

YET ANOTHER ZETA FUNCTION AND LEARNING

IGOR RIVIN

ABSTRACT. We analyze completely the convergence speed of the batch learning algorithm, and compare its speed to that of the memoryless learning algorithm and of learning with memory (as analyzed in [KR2001]). We show that the batch learning algorithm is never worse than the memoryless learning algorithm (at least asymptotically). Its performance vis-a-vis learning with full memory is less clearcut, and depends on certain probabilistic assumptions. These results necessitate the introduction of the moment zeta function of a probability distribution and the study of some of its properties.

INTRODUCTION

The original motivation for the work in this paper was provided by research in learning theory, specifically in various models of language acquisition (see, for example, [KNN2001, NKN2001, KN2001]). In the paper [KR2001H], we had studied the speed of convergence of the memoryless learner algorithm, and also of learning with full memory. Since the batch learning algorithm is both widely known, and believed to have superior speed (at the cost of memory) to both of the above methods by learning theorists, it seemed natural to analyze its behavior under the same set of assumptions, in order to bring the analysis in [KR2001H] and [KR2001] to a sort of closure. It should be noted that the detailed analysis of the batch learning algorithm is performed under the assumption of independence, which was not explicitly present in our previous work. For the impatient reader we state our main result (Theorem 8.1) immediately (the reader can compare it with the results on the memoryless learning algorithm and learning with full memory, as summarized in Theorem 2.1):

Theorem A. Let $N_{\Delta}$ be the number of steps it takes for the student (with probability 1) to have probability $1 - \Delta$ of learning the concept

1991 Mathematics Subject Classification. 60E07, 60F15, 60J20, 91E40, 26C10.
Key words and phrases. learning theory, zeta functions, asymptotics.

The author would like to thank the EPSRC and the NSF for support, and Natalia Komarova and Ilan Vardi for useful conversations.
using the batch learner algorithm. Then we have the following estimates for $N_\Delta$:

- If the distribution of overlaps is uniform, or more generally, the density function $f(1-x)$ at 0 has the form $f(x) = c + O(x^\delta)$, $\delta, c > 0$, then $N_\Delta = |\log \Delta| \Theta(n)$.
- If the probability density function $f(1-x)$ is asymptotic to $x^\beta + O(x^{\beta-\delta})$, $\delta, \beta > 0$, as $x$ approaches 0, then we have $N_\Delta = |\log \Delta| \Theta(n^{1/(1+\beta)})$.
- If the asymptotic behavior is as above, but $-1/2 < \beta < 0$, then $N_\Delta = |\log \Delta| \Theta(n^{1/(1+\beta)})$.

The plan of the paper is as follows: in this Introduction we recall the learning algorithms we study; in Section 1 we define our mathematical model; in Section 2 we recall our previous results, in Section 3 we begin the analysis of the batch learning algorithm, and introduce some of the necessary mathematical concepts; in Sections 4-6 we analyze the three cases stated in Theorem A, and we summarize our findings in Section 7.

**Memoryless Learning and Learning with Full Memory.** The general setup is as follows: There is a collection of concepts $R_0, \ldots, R_n$ and words which refer to these concepts, sometimes ambiguously. The teacher generates a stream of words, referring to the concept $R_0$. This is not known to the student, but he must learn by, at each step, guessing some concept $R_i$ and checking for consistency with the teacher’s input. The memoryless learner algorithm consists of picking a concept $R_i$ at random, and sticking by this choice, until it is proven wrong. At this point another concept is picked randomly, and the procedure repeats. Learning with full memory follows the same general process with the important difference that once a concept is rejected, the student never goes back to it. It is clear (for both algorithms) that once the student hits on the right answer $R_0$, this will be his final answer. We would like to estimate the probability of having guessed the right answer is after $k$ steps, and also the expected number of steps before the student settles on the right answer.

**Batch Learning.** The batch learning situation is similar to the above, but here the student records the words $w_1, \ldots, w_k, \ldots$ he gets from the teacher. For each word $w_i$, we assume that the student can find (in his textbook, for example) a list $L_i$ of concepts referred to by the word. If
we define

\[ \mathcal{L}_k = \bigcap_{i=1}^{k} L_i, \]

then we are interested in the smallest value of \( k \) such that \( \mathcal{L}_k = \{ R_0 \} \). This value \( k_0 \) is the time it has taken the student to learn the concept \( R_0 \). We think of \( k_0 \) as a random variable, and we wish to estimate its expectation.

