# Problem Statement

We are interested in the following stochastic process:

**Algorithm 1 Rotation Random Walk**

Fix $\alpha \in (0, 1)$ to be the step size.

Set $Y_0 = 0$, where $Y_i \in [0, 1)$ is the position of the random walk at time $i$.

for $i = 1, 2, \ldots n$ do

$X_i = \begin{cases} +1, & \text{w.p. } \frac{1}{2} \\ -1, & \text{w.p. } \frac{1}{2} \end{cases}$

$Y_i = \{Y_{i-1} + \alpha X_i\}$ the new position mod 1.

$Z_i = \begin{cases} +1, & \text{if } Y_i \in [0, \frac{1}{2}) \\ -1, & \text{if } Y_i \in [\frac{1}{2}, 1) \end{cases}$

**Question:** The partial sums $S_n = \sum_{i=1}^{n} Z_i$ define a new stochastic process on $Z$. Does this sum $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i$ follow a central limit theorem (the $Z_i$'s are not independent!)?

## Rational Case proof

We have that for $\alpha = \frac{p}{k}$ where $\gcd(p, k) = 1$, the possible positions of $Y_i$ segment $S^1$ into $k$ equidistant portions since $\{n\alpha\} = 0 \iff k | n$. Thus without loss of generality we can consider $\alpha = \frac{1}{k}$.

The idea of our argument is to split up our $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i$ into independent portions and then apply the central limit theorem to the sum of iid random variables.

**Definition 2.1.** Fix $k \in \mathbb{N}$. Suppose we have a random walk on $k$ "pegs" equal length apart along the unit circle. The return time $T_k$ is the random variable which is the number of steps until we return to the starting point, where we take a left and right step with equal probability.

**Lemma 2.1.** The return time has finite expected value:

$$\mathbb{E}[T_k] = k$$

**Proof:**

We do a first step-analysis using the law of total expectation, conditioning on whether we take a left or right step. Let $T_k$ be the random variable that denotes the first return time to the starting point when there are $k$ positions.

$$\mathbb{E}[T_k] = \mathbb{E}[T_k|R]P(R) + \mathbb{E}[T_k|L]P(L) = \frac{1}{2}[\mathbb{E}[T_k|L] + \mathbb{E}[T_k|R]] = \mathbb{E}[T_k|R]$$

$$\mathbb{E}[T_k|R] = \frac{1}{2}(\mathbb{E}[T|R^2] + \mathbb{E}[T|RL]) = \frac{1}{2}\mathbb{E}[T_k|R^2] + 1$$
\[ \mathbb{E}[T_k | R^2] = \frac{1}{2}(\mathbb{E}[T_k | R^3] + \mathbb{E}[T_k | R^2 L]) = \frac{1}{2}(\mathbb{E}[T_k | R^3] + 2 + \mathbb{E}[T_k | R]) \]

Plugging this back in, we get:

\[ \mathbb{E}[T_k] = \frac{1}{2} \mathbb{E}[T_k | R^2] + 1 = \frac{1}{3} \mathbb{E}[T_k | R^3] + 2 \]

From this we get the general pattern:

\[ \mathbb{E}[T_k] = \frac{1}{j} \mathbb{E}[T_k | R^j] + (j - 1) \]

which is equivalent to

\[ j(\mathbb{E}[T_k] - (j - 1)) = \mathbb{E}[T_k | R^j] \]

This pattern is verified by the following inductive argument.

