On the distribution of fractional parts

Vladimir Barzov

March 24, 2009

What happens to the fractional part of the members of an arithmetic series? If the arithmetic series is $nx, n = 1, 2, 3, \ldots$, the fractional parts are denoted as $\{nx\}, n = 1, 2, 3, \ldots$ and reside on the interval $[0, 1)$. If x is rational, that is, $x = \frac{p}{q}$ for some co-prime p and q, the answer to the question is boring: $\{nx\}$ runs over $\frac{1}{q}, \frac{2}{q}, \ldots, \frac{q-1}{q}$ until it reaches 0 for $n = q$, at which point the process repeats.

Not so in the case of irrational x. For one, no two fractional parts coincide. If it were the case that $\{n_1 x\} = \{n_2 x\}$ for some $n_1 \neq n_2$, we would have to have $n_1 x - n_2 x = k$ for some integer k, from where $x = \frac{k}{n_1 - n_2}$ would contradict the irrationality of x. So periodicity is only a privilege of the rational x. Henceforth, x shall be a *fixed positive irrational* number.

Now, if all those infinitely many points are dispersed across $[0, 1)$, there must be very little room to breathe in there. And in fact, they populate every single corner of $[0, 1)$. We shall prove that shortly. However, a much more interesting fact is that the fractional parts are not only *densely* but also *uniformly* spread.

A convenient way to interpret fractional parts of an arithmetic series is like points on a circle of length 1: the top point is 0 (and 1 at the same time) and every point $\{nx\}$ is simply x units down the road from $\{(n-1)x\}$ measured clockwise on the circle.

We need to formalize things a bit and introduce some notation. Let's start with the notion of a number belonging to an interval on the circle.

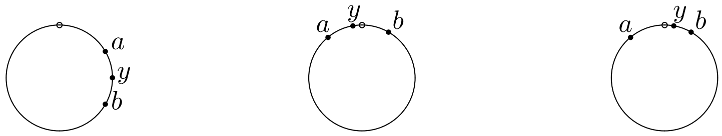
Definition 1. For any $0 \leq a \neq b \leq 1$, we shall call $\langle a, b \rangle$ a *stretch* and shall say that y *belongs* to the stretch (and denote it by $y \in \langle a, b \rangle$) when a, y, b lie in this order clockwise on the circle, that is, when either $a < y < b$, or $y > a > b$, or $a > b > y$ (see Figure 1 below). A stretch $\langle 0, b \rangle$ shall be called *basic.*

The length of a stretch shall naturally be defined as the length of the arc running from a to b clockwise.

Definition 2. The length of $\langle a, b \rangle$, denoted as $l(a, b)$, is defined as:

$$l(a,b) = \begin{cases} b - a & \text{if } b > a \\ b + 1 - a & \text{if } a > b \end{cases}$$

Figure 1: Possible positions for $y \in \langle a, b \rangle$



Here is an important fact that we are going to use throughout the rest of the paper

Proposition 1. For every stretch $\langle a, b \rangle$, there are infinitely many elements of $\{nx\}$ that belong to $\langle a, b \rangle$.

Proof. Let the length of the interval be $l = l(a, b)$ and let $m > \frac{1}{l}$ be an arbitrary integer. Consider the points $\{x\}, \{2x\}, \ldots, \{mx\}$. They break the unit circle into m arcs, and hence one of them must have length of at most $\frac{1}{m}$, which is strictly less than l. Let these be the points $\{n_1 x\}$ and $\{n_2 x\}$, where $n_1 < n_2$. Notice that the distance between them (either clockwise or counterclockwise) is the same as the distance between 0 and $\{(n_2 - n_1)x\}$. Put $t = n_2 - n_1$.

Notice that the points $\{tx\}, \{2tx\}, \{3tx\}, \ldots$ form an arithmetic series over the unit circle with difference less than l. That means that infinitely many of those points would fall in the stretch $\langle a, b \rangle$. $\square$

It seems intuitive that among the points $\{x\}, \{2x\}, \ldots,$ the "share" that belongs to the stretch $\langle a, b \rangle$ should be proportional to $l(a, b)$. This shall be the topic of our further discussion. We need two more definitions of terms that we are going to use later.

Definition 3. For any stretch $\langle a, b \rangle$ and natural n, define

$$F_{a,b}(n, x) = \Big\{ k \in [1, n] \Big| \{kx\} \in \langle a, b \rangle \Big\}.$$

In other words, $F_{a,b}(n, x)$ is the set of positive integers k not exceeding n for which $\{kx\}$ belongs to the stretch.

