m is appended to n, forming the new number $x = m\#n$, when the digits of x read left to right are the digits of m followed by the digits of n. We say n is an initial string of x if there is an integer m such that $x = m\#n$. Note $\mathcal{D}(m\#n) = \mathcal{D}(m)+\mathcal{D}(n)$.*

For example, if $m = 314$ and $n = 271$ then $m\#n = 314271$ and 271 is an initial string of 314271.

Theorem 4.3. *There is an infinite sequence of primes $p_1, p_2, \dots$ such that each prime is a beginning string of the subsequent one.*

Proof. The proof is trivial and follows immediately by Dirichlet's Theorem; the difficulty is in minimizing how fast the sequence grows. We assume we do not start our sequence with 2 or 5, so all terms are relatively prime to 10. Given a prime p_n in our sequence, we find p_{n+1} by appealing to Dirichlet's theorem where the modulus is $10^{\mathcal{D}(p_n)}$ and the residue is p_n. As there are infinitely many primes (since the modulus and the residue are relatively prime) we choose the first one to be p_{n+1}, and note that p_n is an initial string of p_{n+1}. $\square$

There have been a lot of investigations as to how far we must go to find a prime congruent to a modulo q. Linnik proved that there are effectively computable L and c such that the first prime in that progression occurs before cq^L; the current best value for L is 5, due to Xylouris [**?**]. While this gives us an infinite chain, we are increasing the number of digits by a factor of 5 each time and thus our sequence is exponentially growing!

The goal of this note is to explore the following problem and its generalizations: is it possible to walk from a prime to infinity always adding a bounded number of digits at each stage? If not, is it possible to find arbitrarily long sequences of primes? What if we can only add one digit at a time? What if we can add digits at the front or end (or even between existing digits)?

Remark 4.4. *Some quick comments about notes I have on random constructions. Let $\langle d_1, d_2\rangle$ mean an interval of numbers going from d_1 digits to d_2 digits. Consider for a fixed d the sequence of intervals $\langle d, 2d\rangle, \langle 2d, 4d\rangle, \langle 4d, 8d\rangle, \dots$. Want to say the probability we can add something at each stage is like blah, which converges to zero as we multiply. Why is this the case? The probability a random integer x is prime is approximately $1/\log x$, so all numbers with the same number of digits have roughly the same probability of being prime. The probability that we cannot append something to a number when we add $d_2 - d_1$ digits is $\dots$.*

One of the most intriguing aspects of the primes is the interplay between structure and randomness; Dirichlet's theorem and Linnik's theorem are prime examples, illustrating how primes in progressions cannot be too structured and avoid certain sets too often. In this section we show that we can create highly structured sets with large density where the answer to our questions is 'no'. Specifically, we'll construct sets with positive density where we cannot create an infinite chain $s_1, s_2, \dots$ of elements where s_n is an initial string of s_{n+1}.

Theorem 4.5. *There exists an infinite set A and a positive constant $c \in (0,1)$ such that for every positive integer N the cardinality of $A_N := A \cap [0, 10^N - 1)$ is at least $c10^N$ and there is no infinite chain of elements of a where each is an initial string of the next element.*

Proof. If we could construct a set A such that, as we go further and further down, all the elements have initial strings of longer and longer lengths of zeros we would be fine; unfortunately if we were to do that we would violate the assumption that a positive percentage of numbers are in A_N. The difficulty is that if the number of rightmost terms that are zero grows with N, in the limit we cannot have a positive percentage of numbers. The situation would be different if we didn't require A_N to be so large; if we assumed the first $\log_{10}\log_{10}\log_{10} N$ digits of all numbers in A_N are zero then we would still have $10^{N - \log_{10}\log_{10}\log_{10} N} = 10^N/\log_{10}\log_{10} N$ numbers available. Thus we could easily build a counterexample denser than the primes; however, with a bit more work we can construct a set with positive density.