We have already established the base cases of \( i = 1, 2 \). For the inductive argument we again use the law of total expectation

\[ \mathbb{E}[T_k | R^i] = \frac{1}{2} \left( \mathbb{E}[T_k | R^{i+1}] + (2 + \mathbb{E}[T_k | R^{i-1}]) \right) \]

using the inductive hypothesis

\[ i(\mathbb{E}[T_k] - (i - 1)) = \frac{1}{2} \left( \mathbb{E}[T_k | R^{i+1}] + (2 + (i - 1)(\mathbb{E}[T_k] - (i - 1))) \right) \]

\[ 2i\mathbb{E}[T_k] - 2i^2 + 2i = \mathbb{E}[T_k | R^{i+1}] + 2 + i\mathbb{E}[T_k] - \mathbb{E}[T_k] - i^2 + 2i - 1 \]

rearranging the equation yields the desired result

\[ (i + 1)(\mathbb{E}[T_k] - i) = \mathbb{E}[T_k | R^{i+1}] \]

It is the case that \( \mathbb{E}[T_k | R^k] = k \) since taking \( k \) right steps from the origin gets you back to the origin, and thus

\[ \mathbb{E}[T_k] = \frac{1}{k} k + (k - 1) = k \]

**Lemma 2.2.** Let \( N \) be a nonnegative integer-valued random variable with mean \( \gamma \), and \( X_1, ..., X_N \) be a random number of iid rvs each with mean \( \mu \). Further suppose \( X_i \) and \( N \) are independent for all \( i \). Let \( X = \sum_{i=1}^{N} X_i \). We have that

\[ \mathbb{E}[X] = \mathbb{E} \left[ \sum_{i=1}^{N} X_i \right] = \mathbb{E}[N] \mathbb{E}[X_i] = \gamma \mu \]

**Proof:**

First, notice that

\[ \mathbb{E}[X | N = n] = \mathbb{E} \left[ \sum_{i=1}^{N} X_i | N = n \right] = \mathbb{E} \left[ \sum_{i=1}^{n} X_i \right] = n \mu \]

By the law of total expectation,

\[ \mathbb{E}[X] = \sum_{n} \mathbb{E}[X | N = n] \mathbb{P}(N = n) = \sum_{n} n \mu \cdot \mathbb{P}(N = n) = \mu \sum_{n} n \cdot \mathbb{P}(N = n) = \mu \cdot \mathbb{E}[N] = \gamma \mu \]

**Lemma 2.3.** For \( k \in \mathbb{N} \), consider the rotation random walk by \( \alpha = 1/k \) given in algorithm 1, and let \( a \in \{0, ..., k-1\} \). Then, for \( k^2 \) sufficiently large, \( \mathbb{P}(Y_i \neq 0 \ \forall i \leq k^2 \ | \ Y_0 = a) \leq 0.9 \).
Proof:
Reformulate the problem as a random walk starting at $a$, with boundaries $0$ and $k$. Then, the probability we desire is the same as the probability we never leave this range of $[0,k]$ taking steps of size 1 after $k^2$ steps. Let $(X_n)_{n\in\mathbb{N}}$ be iid Rademacher random variables, and $S_n = a + \sum_{i=1}^{n} X_i$ be the random walk. For $n$ large enough, $S_n = a + \sum_{i=1}^{n} X_i \xrightarrow{d} \mathcal{N}(a,n)$.

$$
P(0 < S_i < k \forall 1 \leq i \leq k^2) \leq \mathbb{P}(0 < S_{k^2} < k) \leq \mathbb{P}(a - k < S_{k^2} < a + k) \approx \Phi(1) - \Phi(-1) \approx 0.683 < 0.9$$

The main theorem:

Theorem 2.4. The random walk defined by partial sums of $Z_i$ follows a central limit theorem.

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i - \mu \xrightarrow{\text{d}} \mathcal{N}(0,1)$$

