The Probability that a Matrix of Integers Is Diagonalizable

Andrew J. Hetzel, Jay S. Liew, and Kent E. Morrison

1. INTRODUCTION. It is natural to use integer matrices for examples and exercises when teaching a linear algebra course, or, for that matter, when writing a textbook in the subject. After all, integer matrices offer a great deal of algebraic simplicity for particular problems. This, in turn, lets students focus on the concepts. Of course, to insist on integer matrices exclusively would certainly give the wrong idea about many important concepts. For example, integer matrices with integer matrix inverses are quite rare, although invertible integer matrices (over the rational numbers) are relatively common. In this article, we focus on the property of diagonalizability for integer matrices and pose the question of the likelihood that an integer matrix is diagonalizable. Specifically, we ask: *What is the probability that an $n \times n$ matrix with integer entries is diagonalizable over the complex numbers, the real numbers, and the rational numbers, respectively?*

Probabilistic questions about the integers have a rich history. In 1874, Mertens proved that the probability that two positive integers are relatively prime is $6/\pi^2$, and, in 1885, Gegenbauer proved that the probability that a positive integer is square-free is also $6/\pi^2$. Hardy and Wright [**2**] is a good source for these and related results. The use of the term "probability" in this context by Hardy and Wright needs to be explained. Following Kolmogorov's axiomatization of the foundations of probability in the 1930s, mathematicians have required that a probability measure be countably additive. This means that the process of randomly selecting an integer, with each integer equally probable, is impossible to achieve. The results of Mertens and Gegenbauer are actually statements about limits of probabilities. In particular, for each positive integer k let p_k be the probability that two integers between 1 and k are relatively prime, where the probability measure on $\{1, 2, \dots, k\}$ is normalized counting measure (i.e., $P(\{i\}) = 1/k$ for $i = 1, 2, \dots, k$). Then, as k goes to infinity, the limit of p_k is $6/\pi^2$. In a similar way, the probability that an integer between 1 and k is square-free has the limit $6/\pi^2$ as $k \to \infty$. For this article, we adopt an analogous approach. That is, for $n \times n$ integer matrices and some property of matrices we first consider the probability that such matrices with entries in the range from $-k$ to k have that property. We then define the limit of these probabilities as k goes to infinity to be the "probability" that this property holds among all integer matrices of a given size.

We could choose to avoid the term "probability" in this sense by using "density" in its place. As in [**4**], the (natural) density of a subset S of the positive integers is the limit as $k \to \infty$ (if it exists) of the probability that an integer between 1 and k is in S. However, we prefer to use the more familiar term "probability" with the understanding that it does not arise from a countably additive measure on the sample space of integer matrices. Although the axiom of countable additivity is orthodox, it has not been universally accepted. One notable probabilist opposed to countable additivity was Bruno de Finetti, who advocated the less restrictive axiom of finite additivity for a probability measure. The "probabilities" that we are concerned with can be seen as coming from finitely additive measures on countable sample spaces.

Throughout this article, $|\mathcal{S}|$ signifies the cardinality of a set $\mathcal{S}$. We use the notation $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ to denote the sets of integers, rational numbers, real numbers, and complex numbers, respectively. In addition, the notation $\mathbb{N}$ will be used to designate the natural numbers, that is, the set of positive integers.

We would also like to mention that this work had its genesis in an undergraduate research project. Thus, at the end of section 4, we have provided three open questions based upon this work that may be suitable for other undergraduate research projects.

2. DIAGONALIZABILITY OVER THE COMPLEX NUMBERS. In order to determine the probability of diagonalizability over $\mathbb{C}$, we first begin with the probability that a square matrix with integer entries has a repeated eigenvalue, as stated in Theorem 2.1. This theorem, in turn, gives rise to the most useful result in this article, which is expressed in Corollary 2.2.

Theorem 2.1. *For positive integers n and k let $R_n(k)$ be the number of $n \times n$ matrices with integer entries in the interval $[-k, k]$ that have repeated eigenvalues, and let $T_n(k)$ be the total number of $n \times n$ matrices with integer entries in the interval $[-k, k]$. Then*

$$\lim_{k\to\infty} \frac{R_n(k)}{T_n(k)} = 0.$$

Proof. Since the result is trivially true in the case $n = 1$, we may assume, without loss of generality, that $n \geq 2$. Let A represent an $n \times n$ matrix whose entries are the independent variables $x_1, x_2, \ldots, x_{n^2}$ each of which may take on an integer value uniformly from the interval $[-k, k]$, and let $f(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_0$ be the characteristic polynomial of A. Following [**4**, Definition A.1, p. 487], we define the *discriminant* $D(f)$ of f by

$$D(f) = \prod_{1\leq i<j\leq n} (r_i - r_j)^2,$$

where $r_1, r_2, \ldots, r_n$ are the roots (counted with multiplicity) of f.

