# Empirical Finding of the Random Van der Waerden Numbers

Julian Kovalovsky

August 19, 2025

**Abstract** In Ramsey Theory, many problems ask for the minimum size of a structure such that a smaller substructure is guaranteed to exist. This is the case with Van der Waerden's theorem, which studies monochromatic arithmetic progressions inside of $\{1, 2, \dots, n\}$. Historically, the true values of the Van der Waerden numbers have been difficult to find, with only a few being known despite almost a century passing since the theorem was proven. Consequently, the next-best options are to either find rigorous bounds on the Van der Waerden numbers or find empirical approximations for the numbers. In this paper, we define and empirically estimate the Random Van der Waerden numbers to gain a better understanding of where the true Van der Waerden numbers lie.

## 1 Introduction

In 1926, Bartel Leendert van der Waerden discovered a conjecture made by Pierre Joseph Henry Baudet. A year later, Waerden proved the conjecture to be true, leading to it being named after him [3].

**Theorem 1.1** *Van der Waerden's theorem states that for any $k, c \in \mathbb{Z}^+$, there exists an $n$ such that for all $c$-colorings of $\{1, \dots, n\}$, there must be a monochromatic arithmetic progression with length at least $k$.*

**Def 1.2** Let the Van der Waerden number $W(c, k)$ be the smallest integer $n$ such that for all $c$-colorings of $\{1, 2, \dots, n\}$, there is guaranteed to be a monochromatic arithmetic progression of length $k$.

Since 1927, only 7 non-trivial Van der Waerden numbers have been found. Additionally, the best asymptotic upper and lower bounds are still relatively loose. For example, it was only in 1988 that the first primitive recursive upper bound was found by Shelah [2]. Since then, Gowers used complicated math techniques to find the tighter bound [1]

$$W(c,k) \leq 2^{2^{c^{2^{2^{k+9}}}}}.$$

Due to the current difficulty of finding a new Van der Waerden number or tightening the bounds, the next-best option is to explore similar problems to give more insight on the Van der Waerden numbers.

In this paper, we define the Random Van der Waerden numbers, empirically approximate them, and focus on using them as stand-ins for the unknown Van der Waerden numbers. Specifically, these approximate values of the Random Van der Waerden numbers can be used as stand-ins for the true Van der Waerden numbers when in scenarios which can tolerate error. In addition to finding the empirical values, we found a general lower bound for the Random Van der Waerden numbers, further enabling the use of these numbers in non-rigorous settings.

# 2 Definitions and Notation

**Notation 2.1** Let $[n]$ denote the set $\{1, \ldots, n\}$.

**Notation 2.2** An arithmetic sequence of length $k$ is said to be a $k$-AP.

**Def 2.3** Let the Random Van der Waerden number $W_R(c,k,p)$ be the smallest $n$ such that the probability of a random $c$-coloring of $[n]$ containing a monochromatic $k$-AP is at least $p$. Note that $W_R(c,k,1) = W(c,k)$.

**Def 2.4** Let $A(n,k)$ denote the amount of $k$-AP's in $[n]$.

## 3 Lower Bound Proof

We seek to find a lower bound on the Random Van der Waerden number $W_R(c,k,1)$ for $k \geq 2$. We begin by finding $A(n,k)$.

**Claim 3.1** $A(n,k) = n\left\lfloor \frac{n}{k-1} \right\rfloor - \frac{\left\lfloor \frac{n}{k-1} \right\rfloor \left(\left\lfloor \frac{n}{k-1} \right\rfloor + 1\right)(k-1)}{2}$

**Proof:** We will do a summation over the common difference of the $k$-AP's. Notably, the total amount of $k$-AP's with a valid common difference $d$ is $n - d(k-1)$. Summing over all valid common differences, we get

$$\begin{aligned} A(n,k) &= \sum_{d=1}^{\left\lfloor \frac{n}{k-1} \right\rfloor} n - d(k-1) \\ &= \sum_{d=1}^{\left\lfloor \frac{n}{k-1} \right\rfloor} n - (k-1) \sum_{d=1}^{\left\lfloor \frac{n}{k-1} \right\rfloor} d \\ &= n\left\lfloor \frac{n}{k-1} \right\rfloor - \frac{\left\lfloor \frac{n}{k-1} \right\rfloor \left(\left\lfloor \frac{n}{k-1} \right\rfloor + 1\right)(k-1)}{2} \end{aligned}$$

$\blacksquare$

Since the exact value of $A(n,k)$ is hard to use due to the floors, we bound $A(n,k)$ by using the fact that $\lfloor x \rfloor \leq x$.

