ANNOTATED PUBLICATION LIST

STEVEN J. MILLER

CONTENTS

1.	Number Theory and Random Matrix Theory	2
2.	Classical Random Matrix Theory	3
3.	Additive and General Number Theory	4
4.	Benford's Law (Accounting, Probability, Number Theory,)	5
5.	Economics and Marketing	7
6.	Computational Papers	7
7.	Sabermetrics and Statistics	8
8.	Expository and General Mathematics	9

NOTE: This hasn't been updated since 2010; if anyone is actually reading this and finds it helpful let me know!

While my primary love and research interests are number theory and random matrix theory, I have worked on numerous applied projects in accounting, computer science, economics, marketing, sabermetrics (baseball mathematics) and statistics, as well as several projects in computational number theory. Below is an annotated bibliography, where I have divided my work into the different fields and summarized each paper. Papers and talks are available online at http://www.williams.edu/ Mathematics/sjmiller/public_html/index.htm

I am always looking for new projects and new colleagues to work with, in all fields. I find working on a variety of applied projects satisfying on a number of fronts. First, there is frequently little overlap in skill sets, and each of us is bringing a good but different perspective. I find we often ask different questions and have different techniques, and the ensuing discussions lead to questions and solutions that would have been hard to arrive at individually. For example, some of my best theoretical work on Benford's law is due to questions asked by an accountant colleague in his search for tests to detect tax fraud. Finally, these projects become wonderful additions in the classroom, and frequently excite students about the possibilities of mathematics in their careers.

1. NUMBER THEORY AND RANDOM MATRIX THEORY

My main research is in the distribution of zeros of L-functions near the central point. The Katz-Sarnak philosophy says that to each family of L-functions we can associate a classical compact group such that, as the conductors tend to infinity and the size of the matrices in the classical compact group tends to infinity, then the scaling limit of the statistics in the number theory family and the classical compact ensemble agree.

In my thesis (1) I overcame numerous technical difficulties to uniquely show that only one classical compact group can describe the zeros of elliptic curve L-functions, and it is the expected one. To do so required studying the 2-level density, which involved many technicalities due to the variation in the conductors. The key ingredients were bounded variation arguments and deriving explicit formulas for the conductors in certain large sub-families.

Recently the L-function Ratios Conjecture has gained enormous popularity as it is able to quickly and easily predict numerous quantities (*n*-level correlations and densities, moments, ...) for families of L-functions. The conjecture predicts answers down to essentially square-root cancellation, a remarkably brave thing to do. I have verified the predictions in many cases: (5), (7), (9), (11), (13). My favorite of these papers is (11). The Ratios Conjecture has five steps, three of which are provably false (i.e., they involve adding or subtracting terms of the same size as the terms kept; the miracle is that all of these errors seem to cancel). For the family studied in (11), one of the terms discarded is known to contribute for tests functions with large support; amazingly, another term in the Ratios' prediction surfaces for this large support and gives exactly the missing contribution.

Papers (2), (3), (4), (6), (8), (9), (12) involve studying the zeros near the central point in many families. Paper (3) disproves a folklore conjecture that this theory is essentially one of the sign of functional equations. In (8) my co-author and I return to the subject, and show that the symmetry of the convolution of two families of L-functions is the product of the symmetry constants associated to each constituent family. In this and other papers such as (6), the key ingredient is an analysis of the Satake parameters. Our theory can be recast in a similar manner to the central limit theorem, where the first two moments are essentially normalization and the higher moments control the rate of convergence.

One of the banes of the subject is the difficulty of the ensuing combinatorics. In (4) we get the *n*-level densities up to 1/(n-1) by deriving a more tractable formula than the determinant expansions of Katz-Sarnak, cleverly changing the multi-dimensional Bessel-Kloosterman integrals to the 1-dimensional integral they evaluated. The cost of this is that the ensuing combinatorics look quite different than the answer one finds in random matrix theory, but it is a worthwhile exchange to avoid the *n*-dimensional Bessel-Kloosterman integrals.

Finally paper (14) is a continuation of previous work from other sections, where we have developed a theory for zeros near the central point in families of elliptic curves with finite conductor. The key ingredients are effective matrix size and discretizing the resulting Jacobi ensembles.

