The Stash-Repair Method applied to Partitionable Numbers

Tim Varghese and George Varghese, UCLA

August 8, 2021

Abstract

In the following note, we find it helpful to use a method of *stash-repair* to characterize the density of what we call k-partitionable integers. When $k = 2$, this is the class of numbers that the OEIS calls biquams or "numbers whose digits can be split into two groups with equal sum" (OEIS A064544). Our theoretical results agree with the experimental results for the density of biquams in A065086 (base 10 biquams) and A064686 (base 3 biquams). We show general results for the density of k-partitionable integers written in any base b for $k > 2$, and suggest that the method of stash-repair may be useful for other problems. We explore connections with a result by Borg, Chayes, and Pittel [1] and suggest a small generalization of their result for $k > 2$.

1 Introduction

This note is dedicated in gratitude to Ron Graham whose algorithm for partitioning we use in this note. Ron signed the first author's math textbook as a young boy, and was the second author's colleague at UCSD. Our last fond memories of Ron were at UCSD during a retirement celebration for Ron and Fan, when both authors of this note presented tributes.

The Online Encyclopedia of Integer Sequences (OEIS) [2]) defines a class of numbers called *biquams* which are "numbers whose digits can be split into two groups with equal sum". For example, OEIS A064544 lists the first few biquams written in base 10 as $0, 11, 22, 33, 44, 55, 66, 77, 88, 99, 101, 110, 112, 121, 123, \ldots$. Observe that 123 can be split into two groups 1.2 and 3 both of whose sum is 3.

An interesting theoretical question is how common such biquams are. What is the density of biquams written in any base b? For example, OEIS A065086 lists the "number of n-digit biquanimous numbers in base 10 not allowing leading zeros". This sequence (number of 1 digit biquams, 2-digit biquams, etc.) is $1, 9, 126, 1920, 27190, 347168, 3990467, \ldots$. On the other hand, OEIS A064686 lists the "number of n-digit base-3 biquams" as $0, 2, 7, 23, 73, 227, 697, 2123, 6433, 19427, \ldots$". It is easy to see that both sequences grow rapidly. However, it is unclear whether there is a pattern to this growth and how it differs for different bases.

We will show in this note that the density of biquams approaches $1/2$ as the number of digits n approaches infinity for any base. This seems to contradict the seeming difference in growth patterns for base 10 and base 3 in A065086 and A064686. However, we will show that this apparent contradiction can be explained because A065086 and A064686 assume no leading zeroes, but our asymptotic density of $1/2$ result counts leading zeroes. A simple correction for leading zeroes gives similar density results, and in fact shows a faster growth rate for lower bases like 3.

We will also show lower bounds on the density of biquams for any base b. More generally, we also show that these results generalize when the "number of parts" is $k > 2$. To do so, we start by defining k-partitionable integers.

Definition 1: A decimal integer written in base b is k-partitionable if we can partition its digits into k subsets all of whose sum is the same.

Example: As seen earlier, the biquam 123 is 2-partitionable with the partitions 1.2 and 3. On the other hand 111 is 3-partitionable with the partitions 1, 1, 1. Further, 178286 is 4-partitionable with 1, 7, 8 2, 6 and 8, all adding up to 8.

Remark: Besides the OEIS series mentioned above (A064686, A065086, A064544), the first few elements in the 2-partitionable sequence $11, 22, 33, 44, 55, 66, 77, 88, 99, 101, 112, 123, \ldots$ match A108773 (concatenation of n and the sum of the digits of n. However, several numbers in A108773 are not 2-partitionable (e.g., 3710 which has an odd digit sum) and many numbers that are 2-partitionable are not in A108773 such as 352. Two other superficially similar sequence at low indices are A108773 and A108203 which are also based on concatenation but have no relationship with 2-partitionable integers (which are far more frequent) for large n. In this paper we address the following.

Question: What is the density of k-partitionable numbers for any base b? We know for example that roughly only 10% of the numbers between 0 and 100 are 2-partitionable (the palindromes 11, 22, etc.) What about for arbitrary n, and as n tends to infinity?

2 Related Work

Our question is clearly related to the classical partition problem in computer science [3]. This is the problem of dividing (if possible) an arbitrary set of n numbers (not necessarily integers) into k partitions whose sum is equal. It is well known that while there exists a pseudo-polynomial time exact algorithm for 2-partition, the problem is NP-hard in general for $k > 2$ [3].

