Research Statement: Steven J Miller

Steven.J.Miller@williams.edu

November 16, 2010

My main interest is analytic number theory and random matrix theory (especially the distribution of zeros of L-functions and the eigenvalues of random matrix ensembles); I am also studying equidistribution problems in analysis and probability, and working on applied problems in probability, statistics, graph theory, cryptography, sabermetrics and linear programming where the tools and techniques of number theory can successfully be applied. These have led me to study computational and numerical methods, and I have written numerous programs to investigate and solve the above projects, ranging from zeros of L-functions to constructing elliptic curves with rank to linear programming problems. Papers and talks are available at `http://www.williams.edu/Mathematics/sjmiller/`

I have worked on many of these projects with students in high school (20), college (over 150) and graduate school (over 10). In many instances I ran guided vertically research groups, where the graduate students had their own projects and also helped mentor the undergraduates and high school students. I have been able to maintain my active research program while at a liberal arts college which prides itself on its commitment to teaching.

My research has also been continuously supported by two NSF grants for the past several years.

Contents

1 Research Interests: Summary **1**

2 Summary of Thesis Results **2**

3 Representative Subset of Previous and Current Research **3**

3.1 Previous Research. 3

3.2 Current Work. 9

1 Research Interests: Summary

While my primary training is in analysis and number theory (and I do love working on problems in these fields), I have collaborated with numerous people outside of pure mathematics, writing papers in Accounting, Computer Science, Economics, Engineering, Geology, Marketing and Sabermetrics. I enjoy these projects for many reasons. Frequently these are interesting problems which lead to interesting theoretical issues. Further, I am often able to use these papers in classes I teach, and I find these projects are a great way to interest and excite students.

Since Riemann's investigations 150 years ago, zeros of L-functions have been known to be intimately connected to solutions to many problems in number theory. In the last few decades finer properties of the zeros have helped understand problems such as the observed preponderance of primes congruent to 3 mod 4 over 1 mod 4 as well as the growth of the class number. Random Matrix Theory has become a powerful tool to model the behavior of these zeros, suggesting both the answers as well as new questions to ask. My primary interest is in the distribution of zeros near the central point for families of L-functions, especially families of elliptic curves with rank over $\mathbb{Q}(T)$. Related to this, I am also investigating constructions for moderate to high rank one-parameter families of elliptic curves, lower order corrections to n-level densities of elliptic curves, and the influence of forced zeros at the central point on the distribution of the first zero above the central point. Additionally, I am studying the low zeros of Dirichlet characters with square-free modulus (which have applications to how primes are distributed in arithmetic progressions) and Rankin-Selberg convolutions of GL_n and GL_m families of L-functions (which highlights how the behavior of complex families can be understood in terms of the behavior of the building blocks). Finally, much of my recent work involves determining lower order terms in the behavior of zeros of L-functions. These terms are sadly inaccessible through standard random matrix theory, which misses arithmetic; however, through recent bold conjectures such as the L-functions Ratios Conjecture we have an excellent predictive method. I have studied these lower order terms in a variety of interesting families; the most exciting recent application (joint with several colleagues) a model which explains the observed repulsion of zeros near the central point in families of quadratic twists of elliptic curve L-functions, which we are currently extending to more general families.

I am investigating numerous problems in Random Matrix Theory and Random Graphs, especially ensembles with few degrees of freedom (order N independent matrix elements, instead of order N^2). These provide fascinating windows to see new behavior and have numerous applications (k-regular graphs are used to construct cheap and efficient networks). Along these lines, I am also studying several problems on the boundary of Probability Theory, Number Theory and Analysis, such as proving that the distribution of the first digits of $|L(s, f)|$ near the critical line and iterates of the $3x+1$ map follows Benford's Law of digit bias (the first digit is a 1 about 30% of the time). These problems have led to results ranging from the distribution of digits of order statistics to a generalization of the central limit theorem for random variables modulo 1. With some colleagues and students I am extending these results and working on applications (I have been in contact with the Criminal Investigative Division of the IRS, helped organize the first conference on Benford's law, and am currently editing the first book on the theory and applications of the law). Using recent strong concentration results I have just proved a conjecture on the size of the sumset to the difference set in additive number theory / probability theory, with fascinating behavior at the critical threshold which I am continuing to explore.

I have also worked on and am pursuing several applied projects in Probability, Statistics, Linear Algebra and Cryptography, such as closed-form Bayesian inferences for the multinomial logit model, a binary integer linear programming problem for movie distributors, bounding incomplete multiple exponential sums arising in Computer Science, extreme cases of the Cramer-Rao inequality, modeling baseball games, and determining the security of certain signature schemes in cryptography, as well as studying the propagation of viruses in networks. I am also interested in computational aspects of these problems, writing algorithms to investigate many of these topics, from zeros of elliptic curve L-functions and moments of Dirichlet L-functions over function fields to random matrix theory and graph theory to Bayesian inference and linear programming.

2 Summary of Thesis Results

Following Brumer-Heath-Brown [BH-B], Iwaniec-Luo-Sarnak [ILS], Katz-Sarnak [KaSa1, KaSa2], Rubinstein [Rub] and Silverman [Si], I used the 1- and 2-level densities to study the distribution of low-lying zeros for one-parameter rational families of elliptic curves of rank r over $\mathbb{Q}(T)$; these densities are defined by summing test functions at scaled zeros of the L-functions. Katz and Sarnak [KaSa1, KaSa2] predict that, to each family of L-functions $\mathcal{F}$, there is an associated symmetry group $G(\mathcal{F})$ (a classical compact group) which governs the distribution of the low-lying zeros. In other words, the behavior of zeros in a family of L-functions near the central point is well modeled by the behavior of eigenvalues near 1 of a classical compact group. For families of elliptic curves of rank 0 over $\mathbb{Q}(T)$, we expect $G(\mathcal{F})$ to be $SO(\text{even})$ if all curves have even functional equation, O if half are even, half odd, and $SO(\text{odd})$ if all are odd. If the family has rank r over $\mathbb{Q}(T)$, the densities are trivially modified to take into account the r expected zeros at the central point, and we will still refer to these as $O, SO(\text{even}), SO(\text{odd})$. The 1-level densities for the Unitary and Symplectic are distinguishable from the three Orthogonal groups for test functions of arbitrary small support; unfortunately, the three orthogonal groups all agree for functions supported in $(-1, 1)$. The 2-level densities for the orthogonal groups, however, are distinguishable for functions supported in arbitrarily small neighborhoods of the origin.

Modulo standard conjectures (which can be verified for many specific cases), for small support I showed the densities agree with Katz and Sarnak's predictions [Mil1]. The difficulty is that the logarithm of the analytic conductors must be extremely well controlled or oscillatory behavior drowns out the main term. In particular, it is not enough to confine the logarithms of the conductors to lie in $[\log N^d, \log 2N^d]$ as $N \to \infty$. The conductors are controlled by careful sieving and deriving explicit formulas via Tate's algorithm. Further, the densities confirm that the curves' L-functions behave in a manner consistent with having r zeros at the central point, as predicted by the Birch and Swinnerton-Dyer conjecture. By studying the 2-level densities of some constant sign families, we find the first examples of families of elliptic curves where we can distinguish $SO(\text{even})$ from $SO(\text{odd})$ symmetry.