Now, for fixed values of the variable entries of A the discriminant $D(f)$ vanishes precisely when the corresponding matrix A has a repeated eigenvalue. Furthermore, $D(f)$ is a symmetric polynomial in $r_1, r_2, \ldots, r_n$. The fundamental theorem for symmetric polynomials asserts that there exists a polynomial $P(y_1, y_2, \ldots, y_n)$ such that $D(f) = P(b_1, b_2, \ldots, b_n)$, where $b_k := (-1)^k c_{n-k}$ for $k = 1, 2, \ldots, n$. However, since each of the coefficients c_i of f is expressible as a polynomial in the n^2 variables that constitute the entries of A, the discriminant $D(f)$ can be expressed as a polynomial in these same variables. Let $g_0(x_1, x_2, \ldots, x_{n^2})$ be this (nonzero) polynomial. Our goal is then to determine an upper bound on the number of n^2-tuples $(a_1, a_2, \ldots, a_{n^2})$ such that a_i is an integer in the interval $[-k, k]$ for each i and $g_0(a_1, a_2, \ldots, a_{n^2}) = 0$.

Say that the (total) degree of g_0 is m. Note that g_0 can be viewed as a polynomial in x_1 of degree m_1, where $0 \leq m_1 \leq m$. Let $g_1(x_2, x_3, \ldots, x_{n^2})$ be the leading coefficient of this polynomial (possibly $g_1 = g_0$). Now, g_1 can itself be viewed as a (nonzero) polynomial in x_2 of degree m_2, where $0 \leq m_2 \leq m$. Let $g_2(x_3, x_4, \ldots, x_{n^2})$ be the leading coefficient of this polynomial (again, possibly $g_2 = g_1$). In general, let g_i ($i = 1, 2, \ldots, n^2$) be the leading coefficient of g_{i-1} when g_{i-1} is viewed as a polynomial in x_i (note that g_{n^2} is a nonzero constant). Write m_i for the degree of g_{i-1} viewed as a polynomial in x_i, and observe that $0 \leq m_i \leq m$ for each i.

Now, let $\mathcal{S}$ be the set of all n^2-tuples $(a_1, a_2, \ldots, a_{n^2})$ such that each a_j is an integer in the interval $[-k, k]$ and $g_0(a_1, a_2, \ldots, a_{n^2}) = 0$, and for $i = 1, 2, \ldots, n^2$ let $\mathcal{S}_i$ be the subset of $\mathcal{S}$ whose elements satisfy $g_0(a_1, a_2, \ldots, a_{n^2}) = g_1(a_2, a_3, \ldots, a_{n^2}) =$

$\cdots = g_{i-1}(a_i, a_{i+1}, \ldots, a_{n^2}) = 0$ and $g_i(a_{i+1}, a_{i+2}, \ldots, a_{n^2}) \neq 0$. Then $\mathcal{S} = \cup_{i=1}^{n^2} \mathcal{S}_i$, where the union is, in fact, a disjoint union. Thus, $|\mathcal{S}| = \sum_{i=1}^{n^2} |\mathcal{S}_i|$. However, by the fundamental theorem of algebra, if $g_i(x_{i+1}, x_{i+2}, \ldots, x_{n^2}) \neq 0$ for some values of $x_{i+1}, x_{i+2}, \ldots, x_{n^2}$, then for arbitrary values of $x_1, x_2, \ldots, x_{i-1}$ and these same values of $x_{i+1}, x_{i+2}, \ldots, x_{n^2}$ there are at most m distinct values for x_i such that $g_{i-1}(x_i, x_{i+1}, \ldots, x_{n^2}) = 0$. We infer that $|\mathcal{S}_i| \leq m(2k+1)^{n^2-1}$ for each i, so $|\mathcal{S}| \leq n^2 m(2k+1)^{n^2-1}$. Therefore, it follows that $0 \leq R_n(k) \leq n^2 m(2k+1)^{n^2-1}$ and $T_n(k) = (2k+1)^{n^2}$, whence the assertion of the theorem easily follows from the "squeeze property" of limits. ■

