1. Introduction

The process of factoring polynomials, specifically quadratic trinomials, is taught in nearly every American high school algebra curriculum. Learners devote substantial time and effort to factoring by a number of methods. We claim that tools of technology, specifically computer algebra systems (CAS) and spreadsheets, have deepened high school students' potential thought about factoring. In essence, we believe that technology can provide opportunities for insights into factoring that can build a better conceptual understanding of factoring, as opposed to merely enhancing or supplementing by-hand procedures. Building on an example problems from a high school textbook, we investigated what it means for a quadratic polynomial to be factorable and quantified how rare (or common) factorability is for polynomials over portions of the integers using these tools.

2. Initial Analysis

We began by considering the following problem, adapted from a widely used high school algebra text [1, pp. 738–740]:

There are infinitely many quadratic expressions of the form $y = ax^2 + bx + c$. But what percent of those are factorable over the integers when a, b, and c are each <u>*nonzero*</u> *integers from -10 to 10? Generate 25 random quadratic expressions of this type and determine what percentage are factorable.*

For notational convenience, let:

$$C_M = \{n \in \mathbb{Z} : 1 \leq |n| \leq M\}$$

so that C_{10} is the coefficient domain of interest. Let:

$$\mathcal{P}_M = \{ax^2 + bx + c : (a, b, c) \in C_M^3\}.$$

The question as posed, then, is to find the proportion p of polynomials from $\mathcal{P}_{10}$ that factor over $\mathbb{Z}$.

To approach this problem, we used a CAS (in this case we used *GeoGebra*, but any CAS should do) to generate Table 1, which shows the factorability over $\mathbb{Z}$ for a random sample of 25 polynomials from $\mathcal{P}_{10}$. To create the sample, we generated triples of pseudorandom integers between -10 and 10, discarding any triple with a zero, then used 25 of these triples as coefficients for polynomials as input to *GeoGebra*'s `Factor` function. Note that by default *GeoGebra* factors over $\mathbb{Z}$, which is the domain needed.

While Table 1 addresses the initial question, allowing the finding that $\hat{p}_0 = 16\%$ of our sample is factorable, it seems to raise at least as many questions as it answers. Is our sample representative so that the 16% figure is an accurate estimate of the factorable proportion of $\mathcal{P}_{10}$? What might happen if we expand the coefficient domain? What about more general factorability considerations?

An initial question we considered was: When is a quadratic polynomial of this type factorable? A quadratic is factorable if and only if it can be written as the product of two linear terms with coefficients from some factor domain. In this case, because we are factoring over $\mathbb{Z}$, we want integer coefficients on our factors. Knowing this

Polynomial	Output of `Factor`	Factorable over $\mathbb{Z}$?	Discriminant
$-3x^2-2x+4$	$-3x^2-2x+4$	No	52
$-9x^2+2x-4$	$-9x^2+2x-4$	No	-140
$6x^2-2x-6$	$2(3x^2-x-3)$	No	148
$4x^2+6x-9$	$4x^2+6x-9$	No	180
$-9x^2-x-3$	$-9x^2-x-3$	No	-107
$3x^2-7x+6$	$3x^2-7x+6$	No	-23
$7x^2+10x+7$	$7x^2+10x+7$	No	-96
$-3x^2+6x+10$	$-3x^2+6x+10$	No	156
$-9x^2-x-1$	$-9x^2-x-1$	No	-35
$-9x^2-7x+2$	$-(x+1)(9x-2)$	Yes	$121=11^2$
$9x^2+7x-7$	$9x^2+7x-7$	No	301
$-3x^2+3x+6$	$-3(x-2)(x+1)$	Yes	$81=9^2$
$9x^2+7x+6$	$9x^2+7x+6$	No	-167
$8x^2+9x-6$	$8x^2+9x-6$	No	273
$-x^2-x+6$	$-(x-2)(x+3)$	Yes	$25=5^2$
$-6x^2+3x-8$	$-6x^2+3x-8$	No	-183
$7x^2-x+10$	$7x^2-x+10$	No	-279
$10x^2+2x+7$	$10x^2+2x+7$	No	-276
$9x^2+6x-4$	$9x^2+6x-4$	No	180
$5x^2-5x+5$	$5(x^2-x+1)$	No	-75
$-8x^2-3x-1$	$-8x^2-3x-1$	No	-23
$-8x^2-2x-6$	$-2(4x^2+x+3)$	No	-188
$-3x^2-4x+7$	$-(x-1)(3x+7)$	Yes	$100=10^2$
$4x^2-9x-8$	$4x^2-9x-8$	No	209
$-6x^2-8x+3$	$-6x^2-8x+3$	No	136

Table 1.: Factorability over $\mathbb{Z}$ for 25 random quadratics from $\mathcal{P}_{10}$.

and the conditions of the problem, we see that there are only 8000 unique quadratic expressions that comprise $\mathcal{P}_{10}$.