Let f be a slowly growing function, with $f(x) = o(\mathcal{D}(x))$, and assume g grows slower than f, so $g(x) = o(f(x))$. We construct A_N as follows. We look at the $g(N)$ smallest elements of A_{N-1} and make sure than none of them are initial segments of length at most $f(N)$ of elements in $A_N \setminus A_{N-1}$, and then choose the remaining elements of $A_N \setminus A_{N-1}$ randomly *subject to having first digit non-zero, which ensures we didn't accidentally choose a number in A_{N-1}*. How many possibilities do we have for our choices? There would have been $10^{f(N)}$ choices for the last $f(N)$ digits, but as we are excluding $g(N)$ possibilities we are left with $10^{f(N)} - g(N)$ elements; the key fact is that the loss is not in the exponent, but through subtraction. There are $N - f(N)$ digits remaining and they are free, save the final digit much be at least a 1 and thus there are only 9 options. Thus the number of options here is $9 \cdot 10^{N - f(N) - 1}$. Combining, we see the cardinality of $A_N \setminus A_{N-1}$ is

$$\#(A_N \setminus A_{N-1}) \;=\; 9 \cdot 10^{N-f(N)-1}\left(10^{f(N)} - g(N)\right) \;=\; 9 \cdot 10^{N-1} - \frac{9}{10}\frac{g(N)}{10^{f(N)}}10^N. \qquad (4.1)$$

We thus see that a positive percentage of numbers can be in A (any percentage under .9 is trivial; breaking .9 just requires a little care in noting that there are only $9 \cdot 10^{N-1}$ potential numbers in $A_N \setminus A_{N-1}$ with leading digit non-zero. By construction no initial string is in infinitely many terms (as eventually $g(N)$ is larger than any fixed number), and hence we cannot construct an infinite chain with the desired properties. $\square$

(THE PROOF FEELS A LITTLE FISHY TO ME AS IT FEELS LIKE WE'RE CHEATING AND CIRCUMVENTING THE OBSTRUCTION; THE POINT IS THAT WE'RE NOT REQUIRING ALL THE DIGITS TO BE ZERO IN THE BEGINNING, JUST REQUIRING THAT WE AVOID CERTAIN STRINGS.)

(MAYBE THE TITLE SHOULD BE WHEN MEMBERSHIP IS CLOSED, NOT PRIMALITY, BUT TO ME IT'S THE PRIME CASE THAT'S INTERESTING. WE CAN TALK ABOUT STRUCTURE OF SETS....)

Action Plan.

- Read the problem descriptions in this section and the papers listed in the bullet point above, and then email me which sub-projects you are interested in.

Bibliography for this section.

(1) W. R. Alford, A. Granville, and C. Pomerance, *There are infinitely many Carmichael numbers*, Ann. Math. **139** (1994), 703–722. https://math.dartmouth.edu/~carlp/PDF/paper95.pdf.

(2) T. Wright, *Infinitely Many Carmichael Numbers in Arithmetic Progressions*, to appear in the Bulletin of the London Mathematical Society. http://arxiv.org/pdf/1212.5850v1.

5. Fourth Problem Topic: Zeckendorf Decompositions

Team: TBD. Advisor Professor Miller.

I've become very interested in problems related to Zeckendorf decompositions over the past few years. If you write the Fibonacci numbers as 1, 2, 3, 5, ..., then every integer has a unique decomposition as a sum of non-adjacent Fibonacci numbers. The original proof uses a greedy algorithm approach; I have a much longer one using combinatorics. This perspective is worthwhile as it allows us to solve a multitude of related questions with significantly less work than previous methods (which were heavy on continued fractions). These problems include Gaussian behavior in the number of summands of $m \in [F_n, F_{n+1})$ as $n \to \infty$ (see (3) and then (5) for the generalization to other recurrences). Another problem where progress has been made is studying the spacings between adjacent summands, observing geometric decay (see (2)). A good introduction to the subject is (6).