Corollary 2.2. *For positive integers n and k and for a field K containing $\mathbb{Z}$ let $E_n^K(k)$ be the number of $n \times n$ matrices with integer entries in the interval $[-k, k]$ whose eigenvalues all lie in K, let $D_n^K(k)$ be the number of $n \times n$ matrices with integer entries in the interval $[-k, k]$ that are diagonalizable over K, and let $T_n(k)$ be the total number of $n \times n$ matrices with integer entries in the interval $[-k, k]$. Then*

$$\lim_{k\to\infty} \frac{D_n^K(k)}{T_n(k)} = \lim_{k\to\infty} \frac{E_n^K(k)}{T_n(k)},$$

if either limit exists.

Proof. Let $S_n^K(k)$ be the number of $n \times n$ matrices with integer entries in the interval $[-k, k]$ that have n distinct eigenvalues in K. As in the statement of Theorem 2.1, $R_n(k)$ signifies the number of $n \times n$ matrices with integer entries in the interval $[-k, k]$ that have repeated eigenvalues. Then $S_n^K(k) \leq D_n^K(k) \leq E_n^K(k) \leq S_n^K(k) + R_n(k)$. Thus, we see that

$$\frac{S_n^K(k)}{T_n(k)} \leq \frac{D_n^K(k)}{T_n(k)} \leq \frac{E_n^K(k)}{T_n(k)} \leq \frac{S_n^K(k)}{T_n(k)} + \frac{R_n(k)}{T_n(k)}. \tag{1}$$

Now, if either $\lim_{k\to\infty} D_n^K(k)/T_n(k)$ or $\lim_{k\to\infty} E_n^K(k)/T_n(k)$ exists, then by applying Theorem 2.1 to (1), we can conclude that $\lim_{k\to\infty} S_n^K(k)/T_n(k)$ exists, hence that the other limit exists as well. Moreover, under this existence hypothesis, the desired equality follows. ■

It is interesting to note that if A is a square matrix with integer entries and K is a field containing $\mathbb{Z}$, then A has a Jordan canonical form over K precisely when each eigenvalue of A lies in K. Thus, by Corollary 2.2, the probability (if it exists) that a square matrix of fixed size with integer entries is diagonalizable over K is the same as the probability (if it exists) that a square matrix of the same size with integer entries has a Jordan canonical form over K.

In [**5**], Zhang uses Lebesgue measure to establish that the probability that a square matrix of arbitrary fixed size with real entries is diagonalizable over $\mathbb{C}$ is 1. However, such a result implies nothing concerning the case for matrices with integer entries. Nevertheless, Corollary 2.3 shows that, in fact, the probability is also 1 if one considers square matrices of arbitrary fixed size with integer entries.

Corollary 2.3. *For positive integers n and k let $D_n^{\mathbb{C}}(k)$ be the number of $n \times n$ matrices with integer entries in the interval $[-k, k]$ that are diagonalizable over $\mathbb{C}$, and let $T_n(k)$ be the total number of $n \times n$ matrices with integer entries in the interval $[-k, k]$.*

Then

$$\lim_{k\to\infty} \frac{D_n^{\mathbb{C}}(k)}{T_n(k)} = 1.$$

Proof. For any $n \times n$ matrix with integer entries in the interval $[-k, k]$, the fundamental theorem of algebra guarantees that all of the roots of the corresponding characteristic polynomial lie in $\mathbb{C}$. The result then follows from an appeal to Corollary 2.2. ∎

3. DIAGONALIZABILITY OVER THE REALS AND THE RATIONALS. Unfortunately, our information is far less complete for the probabilities of diagonalizability over $\mathbb{R}$ and $\mathbb{Q}$, respectively. In each of these cases, the complexity of the problem grows rapidly with the increasing size of the matrices considered. Nevertheless, Theorems 3.1 and 3.3 offer definitive results for the case of 2×2 matrices.