1. The mathematical model

We think of the words referring to the concept \( R_0 \) as a probability space \( \mathcal{P} \). The probability that one of these words also refer to the concept \( R_i \) shall be denoted by \( p_i \); the probability that a word refers to concepts \( R_{i_1}, \ldots, R_{i_k} \) shall be denoted by \( p_{i_1 \ldots i_k} \). All the results described below (obviously) depend in a crucial way on the \( p_1, \ldots, p_n \) and (in the case of the batch learning algorithm) also on the joint probabilities. Since there is no a priori reason to assume specific values for the probabilities, we shall assume that all of the \( p_i \) are themselves independent, identically distributed random variables. We shall refer to their common distribution as \( \mathcal{F} \), and to the density as \( f \). It turns out that the convergence properties of the various learning algorithms depend on the local analytic properties of the distribution \( \mathcal{F} \) at 1 – some moments reflection will convince the reader that this is not really so surprising.

To carry out a precise analysis of the batch learning algorithm, we will also need the independence hypothesis:

\[ p_{i_1 \ldots i_k} = p_{i_1} \cdots p_{i_k}. \]

It is again not too surprising that some such assumption on correlations ought to be required for precise asymptotic results, though it is obviously the subject of a (non-mathematical) debate as to whether assuming that the various concepts are truly independent is reasonable from a cognitive science point of view.

2. Previous results

In previous work \[KR2001a\] and \[KR2001b\] we obtained the following result.

**Theorem 2.1.** Let \( N_{\Delta} \) be the number of steps it takes for the student (with probability 1) to have probability \( 1 - \Delta \) of learning the concept. Then we have the following estimates for \( N_{\Delta} \):

\[ \text{...} \]
• if the distribution of overlaps is uniform, or more generally, the density function $f(1-x)$ at 0 has the form $f(x) = c + O(x^\delta)$, $\delta, c > 0$, then $N_\Delta = |\log \Delta| \Theta(n \log n)$ for the memoryless algorithm and $N_\Delta = (1 - \Delta)^2 \Theta(n \log n)$ when learning with full memory;

• if the probability density function $f(1-x)$ is asymptotic to $x^{\beta} + O(x^{\beta-\delta})$, $\delta, \beta > 0$, as $x$ approaches 0, then for the two algorithms we have respectively $N_\Delta = |\log \Delta| \Theta(n)$ and $N_\Delta = (1 - \Delta)^2 \Theta(n)$;

• if the asymptotic behavior is as above, but $-1 < \beta < 0$, then $N_\Delta = |\log \Delta| \Theta(n^{1/(1+\beta)})$ for the memoryless learner and $(N_\Delta = 1 - \Delta)^2 \Theta(n^{1/(1+\beta)})$ for learning with full memory.

Recall that $f(x) = \Theta(g(x))$ means that for sufficiently large $x$, the ratio $f(x)/g(x)$ is bounded between two strictly positive constants. The distribution of overlaps referred to above is simply the distribution $F$. Notice that the theorem says nothing about the situation when $F$ is supported in some interval $[0, a]$, for $a < 1$. That case is (presumably) of scientific interest, but mathematically it is relatively trivial: we replace the arguments of all the $\Theta$s above by 1, though, of course, we are thereby hiding the dependence on $a$.

3. General bounds on the batch learner algorithm

Consider a set of words $w_1, \ldots, w_k$. The probability that they all refer to the concept $R_i$ is, obviously $p_i^k$.

Lemma 3.1. The probability $q_k$ that we still have not learned the concept $R_0$ after $k$ steps is bounded above by $\sum_{i=1}^n p_i^k$, and below by $\max_i p_i^k$.

Proof. Immediate. \square

We will first use these upper and lower bounds to get corresponding bounds on the convergence speed of the batch learner algorithm, and then invoke the independence hypothesis to sharpen these bounds in many cases.

We begin with a trivial but useful lemma.

Lemma 3.2. Let $G$ be a game where the probability of success (respectively failure) after at most $k$ steps is $s_k$ (respectively $f_k = 1 - s_k$). Then the expected number of steps until success is

$$\sum_{k=1}^\infty k(s_k - s_{k-1}) = \sum_{k=1}^\infty s_k = 1 - \sum_{k=1}^\infty f_k,$$

if the corresponding sum converges.
Proof. The proof is immediate from the definition of expectation and
the possibility of rearrangement of terms of positive series.