Consider a typical quadratic polynomial: ax^2+bx+c, with $a \neq 0$.
Its discriminant is: $d = b^2-4ac$, so that its zeros are: $\dfrac{-b \pm \sqrt{d}}{2a}$.
It thus has factored forms:

$$ax^2+bx+c = a\left(x+\frac{b+\sqrt{d}}{2a}\right)\left(x+\frac{b-\sqrt{d}}{2a}\right) = \frac{\left(2ax+b+\sqrt{d}\right)\left(2ax+b-\sqrt{d}\right)}{4a},$$

which will be a factorization over the integers if and only if $(a, b, \sqrt{d}) \in \mathbb{Z}^3$. As all the polynomials in $\mathcal{P}_{10}$ have integer coefficients, the restriction here is reduced to $d = b^2-4ac$ being the square of some integer. This is borne out by the "Discriminant" column in Table 1.

With this in mind, we turned to a spreadsheet to generate all 8000 polynomials in $\mathcal{P}_{10}$. A spreadsheet is well suited this task, because it allows for organization of information and investigation of each individual quadratic. This could also have been approached through programming – one author has done this using both *Mathematica* and a TI-84 graphing calculator. A brute force approach shows that the factorability of all 8000 polynomials in $\mathcal{P}_{10}$ is $p = 892 \div 8000 = 11.15\%$

This result is quite a bit lower than the earlier obtained result in Table 1. This caused us to wonder what someone should reasonably expect to get as a percentage from a random sample of 25 polynomials with the stated criteria.

To generate $\hat{p}_0$, a random sample of $n_0 = 25$ polynomials was taken from a finite population of $N = \#\mathcal{P}_{10} = 8000$. $\hat{p}_0$ is the best available estimate of the population proportion p. Thus, $n_0 p \approx n_0 \hat{p}_0 = 4$. With the best estimate for $n_0 p$ less than 10 and $n_0 = 25 < 30$, these data are not suitable for making inferences, as the sampling distribution of the proportion isn't well characterized until $(np > 10) \wedge [n(1-p) > 0] \wedge (n > 30)$ [2, p. 269]. Therefore, we shouldn't be very confident that $\hat{p}_1$ is a good estimate of p.

To fix this issue, we repeated the experiment 8 more times, adding 200 more random polynomials from $\mathcal{P}_{10}$, 25 of which were factorable. This produced a better estimate of the proportion:

$$\hat{p} = \frac{25+4}{200+25} \doteq 0.129.$$

The data now satisfy all criteria for approximate normality of the sampling distribution of the proportion:

$$(np \approx n\hat{p} = 29 \geq 10) \wedge [n(1-p) \approx n(1-\hat{p}) = 196 \geq 10] \wedge (n = 225 > 30).$$

Therefore, $\hat{p}$ has a standard error [2, pp. 271–272]:

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}\frac{N-n}{N-1}} \approx \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}\frac{N-n}{N-1}} \doteq 0.022.$$

Using this, we can state with 95% confidence [2, pp. 362–364] that:

$$\hat{p} - \sigma_{\hat{p}} z_{0.025} \approx 8.6\% < p < 17.2\% \approx \hat{p} + \sigma_{\hat{p}} z_{0.025}.$$

Finally, with this result in mind, we looked at the distribution of discriminants, reasoning that if discriminants are concentrated near the squares, which are themselves concentrated in the lower positive integers, a correspondingly higher proportion of $\mathcal{P}_{10}$ would factor over $\mathbb{Z}$. Figure 1 shows that the distribution is not uniform, is peaked near zero (the curve being almost normal like), and has increasing gaps between non-zero-frequency entries as the discriminant gets further from zero.