Theorem 3.1. *For each positive integer k let $D_2^{\mathbb{R}}(k)$ be the number of 2×2 matrices with integer entries in the interval $[-k, k]$ that are diagonalizable over $\mathbb{R}$, and let $T_2(k)$ be the total number of 2×2 matrices with integer entries in the interval $[-k, k]$. Then*

$$\lim_{k\to\infty} \frac{D_2^{\mathbb{R}}(k)}{T_2(k)} = \frac{49}{72}\ (\approx 68.056\%)\,.$$

Proof. If $E_2^{\mathbb{R}}(k)$ is the number of 2×2 matrices with integer entries in the interval $[-k, k]$ whose eigenvalues each lie in $\mathbb{R}$, then by Corollary 2.2 it is sufficient to show that

$$\lim_{k\to\infty} \frac{E_2^{\mathbb{R}}(k)}{T_2(k)} = \frac{49}{72}.$$

Let x, y, z, and w be independent real-valued random variables. Define the function F on $\mathbb{R}^4$ by $F(x, y, z, w) = 1$ if each eigenvalue of the matrix $\begin{bmatrix} x & y \\ z & w \end{bmatrix}$ is a real number and $F(x, y, z, w) = 0$ otherwise. Notice that $F(x, y, z, w) = F(x/k, y/k, z/k, w/k)$ everywhere. Accordingly,

$$\begin{aligned} E_2^{\mathbb{R}}(k) &= \sum_{w=-k}^{k} \sum_{z=-k}^{k} \sum_{y=-k}^{k} \sum_{x=-k}^{k} F(x, y, z, w) \\ &= k^4 \sum_{w=-k}^{k} \sum_{z=-k}^{k} \sum_{y=-k}^{k} \sum_{x=-k}^{k} F\left(\frac{x}{k}, \frac{y}{k}, \frac{z}{k}, \frac{w}{k}\right) \Delta\left(\frac{x}{k}\right) \Delta\left(\frac{y}{k}\right) \Delta\left(\frac{z}{k}\right) \Delta\left(\frac{w}{k}\right), \end{aligned}$$

where, for example, $\Delta(x/k)$ represents the change in x/k, that is, $1/k$. Therefore, since $T_2(k) = (2k+1)^4$, it follows that

$$\lim_{k\to\infty} \frac{E_2^{\mathbb{R}}(k)}{T_2(k)} =$$

$$\lim_{k\to\infty} \frac{k^4}{(2k+1)^4} \sum_{w=-k}^{k} \sum_{z=-k}^{k} \sum_{y=-k}^{k} \sum_{x=-k}^{k} F\left(\frac{x}{k}, \frac{y}{k}, \frac{z}{k}, \frac{w}{k}\right) \Delta\left(\frac{x}{k}\right) \Delta\left(\frac{y}{k}\right) \Delta\left(\frac{z}{k}\right) \Delta\left(\frac{w}{k}\right).$$

Now, it turns out that each eigenvalue of the matrix $\begin{bmatrix} x & y \\ z & w \end{bmatrix}$ is a real number precisely when $(x-w)^2+4yz \geq 0$, a fact that can be verified by considering the discriminant of the corresponding characteristic polynomial of the matrix. Hence F is, in fact, Riemann integrable. Moreover, the limit on the right-hand side of the foregoing equality is equal to

$$\frac{1}{16}\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} F(x,y,z,w)\,dx\,dy\,dz\,dw. \tag{2}$$

In order to evaluate (2), consider the inequality $(x-w)^2+4yz \geq 0$ and partition the quadruple integral in (2) according as either (i) $y, z \geq 0$, (ii) $y, z \leq 0$, (iii) $y \geq 0$ and $z \leq 0$, or (iv) $y \leq 0$ and $z \geq 0$. In the first two cases, note that any values of x and w will satisfy $(x-w)^2+4yz \geq 0$. Hence, each of the parts of (2) corresponding to these two cases evaluates to $1/4$. On the other hand, in cases (iii) and (iv), observe that $(x-w)^2+4yz \geq 0$ is equivalent to the assertion that either $x \geq w+\sqrt{-4yz}$ or $x \leq w-\sqrt{-4yz}$. So, for example, if $y \geq 0$, $z \leq 0$, and $x \geq w+\sqrt{-4yz}$, then the part of (2) corresponding to these conditions evaluates to

$$\frac{1}{16}\int_{-1}^{0}\int_{0}^{1}\int_{-1}^{1-\sqrt{-4yz}}\int_{w+\sqrt{-4yz}}^{1} 1\,dx\,dw\,dy\,dz = \frac{13}{288}.$$

