1. Introduction

The process and procedure of factoring polynomials (specifically quadratic trinomials) is one that appears in nearly every American high school algebra textbook. The amount of time and number of methods that exist for factoring is numerous. Technology (specifically CAS) has advanced the potential for how deeply one can think about factoring at the high school level. In essence, we believe that technology can provide opportunities for insight into factoring that will further build a better conceptual understanding of the procedure, rather than enhance by hand methods. Building off of a homework and example problem from an American high school textbook, we investigated what it means for a quadratic polynomial to be factorable and further how rare (or common) it could be for such polynomials to be factorable.

2. Initial Analysis

The problem, as modified from UCSMP Algebra Chapter 12 (pgs. 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:

$$C = \{n \in \mathbb{Z} : |n| \leq 10 \wedge n \neq 0\} = \{-10.-9,-8,-7,-6,-5,-4,-3,-2,-1,1,2,3,4,5,6,7,8,9,10\}$$

and let:

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

Using a CAS (in this case we used *GeoGebra* but any CAS should do), we generated Table 1 which shows the factorability over $\mathbb{Z}$ for a random sample of 25 polynomials from $\mathcal{P}$. We discarded any polynomial that generated with a zero coefficient. We also note that by default *GeoGebra* factors over $\mathbb{Z}$, which is the domain needed.

While Table 1 provides the data to address the initial question (16% factorable), it does not address several related issues related to factorability, expanding the coefficient domain, etc. Are our data even correct?

An initial question we considered was when a quadratic expression of this type would be 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 and the conditions of the problem, we see that there are only 8000 unique quadratic expressions that comprise $\mathcal{P}$.
Now, consider, $f(x)$: $f(x) = ax^2 + bx + c, a \neq 0$.
Its discriminant is: $d = b^2 - 4ac$, so that its roots are: $\dfrac{-b \pm \sqrt{d}}{2a}$.

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 $\mathcal{P}$.

Its factored form is thus:

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

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 $\mathcal{P}$ 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 quadratics. A spreadsheet is an ideal technology for this because it allows for organization of information and a way to investigate 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 $\mathcal{P}$ is $p = 892 \div 8000 = 11.15\%$

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

To generate $\hat{p}$, a random sample of $n_0 = 25$ polynomials was taken from a finite population of $N = \#\mathcal{P} = 8000$. $\hat{p}_0$ is the best available estimate of the population proportion p, i.e. the true percentage of polynomials from $\mathcal{P}$ that factor over $\mathbb{Z}$. 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 $(np > 10) \wedge [n(1-p) > 0] \wedge (n > 30)$ [1, p. 269]. Therefore, one can't be very confident that $\hat{p}$ is a good estimate of p.

To fix this issue, the experiment was repeated for 200 more random polynomials from $\mathcal{P}$, 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 [1, 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 [1, 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 discriminanats, reasoning that if the distribution of discriminants were concentrated near the squares, which are themselves concentrated in the lower positive integers, a correspondingly higher proportion of $\mathcal{P}$ would factor over $\mathbb{Z}$. Figure 1 shows that the distribution is not uniform, has peakness near zero (the curve is almost normal like), and increasing gaps between non-zero-frequency entries as the discriminant gets further from zero.

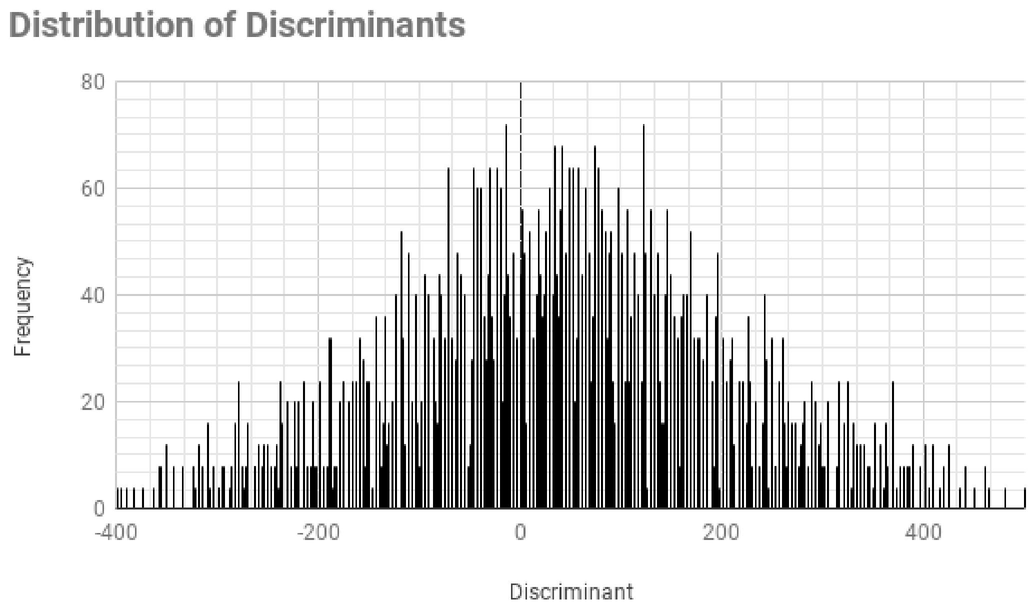

Figure 1.: Distribution of discriminants in $\mathcal{P}$.

3. Further Investigation

With our work on the original problem essentially completed, we turned to considering what happens if we expand the domain on the coefficients. The advantage of technology provides is that we are free to change almost any parameter of the problem and have it assist us in collecting data. We stayed with the original technologies used in the initial investigation.

C	# elements in $\mathcal{P}$	# factorable in $\mathcal{P}$	Percentage
$\{n \in \mathbb{Z} : \lvert n\rvert \leq 11 \wedge n \neq 0\}$	10,648	1084	10.18%
$\{n \in \mathbb{Z} : \lvert n\rvert \leq 12 \wedge n \neq 0\}$	13,824	1404	10.16%
$\{n \in \mathbb{Z} : \lvert n\rvert \leq 13 \wedge n \neq 0\}$	17,576	1640	9.33%
$\{n \in \mathbb{Z} : \lvert n\rvert \leq 14 \wedge n \neq 0\}$	21,952	1972	8.98%
$\{n \in \mathbb{Z} : \lvert n\rvert \leq 15 \wedge n \neq 0\}$	27,000	2344	8.68%
$\{n \in \mathbb{Z} : \lvert n\rvert \leq 20 \wedge n \neq 0\}$	64,000	4628	7.23%
$\{n \in \mathbb{Z} : \lvert n\rvert \leq 25 \wedge n \neq 0\}$	125,000	7800	6.24%
$\{n \in \mathbb{Z} : \lvert n\rvert \leq 30 \wedge n \neq 0\}$	216,000	12,076	5.59%

Table 2.: Number of factorable polynomials in $\mathcal{P}$ over $\mathbb{Z}$ for varying criteria C

Table 2 shows the progression of factorable polynomials as we increase the bounds on the coefficients. At first glance it may be surprising that the net percent of factorable polynomials drops as we increase n. However, seeing that the number of possible polynomials is increasing as a function of x^3, while the number of factorable polynomials gained does not, sheds some light on this.

Figure 2 shows the distribution of discriminates for $n = 30$ and continues to support our arguments regarding their distribution.

4. Conclusion

Mathematical curiosity can lurk anywhere. The use of technology allows one to investigate and pursue that curiosity efficiently and effectively. While the backdrop of this investigation was quadratic factoring, it was the behavior of how those factors result that was interesting and not necessarily the factors themselves. The use of technology, along with the right amount of mathematical sophistication provided 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.

References

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