Proof: Fix $\alpha \in \mathbb{Q} \cap (0,1)$, WLOG we say $\alpha = \frac{1}{k}$ for some $k \in \mathbb{N}$. Let $R_i = \inf\{ j : Y_j = 0, j > R_{i-1} \}$ be the time of the $i^{th}$ return to the origin (let $R_0 = 0$). Notice that $\mathbb{E}[R_i - R_{i-1}] = \mathbb{E}[T_k] = k$, and that $R_0, R_1 - R_0, R_2 - R_1, ...$ are iid. Fix $n \in \mathbb{N}$, and let $r_n = \max\{ m : R_m \leq n \}$ be the number of times we visit the origin after $n$ steps. Then,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{r_n} \sum_{j=R_{i-1}}^{R_i-1} Z_j + \frac{1}{\sqrt{n}} \sum_{i=r_n+1}^{n} Z_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{r_n} W_i + \frac{1}{\sqrt{n}} \sum_{i=R_n}^{n} Z_i$$

where

$$W_i = \sum_{j=R_{i-1}}^{R_i-1} Z_j$$

Let $t_n = \mathbb{E}[r_n]$, so we can further break down the sum as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{t_n} W_i + \frac{1}{\sqrt{n}} \sum_{i=t_n+1}^{r_n} W_i + \frac{1}{\sqrt{n}} \sum_{i=R_n}^{n} Z_i$$

Then $W_1, ..., W_{t_n}$ are iid since each increment $R_i - R_{i-1}$ are return times to the origin, in which the position random variable $(Y_i)$ resets back to zero.

We show that (1) the first term converges in distribution to a normal, (2) the second term converges in distribution/probability to 0, (3) the third term converges in distribution/probability to 0, and (4) By Slutsky's Theorem we can conclude that our sum converges to a normal distribution

(1): We will show that $t_n \to \infty$ as $n \to \infty$, and then we can just apply the CLT.

For large enough $j$, Stirling’s formula says $j! \sim \sqrt{2\pi j} \left(\frac{j}{e}\right)^j$, and hence $(\frac{2j}{j})! = \frac{(2j)!}{(j)!} \sim \sqrt{\frac{4\pi j}{e^2}} \left(\frac{2j}{j}\right)^j = 4^j \sqrt{j \pi}$.

First, we lower bound the probability we are at the origin after $2j$ steps. Notice that if we have exactly $j$ steps in each direction, we are guaranteed to be back at the origin. Hence,

$$\mathbb{P} (\text{at origin after } 2j \text{ steps}) \geq \binom{2j}{j} \left(\frac{1}{2}\right)^{2j} \sim \frac{1}{\sqrt{2\pi j}}$$

For $i = 1, ..., n$, let $V_i$ be 1 if we are at the origin at time $i$, and 0 otherwise. Then, we can write $r_n = \sum_{i=1}^{n} V_i$. Hence

$$t_n = \mathbb{E}[r_n] = \sum_{j=1}^{n} \mathbb{E}[V_j] \geq \sum_{j=1}^{n/2} \mathbb{E}[V_{2j}] = \sum_{j=1}^{n/2} \mathbb{P}(V_{2j} = 1) \geq \sum_{j=1}^{n/2} \binom{2j}{j} \left(\frac{1}{2}\right)^{2j} \sim \sum_{i=1}^{n/2} \frac{1}{\sqrt{2\pi j}} \geq \sum_{i=1}^{n/2} \frac{1}{\sqrt{2\pi j}} \geq \frac{n}{2 \sqrt{2\pi n}} = \frac{\sqrt{n}}{2 \sqrt{2\pi n}}$$

3
So as $n \to \infty$, $t_n \to \infty$, and the CLT applies to this first term.

(2): We are going to show $\frac{1}{\sqrt{n}} \sum_{i=t_n}^{n} W_i \xrightarrow{P} 0$. This quantity is large in absolute value if and only if at least one of the following happens:

1. $|r_n - t_n|$ is large
2. $|W_i|/\sqrt{n}$ is large

We show that these two probabilities can be bounded for $n$ sufficiently large, and then the union bound will imply that the quantity is $< \epsilon$.