We can combine Lemma 3.2 and Lemma 3.1 to obtain:

**Theorem 3.3.** The expected time $T$ of convergence of the batch learner
algorithm is bounded as follows:

\[
\sum_{i=1}^{n} \frac{1}{1 - p_i} \geq T \geq \max_{1 \leq i \leq n} \frac{1}{1 - p_i}.
\]

The leftmost term in equation (1) has been studied at length in
[KR2001a]. We state a version of the results of [KR2001a] below:

**Theorem 3.4.** Let $S = \sum_{i=1}^{n} \frac{1}{1 - p_i}$, where the $p_i$ are independently
identically distributed random variables with values in $[0, 1]$, with prob-
ability density $f$, such that $f(1 - x) = x^\beta + O(x^{\beta - \delta})$, \, $\delta > 0$ for $x \to 0$.
Then If $\beta > 0$, then there exists a mean $m$, such that $\lim_{n \to \infty} \mathbb{P}(|S/n - m| > \epsilon) = 0$, for any $\epsilon > 0$. If $\beta = 0$, then $\lim_{n \to \infty} \mathbb{P}(|S/(n \log n) - 1| > \epsilon) = 0$). Finally, if $-1 \leq \beta < 0$, then $\lim_{n \to \infty} \mathbb{P}(S/n^{1/\beta + 1} - C > a) = g(a)$, where $\lim_{a \to \infty} g(a) = 0$, and $C$ is an arbitrary (but fixed) constant,
and likewise

\[
\mathbb{P}(S/n^{1/(\beta + 1)} < b) = h(b),
\]

where $\lim_{a \to 0} h(a) = 0$.

The right hand side of Eq. (1) is easier to understand. Indeed, let
$p_1, \ldots, p_n$ be distributed as usual (and as in the statement of Theorem
3.4). Then

**Theorem 3.5.** The expected value of $\max_{1 \leq i \leq n} p_i$ equals $1 - Cn^{-1/1+\beta}$,
for some positive constant $C$.

Proof. First, we change variables to $q_i = 1 - p_i$. Obviously, the state-
ment of the Theorem is equivalent to the statement that $E = \mathbb{E}(\min_{1 \leq i \leq n} q_i) =
Cn^{-1/1+\beta}$. We also write $h(x) = f(1 - x)$, and similarly for the prim-
itives $H$ and $F$. Now, the probability of that all of the $q_i$ are greater
than some fixed $y$ equals $1 - (1 - H(y))^n$, so that

\[
E = \int_{0}^{1} td [1 - (1 - H(t))^n] = \int_{0}^{1} (1 - H(t))^n dt.
\]

Perform the change of variables $t = u/n^{1/(1+\beta)}$, to get

\[
E = \frac{1}{n^{1+\beta}} \int_{0}^{n^{1/(1+\beta)}} (1 - H(u/n^{1/(1+\beta)}))^n du.
\]
For $u \ll n^{1/(1+\beta)}$, we can write $H(u/n^{1/(1+\beta)}) \asymp u^{\beta+1}/nH'$, where $H'$ is a constant. We also know that $H$ is a monotonic function so if we break up the integral above as

$$(3) \quad E = \frac{1}{n^{1/(1+\beta)}} \left[ \int_0^{n^{1/(2(1+\beta))}} + \int_{n^{1/(2(1+\beta))}}^\infty \right] (1 - H(u/n^{1/(1+\beta)})) du,$$

we see that the first integral approaches $C = \int_0^\infty \exp(-u^{1/(1+\beta)}) du$, while the second integral goes to 0. Note that the proof also evaluates $C$.

We need one final observation:

**Theorem 3.6.** The variable $n^{1/(1+\beta)} \min_{i=1}^n q_i$ has a limiting distribution with distribution function $G(x) = 1 - \exp(-x^{1+\beta})$.

**Proof.** Immediate from the proof of Theorem 3.5. \qed

We can now put together all of the above results as follows.