Similarly, the parts of (2) corresponding to the three remaining scenarios each evaluate also to $13/288$. Therefore, we have

$$\lim_{k\to\infty}\frac{E_2^{\mathbb{R}}(k)}{T_2(k)} = 2\left(\frac{1}{4}\right)+4\left(\frac{13}{288}\right) = \frac{49}{72},$$

as desired. ■

Remark 3.2. (a) In an analogous fashion to the "natural density approach" for the 2×2 matrices with integer entries in Theorem 3.1, we can define the probability that a 2×2 matrix with real entries is diagonalizable over $\mathbb{R}$. Specifically, let the function F be as in the proof of Theorem 3.1. Then the probability that a 2×2 matrix with real entries is diagonalizable over $\mathbb{R}$ can be given by the expression

$$\lim_{k\to\infty}\frac{1}{(2k)^4}\int_{-k}^{k}\int_{-k}^{k}\int_{-k}^{k}\int_{-k}^{k} F(x,y,z,w)\,dx\,dy\,dz\,dw, \tag{3}$$

provided such a limit exists. However, since $F(x,y,z,w) = F(x/k, y/k, z/k, w/k)$ everywhere, it is not difficult to argue that

$$\begin{aligned}\frac{1}{(2k)^4}&\int_{-k}^{k}\int_{-k}^{k}\int_{-k}^{k}\int_{-k}^{k} F(x,y,z,w)\,dx\,dy\,dz\,dw \\ &= \frac{1}{16}\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1}\int_{-1}^{1} F(x,y,z,w)\,dx\,dy\,dz\,dw.\end{aligned}$$

Since the right-hand side of this equality is $49/72$ (see the proof of Theorem 3.1), it follows that the probability, as defined by (3), that a 2×2 matrix with real entries is diagonalizable over $\mathbb{R}$ is also $49/72$.

(b) In Remark 3.2 (a), the probability of 49/72 was achieved by first considering how often the roots of a particular monic quadratic polynomial, namely, the characteristic polynomial of a random matrix, were real numbers. It is interesting to contrast this probability with the probability that the roots of a random monic quadratic polynomial with real coefficients, say $x^2 + bx + c$, are real numbers. Observe that such a polynomial has real roots precisely when $c \le b^2/4$. Thus, if $A(k)$ is the area beneath the graph of $y = \frac{x^2}{4}$ and within the square $[-k, k] \times [-k, k]$, where k in $\mathbb{N}$ is arbitrary, the probability that a random monic quadratic polynomial with real coefficients has real roots is

$$\lim_{k\to\infty} \frac{A(k)}{(2k)^2} = \lim_{k\to\infty} \left(1 - \frac{2}{3\sqrt{k}}\right) = 1.$$

Theorem 3.3. *For each positive integer k let $D_2^{\mathbb{Q}}(k)$ be the number of 2×2 matrices with integer entries in the interval $[-k, k]$ that are diagonalizable over $\mathbb{Q}$, and let $T_2(k)$ be the total number of 2×2 matrices with integer entries in the interval $[-k, k]$. Then*

$$\lim_{k\to\infty} \frac{D_2^{\mathbb{Q}}(k)}{T_2(k)} = 0.$$

Proof. If $E_2^{\mathbb{Q}}(k)$ is the number of 2×2 matrices with integer entries in the interval $[-k, k]$ whose eigenvalues each lie in $\mathbb{Q}$, then by Corollary 2.2 it is sufficient to show that $\lim_{k\to\infty} E_2^{\mathbb{Q}}(k)/T_2(k) = 0$. Let r, s, t, and u in $\mathbb{Z}$ be such that $-k \le r, s, t, u \le k$, and put

$$A = \begin{bmatrix} r & s \\ t & u \end{bmatrix}.$$

By considering the discriminant of the characteristic polynomial of A, one can see that it is necessary and sufficient that $(r - u)^2 + 4st = x^2$ for some x in $\mathbb{Z}$ in order for each eigenvalue of A to be a rational number. Thus, for fixed values of s and t we consider integer solutions (x, y) to the Diophantine equation

$$x^2 - y^2 = 4st, \tag{4}$$

where $y = r - u$.

We first dispense with the case where either $s = 0$ or $t = 0$. In this case, we may take $x = r - u$ for any choices of r and u. Therefore, each of $2(2k + 1)^3 - (2k + 1)^2$ matrices A where either $s = 0$ or $t = 0$ has only rational eigenvalues.