For the first, we show that the following (equivalent to $r_n - t_n$ being close) holds with high probability:

$$R_{t_n - c\sqrt{n}} \leq n \leq R_{t_n + c\sqrt{n}}$$

We bound the probability of one side not happening, and the other side is equivalent. $t_n = \mathbb{E}[r_n] \approx n/k$ and $\mathbb{E}[T_i] = k$, so $\mathbb{E}[\sum_{i=1}^{t_n-c\sqrt{n}} T_i] = (n/k - c\sqrt{n})(k) = n - kc\sqrt{n}$. Let $\text{Var}(T_i) = \alpha < \infty$, and so $\sigma^2 = \text{Var}(\sum_{i=1}^{t_n-c\sqrt{n}} T_i) = \alpha (n/k - c\sqrt{n})$. Hence

$$\mathbb{P}(R_{t_n - c\sqrt{n}} > n) = \mathbb{P} \left( \sum_{i=1}^{t_n-c\sqrt{n}} T_i > n \right) = \mathbb{P} \left( \sum_{i=1}^{t_n-c\sqrt{n}} T_i - (n - kc\sqrt{n}) > kc\sqrt{n} \right)$$

Applying Chebyshev’s inequality gives

$$\frac{\alpha(n/k - c\sqrt{n})}{k^2\sigma^2 n} < \frac{\alpha}{k^2c^2} < \epsilon$$

Hence we can always find $c$ such that this probability is $< \epsilon$.

For the second, we need to bound the probability of $|W_i|/\sqrt{n}$ being large. We bound the probability of $W_i/\sqrt{n}$ being large, and the same bound holds for $-W_i/\sqrt{n}$ being largely negative by symmetry.

$$\mathbb{P}(W_i/\sqrt{n} > \epsilon) = \mathbb{P}(W_i > \epsilon\sqrt{n}) = \mathbb{P} \left( \sum_{i=1}^{R_1} Z_i > \epsilon\sqrt{n} \right) = \sum_{j=1}^{n} \mathbb{P} \left( \sum_{i=1}^{R_1} Z_i > \epsilon\sqrt{n} \mid R_1 = j \right) \mathbb{P}(R_1 = j)$$

If $j < \epsilon/\sqrt{n}$, the probability is 0, so

$$= \sum_{j = \epsilon/\sqrt{n}}^{n} \mathbb{P} \left( \sum_{i=1}^{R_1} Z_i > \epsilon\sqrt{n} \mid R_1 = j \right) \mathbb{P}(R_1 = j) = \sum_{j = \epsilon/\sqrt{n}}^{n} \mathbb{P} \left( \sum_{i=1}^{j} Z_i > \epsilon\sqrt{n} \right) \mathbb{P}(R_1 = j)$$

By Markov’s inequality, $\mathbb{P}(R_1 \geq u) \leq \frac{\mathbb{E}[R_1]}{u} = \frac{k}{u}$ Hence the probability $R_1 = k$ for large $k$ decays linearly since the probability you don’t return after such a long time is 0 by the inequality. So for $j > \epsilon/\sqrt{n}$ as in our sum, $\mathbb{P}(R_1 = j) \leq \mathbb{P}(R_1 \geq j) \leq \frac{k}{j}$. So we have

$$\leq k \sum_{j = \epsilon/\sqrt{n}}^{n} \mathbb{P} \left( \sum_{i=1}^{j} Z_i > \epsilon\sqrt{n} \right) / j$$

But this probability term is going to 0 by Chebyshev’s inequality; this is the probability $Z_i$ is getting far from its mean of 0. Hence this goes to 0.