Here are some possible problems. There are almost surely more, but this should be a good starting point. Needed pre-requisites are elementary combinatorics, basic probability, and recurrence relations and generating functions.

- One way to prove Gaussian behavior for the number of summands is to calculate the moments of the distribution and show convergence to the Gaussian moments. This leads to combinatorial sums involving the Fibonacci numbers and binomial coefficients. George Andrews and I are able to give a nice interpretation to the mean; it would be nice to be able to do all the moments elementarily like this. If you are interested I will pass along a copy of his paper.

- Is it possible to generalize these results to a number field, say the Gaussian integers $\mathbb{Z}[i]$? What about a function field? What's known?

- Generalize the theory developed to handle more general recurrences. There are a multitude of ways to do this. What happens if some of the coefficients are negative? What happens if the leading coefficient in the recurrence is zero? Can non-linear recurrence be studied? What happens when we have recurrences where there is no longer a unique 'legal' decomposition – what can be said about the multiplicities of representations? For additional reading see some recent work, some of it joint with CMU students from 2016: https://arxiv.org/pdf/1606.09312, https://arxiv.org/pdf/1607.04692 and https://arxiv.org/pdf/1606.09309 (the last two involve CMU students).

- In work from a few summers ago we prove results for the distribution of the longest gap between summands (the papers (7) and (8) might be useful). Generalize this to the longest k gaps where k is fixed and we're studying decompositions of integers in $[F_n, F_{n+1})$, and then allow k to grow with n. Results on order statistics might be helpful; see (4) for an introduction.

- What would the 2-dimensional analogue of Zeckendorf be? For example, imagine a square or rectangle where the (i, j) entry is $2^i 3^j$, and we want decompositions so that no two summands are next to each other (horizontally or vertically). What kind of 'legal' decompositions can we get? What about other sequences? Is there any natural way to extend our 1-dimensional result to 2 dimensions? Some work on this was done with a CMU student in 2016 and gives rice to lattice results in d-dimensions. What can one say about the distribution of gaps here? See https://arxiv.org/pdf/1809.05829.

Action Plan.

- Read (3) and (6) (the more technical version is (5)), and (2). For those considering far-difference representations (i.e., representations with positive and negative summands), also read (1).

- Read the problem descriptions in this section and the papers listed in the bullet point above, and then email me which sub-projects you are interested in. Once you choose a sub-project, read the referenced papers.

Bibliography for this section.

(1) H. Alpert, *Differences of multiple Fibonacci numbers*, INTEGERS: Electronic Journal of Combinatorial Number Theory **9** (2009), 745–749.
http://www.emis.de/journals/INTEGERS/papers/j57/j57.pdf

(2) Beckwith, Bower, Gaudet, Insoft, Li, Miller and Tosteson: *The Average Gap Distribution for Generalized Zeckendorf Decompositions*, to appear in the Fibonacci Quarterly.
http://arxiv.org/abs/1208.5820

(3) Kologlu, Kopp, Miller and Wang: *Gaussianity for Fibonacci case*, Fibonacci Quarterly **49** (2011), no. 2, 116–130.
http://arxiv.org/pdf/1008.3204

(4) A. Merberg and S. J. Miller, *Course Notes for Math 162: Mathematical Statistics: The Sample Distribution of the Median.*
http://web.williams.edu/Mathematics/sjmiller/public_html/BrownClasses/162/Handouts/MedianThm04.pdf

(5) S. J. Miller and Y. Wang, *From Fibonacci numbers to Central Limit Type Theorems*, Journal of Combinatorial Theory, Series A **119** (2012), no. 7, 1398–1413.
http://arxiv.org/pdf/1008.3202

(6) S. J. Miller and Y. Wang, *Gaussian Behavior in Generalized Zeckendorf Decompositions* (with Yinghui Wang), to appear in the conference proceedings of the 2011 Combinatorial and Additive Number Theory Conference.
http://arxiv.org/pdf/1107.2718