There are, however, good approximations including a simple greedy algorithm due to Graham [4] whose run time is either $O(n)$ (without sorting) or $O \log n$ (with sorting), and whose approximation ratio is 3/2 (without sorting) and 7/6 (with sorting). Such algorithms focus on minimizing the *discrepancy* – the difference between the largest and the smallest partition sums. The approximation ratio is defined as the ratio between the largest partition sum produced by the algorithm and the largest sum produced by an optimal algorithm.

A better algorithm than Graham's algorithm is the Karmarkar-Karp algorithm which "sorts the numbers in descending order and repeatedly replaces numbers by their differences" [3]. While its worst case approximation ratio is only 7/6, in the average case it performs much better than Graham's algorithm. In particular, when numbers are distributed uniformly in $[0, 1]$, its approximation ratio is at most $1 + 1/n^{\Theta(\log n)}$.

The Karmarkar-Karp result led to a number of papers that focused on bounding the expected *discrepancy* of partition sums. For example, a later paper by Karmarkar, Karp, Lueker, and Odlyzko (KKLO) [5] bounded discrepancies using the method of second moments (due to Erdos and Spencer).

The KKLO [5] result formalizes the following intuition. First, the sum of n integers drawn IID from a distribution with random signs (where + represents left partition, - represents the right partition) is $O(\sqrt{(}n)$ because it is a simple random walk. However, there are 2^n possible partitions. Thus they show that the discrepancy for most well-behaved distributions can be as small as $O(\sqrt{(}n)/2^n)$ which quickly goes to 0.

The method can seemingly be applied for the special case that each element of the set is drawn uniformly from the set $0, 1, b$ as in our case. Unfortunately, for our particular problem we found two obstacles:

- The KKLO result bounds the discrepancy but does not estimate the probability that the discrepancy is equal to 0.
- The method is not in the form If Property P. then Property Q holds with high probability. For example, even for large n, we know that if the sum of the digits of n is odd, then n is not 2-partitionable. Thus we know achieving a discrepancy of 0 is impossible for at least half the integers.

More useful for our purposes is a very general theorem due to Borg, Chayes and Pittel (BCP) [1]. In fact, the proof of our density result for $k = 2$ follows immediately from the BCP paper. The BCP paper goes much further than our paper and characterizes phase transitions and many deep structural theorems for the case of $k = 2$. It also holds for arbitrary sets of integers, and not just integer digits. However, it does not appear to generalize to $k > 2$ and does not have lower bounds. Further, the BCP results use fairly advanced methods. By contrast, we will use elementary methods to prove more general results for $k > 2$.

3 The Stash-Repair Method

For proving results about the density of k-partitionable numbers, we found the following elementary method sufficed.

Stash-Repair: We use this to help prove facts of the following form. If a sufficiently large integer has property P, then it has property Q with high probability. Intuitively, the stash (whose density we establish with simple counting arguments) is a small collection of integers that can be used to show that simple algorithms like Graham's that come close to establishing Q can be repaired by an explicit construction to make Q hold. Slightly less informally it consists of three steps:

- *Stash existence*: We show that for large n, there is a particular multiset of integers Q of size $O(1)$ we fix in advance that occurs with high probability.
- *Zeroing In*: We then use a simple construction (probabilistic or constructive) that comes close to showing that numbers that have the stash satisfy property Q, where the closeness is measured by a discrepancy measure that we show is size $O(1)$.
- *Final Repair*: Finally, In the third step, we give explicit construction to show how simple repairs (assuming property P) can be done to make the discrepancy 0 using the stash defined in Step 1.

For example, for 2-partitioning and base 10 the stash we use is a multiset of 8 ones; we zero in by showing that using simple greedy placement strategy of the digits of n (due to Graham [4]) leaves a discrepancy of at most 8; the simple final repair is stealing as many of the 1s as needed, taking only an even number from each partition and placing it in the smaller partition to make the discrepancy zero.

4 The Density of k-partitionable numbers

We prove the following results for $k = 2$:

Density for $k = 2$: The fraction of n digit 2-partitionable numbers written in any base b:

- *Lower Bound:* is 0 for $n = 1$ but at least $1/b - 1/b^2$ for $n >= 2$.
- *Upper Bound:* cannot exceed 1/2.
- *Asymptotic Bound:* Approaches 1/2 exponentially fast as n grows.

For arbitrary k-partitioning, the theorem can be generalized as follows (we will only describe proofs for $k = 2$):

Density for any k: The fraction of n digit k-partitionable numbers in any base b:

- *Lower Bound:* is 0 for $n <= k - 1$ but at least $1/b^{k-1} - 1/b^k$ for all $n >= 2$
- *Upper Bound:* cannot exceed $1/k$.
- *Asymptotic Bound:* Approaches $1/k$ exponentially as n increases as long as $k \cdot b = O(\log n)$

The first two results are immediate (and so are their generalizations to k-partitionable numbers):

Theorem 1 Lower Bound *The fraction of 2-partitionable numbers written in any base is* 0 *for* $n = 1$ *but at least* $1/b - 1/b^2$ *for* $n >= 2$.