Similar to the GUE universality Rudnick and Sarnak [RS] found in studying n-level correlations of L-functions, our universality follows from the sums of $a_t^2(p)$ in our families. For non-constant $j(T)$, this follows from a Sato-Tate law proved by Michel [Mic]; however, for many of our families we are able to show this by a direct calculation. The effect of the rank over $\mathbb{Q}(T)$ surfaces through sums of $a_t(p)$.

Finally, while the n-level densities for these families are universal, potential lower order correction terms have been observed in several families. These family dependent corrections are of size $1/\log N$; unfortunately, trivial estimation of the errors lead to terms of size $\log\log N/\log N$. I have recently completed a detailed analysis for many families [Mil5], where these corrections can be isolated. These corrections show how the arithmetic of the family enters as lower order corrections; in particular, families with and without complex multiplication behave differently. I am currently exploring the relation between these lower order terms and finer properties of the behavior of low zeros for small conductors.

3 Representative Subset of Previous and Current Research

3.1 Previous Research.

Below is a representative sample of some of my previous work. Please see my CV or my publication list for all previously completed projects.

(i) Elliptic curves and additive number theory

(a) While most of my work in elliptic curves is related to the behavior of zeros near the central point, with graduate students Scott Arms and Alvaro Lozano-Robledo [ALM] I also proved results on the geometric side. Specifically, generalizing an idea from my thesis, we gave a new, novel construction of one-parameter families of elliptic curves of moderate rank over $\mathbb{Q}(T)$. The key idea is to use Rosen and Silverman's [RSi] proof of Nagao's conjecture for rational surfaces to interpret averages of the Fourier coefficients as the rank. These averages are sums of Legendre sums; while these are impossible to evaluate in general, by careful construction these can be determined for special families. One interesting consequence is that families of moderate rank are constructed *without* having to enumerate rational solutions and showing that they are linearly independent (in other words, no height matrices are needed). The families constructed here are of great use in other investigations, as their Legendre sums are more tractable than the general family, and thus it is possible to isolate lower order correction terms (which depend on the arithmetic).

(b, c) Let A be a finite set of integers, and set $A + A$ and $A - A$ to be the set of all sums (respectively differences) of elements in A. As addition is commutative but subtraction is not, it was believed that 'most' of the time $|A + A| < |A - A|$. It thus came as a surprise when Martin and O'Bryant [MO] proved that if each element from $\{0, 1, \dots, N\}$ is in A with positive probability p then as $N \to \infty$ a positive percentage of A are sum dominated (i.e., $|A + A| > |A - A|$). I and Peter Hegarty [HeMi] investigated the relationship between the sizes of the sum and difference sets attached to a subset of $\{0, 1, ..., N\}$ chosen randomly according to a binomial model with parameter $p(N)$, with $N^{-1} = o(p(N))$. We showed that the random subset is almost surely difference dominated, as $N \to \infty$, for any choice of $p(N)$ tending to zero, thus confirming a conjecture of Martin and O'Bryant. The proofs involve applying recent strong concentration results to our situation. It is worth noting that interesting transitional behavior is observed when $p(N)$ crosses the threshold of $N^{1/2}$. More precisely, the main result is

Theorem Let $p : \mathbb{N} \to (0,1)$ be any function such that $N^{-1} = o(p(N))$ and $p(N) = o(1)$. For each $N \in \mathbb{N}$ let A be a random subset of I_N chosen according to a binomial distribution with parameter $p(N)$. Then, as $N \to \infty$, the probability that A is difference dominated tends to one. More precisely, let $\mathscr{S}, \mathscr{D}$ denote respectively the random variables $|A+A|$ and $|A-A|$. Then the following three situations arise:
(i) $p(N) = o(N^{-1/2})$: Then $\mathscr{S} \sim \frac{(N \cdot p(N))^2}{2}$ and $\mathscr{D} \sim 2\mathscr{S} \sim (N \cdot p(N))^2$.
(ii) $p(N) = c \cdot N^{-1/2}$ for some $c \in (0, \infty)$: Define the function $g : (0,\infty) \to (0,2)$ by $g(x) := 2\left(\frac{e^{-x}-(1-x)}{x}\right)$. Then $\mathscr{S} \sim g\left(\frac{c^2}{2}\right) N$ and $\mathscr{D} \sim g(c^2)N$.
(iii) $N^{-1/2} = o(p(N))$: Let $\mathscr{S}^c := (2N+1) - \mathscr{S}$, $\mathscr{D}^c := (2N+1) - \mathscr{D}$. Then $\mathscr{S}^c \sim 2 \cdot \mathscr{D}^c \sim \frac{4}{p(N)^2}$.

In a related project, with my undergraduate student Dan Scheinerman I developed a new construction for families of more sum than difference sets; at the time of their construction, they were the densest known. The construction is currently being generalized to more general linear combinations of A with Brooke Orvosz and David Newman, as well as two students in my number theory class (Sean Pegado and Luc Robinson) at Williams.

(d) A beautiful theorem of Zeckendorf states that every positive integer can be written uniquely as a sum of non-consecutive Fibonacci numbers. Once this has been shown, it is natural to ask how many Fibonacci numbers are needed. Lekkerkerker proved that the average number of such summands needed for integers in $[F_n, F_{n+1})$ is $n/(\phi^2+1)$, where ϕ is the golden mean. With summer students Murat Kologlu, Gene Kopp and Yinghui Wang [KKMW, MW], we showed that the fluctuation about the mean is Gaussian for the Fibonacci numbers. Unlike previous number-theoretic approaches, we attack the problem through combinatorics and generating functions. Our techniques can be generalized to apply to a large class of linear recurrence relations.

(ii) L-functions and the Ratios Conjecture

Much of my previous research on L-functions splits naturally into two cases, determining the main term and determining lower order corrections for the 1-level density, defined by:

$$\lim_{N\to\infty} \frac{1}{|\mathcal{F}_N|} \sum_{f\in\mathcal{F}_N} \sum_j \phi\left(\widetilde{\gamma}_{j;f}\right).$$

Here ϕ is an even Schwartz function, $\mathcal{F} = \cup \mathcal{F}_N$ is a family of L-functions ordered by conductor, and (assuming GRH) we may write the zeros of $L(s,f)$ as $1/2 + i\gamma_{j;f}$ (the $\widetilde{\gamma}_{j;f}$ are the normalized imaginary parts of the non-trivial zeros). The Katz and Sarnak conjectures [KaSa1], [KaSa2] state that as the conductors tend to infinity, the behavior of the zeros near the central point agree with the scaling limit of normalized eigenvalues of a classical compact group.