(1) 1- and 2-level densities for families of elliptic curves: evidence for the underlying group symmetries, Compositio Mathematica **140** (2004), 952–992.

- (2) 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.
- (3) *The low lying zeros of a GL*(4) *and a GL*(6) *family of L-functions* (with Eduardo Dueñez), Compositio Mathematica **142** (2006), no. 6, 1403–1425.
- (4) Low lying zeros of L-functions with orthogonal symmetry (with Christopher Hughes), Duke Mathematical Journal **136** (2007), no. 1, 115–172.
- (5) 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.
- (6) Lower order terms in the 1-level density for families of holomorphic cuspidal newforms, Acta Arithmetica **137** (2009), 51–98.
- (7) An orthogonal test of the L-Functions Ratios Conjecture, Proceedings of the London Mathematical Society 2009, doi:10.1112/plms/pdp009.
- (8) The effect of convolving families of L-functions on the underlying group symmetries (with Eduardo Dueñez), Proceedings of the London Mathematical Society, 2009; doi: 10.1112/plms/pdp018.
- (9) A unitary test of the L-functions Ratios Conjecture (with John Goes, Steven Jackson, David Montague, Kesinee Ninsuwan, Ryan Peckner and Thuy Pham), Journal of Number Theory **130** (2010), no. 10, 2238–2258.
- (10) Towards an 'average' version of the Birch and Swinnerton-Dyer Conjecture (with John Goes), Journal of Number Theory **130** (2010), no. 10, 2341–2358.
- (11) An Orthogonal Test of the L-functions Ratios Conjecture, II (with David Montague), Acta Arith. **146** (2011), 53–90.
- (12) *Low-lying zeros of number field L-functions* (with Ryan Peckner), submitted September 2010 to Compositio Mathematica.
- (13) An elliptic curve family test of the Ratios Conjecture (with Duc Khiem Huynh and Ralph Morrison), submitted November 2010 to the Journal of Number Theory.
- (14) *Models for zeros at the central point in families of elliptic curves* (with Eduardo Dueñez, Duc Khiem Huynh, Jon Keating and Nina Snaith), preprint.

2. CLASSICAL RANDOM MATRIX THEORY

I have worked on numerous projects in classical random matrix theory, studying highly structured matrices. While not immediately applicable for number theory, it is useful to see how the structure effects the distribution of the eigenvalues. All of these results are joint with undergraduates. The proofs require combinatorics, Diophantine equations and algebraic topology.

Paper (1) highlights the danger of numerical exploration. Both we and another research group conjectured that the density of the normalized eigenvalues of these Toeplitz ensembles was a Gaussian after preliminary simulations in Mathematica; however, theoretical investigations then showed that the normalized fourth moment was $2 \ 2/3^{rds}$ and not 3 (one may interpret this in terms of obstructions to solving certain Diophantine equations). We were able to prove many results about this distribution (moments grow significantly slower than the Gaussian moments, but fast enough to ensure unbounded support). We conjectured that if we were to impose additional structure by making the first row a palindrome then these obstructions would vanish; we proved this in (2). We continued to explore the effect of increasing the structure, specifically by increasing the number of palindromes, in (3).

In paper (4) we weaken the structure a bit. We make each diagonal periodic with period m, and are able to derive a closed form expression for the density of eigenvalues by using techniques from complex analysis and algebraic geometry; it is quite unusual in the subject to be able to get a closed form solution.

- Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices (with Christopher Hammond), Journal of Theoretical Probability 18 (2005), no. 3, 537–566.
- (2) Distribution of eigenvalues of real symmetric palindromic Toeplitz matrices and circulant matrices (with Adam Massey and John Sinsheimer), Journal of Theoretical Probability **20** (2007), no. 3, 637–662.
- (3) *Distribution of eigenvalues for highly palindromic real symmetric Toeplitz matrices* (with Steven Jackson and Thuy Pham), to appear in the Journal of Theoretical Probability.
- (4) The Limiting Eigenvalue Density for the Ensemble of Symmetric Period m-Circulant Matrices (with Murat Koloğlu and Gene S. Kopp), submitted October 2010 to the Annals of Probability.