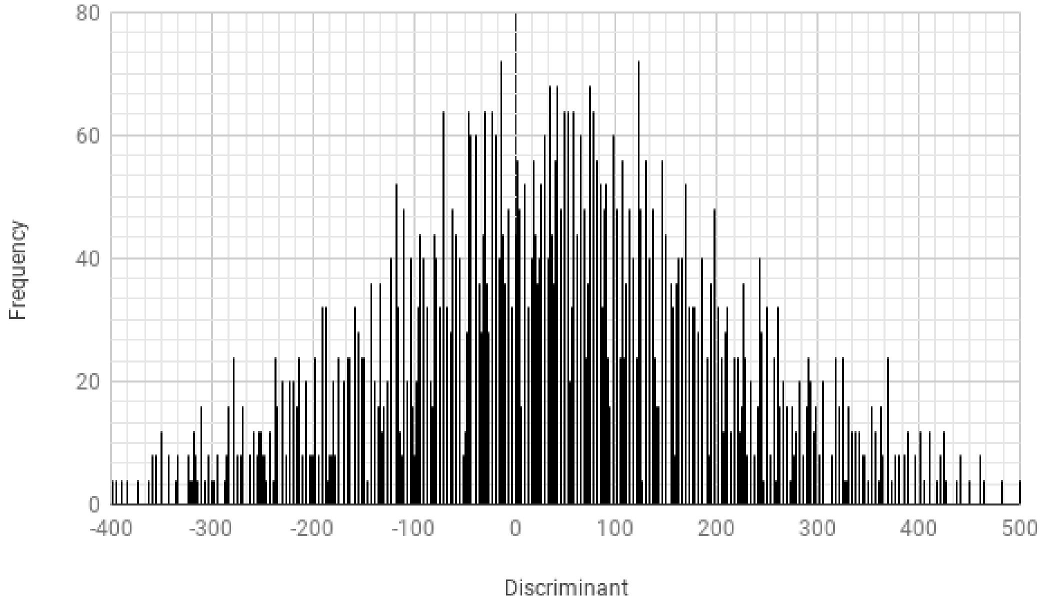

Figure 1.: Distribution of discriminants in $\mathcal{P}_{10}$.

3. Further Investigation

With the original problem exhausted, we turned to considering what happens if we expand the coefficient domain. A considerable advantage of technology in this regard is that we are free to change almost any parameter of the problem and have it assist us in collecting data by almost the exact same procedure. We were thus able to continue to use the spreadsheet used in the initial investigation with only minor adaptations.

M	$\#\mathcal{P}_M$	# factorable	% factorable
10	8000	892	11.15
11	10648	1084	10.18
12	13824	1404	10.16
13	17576	1640	9.33
14	21952	1972	8.98
15	27000	2344	8.68
20	64000	4628	7.23
25	125000	7800	6.24
30	216000	12076	5.59

Table 2.: Proportion of factorable polynomials in $\mathcal{P}_M$ for varying M.

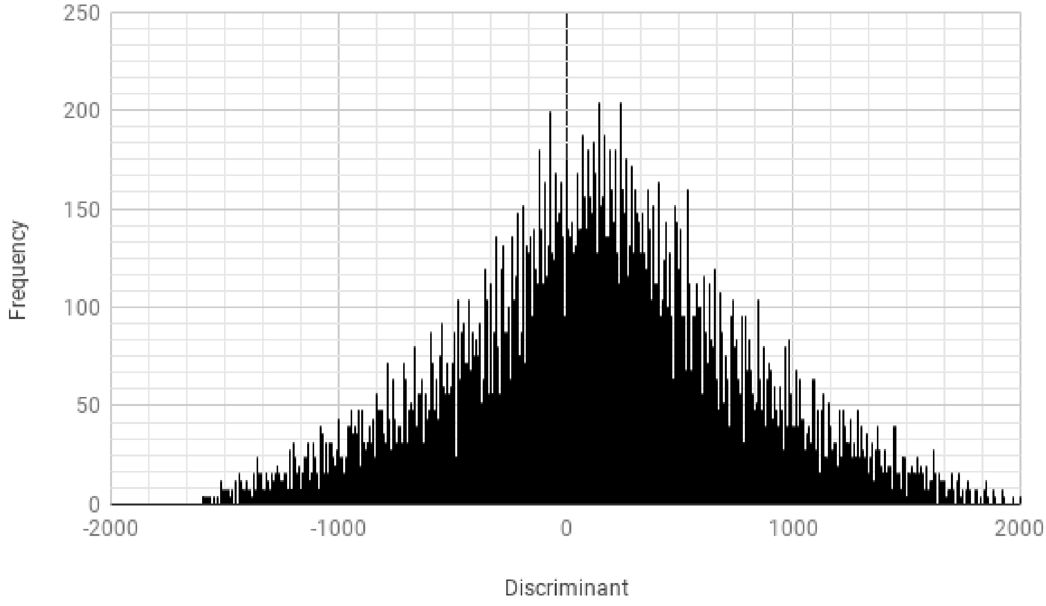

Figure 2.: Distribution of discriminants in $\mathcal{P}_{20}$.