(a) For families where the signs of the functional equations are all even and there is no corresponding family with odd functional equations, a "folklore" conjecture (for example, see page 2877 of [KeSn1]) states that the symmetry is symplectic, presumably based on the observation that SO(even) and SO(odd) symmetries in the examples known to date arise from splitting orthogonal families according to the sign of the functional equations. *A priori* the symmetry type of a family with all functional equations even is either symplectic or SO(even). Let ϕ be a fixed Maass form and let H_k denote the space of cusp forms of full level. All L-functions in the families (formed by Rankin-Selberg convolution) $\{\phi \times H_k\}$ and $\{\phi \times \mathrm{sym}^2 H_k\}$ (by $\mathrm{sym}^2 H_k$ we mean the set $\mathrm{sym}^2 f$ for $f \in H_k$) have even functional equations, and neither family seems to naturally arise from splitting sign within a full orthogonal family. By calculating the 1-level density I and Dueñez proved in [DM1] that the symmetry of the first is symplectic (as predicted); however, the second family has orthogonal symmetry (we cannot distinguish between SO(even), O and SO(odd) due to the small-support restriction on the allowable test functions; however, analyzing the 2-level density allowed us to discard O and SO(odd)). Thus our calculations are only consistent with the symmetry being SO(even). In particular, this work proved that the theory of low-lying zeros is more than just a theory of the distribution of signs of functional equations.

(b) Building on this work, I and Dueñez in [DM2] considered the Rankin-Selberg convolution of more general families. They define a NT-good family of L-functions to be a family where there is good control over the conductors, the cardinality of the family, and a good enough averaging formula to evaluate the needed prime sums to compute the 1-level density; known examples include Dirichlet L-functions, cuspidal newforms, as well as twists and symmetric powers of these. The main result relates how the zeros of compound families formed by Rankin-Selberg convolution are distributed in terms of how the constituent families are distributed. Explicitly,

Thereom. Let $\mathcal{F}$ and $\mathcal{G}$ be NT-good families of unitary automorphic cuspidal representations of $\mathrm{GL}_n(\mathbb{A}_\mathbb{Q})$ and $\mathrm{GL}_m(\mathbb{A}_\mathbb{Q})$ with trivial central character, with symmetry constants $c_\mathcal{F}$ and $c_\mathcal{G}$. Assume $\mathcal{F} \times \mathcal{G}$ is an NT-good family. Then the family $\mathcal{F} \times \mathcal{G}$ (which is the limit of $\mathcal{F}_N \times \mathcal{G}_M$, where N and M tend to infinity together) has symmetry constant $c_{\mathcal{F}\times\mathcal{G}} = c_\mathcal{F} \cdot c_\mathcal{G}$.

The proof follows from understanding the moments of the Satake parameters, with the main terms determined from the first and second moments. It is the universality of these moments that is responsible for the universality of the results. Numerous specific examples are given, in particular families of elliptic curves with rank. The difficulty there is in controlling the conductors.

(c) The last result involving the main term of the zeros near the central point is for the family of cuspidal newforms of prime level N split by sign of the functional equation. Iwaniec, Luo and Sarnak [ILS] showed that the main term of the 1-level density in these families agrees with random matrix theory when the Fourier transform of the test function is supported in $(-2,2)$. This impressive calculation depends on a difficult analysis of the non-diagonal term in the Petersson formula, namely the Bessel-Kloosterman sums. With Chris Hughes [HuMi] generalized these results to the n-level density (which improves estimates of high order of vanishing), avoiding analyzing n-dimensional analogues of the Bessel-Kloosterman piece by cleverly changing variables. The cost of this switch was difficult combinatorics needed to show agreement with random matrix theory; this is because random matrix theory was expecting an n-dimensional integral whereas we changed test functions to a one-dimensional integral which can be evaluated by the techniques in [ILS]. (It is not the case that the combinatorics can always be analyzed; in Gao's thesis [Gao] the number theory computations are thought to agree with random matrix theory throughout his range, but this can only be shown for a restricted window.) This led to the discovery of a new formula for the n-level density, which after some combinatorics was shown to agree with the determinantal expansions of Katz-Sarnak [KaSa1, KaSa2]. This formula is significantly easier for comparison purposes in restricted ranges. We proved

Theorem Let $n \geq 2$, $\mathrm{supp}(\widehat{\phi}) \subset (-\frac{1}{n-1}, \frac{1}{n-1})$, and define

$$R_n(\phi) \ = \ (-1)^{n-1}2^{n-1}\left[\int_{-\infty}^{\infty} \phi(x)^n \frac{\sin 2\pi x}{2\pi x}\,\mathrm{d}x - \frac{1}{2}\phi(0)^n\right], \quad \sigma_\phi^2 \ = \ 2\int_{-1}^{1} |y|\widehat{\phi}(y)^2\,\mathrm{d}y.$$

Assume GRH for $L(s,f)$ and for all Dirichlet L-functions. As $N \to \infty$ through the primes, the centered n^{th} moment for $H_k^{\pm}(N)$ (the family of weight k cuspidal newforms of level N and functional equation either even or odd) agrees with RMT and equals $(2m-1)!!\sigma_\phi^{2m} \pm R_{2m}(\phi)$ if $n = 2m$ is even and $\pm R_{2m+1}(\phi)$ if $n = 2m+1$ is odd. One application of these results (where it is essential that we are able to evaluate the relevant sums without using the Petersson weights) is to bound high vanishing in the family. Specifically, we prove the following (which for large r provides better bounds than the previous records):

Theorem: Consider the families of weight k cuspidal newforms split by sign, $H_k^{\pm}(N)$. Assume GRH for all Dirichlet L-functions and all $L(s,f)$. For each n there are constants r_n and c_n such that as $N \to \infty$ through the primes, for $r \geq r_n$ the probability of at least r zeros at the central point is at most $c_n r^{-n}$; equivalently, the probability of fewer than r zeros at the central point is at least $1 - c_n r^{-n}$.

(d) In addition to studying the main term, I've also investigated the arithmetic dependence of lower order terms in various families. These investigations are important as the main term is independent of the arithmetic. In [Mil3] I studied families of elliptic curves (using many of the families constructed in [ALM] and discussed earlier). By isolating the first correction term to the 1-level density, we csn see differences depending on whether or not the families had complex multiplication or what the torsion group was.

(e, f) Building on analogies with Random Matrix Theory and previous number theory conjectures, recently Conrey, Farmer and Zirnbauer [CFZ1], [CFZ2] created the L-functions Ratios Conjecture (or recipe) to predict sums of ratios of L-functions in a family. These quantities allow us to compute almost any desired statistic, from n-level correlations and densities to mollifiers and moments, to name a few. The predictions are believed to be correct to square-root cancelation in the family's cardinality. To put this in context, a typical family of one-parameter elliptic curves over $\mathbb{Q}(T)$ with T specialized to be in $[N, 2N]$ has an error term of size $O(\log\log N/\log N)$, much larger than $N^{-1/2+\epsilon}$! There is little basis for such a small conjectured error term, other than the philosophy of square-root cancelation. The recipe is as follows: let

$$R_{\mathcal{F}_N}(\alpha,\gamma) \ := \ \sum_{f\in\mathcal{F}_N} \frac{L(1/2+\alpha,f)}{L(1/2+\gamma,f)}.$$

1. Use the approximate functional equation to expand the numerator into two sums plus a remainder. The first sum is over m up to x and the second over n up to y, where xy is of the same size as the analytic conductor (typically one takes $x \sim y$). We ignore the remainder term.