3. Additive and General Number Theory

I have written numerous papers in various parts of number theory, as well as the well-received textbook (12) which gives students a somewhat connected view of parts of modern analytic number theory, including being the first undergraduate book to cover random matrix theory. Of the papers below, the following are a representative sample of my work.

In (1), working with two graduate students to expand an idea in my thesis, we derive ways to construct one-parameter families of elliptic curves over $\mathbb{Q}(T)$ with moderate rank. Unlike other methods, ours does not require us to compute the associated height matrix and show that the resulting determinant is non-zero, bypassing this by applying results from Rosen and Silverman to show that the determination of certain Legendre sums specifies the rank.

In (3), (4), (5) I study how often A + A has greater cardinality than A - A for finite sets of integers A chosen with a given probability. As addition is commutative and subtraction is not, Nathanson conjectured that in the limit 100% of the time |A - A| > |A + A|; however, Martin and O'Bryant proved this is not the case. In (3) we salvaged the conjecture by showing that, while the claim fails if the probability of a k in $\{1, \ldots, n\}$ equals p with p independent of n, it is correct if instead we have p(n) which decays to 0. For some p(n) the results follow by Chebyshev, though in some ranges we need recent strong concentration theorems. There is a nice phase transition in behavior of the cardinalities as a function of p(n). In (4) and (5) we discuss constructions of sets with |A + A| > |A - A|, A chosen uniformly from $\{1, \ldots, n\}$. The previous records had densities of such examples on the order of $p(n)/2^{n/2}$ with p a polynomial; we're able to get densities of the size $1/n^4$, a significant improvement.

In (9), (10) we study an interesting generalization of results of Zeckendorf and Lekkerkerker that surprisingly seems to have not been asked before. Zeckendorf showed every integer can be written as a sum of non-adjacent Fibonacci numbers; Lekkerkerker then proved that on average numbers between the n^{th} and $(n + 1)^{\text{st}}$ Fibonacci numbers requires approximately $n/(\varphi^2 + 1)$ (with φ the golden mean). We prove that not just for the Fibonacci numbers, but for many recurrence relations, the distribution of the number of summands is a Gaussian as n grows. Interestingly, the proof involves forgetting about the number theory (previous approaches used continued fractions, for example), and recasting as a combinatorial problem.

- (1) Constructing one-parameter families of elliptic curves over $\mathbb{Q}(T)$ with moderate rank (with Scott Arms and Álvaro Lozano-Robledo), Journal of Number Theory **123** (2007), no. 2, 388–402.
- (2) An identity for sums of polylogarithm functions, Integers: Electronic Journal Of Combinatorial Number Theory **8** (2008), #A15.
- (3) When almost all sets are difference dominated (with Peter Hegarty), Random Structures and Algorithms **35** (2009), no. 1, 118–136.
- (4) Explicit constructions of infinite families of MSTD sets (with Brooke Orosz and Dan Scheinerman), Journal of Number Theory **130** (2010) 1221–1233.
- (5) Explicit constructions of infinite families of MSTD sets (with Dan Scheinerman), Additive Number Theory: Festschrift In Honor of the Sixtieth Birthday of Melvyn B. Nathanson (David Chudnovsky and Gregory Chudnovsky, editors), Springer-Verlag, 2010.
- (6) *Effective equidistribution and the Sato-Tate law for families of elliptic curves* (with M. Ram Murty), Journal of Number Theory **131** (2011), no. 1, 25–44.
- (7) A combinatorial identity for studying Sato-Tate type problems (with M. Ram Murty and Frederick Strauch), to appear in Rendiconti del Seminario Matematico.
- (8) *Moments of the rank of elliptic curves* (with Siman Wong), to appear in the Canadian Journal of Mathematics.
- (9) On the number of summands in Zeckendorf decompositions (with Murat Koloğlu, Gene S. Kopp and Yinghui Wang), submitted September 2010 to the Fibonacci Quarterly.
- (10) *From Fibonacci numbers to Central Limit Type Theorems* (with Yinghui Wang), submitted October 2010 to Transactions of the AMS.
- (11) *Quadratic fields with cyclic 2-class groups* (with Carlos Dominguez and Siman Wong), submitted November 2010 to Acta Arithmetica.
- (12) An Invitation to Modern Number Theory (with Ramin Takloo-Bighash), Princeton University Press, Princeton, NJ, 2006, 503 pages.
 - 4. BENFORD'S LAW (ACCOUNTING, PROBABILITY, NUMBER THEORY, ...)