(7) M. F. Schilling, *The longest run of heads*, The College Mathematics Journal **21** (1990), no. 3, 196–207.
http://mathdl.maa.org/images/upload_library/22/Polya/07468342.di020742.02p0021g.pdf

(8) J. Turi, *Limit theorems for the longest run*, Annales Mathematicae et Informaticae **36** (2009), 133–141.
http://www.emis.de/journals/AMI/2009/ami2009-turi.pdf

(9)

6. Fifth Problem Topic: Random Matrix Theory

Team: TBD. Advisor Professor Miller.

Random matrix theory is one of my favorite subjects. For historical introductions see (3) and (6), and see (2) for connections with L-functions (see also the recent paper at https://arxiv.org/pdf/1505.07481). For additional reading, see the papers on my webpages:

http://web.williams.edu/Mathematics/sjmiller/public_html/ntprob13/
http://web.williams.edu/Mathematics/sjmiller/public_html/ntandrmt/index.htm

Here are some possible problems. There are almost surely more, but this should be a good starting point. Needed pre-requisites are linear algebra, combinatorics and basic probability.

- There are a lot of interesting 'thin' ensembles of matrices that I've studied with my students: (1), (4), (5), (7), (8), (9), (10). Find an interesting family of matrices where there is some structure and investigate. Papers by Arup Bose will probably be useful. For some recent ensembles with very different behavior (non-connected density of states) see https://arxiv.org/pdf/1710.00240 and https://arxiv.org/pdf/1803.08127.

- Find an analogue to Rankin-Selberg convolution of L-functions for matrices. This is very hard; I have tried with several people but no natural analogue has been found to date, though there is a very promising possibility I found with some summer students in 2018; ***I think this is one of the more exciting projects to investigate***. The number theory process is described here:
http://arxiv.org/pdf/math/0607688

- *New problem:* I recently completed an analysis of weighted d-regular graphs, where the eigendistribution for the weights is almost, but provably not, the semi-circle. This leads to several interesting problems. I don't think it's possible to find a simple, similar system where the semi-circle is an eigendistribution, but if so that would be interesting. It might be worthwhile computing explicitly the lower order correction terms further than we've done, and attempting to find combinatorial meaning.

Action Plan.

- Read (3), (5) and (8).

- Think about how the structure of the ensemble affects the distribution of the eigenvalues, and what ensemble you would like to study.

Bibliography for this section.

(1) O. Beckwith, V. Luo, S. J. Miller and N. Triantafillou, *Distribution of eigenvalues of weighted, structured matrix ensembles*, preprint.
http://arxiv.org/abs/1112.3719
(2) B. Conrey, *L-functions and random matrix theory*, arXiv version.
http://arxiv.org/pdf/math/0005300v1
(3) F. W. K. Firk and S. J. Miller, *Nuclei, Primes and the Random Matrix Connection*, Symmetry **1** (2009), 64–105; doi:10.3390/sym1010064.
http://arxiv.org/abs/0909.4914
(4) L. Goldmahker, C. Khoury, S. J. Miller and K. Ninsuwan, *The expected eigenvalue distribution of large, weighted d-regular graphs*, Random Matrices: Theory and Applications **3** (2014), no. 4, 1450015 (22 pages). https://arxiv.org/pdf/1306.6714
(5) C. Hammond and S. J. Miller, *Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices*, Journal of Theoretical Probability **18** (2005), no. 3, 537–566.
http://arxiv.org/abs/math/0312215
(6) B. Hayes, *The Spectrum of Riemannium*, American Scientist July-August 2003.
http://www.americanscientist.org/issues/pub/the-spectrum-of-riemannium
(7) S. Jackson, S. J. Miller and T. Pham, *Distribution of eigenvalues for highly palindromic real symmetric Toeplitz matrices*, Journal of Theoretical Probability **25** (2012), 464–495.
http://arxiv.org/abs/1003.2010
(8) M. Koloğlu, G. S. Kopp, S. J. Miller, F. W. Strauch and W. Xiong, *The Limiting Spectral Measure for Ensembles of Symmetric Block Circulant Matrices*, to appear in the Journal of Theoretical Probability.
http://arxiv.org/abs/1008.4812
(9) A. Massey, S. J. Miller and J. Sinsheimer, *Distribution of eigenvalues of real symmetric palindromic Toeplitz matrices and circulant matrices*, Journal of Theoretical Probability **20** (2007), no. 3, 637–662.
http://arxiv.org/abs/math/0512146
(10) S. J. Miller, T. Novikoff and A. Sabelli, *The distribution of the second largest eigenvalue in families of random regular graphs*, Experimental Mathematics **17** (2008), no. 2, 231–244.
http://arxiv.org/abs/math/0611649
(11) S. J. Miller and R. Takloo-Bighash, *An Invitation to Modern Number Theory*, Princeton University Press, Princeton, NJ, 2006, 503 pages.
http://press.princeton.edu/chapters/s15_8220.pdf (a useful chapter).