2. Expand the denominator by using the generalized Mobius function.

3. Execute the sum over $\mathcal{F}_N$, keeping only main (diagonal) terms; however, before executing these sums replace any product over epsilon factors (arising from the signs of the functional equations) with the average value of the sign of the functional equation in the family.

4. Extend the m and n sums to infinity (i.e., complete the products).

5. Differentiate with respect to the parameters, and note that the size of the error term does not significantly change upon differentiating.

6. A contour integral involving $\frac{\partial}{\partial\alpha} R_{\mathcal{F}_N}(\alpha,\gamma)\Big|_{\alpha=\gamma=s}$ yields the 1-level density.

What is remarkable is that in many of the steps above we throw away error terms that are the same size as main terms, yet at the end we obtain (conjecturally) perfect agreement! I've studied families of quadratic characters, cuspidal newforms (not split by sign) and number field L-funtions; for suitably restricted test functions, we've [GJMMNPP, Mil1, Mil4, MilMo, MilPe] proved the Ratios' prediction (complete with square-root cancelation) is correct. For example, for quadratic characters the Ratios Conjecture predicts one of the lower order terms is

$$\frac{2}{\log X}\int_{-\infty}^{\infty} \phi(\tau)\frac{\zeta'}{\zeta}\left(1+\frac{4\pi i\tau}{\log X}\right) d\tau;$$

which implies that the lower order term depends on the zeros of the Riemann zeta function. This was numerically observed by Rubinstein, and the improvement in the fit by incorporating these terms was powerfully demonstrated by Stopple's computations. Additional applications of knowing these lower order terms is to finding $N_{\text{effective}}$, the optimal size matrices to model behavior of zeros at a finite conductor (and not in the limit).

(iii) Random Graphs and Random Matrix Ensembles

Let A be a real $N \times N$ symmetric matrix with eigenvalues $\lambda_i(A)$. We can form a measure by placing a mass of size $1/N$ at each (normalized) eigenvalue. If our ensemble is large enough, we can average over the family and a generic matrix will have behavior close to the system average. While these ensembles typically do not directly correspond to families of L-functions, they are nevertheless useful in building intuition as to how small sub-families can behave, as well as being interesting in their own right.

(a) Building on joint work with Chris Hammond [HaMi], I studied the distribution of eigenvalues of real symmetric palindromic Toeplitz matrices with undergraduates John Sinsheimer and Adam Massey, both of whom have continued to graduate school. Many authors had noticed that the density of normalized eigenvalues of real symmetric Toeplitz matrices was close to, but not equal to, the standard normal. In [HaMi] we interpreted the discrepancy in terms of Diophantine obstructions to systems of equations, and conjectured that forcing the first row to be a palindrome would remove these obstructions. We proved this in [MMS]. The difficulty in this project, as is frequently the case in studying random matrix ensembles, was developing the combinatorics to obtain closed form expressions for the moments. Similar to many other problems in the field, it is straightforward to show the averages of the moments of the measures converge; however, it is quite difficult to determine the precise value of the average moments. One reason this ensemble is so difficult to study is that it has of the order N degrees of freedom, far less than the full family of all real symmetric matrices (which is of order $N^2/2$). Through Cauchy's interlacing formula, there are explicit formulas for the eigenvalues of these matrices, and as one application we obtain a central limit theorem for weighted sums of random variables.

(b) Joint with undergraduates Tim Novikoff and Anthony Sabelli [MNS] (who are now graduate students in applied math at Cornell), I studied the distribution of the second largest eigenvalue in families of d-regular graphs. Recently Friedman [Fr] proved Alon's conjecture [Al] for many families of d-regular graphs, namely that given any $\epsilon > 0$ "most" graphs have their largest non-trivial eigenvalue at most $2\sqrt{d-1} + \epsilon$ in absolute value; if the absolute value of the largest non-trivial eigenvalue is at most $2\sqrt{d-1}$ then the graph is said to be Ramanujan. These graphs have important applications in communication network theory, allowing the construction of superconcentrators and nonblocking networks, coding theory and cryptography. As many of these applications depend on the size of the largest non-trivial positive and negative eigenvalues, it is natural to investigate their distributions. We showed these are well-modeled by the $\beta = 1$ Tracy-Widom distribution for several families. If the observed growth rates of the mean and standard deviation as a function of the number of vertices holds in the limit, then in the limit approximately 52% of d-regular graphs from bipartite families should be Ramanujan, and about 27% from non-bipartite families (assuming the largest positive and negative eigenvalues are independent). The key difficulty in interpreting the numerical investigations is that, appropriately normalized, the three Tracy-Widom distributions and the standard normal are very close to each other; the best test was looking at the percentage of eigenvalues to the right of the mean.

(iv) Benford's law

Many mathematical and natural phenomena satisfy Benford's law (or a close approximation), where the probability of a leading digit of d is $\log_{10}(1 + \frac{1}{d})$. While the proofs are typically related to equidistribution theory, and thus fall in the province of standard number theoretic investigations, the numerous applications (especially for fraud detection and data integrity tests) ensure that the subject is of interest to many.

(a) Funded in part by NSF Grant DMS0753043, I helped organize the *Conference on the Theory and Applications of Benford's Law* in 2007, and am currently editing an introductory book on the subject with the participants. The first conference on the subject, participants came from mathematics, statistics, biology, engineering, computer science, accounting and industry, to name a few. The discussions there have led to many projects.

(b, c) One of the most important questions is why Benford's law is so prevalent, and how rapidly it sets in (such estimates are essential in prosecuting fraud). In [JKKKM, MN1] we answer these questions for data that is the product of a growing number of independent random variables. The key ingredients are Fourier analysis and the Mellin transform, which allow us to prove a version of the Central Limit Theorem for sums of independent random variables modulo 1 with (at times) exponentially decaying error term. For example, if we consider a product of n independent uniform random variables on $[0, k]$, the difference between the cumulative distribution function here and that of Benford's law is bounded by

$$\frac{k}{s}\frac{(\log k)^{n-1}}{\Gamma(n)} + \left(\frac{1}{2.9^n} + \frac{\zeta(n)-1}{2.7^n}\right) 2\log_{10} s,$$

where $\zeta(n)$ is the Riemann zeta function and $\log_{10} s$ is the probability of a Benford random variable having mantissa of s or less. These results have been converted to new tests for fraud by Nigrini and Miller [MN2], [NM2], which have been shared at an invited address at the Boston headquarters of the IRS.