In many mathematical, man-made and natural data sets, not all digits 1 through 9 are equally likely to occur as leading digits. Now known as Benford's law, for many sets of data the probability of having a first digit of d is $\log_B (1 + \frac{1}{d})$; thus we see a first digit of 1 approximately 30% of the time, significantly more than 11% (or 1/9th).

This phenomenon is not just of theoretical interest, but has powerful applications in data fraud. For example, the IRS uses it to detect tax fraud, and current work indicates it may also detect whether or not an image file has been doctored.

I am one of the experts in the field. I have written and refereed numerous papers in the subject, as well as co-organized the first conference in the field and am currently the senior editor of the first book on Benford's law (under contract with Princeton University Press). My work is both theoretical as well as applied. I have been extremely successful in getting undergraduates and graduate students involved, writing papers with six undergraduates and one graduate student.

Papers (1), (3), (4), (5), (7) advance the theory of the subject, deriving various sets of conditions that ensure Benford's law holds. The main techniques in (1) are Fourier analysis and quantified rates of equidistribution of $n\alpha \mod 1$ for irrational α , which involves the irrationality exponent and Baker's theory of linear forms of logarithms. The difficulty in showing the Benfordness of the 3x + 1 problem is that the key input is the structure theorem, which leads to a lattice supported distribution, and the discreteness causes numerous technical issues. Paper (3) provides sufficient conditions for when a product converges to Benford behavior, and when it does not. This is an extremely important case, as many observations can be regarded as the product of independent measurements. The main ingredient here is Fourier analysis.

Paper (2) involves the largest natural data set analyzed to date, hydrology data (almost half a million records spanning a century of measurements). The size of the data set introduces interesting complications in the analysis. Other data sets are studied in (8).

Paper (6) involves a new test to detect data fraud (or simply errors in data filing), building on the theory developed in (4). I also have a paper with a new test to detect tax fraud. This is currently being reviewed by the IRS (I have corresponded with several of their agents for years now, and have given a talk at the Boston headquarters). I am also working on a method to detect image fraud, which will be a chapter in the Benford book (9).

- (1) Benford's Law, values of L-functions and the 3x + 1 problem (with Alex Kontorovich), Acta Arithmetica **120** (2005), no. 3, 269–297.
- (2) Benford's Law applied to hydrology data results and relevance to other geophysical data (with Mark Nigrini), Mathematical Geology **39** (2007), no. 5, 469–490.
- (3) *The Modulo* 1 *Central Limit Theorem and Benford's Law for Products* (with Mark Nigrini), International Journal of Algebra **2** (2008), no. 3, 119–130.
- (4) Order statistics and Benford's law (with Mark Nigrini), International Journal of Mathematics and Mathematical Sciences, Volume 2008 (2008), Article ID 382948, 19 pages. doi:10.1155/2008/382948
- (5) Chains of distributions, hierarchical Bayesian models and Benford's Law (with D. Jang, J. U. Kang, A. Kruckman and J. Kudo), Journal of Algebra, Number Theory: Advances and Applications, volume 1, number 1 (March 2009), 37–60.
- (6) Data diagnostics using second order tests of Benford's Law (with Mark Nigrini), Auditing: A Journal of Practice and Theory 28 (2009), no. 2, 305–324. doi: 10.2308/aud.2009.28.2.305
- (7) *The Weibull distribution and Benford's law* (with Victoria Cuff and Allie Lewis), preprint.
- (8) *Climate, hydrology and election data and Benford's law* (with Victoria Cuff and Allie Lewis), preprint.

(9) *Theory and Applications of Benford's Law* (senior editor; co-editors Arno Berger and Ted Hill), Princeton University Press, under contract.