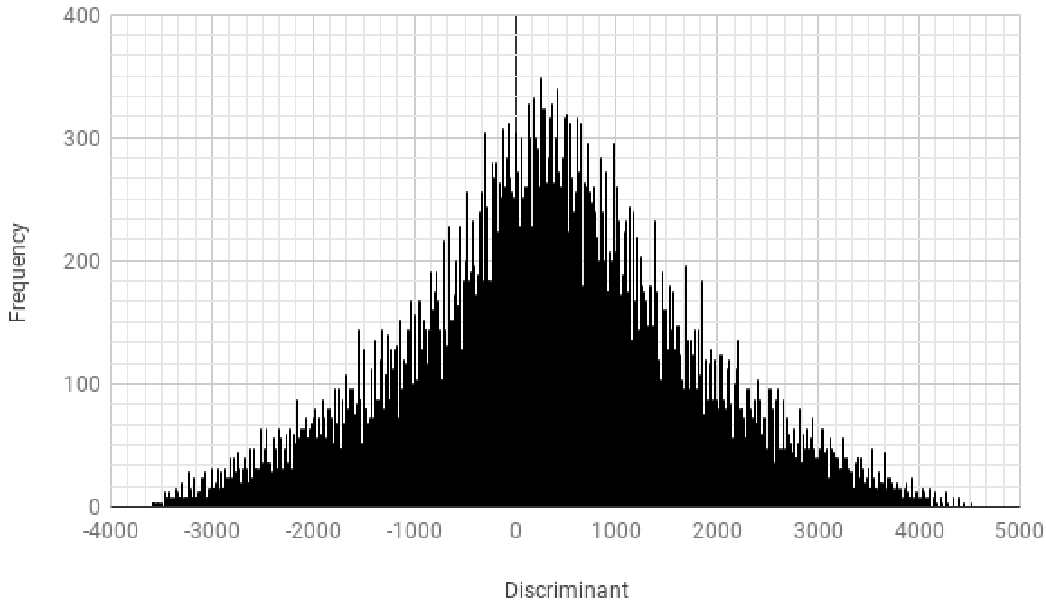

Figure 3.: Distribution of discriminants in $\mathcal{P}_{30}$.

Table 2 shows the progression of factorable polynomials the bound on the coefficients increases. It may seem surprising that the proportion of factorable polynomials drops as M increases. However, consider that, as M increases, the number of polynomials, $8M^3$, will tend to outgrow the number that are factorable, $\sim M^{2.38}$ in these data. To consider this second growth rate further, note that, as M increases, so does the average magnitude of the discriminant of a polynomial from $\mathcal{P}_M$. That, in turn, decreases the probability that the discriminant is a square – the range of positive discriminants is $\left[1 - 4M^2, 5M^2\right]$, which contains $\left\lfloor M\sqrt{5}\right\rfloor$ squares, meaning that the proportion of squares falls off like $1/M$. Consequently, we should expect that proportion of factorable polynomials will continue to drop, so long as the distribution of discriminants does not change too much in shape.

To examine this further, we generated figures 2 and 3, which show the distributions of discriminants for $M = 20$ and 30. We can see that the shapes of these distributions are similar to that in figure 1 – all are highly non-uniform, peaked slightly to the right of 0, and normal-like. The stability of this shape tends to lead to the conclusion that the factorable proportion will continue to decrease with M[1].

4. Conclusion

Even the most "done and dusted"-seeming mathematics can hide truly curiosity-inspiring subtleties. The use of technology allows investigation and pursuance of that curiosity efficiently and effectively. While the backdrop of this investigation was quadratic factoring, technology helped us to consider the interesting behavior of those factors under changes in domain, which may be a very unfamiliar face of this well-known body of procedure. The use of technology, along with the an appropriate amount of mathematical sophistication, enabled an investigation that cut across mathematical technologies and ideas. It is our hope that technology can continue to be a conduit for these types of investigations.

[1] However, the peakedness of these distributions may be increasing with M, which, if significant, would tend to offset this trend.

References

[1] John W. McConnell et al. *Algebra*. 2nd ed. University of Chicago School Mathematics Project. Prentice Hall, 2002.

[2] Irwin Miller and Marylees Miller. *Mathematical Statistics*. 7th ed. Pearson, 2003.