7. Sixth Problem Topic: L-functions and Random Matrix Theory

Team: TBD. Advisor Professor Miller.

There is now a rich theory connecting number theory and random matrix theory; for some of the history and problems see (2), (5) and (6). For additional reading, see the papers on my webpages:

http://web.williams.edu/Mathematics/sjmiller/public_html/ntprob13/
http://web.williams.edu/Mathematics/sjmiller/public_html/ntandrmt/index.htm

An absolute must read to work on these problems is (8).

Here are some possible problems. There are almost surely more, but this should be a good starting point. Needed pre-requisites are number theory, combinatorics and basic probability, though additional courses (such as complex analysis) are helpful. For some of the projects, you need to be *very* good and interested in combinatorics (as you go further in number theory, you'll see just how much of the subject is combinatorics).

- Rubinstein and Gao were able to calculate the n-level densities for families of quadratic Dirichlet L-functions. Interestingly, both were computed in a certain range but becomes the combinatorics are so difficult, we could only see that they were equal for $n \leq 3$. Jake Levinson, in his undergraduate thesis with me (10), was able to show agreement for $n \leq 7$ by developing a new combinatorial vantage. We are one identity away from a proof of agreement for all n in the ranges both are known. Unfortunately we are stuck. Help us! Note: this problem is still important as it is a direct assault on the problem, but Ze'ev Rudnick and two of his students were able to prove agreement via a clever trick where they invoked the function field result. That said, this problem is also exceptionally difficult.

- A major application of the n-level densities are to bound the order of vanishing of families of L-functions at the central point; there is much discussion on this in (8), including an appendix where the optimal test function is determined when $n = 1$ for given ranges of support. Find the optimal test functions for other n and various ranges of support. I have work with some colleagues (see (7) and (9)) where we calculate the n-level density for families of cuspidal newforms. Perhaps (9) has now gone far enough so that you can use that plus a better test function to beat the vanishing estimates from (8). I have some preprints with a recent thesis student and others (including some CMU students) with some partial results. ***This is another very important problem.***

- Continue the program started in (7) and currently stopped at (9). This program finds an alternative to the Katz-Sarnak determinant expansion by converting an n-dimensional integral to a 1-dimensional integral of a new test function (which is an n-fold convolution of the old test function), and then requires combinatorics to show agreement between this and the determinantal expansion. The reason this approach is done is that the new 1-dimensional integral is significantly easier to handle than the

n-dimensional integral, and thus we think this is a good trade. Then again, maybe we're wrong and it would be a good exercise to try the integral directly.