(3): When considering $\frac{1}{\sqrt{n}} \sum_{i=R_n}^{n} Z_i$, we wish to show

$$\frac{1}{\sqrt{n}} \sum_{i=R_n}^{n} Z_i \xrightarrow{d} 0$$
This is equivalent to showing for any $\epsilon > 0$,

$$
\lim_{n \to \infty} \mathbb{P}\left( \left| \frac{1}{\sqrt{n}} \sum_{i=R_{rn}}^{n} Z_i \right| > \epsilon \right) = 0
$$

since $|Z_i| = 1$ we have that

$$
\left| \frac{1}{\sqrt{n}} \sum_{i=R_{rn}}^{n} Z_i \right| \leq \frac{1}{\sqrt{n}} \sum_{i=R_{rn}}^{n} |Z_i| = \frac{n - R_{rn}}{\sqrt{n}}
$$

In other words

$$
\left| \frac{1}{\sqrt{n}} \sum_{i=R_{rn}}^{n} Z_i \right| > \epsilon \Rightarrow \frac{n - R_{rn}}{\sqrt{n}} > \epsilon
$$

so

$$
\mathbb{P}\left( \left| \frac{1}{\sqrt{n}} \sum_{i=R_{rn}}^{n} Z_i \right| > \epsilon \right) \leq \mathbb{P}\left( \frac{n - R_{rn}}{\sqrt{n}} > \epsilon \right) = \mathbb{P}(R_{rn} < n - \epsilon \sqrt{n})
$$

$$
= \sum_{j=1}^{\infty} \mathbb{P}(R_{rn} = j)
$$

Since $R_{rn}$ was defined as the last return time of the $Y_i$s,

$$
\mathbb{P}(R_{rn} = j) = \mathbb{P}(\{Y_j = 0\} \cap \{Y_i \neq 0 \forall i > j\}) = \mathbb{P}(Y_i \neq 0 \forall i > j \mid Y_j = 0)\mathbb{P}(Y_j = 0) \leq \mathbb{P}(Y_i \neq 0 \forall i > j \mid Y_j = 0)
$$

This probability is equivalent to the probability of a new random walk on the circle not returning after $n$ steps. From Lemma 2.3, for sufficiently large $n$ we get

$$
\mathbb{P}(R_{rn} = j) \leq 0.9^{(n-j)/k^2-1}
$$

Thus

$$
\mathbb{P}\left( \left| \frac{1}{\sqrt{n}} \sum_{i=R_{rn}}^{n} Z_i \right| > \epsilon \right) \leq \sum_{j=1}^{\infty} 0.9^{(n-j)/k^2-1}
$$

This the tail of a geometric series with some reindexing

$$
= \sum_{j=\epsilon \sqrt{n}}^{n} 0.9^{(j)/k^2-1} \leq \sum_{j=\epsilon \sqrt{n}}^{\infty} 0.9^{(j)/k^2-1} \to 0 \text{ as } n \to \infty
$$

(4): (Slutsky) Suppose $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} c$, with $(X_n)$ and $(Y_n)$ independent sequences of random variables. Then, $X_n + Y_n \xrightarrow{d} X + c$.

We are given that

$$
\lim_{n \to \infty} \mathbb{P}(X_n \leq x) = \mathbb{P}(X \leq x) = F_X(x)
$$

and

$$
\lim_{n \to \infty} \mathbb{P}(Y_n \leq y) = \mathbb{P}(c \leq y) = 1_{(c \leq y)}
$$

We wish to show that

$$
\lim_{n \to \infty} \mathbb{P}(X_n + Y_n \leq z) = \mathbb{P}(X + c \leq z) = \mathbb{P}(X \leq z - c)
$$

We suppose $X_n$ is discrete, but a similar argument holds if $X_n$ is continuous by replacing sums with integrals and $\mathbb{P}(X_n = x)$ by $f_{X_n}(x)dx$.  

5
\[ P(X_n + Y_n \leq z) = \sum_x P(X_n + Y_n \leq z | X_n = x) P(X_n = x) \]
\[ = \sum_x P(Y_n \leq z - x) P(X_n = x) \]
\[ \rightarrow \sum_x \mathbb{1}_{c \leq z - x} P(X_n = x) \]
\[ = \sum_{x : x \leq z - c} P(X_n = x) \]
\[ = P(X_n \leq z - c) \]
\[ \rightarrow P(X \leq z - c) \]

as desired. A similar argument holds replacing sums with integrals.

\[ \blacksquare \]