Definition 4. We shall say that $\langle a', b' \rangle$ is a *translation* of $\langle a, b \rangle$ by tx if $\{a' - a\} = \{b' - b\} = \{tx\}$ for some integer t.

Let us start with some auxiliary claims.

Proposition 2. If $\langle a', b' \rangle$ is a translation of $\langle a, b \rangle$ by tx, then

$$-t \le |F_{a',b'}(n, x)| - |F_{a,b}(n, x)| \le t.$$

Proof. For every integer $m \in [1, n-t]$, if $\{mx\} \in \langle a, b\rangle$ then $\{(m+t)x\} \in \langle a', b'\rangle$, so $|F_{a',b'}(n,x)| \geq |F_{a,b}(n,x)| - t$. Similarly, for every integer $m \in [t+1, n]$, $\{mx\} \in \langle a', b'\rangle$ implies $\{(m-t)x\} \in \langle a, b\rangle$, hence $|F_{a,b}(n,x)| \geq |F_{a',b'}(n,x)| - t$. The result follows. □

Proposition 3. For any given stretch $\langle 0, b\rangle$, for all large enough n holds

$$\frac{|F_{0,b}(n,x)|}{n} < 3b.$$

Proof. Let $m \geq 1$ be the largest integer such that $mb < 1$; We have $(m+1)b \geq 1$ and, in particular, $2bm \geq bm + b \geq 1$.

Define $\delta = \dfrac{1-mb}{m} > 0$. From proposition 1, we know that there exists a positive integer t such that $\{tx\} \in \langle b, b+\delta\rangle$, that is, $b < \{tx\} < b + \delta$. Now, for $k = 0, 1, 2, \ldots, m-1$, consider the stretches $\langle a_k, b_k\rangle$ defined as translations of $\langle 0, b\rangle$ by ktx, that is, $a_k = \{ktx\}$ and $b_k = a_k + b$. Notice that these stretches are non-overlapping as

$$0 = a_0 < b_0 < a_1 < b_1 < a_2 < b_2 < \cdots < a_{m-1} < b_{m-1} < 1.$$

For convenience, denote $f_k(n) = |F_{a_k,b_k}(n,x)|$ and notice that the non-overlaping fact implies that for all n

$$n \geq f_0 + f_1 + \cdots + f_{m-1}.$$

As $\langle a_k, b_k\rangle$ is a ktx translation of $\langle 0, b\rangle$, from proposition 2 we get $f_k \geq f_0 - kt > f_0 - mt$. Therefore,

$$n \geq f_0 + f_1 + \cdots + f_{m-1} > m(f_0 - mt).$$

From $2bm \geq 1$, we have

$$\frac{f_0}{n} \leq \frac{1}{m} + \frac{mt}{n} \leq 2b + \frac{mt}{n}.$$

For all large enough n, $\dfrac{mt}{n} < b$, so

$$\frac{|F_{0,b}(n,x)|}{n} = \frac{f_0}{n} < 3b$$

and the proof is complete. □

We are going to extend this upper bound to non-basic stretches, that is, stretches that don't necessarily start at 0.

Proposition 4. For any stretch $\langle a, b\rangle$, for all large enough n

$$\frac{|F_{a,b}(n,x)|}{n} < 7l(a,b).$$

Proof. Let $l = l(a,b)$. Consider the stretch $\langle 0, 2l\rangle$. From proposition 1, there is an integer t such that $\{tx\} \in \langle a-l, a\rangle$. That means that the translation $\langle a', b'\rangle$ of $\langle 0, 2l\rangle$ by tx contains $\langle a, b\rangle$ entirely, hence $F_{a,b}(n,x) \subseteq F_{a',b'}(n,x)$ for all n.

From proposition 2,

$$|F_{0,2l}(n,x)| + t \geq |F_{a',b'}(n,x)| \geq |F_{a,b}(n,x)|.$$

Next, the result of proposition 3 tells us that $3 \times 2ln > |F_{0,2l}(n,x)|$ for large enough n. Picking n large enough to also satisfy $nl > t$, we get

$$7ln = 6ln + ln > |F_{0,2l}(n,x)| + t \geq |F_{a,b}(n,x)|$$

and the proof is complete. □

Proposition 5. For any two stretches $\langle a, b\rangle$ and $\langle a', b'\rangle$ with the *same length* and for any $\epsilon > 0$,

$$\left| \frac{|F_{a,b}(n,x)|}{n} - \frac{|F_{a',b'}(n,x)|}{n} \right| < \epsilon$$