**Theorem 3.7.** Let $p_1, \ldots, p_k$ be independently distributed with common density function $f$, such that $f(1 - x) = cx^\beta + O(x^{\beta+\delta})$, $\delta > 0$. Let $T$ be the expected time of the convergence of the batch learning algorithm with overlaps $p_1, \ldots, p_k$. Then, if $\beta > 0$, then there exist $C_1, C_2$, such that $C_1 n^{1/(1+\beta)} \leq T \leq C_2 n$, with probability tending to 1 as $n$ tends to $\infty$. If $\beta = 0$, then there exist $C_1, C_2$, such that $C_1 n \leq T \leq C_2 n \log n$, with probability tending to one as $n$ tends to $\infty$. If $\beta > 0$, then $C^{-1} n^{1/(\beta+1)} \leq T \leq C n^{1/(\beta+1)}$ with probability tending to 0 as $C$ goes to infinity.

The reader will remark that in the case that $\beta > 0$, the upper and lower bounds have the same order of magnitude as functions of $n$.

4. **INDEPENDENT CONCEPTS**

independence hypothesis, whereby an application of the inclusion-exclusion principle gives us:

**Lemma 4.1.** The probability $l_k$ that we have learned the concept $R_0$ after $k$ steps is given by

$$l_k = \prod_{i=1}^n (1 - p_i^k).$$

Note that the probability $s_k$ of winning the game on the $k$-th step is given by $s_k = l_k - l_{k-1} = (1 - l_{k-1}) - (1 - l_k)$. Since the expected
number of steps $T$ to learn the concept is given by

$$T = \sum_{k=1}^{\infty} ks_k,$$

we immediately have

$$T = \sum_{k=1}^{\infty} (1 - l_k).$$

**Lemma 4.2.** The expected time $T$ of learning the concept $R_0$ is given by

$$T = \sum_{k=1}^{\infty} \left( 1 - \prod_{i=1}^{n} (1 - p^k_i) \right).$$

Since the sum above is absolutely convergent, we can expand the products and interchange the order of summation to get the following formula for $T$:

$$T = \sum_{s \subseteq \{1, \ldots, n\}} (-1)^{|s|-1} \sum_{k=1}^{\infty} p^k_{s} = \sum_{s \subseteq \{1, \ldots, n\}} (-1)^{|s|-1} \left( \frac{1}{1 - p_{s}} - 1 \right),$$

where we have identified subsets of $\{1, \ldots, n\}$ with the corresponding multindexes.

The formula (4) is useful in and of itself, but we now use it to attempt to get the expectation of the expected time of success $T$ under our distribution and independence assumption. For this we shall need the following:

**Definition 4.3.** Let $\mathcal{F}$ be a probability distribution on an interval $I$, and let $m_k(\mathcal{F}) = \int_I x^k \mathcal{F}(dx)$ be the $k$-th moment of $\mathcal{F}$. Then the moment zeta function of $\mathcal{F}$ is defined to be

$$\zeta_\mathcal{F}(s) = \sum_{k=1}^{\infty} m^s_k(\mathcal{F}),$$

whenever the sum is defined.

**Lemma 4.4.** Let $\mathcal{F}$ be a probability distribution as above, and let $x_1, \ldots, x_n$ be independent random variables with common distribution $\mathcal{F}$. Then

$$\mathbb{E} \left( \frac{1}{1 - x_1 \ldots x_n} \right) = \zeta_\mathcal{F}(n).$$

In particular, the expectation is undefined whenever the zeta function is undefined.
Proof. Expand the fraction in a geometric series and apply Fubini’s theorem.

Example 4.5. For \( F \) the uniform distribution on \([0, 1]\), \( \zeta_F \) is the familiar Riemann zeta function. Notice that this is not defined for \( n = 1 \) – this will be important in the sequel.

It should be noted that in the case we are interested in (distributions supported in \([0, 1]\)), the asymptotics of the moments are determined by the local properties of the distribution at 1, up to exponentially decreasing error terms. So, if \( f(1 - x) \asymp x^\beta \) (recall that \( f \) is the density), we see that the \( k \)-th moment of \( F \) is asymptotic to \( C k^{-(1+\alpha)} \), for some constant \( C \). To show this, we first define the Mellin transform of \( f \) to be

\[
\mathcal{M}(f)(s) = \int_0^1 f(x) x^{s-1} dx.
\]

We see that \( m_k(F) = \mathcal{M}(f)(k+1) \). Mellin transform is very closely related to the Laplace transform. Indeed, making the substitution \( x = \exp(-u) \), we see that

\[
\mathcal{M}(f) = \int_0^\infty f(\exp(-u)) \exp(-su) du,
\]

so the Mellin transform of \( f \) is equal to the Laplace transform of \( f \circ \exp \).