- **Important:** Building on the previous problem, an interesting issue has developed which should be studied. In (8) the authors assume what they call Hypothesis S, which allows them to extend the support. They remark that these are delicate sums and one must be careful, and give an example of a similar exponential sum which does not have the cancellation assumed in Hypothesis S. In trying to extend the support for the 2-level density, Triantafillou and I assumed the most natural analogue of Hypothesis S. It is *not* consistent with the Katz-Sarnak density conjectures, so we believe it is not the right hypothesis to assume. Find the right hypothesis.

- I have looked at low-lying zeros for many families of L-functions, finding the main terms and sometimes the lower order terms: (1), (3), (4), (7), (9), (10), (11) and (12). Find a family of L-functions that has yet to be studied and is approachable, and compute the 1-level density. Possible examples would be families of elliptic curves, or maybe standard L-functions but over a number field. ***Warning:*** *we have to make sure that such families have not been done.*

Action Plan.

- This is a very rich field. You must read (8). After that, you should read some of the family-specific papers to get a sense of the methods, say (4), (7) and (12). It's fine to just skim these last three papers.
- Read the problem descriptions in this section and the papers listed in the bullet point above, and then email me which sub-projects you are interested in. Once you choose a sub-project, read the referenced papers.

Bibliography for this section.

(1) L. Alpoge and S. J. Miller, *The low-lying zeros of level 1 Maass forms*, preprint 2013.
http://arxiv.org/abs/1301.5702

(2) B. Conrey, *L-functions and random matrix theory*, arXiv version.
http://arxiv.org/pdf/math/0005300v1

(3) E. Dueñez and S. J. Miller, *The low lying zeros of a* $GL(4)$ *and a* $GL(6)$ *family of L-functions*, Compositio Mathematica **142** (2006), no. 6, 1403–1425.
http://arxiv.org/abs/math/0506462

(4) E. Dueñez and S. J. Miller, *The effect of convolving families of L-functions on the underlying group symmetries*, Proceedings of the London Mathematical Society, 2009; doi: 10.1112/plms/pdp018.
http://arxiv.org/pdf/math/0607688.pdf

(5) F. W. K. Firk and S. J. Miller, *Nuclei, Primes and the Random Matrix Connection*, Symmetry **1** (2009), 64–105; doi:10.3390/sym1010064.
http://arxiv.org/abs/0909.4914

(6) B. Hayes, *The Spectrum of Riemannium*, American Scientist July-August 2003.
http://www.americanscientist.org/issues/pub/the-spectrum-of-riemannium

(7) C. Hughes and S. J. Miller, *Low lying zeros of L–functions with orthogonal symmetry*, Duke Mathematical Journal **136** (2007), no. 1, 115–172.
http://arxiv.org/abs/math/0507450

(8) H. Iwaniec, W. Luo, and P. Sarnak, *Low lying zeros of families of L-functions*, Inst. Hautes Études Sci. Publ. Math. **91** (2000), 55–131.
http://link.springer.com/content/pdf/10.1007%2FBF02698741

(9) G. Iyer, S. J. Miller and N. Triantafillou, *Moment Formulas for Ensembles of Classical Compact Groups* (with Geoffrey Iyer and Nicholas Triantafillou), preprint.

(10) J. Levinson and S. J. Miller, *The n-level density of zeros of quadratic DirichletL-functions*, to appear in Acta Arithmetica.
http://arxiv.org/abs/1208.0930

(11) S. J. Miller, *Lower order terms in the 1-level density for families of holomorphic cuspidal newforms*, Acta Arithmetica **137** (2009), 51–98.
http://arxiv.org/abs/0704.0924

(12) S. J. Miller and R. Peckner, *Low-lying zeros of number field L-functions*, Journal of Number Theory **132** (2012), 2866–2891.
http://arxiv.org/abs/1003.5336

E-mail address: sjm1@williams.edu

Professor of Mathematics, Department of Mathematics and Statistics, Williams College, Williamstown, MA 01267