Any $n - 1$ digit number can be extended to make at least one n-digit 2-partitionable number by taking the number and applying the simply greedy algorithm due to Graham [3] that places the next digit (in any order) in the partition with lowest sum (breaking ties arbitrarily). It is easy to see that at the end of this process the discrepancy is at most $b - 1$ (e.g., 9 for base 10). If the absolute value of the discrepancy is placed as digit n, the resulting n-digit number is 2-partitionable. We choose not to count the n digit number with all 0's as 2-partitionable. Thus the fraction is at least $1/b - 1/b^n$. Since $n >= 2$, the result follows. A similar argument holds for k-partitioning but this time we need to repair the discrepancies for $k - 1$ partitions and so we need $k - 1$ more digits. This leads to a lower bound of $1/b^{k-1}$.

Theorem 2 Upper Bound *The fraction of k-partitionable numbers less than n cannot exceed* $1/k$.

The sum of the digits of any 2-partitionable n digit number is even because, by definition, its digits can be put into two groups whose sum is equal. Since every second n digit number has an odd digit sum, the result follows. More generally, for any n digit number, if the sum of its digits is not divisible by k, the number cannot be k-partitionable. Hence, the fraction of k-partitionable numbers less than n cannot exceed $1/k$ for any n.

We now prove the main result for $k = 2$ using two theorems:

Theorem 3 Stash Repair *Any integer written in any base b whose digit sum is even, and that contains at least r ones (where r is $b - 1$ if b is odd and $b - 2$ if b is even) is 2-partitionable.*

This is the "repair" part of stash-repair given a stash of r ones. To do the repair, take all the digits except the r ones and apply Grahams Algorithm [4], putting digits in any order into the two partitions, such that at each stage we put the next digit in the partition that has the smaller sum. At the end of this process, the discrepancy is at most $b-1$, but since the digit sum must be even, the discrepancy can be at most $b-2$ if b is even (e.g.,8 for base 10). Since the discrepancy is even, assume the discrepancy is $2m$. Then we start the repair process by putting $2m$ ones in the smaller partition. This must be possible because $2m <= r$. The remaining $r-2m$ ones are even. So, we distribute them evenly in the two partitions, maintaining a discrepancy of 0. Thus, the number is 2-partitionable.

Next, we prove the main asymptotic result by showing that the density of n digit integers that contain a stash of r ones tends to 1 as n tends to infinity.

Theorem 4 Asymptotic Bound *The fraction of n-digit numbers in any base b that are 2-partitionable tends to $1/2$ as n tends to infinity as long as $b = O(\log n)$.*

Consider the fraction of n-digit numbers that have less than r ones, where as in the last theorem, r is $b-1$ if b is odd and $b-2$ if b is even. We will show that this fraction tends to 0 in the limit as n tends to infinity. The theorem follows because of the last theorem (Theorem 4) and the fact that half the n-digit numbers have an even sum.

The fraction of n-digit numbers that have exactly m ones is $\frac{\binom{n}{m}(b-1)^{n-m}}{b^n}$. By Stirlings approximation, this is $\frac{n^m}{m!(b-1)^m(1+1/(b-1))^n}$. This fraction tends to zero for large n because the exponential in the numerator swamps the polynomial in the denominator as long as m is a constant. Thus, the fraction of n-digit numbers that have less than r ones is the sum of this fraction from $m=0$ to $b-1$ or $b-2$, which also tends to 0 as long as b is constant. This remains true if the base $b = O(\log n)$. This also shows that the convergence to 0 grows exponentially with n, although the underlying constants are somewhat weaker than the experimental data we present later.

The generalization to k partitioning requires a larger stash of size $r(k-1)$ because after applying Graham's algorithm we have to repair $k-1$ partitions, and thus requires that $k \cdot b = O(\log n)$.

5 Numbers of unbounded size with even digit sum that are not 2-partitionable

One might wonder if all n-digit numbers with even digit sum are 2-partitionable if n exceeds some bound. This is not true. Many numbers whose digit sum divides 2 are not 2-partitionable, although the density of such numbers goes to 0 in the limit as n tends to infinity by Theorem 4.