5. ECONOMICS AND MARKETING

I have worked on many problems in economics and marketing. My wife was a graduate student at Wharton, and through her I met and began collaborations with several faculty members there.

The first paper involves mathematics very similar to what I have used in my studies in random matrix theory and combinatorics. The goal was to derive closed form expressions for Bayesian inferences, as it is highly desirable to be able to see the parameter dependence in the solutions theoretically and not have to resort to numerical simulations. In addition to theoretical considerations, this project required computational expertise as well to demonstrate the efficacy of our series expansions.

The second paper applies linear programming to help a movie theater optimize its revenue subject to hard constraints (such as two movies cannot simultaneously play on the same screen) and soft constraints (such as the manager's preference that there is never a 20 minute window when no movie starts playing). Our algorithm has been implemented by the Pathe theater in the Netherlands, leading to a significant improvement in their revenue. We only have two hours from when we receive the list of candidate movies to when we must inform the newspapers of the schedule; this forces us to be extremely efficient in setting up the linear program. Our paper won the award for best paper in the International Journal of Research in Marketing in 2009, leading to the short follow-up note (3).

The last paper (4) deals with various models of social learning.

- (1) *Closed-form Bayesian inference for the logit model via polynomial expansions* (with Eric T. Bradlow and Kevin Dayaratna), Quantitative Marketing and Economics **4** (2006), no. 2, 173–206.
- (2) Silver Scheduler: a demand-driven modeling approach for the construction of micro-schedules of movies in a multiplex (with Jehoshua Eliashberg, Quintus Hegie, Jason Ho, Dennis Huisman, Sanjeev Swami, Charles B. Weinberg and Berend Wierenga), Intern. J. of Research in Marketing (2009), doi:10.1016/j.ijresmar.2008.09.004. Award for best paper in IJRM, 2009.
- (3) *Demand-driven scheduling of movies in a multiplex* (with Jehoshua Eliashberg and Charles B. Weinberg), newsletter of the European Marketing Academy, October 2010 (requested summary of Silver-Scheduler paper in honor of it receiving the IJRM Best Paper Award for 2009).
- (4) *Social Learning, Opinion Leaders and Herding* (with Daniel Stone), submitted October 2010 to Economica.

6. COMPUTATIONAL PAPERS

Many of my number theory projects require extensive programming, frequently requiring integration of special environments such as PARI with C.

Paper (1) is mostly theoretical, and involves solving some special cases of a conjecture in computer science that certain very efficient circuits cannot be built by reducing this to a problem in counting the number of solutions in certain sub-varieties.

Paper (2) studies the distribution of zeros of elliptic curve L-functions near the central point. Many authors (including myself in my thesis) have studied these zeros. While we know their behavior as the conductors tend to infinity, the data in the case of finite conductors is quite mysterious and very different from what we know is true in the limit. In the course of studying the data, I noticed interesting patterns that led us to what we believe is the correct conjecture to describe this behavior.

Paper (3) looks at the distribution of the second largest eigenvalue in many families of d-regular graphs. These graphs are extremely important in modern communication theory, as they led to cheap graphs to build that are very well connected. In studying the observed growth rates, we were led to certain conjectures as to the probability that certain graphs are Ramanujan.

Paper (4) is a continuation of (2). Using SAGE and Matlab, we solve a Painlevé VI equation. This is needed to numerically compute eigenvalues from our Jacobi random matrix ensembles.

Paper (5) is a work in progress on virus propagation in certain graph networks. One of the most important applications is in understanding the spread of viruses (such as the avian bird flu).

- (1) *Incomplete quadratic exponential sums in several variables* (with Eduardo Dueñez, Amitabha Roy and Howard Straubing), Journal of Number Theory **116** (2006), no. 1, 168–199.
- (2) Investigations of zeros near the central point of elliptic curve L-functions, Experimental Mathematics **15** (2006), no. 3, 257–279.
- (3) The distribution of the second largest eigenvalue in families of random regular graphs (with Tim Novikoff and Anthony Sabelli), Experimental Mathematics 17 (2008), no. 2, 231–244.
- (4) *The lowest eigenvalue of Jacobi Random Matrix Ensembles and Painlevé VI*, (with Eduardo Dueñez, Duc Khiem Huynh, Jon Keating and Nina Snaith), Journal of Physics A: Mathematical and Theoretical **43** (2010) 405204 (27pp).
- (5) *Virus propagation in certain types of networks* (with Leo Kontorovich and Amitabha Roy), in preparation.