(v) Applied Projects

(a: Incomplete Exponential Sums) Exponential sums have a rich history, and estimates of their size have numerous applications, ranging from uniform distribution to solutions to Diophantine equations to L-functions to the Circle Method, to name a few. Consider the following incomplete exponential sum:

$$S(f, n, q) \ = \ \sum_{x_1=\pm 1} \cdots \sum_{x_n=\pm 1} x_1 \cdots x_n e^{2\pi i f(x_1,\dots,x_n)/q},$$

with f a non-homogenous quadratic. Proving non-trivial exponentially decreasing upper bounds for $S(f, n, m)$ will provide insight into the computational complexity of a class of boolean circuits needed to compute the parity of n binary inputs. A theorem that shows that the norm of $S(f, n, m)$ is c^n, where $c < 1$ will show that the size (number of binary gates) of these circuits computing parity has to grow exponentially fast. These lower bound results are of great interest to the theoretical computer science community. Using Ramsey-theoretic techniques, Alon and Beigel [AB] proved that for each fixed n, d and m there exists a positive constant $b_{d,m,n}$ such that $|S(f, n, m)| \ < \ b_{d,m,n}$ and $\lim_{n\to\infty} b_{d,m,n} = 0$; note the resulting sequences converge very slowly to 0. In terms of computational complexity, this only tells us that the minimum circuit size required to compute parity of n bits tends to infinity with n. It is of far more interest, from the computational point of view, to show exponentially fast growth in minimum circuit size. This is generally interpreted as showing that parity circuits of the required kind cannot feasibly be built. We have currently obtained sharp bounds on average, and have solved the problem for n small; this project is joint with Eduardo Dueñez and Amitabha Roy and Howard Straubing [DMRS].

(b: Empirical Bayes Inference in the Multinomial Logit Model) Whether it's the 20,000+ hits based on a www.google.com search or the 1000+ hits on www.jstor.org, the multinomial logit (MNL) model plays a very prominent role in many literatures as a basis for probabilistic inferences. One of the recent advances regarding the MNL model is the ability to incorporate heterogeneity into the response coefficients; unfortunately, this leads to increased numerical computation. Once one combines the MNL kernel, a Bernoulli random variable with logit link function, with a heterogeneity distribution, closed-form inference is unavailable due to the non-conjugacy of the product Bernoulli likelihood and the heterogeneity distribution (prior).

Eric Bradlow (professor of marketing and statistics, University of Pennsylvania). Kevin Dayaratna (graduate student in marketing at the University of Maryland) and I [MBD] derived a closed-form solution to the heterogeneous MNL problem; unfortunately, the closed-form expansion requires too many terms to be computationally feasible at present. We reduce the number of computations by several orders of magnitude by rewriting the expansion in terms of the number of solutions to systems of Diophantine equations. This allows us to have one very long initial calculation, with all subsequent calculations involving only 10 or 20 terms (instead of 10^8 and higher), and is now applicable for some problems.

(c: Binary Integer Linear Programming) With Joshua Eliashberg (professor, Wharton), Sanjeev Swami (professor, India Institute of Technology Kanpur) Chuck Weinberg (professor, University of British Columbia) and Berend Wierenga (professor, Erasmus University Rotterdam), I solved a binary integer linear programming problem to allow movie theaters to optimally schedule movies each day, taking into account a variety of managerial constraints in reel time. We are currently expanding our model to include additional constraints, and it is being implemented at movie theaters in Amsterdam. This work [EHHHMSWW] won the best paper award for 2009 for the International J. of Research in Marketing.

(d: Dynamical Systems) Leo Kontorovich (postdoc, Weizmann Institute), Amitabha Roy and I are studying various models for the propagation of viruses in different systems. We are primarily concerned with how infections are transmitted in various networks. For certain configurations we have derived a differential equation whose fixed points answer the problem, and numerical and theoretical investigations suggest what the limiting behavior should be. We have developed techniques to analyze the resulting equations and can determine the limiting behavior in most cases.

(e: Sabermetrics) Sabermetrics is the application of mathematical tools to analyze baseball. The field really took off with Bill James' (who later helped build the Red Sox championship teams of '04 and '08) work in the 70's and 80's, and has since become an major force in the baseball industry [Le]. The subject poses numerous questions of both theoretical and practical interest, and it is often highly non-trivial to derive a mathematically tractable model for a baseball event which captures the essential features. It has been noted that in many professional sports leagues a good predictor of a team's end of season won-loss percentage is Bill James' Pythagorean Formula ${\rm RS_{obs}}^\gamma/({\rm RS_{obs}}^\gamma + {\rm RA_{obs}}^\gamma)$, where ${\rm RS_{obs}}$ (resp. ${\rm RA_{obs}}$) is the observed average number of runs scored (allowed) per game and γ is a constant for the league; for baseball the best agreement is when γ is about 1.82. This formula is often used in the middle of a season to determine if a team is performing above or below expectations, and estimate their future standings (the principles involved, however, have enormous application, as the question could just as easily be asked about which mutual funds or stocks are over- or underperforming, and thus determine when to buy or sell). In [Mil4] I showed how this formula is a consequence of a reasonable model of a baseball game. I have presented this result at numerous conferences, discussed my model with Bill James at the Boston Red Sox, and am running an undergraduate research class on improving the model (and related questions) in the Spring of 2008.

3.2 Current Work.

Below is a representative sample of some work in progress.

Zeros of Elliptic Curve L-functions

My main ongoing research project involves the observed repulsion of zeros near the central point by zeros at the central point. By the Birch and Swinnerton-Dyer Conjecture, if an elliptic curve has geometric rank r, its L-function should vanish to order r at the central point, and these curves offer an exciting laboratory to test the conjectures of Random Matrix Theory. I have introduced two 'natural' models for the random matrix analogues into the literature, what I call the independent model (where the forced zeros do not interact with the remaining zeros) and the interaction model (where they do) [Mil3]. In my thesis I proved that as the conductors tend to infinity the distribution of zeros agrees with the independent model (which is the same as the interaction model with no forced zeros). The interaction model is related to the classical Bessel kernels of RMT, and gives a very different prediction as to the behavior of the first few zeros when there are forced zeros. I and my students (at Princeton, AIM and Ohio State) wrote code to construct large numbers of elliptic curves and study the effect on the location of the first zero above the central point. Extensive calculations and theoretical modeling are in progress; this project is joint with Eduardo Dueñez, Jon Keating and Nina Snaith (professors, University of Bristol) and their student Duc Khiem Huynh (in fact, work related to this become Duc Khiem's thesis). Unlike the excess rank investigations, however, as we increase the conductor we see a marked change in data; specifically, the repulsion decreases. The detailed numerical investigations I have run [Mil3] provide several promising clues as to what the correct theory is. In particular, we observe that the repulsion increases with rank, decreases with the size of the conductor, and all the zeros are shifted by the same amount. This suggests the right model for *finite* conductors is the interaction model, with parameters a