Now, the asymptotics of the Laplace transform are easily computed by Laplace’s method, and in the case we are interested in, Watson’s lemma (see, eg, [BenOrsz]) tells us that if \( f(x) \asymp c(1 - x)^\beta \), then \( \mathcal{M}(f)(s) \asymp c \Gamma(\beta) x^{-(\beta+1)} \). In particular, \( \zeta_F(s) \) is defined for \( s > 1/(1 + \beta) \). Below we shall analyze three cases (though the analysis is almost the same in the three cases, there are some important variations). In the sequel, we set \( \alpha = \beta + 1 \).

5. \( \alpha > 1 \)

In this case, we use our assumptions to rewrite Eq. (4) as

\[
T = -\sum_{k=1}^{n} \binom{n}{k} (-1)^k \zeta_F(k).
\]

This, in turn, can be rewritten (by expanding the definition of zeta) as

\[
T = -\sum_{j=1}^{\infty} [(1 - m_j(F))^{n} - 1]
\]

Since the term in the sum is monotonically decreasing, the sum in Eq. (6) can be approximated by an integral (of any monotonic interpolation
m of the sequence \( m_j(\mathcal{F}) \); however there is no reason not to set \( m(x) = \mathcal{M}(f)(x+1) \), with error bounded by the first term, which is, in term, bounded in absolute value by 2, to get

\[
T = - \int_1^\infty [(1 - m(x))^n - 1] \, dx + O(1),
\]

where the error term is bounded above by 2.

Now, let us assume that \( m(x) \) is of order \( x^{-\alpha} \) for some \( \alpha > 1 \). We substitute \( x = n^{1/\alpha}/u \), to get

\[
T = n^{1/\alpha} \int_0^{n^{1/\alpha}} \frac{1 - (1 - m(n^{1/\alpha}/u)^n)}{u^2} \, du + O(1)
\]

\[
= n^{1/\alpha} \int_0^{n^{1/\alpha}} \frac{1 - (1 - m'(u)u^{\alpha}/n)^n}{u^2} \, du + O(1)
\]

\[
= n^{1/\alpha} \left( \int_0^{n^{1/2\alpha}} \frac{1 - (1 - m'(u)u^{\alpha}/n)^n}{u^2} \, du + O(1) \right),
\]

where \( m' \) is a bounded (asymptotically constant) function. In the second integral the integrand is bounded above by \( 1/u^2 \), so the contribution from that integral goes to 0, while in the first integral we can approximate \( (1 - m'(u)u^{\alpha}/n)^n \) by \( \exp(-m'(u)u^{\alpha}) \), and the contribution from that integral goes to

\[
T = n^{1/\alpha} \int_0^\infty \frac{1 - \exp(-m'(u)u^{\alpha})}{u^2} \, du + O(1) \approx C n^{1/\alpha}.
\]

6. \( \alpha = 1 \)

In this case, \( f(x) = c + o(1) \) as \( x \) approaches 1. It is not hard to see that \( \zeta_F(n) \) is defined for \( n \geq 2 \). We break up the expression in Eq. (4) as

\[
T = \sum_{j=1}^n \frac{1}{1 - p_j} - 1 + \sum_{s \subseteq \{1, \ldots, n\}, \ |s| > 1} (-1)^{|s|-1} \left( \frac{1}{1 - p_s} - 1 \right).
\]

Let

\[
T_1 = \sum_{j=1}^n \frac{1}{1 - p_j} - 1,
\]

\[
T_2 = \sum_{s \subseteq \{1, \ldots, n\}, \ |s| > 1} (-1)^{|s|-1} \left( \frac{1}{1 - p_s} - 1 \right).
\]
The first sum $T_1$ has no expectation, however $T_1/n$ does have a stable distribution centered on $c \log n + c_2$. We will keep this in mind, but now let us look at the second sum $T_2$. It can be rewritten as

\begin{equation}
T_2 = - \sum_{j=1}^{\infty} [(1 - m_j(F))^n - 1 + nm_j].
\end{equation}