7. SABERMETRICS AND STATISTICS

As a mathematician and a lifelong Red Sox fan (who also roots for the Phillies after going to graduate school in Princeton and moving to Philadelphia during my postdoc there), I've always been fascinated by the power of mathematics in baseball. I have become active in the sabermetrics (studying baseball through mathematics and statistics) community. In addition to writing papers, I have taught several independent studies in the subject, which are wonderful ways to channel student enthusiasm into mathematics.

I have been corresponding with the San Diego Padres for several years, and have had students work on some projects for them during this time (which I am not at liberty to discuss further). My main result in the field is the first paper, where I show Bill James' Pythagorean Won-Loss statistic (which he was led to from looking at years of results) can be derived from some (mostly) reasonable assumptions about how a baseball game is played. Given a sports league, if the observed average number of runs a team scores and allows are RS_{obs} and RA_{obs}, then the Pythagorean Formula predicts the team's won-loss percentage should be $\frac{RS_{obs}\gamma}{RS_{obs}\gamma+RA_{obs}\gamma}$ for some γ which is constant for the league. Initially in baseball the exponent γ was taken to be 2 (which led to the name), though fitting γ to the observed records from many seasons lead to the best γ being about 1.82. This statistic is widely used, and is shown in the standings on sites such as ESPN and MLB. I was led to model the runs scored and allowed by Weibulls due to my physics experience; these distributions do a much better job than others that had been tried before in the literature.

The second paper is a short note on a distribution where the Cramér-Rao inequality provides no information. The key input here is the dominated convergence theorem and the modification of a density closely associated to Benford problems.

- 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).
- (2) When the Cramér-Rao Inequality provides no information, Communications in Information and Systems 7 (2007), no. 3, 265–272.

8. EXPOSITORY AND GENERAL MATHEMATICS

I have written numerous expository articles. My favorite is (2), the only history of the connections between number theory and random matrix theory. This paper was written jointly with my physics mentor from my undergraduate days, who was one of the key experimenters in gathering the nuclear physics data leading to the birth of random matrix theory in physics. In (3) we generalize Tennenbaum's beautiful geometric proof of the irrationality of $\sqrt{2}$ to handle several other integers, and describe the obstructions that prevent it from handling \sqrt{n} for a larger class of square-free integers. In (5) (recently posted on the arxiv, and to be submitted after a few more days wait for additional comments) we explore yet another difference between real and complex analysis.

Additionally, I am working on two book projects. The first (6) (under contract with Princeton University Press) is a general probability book designed to supplement any course to both help struggling students as well as encourage students wishing to know more. This book is being jointly written with one of my undergraduates. The second (7) is a cryptography book. We have a very extensive draft, which my colleagues have used successfully at Rutgers and I at Williams.

- (1) A probabilistic proof of Wallis' formula for π , Amer. Math. Monthly **115** (2008), no. 8, 740–745.
- (2) *Nuclei, Primes and the Random Matrix Connection* (with Frank W. K. Firk), invited submission to Symmetry **1** (2009), 64–105; doi:10.3390/sym1010064.
- (3) *Rational irrationality proofs* (with David Montague), to appear in Mathematics Magazine.
- (4) Isoperimetric Sets of Integers (with Frank Morgan, Edward Newkirk, Lori Pedersen and Deividas Seferis), to appear in Mathematics Magazine.

- (5) The real analogue of the Schwarz lemma (with David Thompson), to appear in the American Mathematical Monthly. http://arxiv.org/abs/1012. 0585
- (6) If a prime divides a product... (with Cesar Silva).
- (7) *The Probability Lifesaver* (with David Thompson), Princeton University Press, under contract.
- (8) *The Mathematics of Encryption* (with Midge Cozzens and Wesley Pegden), general cryptography textbook.

E-mail address: Steven.J.Miller@williams.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267

10