Theorem 5 Counterexample *There exists integers of unbounded size whose digit sum divides k that are* **not** *k-partitionable.*

To prove this theorem, we limit ourselves to base 10 and to $k = 2$. Clearly every integer whose digit sum is odd is not 2-partitionable. Consider any number of the form 916^m (i.e, 9 and 1 followed by any number of 6s) and all its permutations, for any m. The digit sum is even but if the 9 and 1 are in the same partition, then there is no solution to $6x = 10$ for any integer x. On the other hand, if 9 and 1 are in different partitions, there is no way to make $6x = 8$ for any integer x. There are many such examples including $71(5)^{2m}$ for $k = 2$ and $b = 10$.

6 Comparison with OEIS Data

Recall that n stands for the number of digits. Define f to be the fraction of 2-partitionable n digit numbers *allowing leading* zeros, and g to be the fraction of 2-partitionable n digit numbers *not allowing* leading zeros, For example, if $n = 3$, f counts 011 in the numerator, but g does not. Thus $f >= g$ for all n. So far, we have been talking about f and have shown that f approaches $1/2$ in the limit as n goes to infinity. However, the OEIS Data, in A065086 (base 10) and A064686 (base 3) give the number of biquams that do not allow leading zeros.

Table 1 shows the data for f and g computed from the OEIS table A065086 for base 10. The first three terms in A065086 are $1, 9, 126$. Ignoring the first term, the number of 2 digit biquams without leading zeroes is 9. Since there are $10 * 10$ 2-digit numbers in base 10, f and g are $9/100$. For $n = 3$, however, we have to add the number of two digit biquams without leading zeros (9) to the number of three digit biquams (126) to get the total number of three digit biquams, 135. Thus, f becomes $135/1000$, and g is $126/1000$. The remaining rows are built up in the same way with f adding up the counts of the previous rows.

n	f	g
2	0.09	0.09
3	0.135	0.126
4	0.2055	0.192
5	0.2925	0.2719
6	0.3764	0.3417
7	0.4368	0.3990
8	0.4711	0.4274
9	0.4879	0.4407
10	0.4952	0.4464
11	0.4982	0.4486
12	0.4993	0.4494
13	0.4997	0.4498

Table 1: Fraction of 2-partitionable integers in base 10 with leading zeros (f) and without leading zeroes (g) as the number of digits n grows. The data is taken from OEIS A065086. Note that f appears to converge to 0.5 as predicted by the theory, but g appears to converge to 0.45.

By contrast, Table 2 shows the corresponding data for f and g computed from the OEIS table A064686 for base 3. The first three terms in A064686 are $0, 2, 7$. Ignoring the first term, the number of 2 digit base 3 biquams without leading zeroes is 2. This time since there are $3 * 3$ 2-digit numbers, f and g are $2/9$. Again, for $n = 3$, we have to add the number of two digit biquams without leading zeros (2) to the number of three digit biquams (7) to get the total number of three digit biquams, 9. Thus, f becomes $9/27$, while g is $7/27$. The remaining rows are built up similarly.

n	f	g
2	0.2222	0.2222
3	0.3333	0.2593
4	0.3951	0.2840
5	0.4321	0.3000
6	0.4554	0.3114
7	0.4705	0.3187
8	0.4804	0.3236
9	0.4870	0.3268
10	0.4913	0.3290
11	0.4942	0.3304
12	0.4993	0.3314
13	0.4997	0.3320

Table 2: Fraction of 2-partitionable integers in base 3 with leading zeros (f) and without leading zeroes (g) as the number of digits n grows. The data is taken from OEIS A064686. Note that f appears to converge to 0.5 as predicted by the theory, but g appears to converge to 1/3.

First, notice that Table 1 and Table 2 are consistent with the lower bound in Theorem 1, with a lower bound of 0.09 for base 10, and $0.22\ldots$ for base 3. Next observe that in both Table1 and Table2, f appears to converge to 0.5, as predicted by Theorem 4.

However, g the fraction without leading zeroes, appears to converge to $9/20$ for base 10, and to $1/3$ for base 3. Is there any reason for this? The following heuristic argument suggests that g should converge to $(b - 1)/2b$ which is $9/20$ for $b = 10$ and $1/3$ for $b = 3$.

Although Table 1 and Table 2 show that g is not constant as n grows, let us assume that it is, and argue later that the error introduced by this approximation is small. Given this assumption, the number of m digit biquams in base b with no leading zeroes is gb^m. Then the total number of n digit biquams including leading zeroes is the sum of the m digit biquams with no trailing zeroes from $m = 2$ to $m = n$. Thus the fraction of n digit biquams (counting leading zeroes) becomes $g\sum_{m=2}^{n} b^m/b^n = g\sum_{m=0}^{n-2} 1/b^m$.