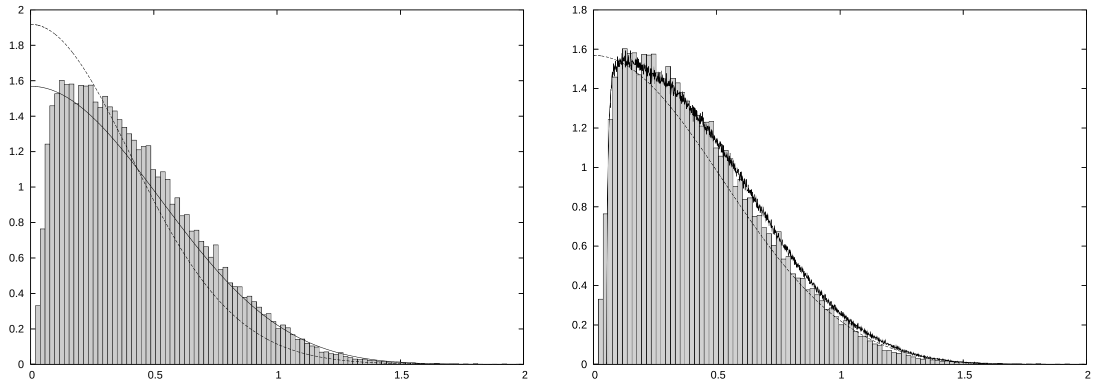

Figure 1: (left) Distribution of lowest zeros for $L_{E_{11}}(s, \chi_d)$ with $0 < d < 400,000$ (bar chart), distribution of lowest eigenvalue of SO(2N) with N_{eff} (solid), standard N_0 (dashed). (right) Distribution of lowest zeros for $L_{E_{11}}(s, \chi_d)$ with $0 < d < 400,000$ (bar chart), distribution of lowest eigenvalue of SO(2N) effective N of N_{eff} = 2 (solid) with discretisation, distribution of lowest eigenvalue of SO(2N) effective N of $N_{\text{eff}} = 2.32$ (dashed) without discretisation.

function of the average rank for conductors of a given size. As excess rank has long been observed for finite conductors, this explains how we can have repulsion in rank 0 families over $\mathbb{Q}(T)$ (where there are no zeros at the central point to repel other zeros!).

Amazingly, all the zeros appear to be repelled equally. There is a random matrix ensemble (Jacobi ensembles) with very similar properties, namely there is a variable parameter (which may be thought of as corresponding to the number of zeros at the central point), and as that parameter is increased the remaining zeros are all equally repelled. We call this the interaction model, and expect it to model the low-lying zeros for small conductors. We need two pieces of information for the analysis. The first is related to discretizing the random matrix ensemble. The second is the matrix size. This corresponds to the problem of finding $N_{\text{effective}}$ for modeling zeros of $\zeta(s)$ at finite heights, done by Keating-Snaith and others. We are constructing a rigorous theory of these low-lying zeros, especially the first zero above the central point. There are numerous technical difficulties, ranging from the fact that values of elliptic curve L-functions are discretized at the central point (with the discretization depending on arithmetic) to the difficulty in determining the lower-order terms in the one-level density to determine $N_{\text{effective}}$. We are using the Ratios Conjecture to predict these terms, and then using that as a guide for the number theory computations. Much progress has been made (including writing code to solve a Painleve VI equation that arises), and our prediction does an outstanding job of describing the data for the family of quadratic twists (*see Figure 1*). In particular, incorporating discretization and the lower order arithmetic terms captures the repulsion. We are extending our theory to general families of elliptic curve L-functions.

Using these observations, we are currently working on finding the correct model for finite conductors. This is similar to Keating and Snaith's observation [KeSn1, KeSn2] that zeros of L-functions at height T should not be modeled by the infinite scaling limits of matrices, but by $N \times N$ matrices with $N \approx \log T$. (I am also working on the analogous problem for function fields with Salman Butt (graduate student, University of Texas at Austin) and Chris Hall (postdoc, Michigan)). To date we have written one paper on the algorithms needed to solve related Painleve VI differential equations [DHKMS], and are finishing up two others related to the number theory.

Random Matrix Ensembles

(a) Given an ensemble of $N \times N$ random matrices, the first question to ask is whether or not the empirical spectral measures of typical matrices converge to a limiting spectral measure as $N \to \infty$. While this has been proved in many thin patterned ensembles sitting inside all real symmetric matrices, frequently there is no nice closed form expression for the limiting measure. Further, current theorems provide few pictures of transitions between ensembles. Building on earlier work with my students [HaMi, MMS, JMP], I continued to explore highly patterned matrices this past summer with two REU students, Murat Kologlu and Gene Kopp

[KKM]. We considered the ensemble of symmetric period m–circulant matrices with entries i.i.d.r.v. These matrices have toroidal diagonals periodic of period m. We view m as a "dial" we can "turn" from the highly structured symmetric circulant matrices, whose limiting eigenvalue density is a Gaussian, to the ensemble of all real symmetric matrices, whose limiting eigenvalue density is a semi-circle. The limiting eigenvalue densities f_m show a visually stunning convergence to the semi-circle as $m \to \infty$, which we prove.

In contrast to most studies of patterned matrix ensembles, we find explicit closed form expressions for the densities. We prove that f_m is the product of a Gaussian and a certain even polynomial of degree $2m-2$. The proof is by derivation of the moments from the eigenvalue trace formula. The new feature, which allows us to obtain closed form expressions, is converting the central combinatorial problem in the moment calculation into an equivalent counting problem in algebraic topology. The explicit formula is then obtained using topology (especially Euler characteristic), generating functions, and complex analysis.

I am currently exploring the effect of changing the structure on the answer with one of my thesis students, Wentao Xiong, who is also being mentored by Murat and Gene.

(b) Continuing the work I did with Chris Hughes, one of my thesis students (Jake Levinson) and I are trying to extend that analysis to derive alternatives to the determinantal formuls of Katz and Sarnak for the n-level densities for larger support than done in [HuMi]. The motivation for this is Peng Gao's thesis [Gao]. Gao was able to determine the number theory for the n-level density for families of Dirichlet L-functions, unfortunately, due to the difficulty of the combinatorics, he was not able to show that his answer always agrees with the random matrix theory prediction (though we do believe they always agree). We are currently attempting to derive more tractable formulas for these expansions to facilitate these comparisons.

Random Graphs

I hope to continue investigations begun with Leo Goldmakher years ago at AIM and extended with undergraduate Eve Ninsuwan at Williams on the density of eigenvalues of weighted d-regular graphs. McKay [McK] computed the density of normalized eigenvalues (it is different from the semi-circle seen for the GOE ensemble, but does converge to the semi-circle as $d \to \infty$). Goldmakher and I noticed that if the adjacency matrix elements are weighted by multiplying by independent identically distributed random variables, then if the resulting distribution is to be the semi-circle then the first 9 moments of the weighting distribution must agree with the semi-circleŠs distribution. This suggests that the semi-circle is a fixed point of this weighting procedure. The combinatorics become quite involved, and we are currently unable to handle the general case (and thus have had to resort to brute force computation). I plan on developing combinatorial results in graph theory to extend the analysis to the general case.

Benford's Law

I am currently editing the first book on Benford's law. In addition to survey articles, this includes some new research. I am writing chapters on some new results in various number-theoretic situations. Further, with some engineering colleagues I am writing a chapter on applying Benford's law to detect image fraud.

This past summer I worked with two summer students on three projects on Benford's law. One involves a new test for tax fraud, which the IRS is currently reviewing. Another project was trying to detect fraud in some of the ClimateGate data sets as well as the Iranian election. The final result concerns quantifying how close various Weibull distributions are to Benford's law.