The same method as in section 5 under the assumption that the $k$-th moment is asymptotic to $k^\alpha$ (this time for $\alpha \leq 1$) can be used to write

\begin{equation}
T_2 = n \int_{n^{1/2}}^{n} \frac{[1 - nm(u/n) - (1 - m(u/n))^n]}{u^2} du + O(1)
= n \left( \int_{0}^{n^{1/2}} + \int_{n^{1/2}}^{n} \right) \frac{[1 - m'(u)u - (1 - m'(u)u/n)^n]}{u^2} du + O(1).
\end{equation}

The conclusion differs somewhat from that of section $\S$ in that we get an additional term of $cn \log n$, where $c = \lim_{x \to 1} f(x) = \lim_{j \to \infty} jm_j$. This term is equal (with opposing sign) to the center of the stable law satisfied by $T_1$, so in case $\alpha = 1$, we see that $T$ has no expectation but satisfies a law of large numbers, of the following form:

**Theorem 6.1** (Law of large numbers). *There exists a constant $C$ such that $\lim_{y \to \infty} P(|T/n - C| > y) = 0$.*

7. $\alpha < 1$

In this case the analysis goes through as in the preceding section when $\alpha > 1/2$, but then runs into considerable difficulties. However, in this case we note that Theorem $\S$ actually gives us tight bounds.

8. The inevitable comparison

We are now in a position to compare the performance of the batch learning algorithm with that of the memoryless learning algorithm and of learning with full memory, as summarized in Theorem $\S$. We combine our computations above with the observation that the batch learner algorithm converges geometrically (Lemma 4.1), to get:

**Theorem 8.1.** *Let $N_\Delta$ be the number of steps it takes for the student (with probability 1) to have probability $1 - \Delta$ of learning the concept using the batch learner algorithm. Then we have the following estimates for $N_\Delta$:

- If the distribution of overlaps is uniform, or more generally, the density function $f(1 - x)$ at 0 has the form $f(x) = c + O(x^\delta)$, $\delta, c > 0$, then $N_\Delta = |\log \Delta| \Theta(n)$.*
• If the probability density function \( f(1-x) \) is asymptotic to \( x^{\beta} + O(x^{3-\delta}) \), \( \delta, \beta > 0 \), as \( x \) approaches 0, then we have \( N_\Delta = |\log \Delta| \Theta(n^{1/(1+\beta)}) \);

• If the asymptotic behavior is as above, but \( -1 < \beta < 0 \), then \( N_\Delta = |\log \Delta| \Theta(n^{1/(1+\beta)}) \).

Comparing Theorems 2.1 and 8.1, we see that batch learning algorithm is uniformly superior for \( \beta \geq 0 \), and the only one of the three to achieve sublinear performance whenever \( \beta > 0 \) (the other two never do better than linearly, unless the distribution \( F \) is supported away from 1.) On the other hand, for \( \beta < 0 \), the batch learning algorithm performs comparably to the memoryless learner algorithm, and worse than learning with full memory.

References

[BenOrsz] C. M. Bender and S. Orszag (1999) Advanced mathematical methods for scientists and engineers, I, Springer-Verlag, New York.

[KNN2001] Komarova, N. L., Niyogi, P. and Nowak, M. A. (2001) The evolutionary dynamics of grammar acquisition, J. Theor. Biology, 209(1), pp. 43-59.

[KN2001] Komarova, N. L. and Nowak, M. A. (2001) Natural selection of the critical period for grammar acquisition, Proc. Royal Soc. B, to appear.

[KR2001a] Komarova, N. L. and Rivin, I. (2001) Harmonic mean, random polynomials and stochastic matrices, preprint.

[KR2001b] Komarova, N. L. and Rivin, I. (2001) On the mathematics of learning.

[Niyogi1998] Niyogi, P. (1998). The Informational Complexity of Learning. Boston: Kluwer.

[NKN2001] Nowak, M. A., Komarova, N. L., Niyogi, P. (2001) Evolution of universal grammar, Science 291, 114-118.

Mathematics department, University of Manchester, Oxford Road, Manchester M13 9PL, UK

Mathematics Department, Temple University, Philadelphia, PA 19122

Mathematics Department, Princeton University, Princeton, NJ 08544

E-mail address: irivin@math.princeton.edu