Consider now the case where both s and t are positive. Observe that (4) can be reexpressed as $4st = (x + y)(x - y)$. Thus, if (x, y) is an integer solution to (4), then $x + y$ must divide $4st$. Conversely, if v is a divisor of $4st$, then there exist unique values of x and y such that $x + y = v$ and $x - y = 4st/v$ (although (x, y) may not represent an integer solution to (4)). As in [**2**] or [**4**], let d be the number theoretic divisor function (i.e., for each n in $\mathbb{N}$, $d(n)$ is the number of positive divisors of n). Then there are at most $d(4st)$ distinct integer values for y such that (x, y) is a solution to (4) for some x in $\mathbb{Z}$. Hence, for each choice of u there are at most $d(4st)$ integers r in the interval $[-k, k]$ such that $(r - u)^2 + 4st = x^2$ for some x in $\mathbb{Z}$. Therefore, there are at most $\sum_{s=1}^{k} \sum_{t=1}^{k} (2k + 1)\, d(4st)$ matrices A such that $s > 0$, $t > 0$, and each eigenvalue of A is a rational number. However, since $d(4st) \le d(4)d(s)d(t)$ for any values of s and t in the interval $[1, k]$, we arrive at the inequality

$$\sum_{s=1}^{k}\sum_{t=1}^{k}(2k+1)\,d(4st) \le 3(2k+1)\left(\sum_{s=1}^{k} d(s)\right)^2.$$

The three remaining cases—namely, where $s > 0$ and $t < 0$, where $s < 0$ and $t > 0$, and where $s < 0$ and $t < 0$—are treated in similar fashion. The upshot is that

$$E_2^{\mathbb{Q}}(k) \le 2(2k+1)^3 - (2k+1)^2 + 4\left(3(2k+1)\left(\sum_{s=1}^{k} d(s)\right)^2\right).$$

However, we note that

$$\sum_{s=1}^{k} d(s) = \sum_{j=1}^{k}\left\lfloor \frac{k}{j}\right\rfloor,$$

where $\lfloor\cdot\rfloor$ is the floor function (i.e., $\lfloor a\rfloor$ signifies the greatest integer that is less than or equal to a), since j divides exactly $\lfloor k/j\rfloor$ integers in the interval $[1, k]$. As a result,

$$\sum_{s=1}^{k} d(s) \le k\left(1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{k}\right) \le k(1+\ln k).$$

Thus, for each $\epsilon > 0$ there exists a positive constant C_ϵ such that $\sum_{s=1}^{k} d(s) < C_\epsilon k^{1+\epsilon}$ whenever $k \ge 1$. Choose ϵ such that $0 < \epsilon < 1/2$. Then there exists a positive constant C such that $\sum_{s=1}^{k} d(s) < Ck^{1+\epsilon}$ for each positive integer k. Therefore, since $T_2(k) = (2k+1)^4$, it follows that

$$0 \le \frac{E_2^{\mathbb{Q}}(k)}{T_2(k)} < \frac{2(2k+1)^3 - (2k+1)^2 + 12C^2(2k+1)k^{2+2\epsilon}}{(2k+1)^4},$$

whence an application of the "squeeze property" of limits produces the desired result. ■

Remark 3.4. By the "rational roots theorem," any rational eigenvalue of a square matrix with integer entries must be an integer. Hence, in view of Theorem 3.3 and Corollary 2.2, the probability that a 2×2 matrix with integer entries has only integral eigenvalues is 0. Moreover, Kowalsky [**3**] has demonstrated that for each $\epsilon > 0$ there are $\mathcal{O}(k^{3+\epsilon})$ 2×2 matrices with integer entries in the interval $[-k, k]$ ($k \in \mathbb{N}$) that have integer eigenvalues. In fact, Kowalsky's result coupled with the observation concerning the rational eigenvalues of a square matrix can be used to recover Theorem 3.3.

