

Inverting a Batting Average -
an Application of Continued Fractions
(Preliminary Version)

Allen Back

August 30, 2000

Contents

1	Introduction and Motivation	1
2	Continued Fractions	2
2.1	Comparison of Two Numbers from their Continued Fraction Expansions	4
2.2	Lower Bounds on Convergent Denominators	4
3	Application to Batting Averages	4
4	References	5

1 Introduction and Motivation

It's the first month of the baseball season and Colorado Rockies announcer Dave Armstrong reports that Todd Helton is hitting .393 so far this year. Todd lines the first pitch over the shortstop's head and arrives safely at first base. You wonder: Has this hit lifted Todd's batting average over .400?

Of course there are infinitely many batting records which lead to a batting average which rounds to .393. But only a few with a small number of at bats. If we knew the number of at bats, it would be trivial to decide how the extra hit would affect the batting average. The purpose of this paper is to point out an efficient algorithm for enumerating all batting records whose batting average rounds to a given one, and for which the total number of at bats is at most some given number N .

By *batting record*, we mean a pair of non-negative integers (h, n) with $0 \leq h \leq n$. Here h is the number of hits and n the number of at bats. By *batting*

average we mean the result of rounding the fraction $\frac{h}{n}$ to three decimal places. By *inverting* a batting average, we mean the problem of enumerating all batting records with at most N at bats that result in the given batting average. We use the term *inversion* here because of its relation to finding partial right inverses of the map $f : N \times N \rightarrow R$ given by

$$f(h, n) = \text{round}_3\left(\frac{h}{n}\right)$$

Our rounding procedure round_3 will treat a batting record whose decimal representation is $.a_1a_2a_35$ (i.e. midway between two three decimal numbers) by rounding upwards. Thus Todd Elton's .393 means that $.3925 \leq \frac{h}{n} < .3935$. ($\text{Round}_3(x)$ could be defined as

$$\frac{[1000x]}{1000} + \frac{[2000x - 2[1000x]]}{1000}$$

where $[x]$ denotes the greatest integer less than or equal to x .)

Note that the total number of batting records with at most N at bats is given by the triangular number

$$1 + 2 + 3 + \dots + N + (N + 1) = \frac{(N + 1)(N + 2)}{2}$$

So when, say $N = 100$, there are 5151 batting records to distribute among the 1001 possible batting averages (the three decimal numbers from .000 to 1.000.) On average, there are a little more than 5 which lead to a particular batting average.

To efficiently enumerate the batting records which lead to a given batting average, we will use continued fractions.

2 Continued Fractions

Continued fractions (finite simple ones actually) are expressions

$$a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \dots + \frac{1}{a_{n-1} + \frac{1}{a_n}}}}$$

where a_0 is an integer, and the other a_i are positive integers. We will follow Hardy and Wright's *An Introduction to the Theory of Numbers* in notation and basic facts about continued fractions. The above general continued fraction will be denoted $[a_0, a_1, \dots, a_n]$. For example

$$.395 = \frac{79}{200} = 0 + \frac{1}{2 + \frac{1}{1 + \frac{1}{1 + \frac{1}{7 + \frac{1}{2 + \frac{1}{2}}}}}}}$$

So $.395 = [0, 2, 1, 1, 7, 2, 2]$.

The k' th convergent of a continued fraction $[a_0, a_1, \dots, a_n]$ is the value of the continued fraction $[a_0, a_1, \dots, a_k]$. For example the 3rd convergent of $.395$ would be

$$[0, 2, 1, 1] = 0 + \frac{1}{2 + \frac{1}{1 + \frac{1}{1}}} = \frac{2}{5}$$

The value of the k' th convergent can be written as

$$\frac{p_k}{q_k}$$

where p_k and q_k are determined recursively for $k \geq 2$ by the relations

$$\begin{aligned} p_k &= a_k p_{k-1} + p_{k-2} \\ q_k &= a_k q_{k-1} + q_{k-2} \end{aligned}$$

where

$$\begin{aligned} p_0 &= a_0 \\ q_0 &= 1 \\ p_1 &= a_1 a_0 + 1 \\ q_1 &= a_1 \end{aligned}$$

The applicability of continued fractions to our inversion problem depends on the following facts:

- Every non-integral rational number has a unique finite simple continued fraction representation $[a_0, a_1, \dots, a_n]$ with $a_n > 1$. We will call this the *canonical continued fraction representation (CCFR)* of the rational number. (This is *not* a standard term.)
- It is easy to calculate continued fraction expansions by a variant of the Euclidean algorithm.
- It is easy to compare two continued fractions in size.
- The fraction $\frac{p_k}{q_k}$ described above as representing the k' th convergent will be in lowest terms.
- The convergent denominators q_k form an increasing sequence ($q_k \geq q_{k-1}$ for $k \geq 1$), strictly increasing for $k > 1$.
- There are some easy lower bounds on the size of the convergent denominator q_k .