**Claim 3.2** $A(n,k) \leq \frac{n^2}{2(k-1)}$

**Proof:**

$$\begin{aligned} n\left\lfloor \frac{n}{k-1} \right\rfloor - \frac{\left\lfloor \frac{n}{k-1} \right\rfloor \left(\left\lfloor \frac{n}{k-1} \right\rfloor + 1\right)(k-1)}{2} &= \frac{\left\lfloor \frac{n}{k-1} \right\rfloor \left(2n - \left(\left\lfloor \frac{n}{k-1} \right\rfloor + 1\right)(k-1)\right)}{2} \\ &\leq \frac{\left\lfloor \frac{n}{k-1} \right\rfloor (n+1-k)}{2} \\ &\leq \frac{n^2}{2(k-1)} - \frac{n}{2} \\ &\leq \frac{n^2}{2(k-1)} \end{aligned}$$

$\blacksquare$

Since the probability of a $k$-AP being monochromatic is $c^{1-k}$, we use the union bound to find an upper bound on $p$ in terms of some arbitrary $n$, $c$, and $k$. From that, we rearrange terms to find a lower bound of $n$.

$$p \le A(n,k)c^{1-k} \le \frac{n^2}{k}c^{1-k}$$
$$pkc^{k-1} \le n^2$$
$$\sqrt{pkc^{k-1}} \le n = W_R(c,k,p)$$

# 4 Empirical Data

In addition to finding a lower bound for the Random Van der Waerden numbers, we also ran random simulations in an attempt to estimate the Random Van der Waerden numbers for $p$ =1, 0.95, and 0.5.

To empirically find $W_R(c,k,p)$, we binary searched to find the smallest $n$ in which the observed fraction of $k$-AP's which are monochromatic was greater than $p$. We also did not compute any values with $n \ge 2000$, as it would take over a second to check if there was a monochromatic $k$-AP in the randomly colored set. While binary searching, we ran as many simulations as possible in 30 seconds for every $(n,c,k)$ tuple to ensure that we adequately represented the true $p$ value for all values of $n,c,k$ that we are computing.

Below are the estimates empirically determined for $p$ =1, 0.95, and 0.5.

Table 1: Data Values for Different $p$ Values

(a) $p = 1.0$

| $c \backslash k$ | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2** | 3 | 9 | 29 | 49 | 82 | 123 | 193 | 273 | 426 | 619 | 768 | 1213 | 1635 |
| **3** | 4 | 22 | 52 | 104 | 190 | 342 | 616 | 998 | 1797 | | | | |
| **4** | 5 | 30 | 74 | 188 | 373 | 712 | 1510 | | | | | | |
| **5** | 6 | 40 | 104 | 248 | 573 | 1322 | | | | | | | |

(b) $p = 0.95$

| $c \backslash k$ | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2** | 3 | 8 | 16 | 26 | 43 | 69 | 104 | 159 | 235 | 352 | 516 | 769 | 1096 | 1677 | | |
| **3** | 4 | 12 | 26 | 55 | 105 | 201 | 370 | 692 | 1261 | 2177 | | | | | | |
| **4** | 5 | 16 | 39 | 93 | 204 | 444 | 928 | 2046 | | | | | | | | |
| **5** | 5 | 20 | 53 | 139 | 342 | 841 | 1991 | | | | | | | | | |

(c) $p = 0.5$

| $c \backslash k$ | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** | **15** | **16** | **17** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2** | 2 | 5 | 9 | 15 | 23 | 36 | 53 | 80 | 118 | 173 | 253 | 375 | 538 | 800 | 1125 | 1695 |
| **3** | 3 | 7 | 14 | 28 | 53 | 98 | 181 | 334 | 601 | 1104 | 1922 | | | | | |
| **4** | 3 | 9 | 20 | 46 | 99 | 216 | 458 | 992 | 2088 | | | | | | | |
| **5** | 3 | 10 | 27 | 68 | 166 | 399 | 962 | | | | | | | | | |

Despite the known Van der Waerden numbers growing rapidly, the numbers from these simulations don't grow nearly as quickly. Consequently, if the objective is just to have high probability that no monochromatic $k$-AP exists, then a lower value could be used.

If the Van der Waerden numbers become applicable in real-life environments, such as architecture or general infrastructure, knowing approximate Random Van der Waerden numbers would be useful, as significantly lowering $n$ many times only slightly increases the probability of a monochromatic $k$-AP appearing.

# 5 Conclusion

In this paper, we found a lower bound for the Random Van der Waerden numbers and empirically estimated the Random Van der Waerden numbers. While Ramsey Theory and the Van der Waerden numbers are currently not heavily used, in the future, if a researcher needs to have high certainty that a monochromatic $k$-AP will not appear in a $c$-color of $[n]$, they can use a much smaller value than the true Van der Waerden number to have similar results.

Additionally, depending on the tolerance for error required, different values of $p$ can be used. For example, if a researcher needs to be relatively confident that a monochromatic $k$-AP won't appear, they could use the loose $W_R$ values that we empirically found for $p = 0.95$. On the other hand, if the researcher could tolerate more frequent occurences of a monochromatic $k$-AP, they could use our values for $p = 0.5$ instead. By generating these values, others can now use them as loose estimates for the true Van der Waerden numbers depending on the level of precision needed.

## References


[1] W. Gowers. A new proof of Szemerédi's theorem. *Geometric and Functional Analysis*, 11:465–588, 2001.

[2] S. Shelah. Primitive recursive bounds for van der Waerden numbers. *Journal of the American Math Society*, 1:683–697, 1988.

[3] B. van der Waerden. Beweis einer Baudetschen Vermutung (in dutch). *Nieuw Arch. Wisk.*, 15:212–216, 1927.