Hence, $\lim_{n\to\infty} g\sum_{m=0}^{n-2} 1/b^m = g(b-1)/b$ using the result for the limit of a geometric series. But this limit should be equal to $\lim_{n\to\infty} f = 1/2$ by Theorem 4. Hence $g(b-1)/b = 1/2$. Solving for g, this suggests that g should converge to $(b-1)/2b$ in the limit.

Note that the approximation (assuming g to be constant for all n) is only bad for the first few terms (see first few rows of Table 1 and Table 2). But the contribution of these terms to the series is negligible because they get divided by the highest powers of b. Formalizing this intuition (if true) will take some work.

7 Connection to the Borg-Chayes-Pittel result [1]

As we said earlier, Theorem 4 for $k = 2$ follows from the main result of Borg, Chase and Pittel which states [1]:

> *"Consider the problem of partitioning n randomly chosen integers between* 1 *and* 2^m *into* 2 *subsets such that the discrepancy, the absolute value of the difference of their sums, is minimized. A partition is called perfect if the optimum discrepancy is* 0 *when the sum of all n integers in the original set is is even, or 1 when the sum is odd. Parameterizing the random problem in terms of* $\kappa = m/n$*, we prove that the problem has a phase transition at* $\kappa = 1$*, in the sense that for* $\kappa < 1$*, there are many perfect partitions with probability tending to 1 as* $n \to \infty$*, while for* $\kappa > 1$*, there are no perfect partitions with probability tending to 1.*

Then Theorem 4 for $k = 2$ follows by setting $m = \log_2 b$.

However, since our result about k-partitionable numbers for $k > 2$ apply to digits in any base $b = 2^m$, the elementary methods used in this paper appear to weakly generalize a small part of the Borgs-Chase-Pittel result to k-way partitioning. However, there is a considerable difference in power. Our elementary methods for $k = 2$ require that $b = O(\log n)$, which is a much stronger condition. Further, the Borg-Chase-Pittel provides tremendous insight into the structure of the phase transition as κ crosses 1 that our elementary methods have no hope of addressing. However, it does suggest the following

Question: Consider the problem of partitioning n randomly chosen integers between 1 and 2^m into k subsets such that the discrepancy, the absolute value of the difference of their sums, is minimized. A partition is called perfect if the optimum discrepancy is 0 when the sum of all n integers in the original set is is even, or 1 when the sum is odd. Then for some function f, if $f(m, k, n) < 1$, can we prove that there exists at least 1 perfect partition with probability 1. (For $k = 2$, the BCP result [1] used $f(m, n, k) = m/n$). Does such an f exist for $k > 2$ and what should it be?

8 Remaining Questions

We end with the following questions beyond the possible generalization of the BCP result we suggested in Section 7:

- Can the asymptotics in Theorem 4 be sharpened to match the growth rate seen experimentally in Table 1 and Table 2? We have used only a particular stash of r ones. Many other stashes work. For example, for base 10, a stash of 8 ones suffices, but a stash of 6 ones and a 2 also works, as do several others. This observation could potentially be used to improve the constants.

- Are there other applications of the random stash method for number theoretic properties of the form If the number satisfies Property P, then it satisfies Property Q with high probability?

- Can the results in this paper be used to make more efficient algorithms to enumerate k-partitionable numbers? After all, even for checking whether a number is 2-partitionable, we know all numbers with odd digit sums are not and almost all the even sums are 2-partitionable. A suitable partition for numbers with even sum of digits can often be found often by the simple algorithms underlying

Theorem 4. Further, if the number contains a suitable set of stashes (like 8 ones in base 10), the number can be immediately determined to be partitionable. If this linear time algorithm fails, one can still fall back on the classical pseudo-polynomial dynamic programming approach (for $k = 2$) or the exponential algorithm (for $k > 2$). It would be nice to add an OEIS series for "triquams" but this will require a more efficient algorithm than the brute force exponential algorithm for $k = 3$.

9 Acknowledgements

We thank Raghu Meka, Amit Sahai, and Terry Tao of UCLA for helpful suggestions and references.

References

[1] Borgs, C., Chayes, J., and Pittel, B. Phase transition and finite-size scaling for the integer partitioning problem. *Random Structures and Algorithms 19*, January.

[2] Online encyclopedia of integer sequences. https://oeis.org/.

[3] Partition problem. https://en.wikipedia.org/wiki/Partition_problem.

[4] Graham, R. L. Bounds on multiprocessing timing anomalies. *SIAM Journal on Applied Mathematics*, **2**, 416–429.

[5] Karmarkar, N., Karp, R. M., Lueker, G. S., and Odlyzko, A. M. Probabilistic analysis of optimum partitioning. *Journal of Applied Probability*, **23**, 626–645.