for all large enough n holds

Proof. Let $\delta = \dfrac{\epsilon}{8}$. From proposition 1, there exists a positive integer t such that $\{tx\} \in \langle a, a+\delta\rangle$, the translation $\langle a'', b''\rangle$ of $\langle a', b'\rangle$ by tx contains $\langle a+\delta, b\rangle$. That means that

$$F_{a,b}(n,x) \subseteq F_{a,a+\delta}(n,x) \cup F_{a'',b''}(n,x). \tag{1}$$

Propositoin 2 tells us that for all large enough n

$$|F_{a'',b''}(n,x)| < |F_{a',b'}(n,x)| + t < |F_{a',b'}(n,x)| + n\delta$$

as long as $n\delta > t$. Also, from proposition 4, $|F_{a,a+\delta}(n,x)| < 7n\delta$ for all large enough n. Combining with (1), we get

$$\begin{aligned} |F_{a,b}(n,x)| &< |F_{a,a+\delta}(n,x)| + |F_{a'',b''}(n,x)| \\ &< 7n\delta + |F_{a',b'}(n,x)| + n\delta \\ &= |F_{a',b'}(n,x)| + n\epsilon, \end{aligned}$$

Similarly, we can show that

$$|F_{a',b'}(n,x)| < |F_{a,b}(n,x)| + n\epsilon.$$

Dividing by n, we get the desired result. □

We have enough ammunitions to attack the main goal of this paper, summarized in the following theorem.

Theorem 1. *For any stretch* $\langle a, b\rangle$,

$$\lim_{n\to\infty} \frac{|F_{a,b}(n,x)|}{n} = l(a,b).$$

Proof. Let $l = l(a,b)$ and let r be an arbitrary positive integer. Put $m = \left\lceil \frac{r}{l} \right\rceil$, so $r \leq ml < r+1$. For $k = 0, 1, 2, \ldots, m$, consider the stretches $\langle a_k, b_k\rangle$ which are translations of $\langle 0, l\rangle$ by kl. Notice that the stretches cover the unit circle r full times, so each of the points $\{x\}, \{2x\}, \ldots, \{nx\}$ belongs to at least r of the stretches, with the exception possibly of the $m+1$ points $0, \{l\}, \{2l\}, \ldots, \{ml\}$ which might not belong to any stretch. Therefore, if we denote $f_k(n) = |F_{a_k,b_k}(n,x)|$, we get

$$f_0 + f_1 + \cdots + f_m \geq r(n - (m+1)).$$

At the same time, none of these n points belongs to more than $r+1$ stretches, whence

$$f_0 + f_1 + \cdots + f_m \leq (r+1)n.$$

If $f(n) = |F_{a,b}(n,x)|$, since $l(a_k, b_k) = l = l(a,b)$, proposition 5, applied $m+1$ times with $\epsilon = \frac{1}{m+1}$, guarantees that for all large enough n, $|f - f_k| < \frac{n}{m+1}$ for $k = 0, 1, \ldots, m$. Therefore,

$$(r+1)n \geq f_0 + f_1 + \cdots + f_m > (m+1)\left(f - \frac{n}{m+1}\right)$$

and

$$(m+1)\left(f + \frac{n}{m+1}\right) > f_0 + f_1 + \cdots + f_m \geq r(n - (m+1)).$$

From the last two inequalities, as long as $n > r(m+1)$, we have

$$\frac{r+2}{m+1} > \frac{f}{n} > \frac{r-1}{m+1} - \frac{r}{n} > \frac{r-2}{m+1}$$

for all large enough n. Therefore,

$$A(r) = \frac{r+2}{m+1} - l > \left(\frac{f}{n} - l\right) > \frac{r-2}{m+1} - l = B(r)$$

for any given r and all large enough n. Remember that r was arbitrary and that $m = \left\lceil \frac{r}{l} \right\rceil$. If we let $r \to \infty$, we have $A(r) \to 0 \leftarrow B(r)$, so for every $\epsilon > 0$ there is an r such that $-\epsilon < B(r) < A(r) < \epsilon$, hence

$$\left| \frac{f(n)}{n} - l \right| < \epsilon$$

for all large enough n. But that is equivalent to

$$\lim_{n\to\infty} \frac{|F_{a,b}(n,x)|}{n} = l$$

and the proof is complete. □

The significance of this result is that the fractional parts of an arithmetic series with irrational difference are spread uniformly over the interval $[0, 1)$.

Here is an interesting problem that can serve as an exercise of the aforementioned concepts.

Problem 1. Prove that for any real $x_1, x_2, \ldots, x_k$ and $\epsilon > 0$ there exists an integer n such that $\{nx_i\} \in \langle 1 - \epsilon, \epsilon \rangle$ for $i = 1, 2, \ldots, k$.