4. EVIDENCE FOR HIGHER DIMENSIONAL MATRICES. In the final section of this article, we provide some numerical evidence for the probability that an $n \times n$ matrix with integer entries is diagonalizable over $\mathbb{R}$ (respectively, $\mathbb{Q}$) when $n > 2$. Thanks to Corollary 2.2, it is enough to develop data that indicate approximately how often a square matrix with integer entries has eigenvalues each of which lies in $\mathbb{R}$

Table 1. Probability of having all eigenvalues in $\mathbb{R}$ (respectively, $\mathbb{Q}$), based upon 100,000 randomly generated $n \times n$ matrices with integer entries taken uniformly from the interval $[-1000, 1000]$.

n	$\mathbb{R}$	$\mathbb{Q}$
3	0.32061	0.00000
4	0.10526	0.00000
5	0.02553	0.00000
6	0.00435	0.00000
7	0.00050	0.00000
8	0.00003	0.00000
9	0.00000	0.00000
10	0.00000	0.00000

(respectively, $\mathbb{Q}$). Such data are given in Table 1 and were generated using Release 8.2 of SAS with 100,000 random matrices with integer entries taken uniformly from the interval $[-1000, 1000]$.

It is interesting to note how close the probabilities in the $\mathbb{R}$-column of Table 1 are to the value $2^{-n(n-1)/4}$ for each n, which Edelman gives in [**1**] as the probability that an $n \times n$ matrix whose entries are independent random variables, each with a standard normal distribution, has all real eigenvalues.

In conclusion, we pose three questions. First, if $n > 2$, can one find the exact probability, as in Theorem 3.1, that an $n \times n$ matrix with integer entries is diagonalizable over $\mathbb{R}$? Second, can one prove, à la Theorem 3.3, that the probability that an $n \times n$ matrix with integer entries is diagonalizable over $\mathbb{Q}$ is 0 when $n > 2$? Third, in view of Theorem 3.1 and Table 1, is there a correspondence between 2×2 matrices with integer entries that have complex, nonreal eigenvalues and 3×3 matrices with integer entries that have all real eigenvalues that justifies the nearly complementary probabilities of diagonalizability (that is, approximately 0.68056 and 0.32061, respectively)?

ACKNOWLEDGMENT. We wish to express our gratitude to Kristi Hetzel and A. Dale Magoun for their individual contributions to this project.

REFERENCES

1. A. Edelman, The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law, *J. Multivariate Anal.* **60** (1997) 203–232.
2. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5th ed., Clarendon Press, Oxford, 1979.
3. H.-J. Kowalsky, Ganzzahlige Matrizen mit ganzzahligen Eigenwerten, *Abh. Braunschweig. Wiss. Ges.* **34** (1982) 15–32.
4. I. Niven, H. S. Zuckerman, and H. L. Montgomery, *An Introduction to the Theory of Numbers*, 5th ed., John Wiley, New York, 1991.
5. Z. N. Zhang, The Jordan canonical form of a real random matrix, *Numer. Math. J. Chinese Univ.* **23** (2001) 363–367.

ANDREW J. HETZEL received his B.S. degree in mathematics from the University of Dayton in 1998 and his M.S. and Ph.D. degrees in mathematics from the University of Tennessee in 2000 and 2003, respectively.

He is an Orange Dot in the Project NExT program whose research interests include commutative ring theory and number theory. He has been married to his wife Kristi for four wonderful years and, in his free time, likes to watch the Food Network.
Department of Mathematics, Box 5054, Tennessee Technological University, Cookeville, TN 38505
ahetzel@tntech.edu

JAY S. LIEW received his B.S. degree in computer science from the University of Louisiana at Monroe in 2004 with a minor in mathematics. He is currently a software engineer for Websense, Inc., where he reverse engineers proprietary protocols and conducts quality assurance testing. His research interests include nonlinear dynamical systems and the theory of computability and complexity. He believes in attempting the impossible because it is only absurd until someone achieves it, and he wants to change the world through the use of technology. In his spare time, he sleeps when necessary.
12404 Nonie Terrace, San Diego, CA 92129
liew@jaysern.org

KENT E. MORRISON is chair of the Mathematics Department at California Polytechnic State University in San Luis Obispo, where he has taught for over twenty-five years. He has also taught at Utah State University, Haverford College, and the University of California at Santa Cruz, where he received his Ph.D. and B.A. degrees. In recent years his research interests have centered on enumerative and algebraic combinatorics.
Department of Mathematics, California Polytechnic State University, San Luis Obispo, CA 93407
kmorriso@calpoly.edu

Back to the Drawing Board Lime

You've completed a beautiful proof,
But your colleague has spotted a goof.
So that one *minus* one
Isn't two, it is none,
And the beautiful proof just went *poof*!

—Submitted by Bob Scher, Mill Valley, CA
(By the author's definition, a "lime" is a "clean limerick.")