References

[Al] N. Alon, *Eigenvalues and expanders*, Combinatorica **6** (1986), no. 2, 83–96.

[AB] N. Alon and R. Beigel, *Lower bounds for approximations by low degree polynomials over* $\mathbb{Z}_m$. In *Sixteenth Annual IEEE Conference on Computational Complexity*, IEEE Computer Society Press (2001), 184-187.

[ALM] S. Arms, A. Lozano-Robledo and S. J. Miller, *Constructing One-Parameter Families of Elliptic Curves over $\mathbb{Q}(T)$ with Moderate Rank*, Journal of Number Theory **123** (2007), no. 2, 388–402.

[Ben] F. Benford, *The law of anomalous numbers*, Proceedings of the American Philosophical Society **78** (1938), 551-572.

[BBH] A. Berger, Leonid A. Bunimovich and T. Hill, *One-dimensional dynamical systems and Benford's Law*, Trans. Amer. Math. Soc. **357** (2005), no. 1, 197–219.

[BH] A. Berger and T. Hill, *Newton's method obeys Benford's law*, The Amer. Math. Monthly **114** (2007), no. 7, 588-601.

[Be] M. V. Berry, *Semiclassical formula for the number variance of the Riemann zeros*, Nonlinearity **1** (1988), 399–407.

[BeKe] M. V. Berry and J. P. Keating, *The Riemann zeros and eigenvalue asymptotics*, Siam Review **41** (1999), no. 2, 236–266.

[BBLM] E. Bogomolny, O. Bohigas, P. Leboeuf and A. G. Monastra, *On the spacing distribution of the Riemann zeros: corrections to the asymptotic result*, Journal of Physics A: Mathematical and General **39** (2006), no. 34, 10743–10754.

[BoKe] E. B. Bogomolny and J. P. Keating, *Gutzwiller's trace formula and spectral statistics: beyond the diagonal approximation*, Phys. Rev. Lett. **77** (1996), no. 8, 1472–1475.

[BrDu] J. Brown and R. Duncan, *Modulo one uniform distribution of the sequence of logarithms of certain recursive sequences*, Fibonacci Quarterly **8** (1970) 482–486.

[BH-B] A. Brumer and R. Heath-Brown, *The average rank of elliptic curves V*, preprint.

[CFKRS] B. Conrey, D. Farmer, P. Keating, M. Rubinstein and N. Snaith, *Integral moments of L-functions*, Proc. London Math. Soc. (3) **91** (2005), no. 1, 33–104.

[CFZ1] J. B. Conrey, D. W. Farmer and M. R. Zirnbauer, *Autocorrelation of ratios of L-functions*, preprint. `http://arxiv.org/abs/0711.0718`

[CFZ2] J. B. Conrey, D. W. Farmer and M. R. Zirnbauer, *Howe pairs, supersymmetry, and ratios of random characteristic polynomials for the classical compact groups*, preprint. `http://arxiv.org/abs/math-ph/0511024`

[CS] J. B. Conrey and N. C. Snaith, *Applications of the L-functions Ratios Conjecture*, to appear in the Proc. London Math. Soc. `http://arxiv.org/abs/math/0509480`

[Dia] P. Diaconis, *The distribution of leading digits and uniform distribution mod 1*, Ann. Probab. **5** (1979), 72–81.

[DHKMS] E. Dueñez, D. K. Huynh, J. Keating, S. J. Miller and N. Snaith), *The lowest eigenvalue of Jacobi Random Matrix Ensembles and Painlevé VI*, Journal of Physics A: Mathematical and Theoretical **43** (2010) 405204 (27pp).

[DM1] 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.

[DM2] 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.

[DMRS] E. Dueñez, S. J. Miller, A. Roy and H. Straubing, *Incomplete quadratic exponential sums in several variables*, Journal of Number Theory **116** (2006), no. 1, 168–199.

[EHHHMSWW] J. Eliashberg, Q. Hegie, J. Ho, D. Huisman, S. J. Miller, S. Swami, C. B. Weinberg and B. Wierenga, *Silver Scheduler: a demand-driven modeling approach for the construction of micro-schedules of movies in a multiplex*, Intern. J. of Research in Marketing (2009), doi:10.1016/j.ijresmar.2008.09.004 (lead article, and won the IJRM Best Paper Award for 2009).

[FI] E. Fouvry and H. Iwaniec, *Low-lying zeros of dihedral L-functions*, Duke Math. J. **116** (2003), no. 2, 189-217.

[Fr] J. Friedman, *A proof of Alon's second eigenvalue conjecture*, Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing, 720–724 (electronic), ACM, New York, 2003.

[Gao] P. Gao, *N-level density of the low-lying zeros of quadratic Dirichlet L-functions*, Ph. D thesis, University of Michigan, 2005.

[GJMMNPP] (with J. Goes, S. Jackson, S. J. Miller, D. Montague, K. Ninsuwan, R. Peckner and T. Pham, *A unitary test of the L-functions Ratios Conjecture*, Journal of Number Theory **130** (2010), no. 10, 2238–2258.

[GV] D. A. Goldston and R. C. Vaughan, *On the Montgomery-Hooley asymptotic formula*, Sieve methods, exponential sums and their applications in number theory (ed. G. R. H. Greaves, G. Harman and M. N. Huxley, Cambridge University Press, 1996), 117-142.

[Gü] A. Güloğlu, *Low-Lying Zeros of Symmetric Power L-Functions*, Internat. Math. Res. Notices 2005, no. 9, 517-550.

[Ha] R. W. Hamming, *On the distribution of numbers*, Bell Syst. Tech. J. **49** (1970), 1609-1625.

[HaMi] 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.

[HeMi] P. Hegarty and S. J. Miller, *When almost all sets are difference dominated*, Random Structures and Algorithms **35** (2009), no. 1, 118–136.

[Hi1] T. Hill, *The first-digit phenomenon*, American Scientist **86** (1996), 358–363.

[Hi2] T. Hill, *A statistical derivation of the significant-digit law*, Statistical Science **10** (1996), 354–363.

[HuMi] C. Hughes and S. J. Miller, *Low-lying zeros of L-functions with orthogonal symmtry*, Duke Mathematical Journal, **136** (2007), no. 1, 115–172.

[HR] C. Hughes and Z. Rudnick, *Linear Statistics of Low-Lying Zeros of L-functions*, Quart. J. Math. Oxford **54** (2003), 309–333.

[Hu] W. Hurlimann, *Benford's Law from 1881 to 2006: a bibliography*, `http://arxiv.org/abs/math/0607168`.

[ILS] 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.

[JMP] S. Jackson, S. J. Miller and V. Pham, *Distribution of eigenvalues for highly palindromic real symmetric Toeplitz matrices*, to appear in the Journal of Theoretical Probability.

[JKKKM] 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.

[KaSa1] N. Katz and P. Sarnak, *Random Matrices, Frobenius Eigenvalues and Monodromy*, AMS Colloquium Publications **45**, AMS, Providence, 1999.

[KaSa2] N. Katz and P. Sarnak, *Zeros of zeta functions and symmetries*, Bull. AMS **36**, 1999, $1-26$.