2.1 Comparison of Two Numbers from their Continued Fraction Expansions

The general idea is that increasing the even indexed entries in a CCFR increases the value of the continued fraction while increasing the odd indexed entries decreases the value.

Suppose $A = [a_0, a_1, \dots, a_n]$ and $B = [b_0, b_1, \dots, b_m]$ are different CCFR's.

Let j be the leftmost position in which these CCFR's differ; i.e. $a_i = b_i$ for $i < j$ and $a_j \neq b_j$. In case $a_i = b_i$ for all $i \leq \min(m, n)$ with $m \neq n$, we set $j = \min(m, n)$.

Comparison of the size of A and B depends on the parity of j . Suppose first that $j < \min(m, n)$. Should j be even, then $A < B$ iff $a_j < b_j$. Should j be odd, the inequality is reversed; $A < B$ iff $a_j > b_j$.

In case this leftmost differing position j takes on the value $\min(n, m)$ with $n \neq m$, let us suppose, by renaming if necessary, that $n < m$. Now $j = n + 1$. Then $A < B$ if $n + 1$ is odd and $A > B$ if $n + 1$ is even. This may be remembered heuristically by thinking of $[a_0, a_1, \dots, a_n]$ as $[a_0, a_1, \dots, a_n, \infty]$ even though the latter is not a CCFR.

2.2 Lower Bounds on Convergent Denominators

Two applications of the basic recursion relation $q_k = a_k q_{k-1} + q_{k-2}$ gives

$$q_k = (a_k a_{k-1} + 1)q_{k-2} + a_k q_{k-3}$$

for $k \geq 3$. In particular we have:

Lemma 1: The convergent denominators q_k satisfy $q_k \geq (a_k a_{k-1} + 1)q_{k-2}$ for $k \geq 3$. And so $q_{p+2r} \geq 2^r q_p$ for $p \geq 1$.

3 Application to Batting Averages

Let's apply the above ideas to Todd Helton's hypothetical .393 batting average early in the season. Let us suppose he has at most 100 at bats. Since .393 is a rounding to three decimal places, we first compute the continued fraction expansions of .3925 and .3935. We obtain

$$.3925 = [0, 2, 1, 1, 4, 1, 2, 1, 3]$$

$$.3935 = [0, 2, 1, 1, 5, 1, 1, 4, 7]$$

Our task is now to enumerate all CCFR's of rational numbers between these two numbers and whose denominator is at most 100.

Note that the two CCFR's differ in an even indexed position (counting begins at 0.) By the basic ordering properties, to be between these two CCFR's, the

continued fraction must start with either $[0, 2, 1, 1, 4, 1]$, or $[0, 2, 1, 1, 5]$. These both have the denominator $q_4 = 28$. (Note the first is not a CCFR.) Since $q_8 \geq 4q_4$, it is immediate that our CCFR's with denominator at most 100 will not have more than 9 entries in their CCFR. More specifically, using Lemma 1, the only possibilities are:

$$\begin{aligned} \frac{11}{28} &= [0, 2, 1, 1, 5] \\ \frac{35}{89} &= [0, 2, 1, 1, 5, 3] \\ \frac{24}{61} &= [0, 2, 1, 1, 5, 2] \\ \frac{13}{33} &= [0, 2, 1, 1, 5, 1] \end{aligned}$$

where we have eliminated possibilities that start with

$$\begin{aligned} [0, 2, 1, 1, 4, 1, 3] &= \frac{42}{107} \\ [0, 2, 1, 1, 4, 2, 2] &= \frac{49}{125} \\ [0, 2, 1, 1, 5, 4] &= \frac{46}{117} \end{aligned}$$

since their denominators are too large.

So the only possible hitting records (h, n) leading to a batting average of .393 are $(11, 28)$, $(22, 56)$, $(33, 84)$, $(35, 89)$, $(24, 61)$, $(13, 33)$, $(26, 66)$, and $(39, 99)$.

And the question of whether or not Todd Helton is hitting .400 after his hit? We don't know for sure since there are 8 possible hitting records.

Roughly speaking (using linear approximation or the binomial theorem), the effect of a hit when your current batting average is u with n at bats is to raise the batting average by $\frac{1-u}{n}$. So to raise the batting average by .007 from .393 requires $n \leq \frac{.693}{.007}$ or $n < 86$. In fact, all these possibilities result in Todd reaching .400 after his hit.

4 References

[HW], Hardy and Wright, "An Introduction to the Theory of Numbers", Fifth Edition, Clarendon Press, Oxford, 1979.