[Ke] J. P. Keating, *Statistics of quantum eigenvalues and the Riemann zeros*, in Supersymmetry and Trace Formulae: Chaos and Disorder, eds. I. V. Lerner, J. P. Keating & D. E Khmelnitskii (Plenum Press), 1–15.

[KeSn1] J. P. Keating and N. C. Snaith, *Random matrix theory and $\zeta(1/2+it)$*, Comm. Math. Phys. **214** (2000), no. 1, 57–89.

[KeSn2] J. P. Keating and N. C. Snaith, *Random matrix theory and L-functions at $s = 1/2$*, Comm. Math. Phys. **214** (2000), no. 1, 91–110.

[KeSn3] J. P. Keating and N. C. Snaith, *Random matrices and L-functions*, Random matrix theory, J. Phys. A **36** (2003), no. 12, 2859–2881.

[KiVu] J. H. Kim and V. H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), 417–434.

[KKM] M. Kologlu, G. Kopp and S. J. Miller, *The Limiting Eigenvalue Density for the Ensemble of Symmetric Period m–Circulant Matrices*, preprint.

[KKMW] M. Kologlu, G. Kopp, S. J. Miller and Y. Wang, *On the number of summands in Zeckendorf decompositions*, preprint.

[KoMi] A. Kontorovich and S. J. Miller, *Benford's Law, Values of L-functions and the $3x + 1$ Problem*, Acta Arith. **120** (2005), 269–297.

[Knu] D. Knuth, *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*, Addison-Wesley, third edition, 1997.

[Le] M. Lewis, *Moneyball: The Art of Winning an Unfair Game*, W. W. Norton & Company, 2004.

[MO] G. Martin and K. O'Bryant, *Many sets have more sums than differences*. To appear in : Proceedings of CRM-Clay Conference on Additive Combinatorics, Montréal 2006.

[MMS] 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.

[McK] B. McKay, *The expected eigenvalue distribution of a large regular graph*, Linear Algebra Appl. **40** (1981), 203–216.

[MS] F. Mezzadri and N. C. Snaith (eds.), *Recent perspectives in random matrix theory and number theory*, LMS Lecture Note Series, 322, Cambridge University Press, Cambridge, 2005.

[Mic] P. Michel, *Rang moyen de familles de courbes elliptiques et lois de Sato-Tate*, Monat. Math. **120**, (1995), 127-136.

[Mil1] S. J. Miller, 1- *and* 2-*level densities for families of elliptic curves: Evidence for the underlying group symmetries*, Compositio Mathematica **140** (2004), no. 4, 952–992.

[Mil2] S. J. Miller, *Variation in the number of points on elliptic curves and applications to excess rank*, C. R. Math. Rep. Acad. Sci. Canada **27** (2005), no. 4, 111–120.

[Mil3] S. J. Miller, *Investigations of zeros near the central point of elliptic curve L-functions*, Experimental Mathematics **15** (2006), no. 3, 257–279.

[Mil4] S. J. Miller, *A derivation of the Pythagorean Won-Loss Formula in baseball*, Chance Magazine **20** (2007), no. 1, 40–48 (an abridged version appeared in The Newsletter of the SABR Statistical Analysis Committee **16** (February 2006), no. 1, 17–22).

[Mil5] S. J. Miller, *Lower order terms in the* 1*-level density for families of holomorphic cuspidal newforms*, Acta Arithmetica **137** (2009), 51–98.

[Mil6] S. J. Miller, *A symplectic test of the* L*-functions Ratios Conjecture*, Int Math Res Notices (2008) Vol. 2008, article ID rnm146, 36 pages, doi:10.1093/imrn/rnm146.

[MBD] S. J. Miller, E. T. Bradlow and K. Dayaratna, *Closed-form Bayesian inference for the logit model via polynomial expansions*, Quantitative Marketing and Economics **4** (2006), no. 2, 173–206.

[MilMo] S. J. Miller and D. Montague, *An Orthogonal Test of the* L*-functions Ratios Conjecture, II*, to appear in Acta Arithmetica.

[MNS] 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.

[MN1] 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.

[MN2] 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

[MilPe] S. J. Miller and R. Peckner, *Low-lying zeros of number field* L*-functions*, preprint.

[MW] S. J. Miller and Y. Wang, *From Fibonacci numbers to Central Limit Type Theorems*, preprint.

[Mon] H. Montgomery, *The pair correlation of zeros of the zeta function*, Analytic Number Theory, Proc. Sympos. Pure Math. **24**, Amer. Math. Soc., Providence, 1973, $181 - 193$.

[Nig1] 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.

[Nig2] 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.

[NM1] M. Nigrini and S. J. Miller, *Benford's Law applied to hydrology data – results and relevance to other geophysical data*, Mathematical Geology **39** (2007), no. 5, 469–490.

[NM2] 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

[OS1] A. E. Özlük and C. Snyder, *Small zeros of quadratic* L*-functions*, Bull. Austral. Math. Soc. **47** (1993), no. 2, 307–319.

[OS2] A. E. Özlük and C. Snyder, *On the distribution of the nontrivial zeros of quadratic* L*-functions close to the real axis*, Acta Arith. **91** (1999), no. 3, 209–228.

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

[RR] G. Ricotta and E. Royer, *Statistics for low-lying zeros of symmetric power* L*-functions in the level aspect*, preprint. `http://arxiv.org/abs/math/0703760`

[RSi] M. Rosen and J. Silverman, *On the rank of an elliptic surface*, Invent. Math. **133** (1998), 43-67.

[Ro] E. Royer, *Petits zéros de fonctions L de formes modulaires*, Acta Arith. **99** (2001), no. 2, 147-172.

[Rub] M. Rubinstein, *Low-lying zeros of L–functions and random matrix theory*, Duke Math. J. **109**, (2001), 147–181.

[RS] Z. Rudnick and P. Sarnak, *Zeros of principal L-functions and random matrix theory*, Duke Journal of Math. **81**, 1996, $269-322$.

[Si] J. Silverman, *The average rank of an algebraic family of elliptic curves*, J. reine angew. Math. **504**, (1998), 227-236.

[TW] C. Tracy and H. Widom, *Distribution functions for largest eigenvalues and their applications*, ICM Vol. I (2002), 587–596.

[Vu1] V. H. Vu, *New bounds on nearly perfect matchings of hypergraphs: Higher codegrees do help*, Random Structures and Algorithms **17** (2000), 29–63.

[Vu2] V. H. Vu, *Concentration of non-Lipschitz functions and Applications*, Random Structures and Algorithms **20** (2002), no.3, 262-316.

[Yo1] M. Young, *Lower-order terms of the 1-level density of families of elliptic curves*, Internat. Math. Res. Notices 2005, no. 10, 587–633.

[Yo2] M. Young, *Low-lying zeros of families of elliptic curves*, J. Amer. Math. Soc. **19** (2006), no. 1, 205–250.

[Za] I. Zakharevich, *A generalization of Wigner's law*, Comm. Math. Phys. **268** (2006), no. 2, 403–414.