

# Assignment 2 \*

Ishaan Kapoor

February 8, 2022

## 1 Choosing $r, b$

A

We know,  $t = 160$

$\tau = 0.70$

Also, we know

$b \approx -\log_{\tau} t$

$\implies b \approx 14.229129075$

Or,

$b = 14$

We also know

$t = r \times b$

$\implies r = \frac{t}{b}$

$\therefore r = 11.42$

Or,

$r = 12$

However, this will not give us a decent similarity so we need to test other values.

From this we can clearly see that  $r=20, b=8$  gives us a good curve.  $\implies r = 20, b = 8$

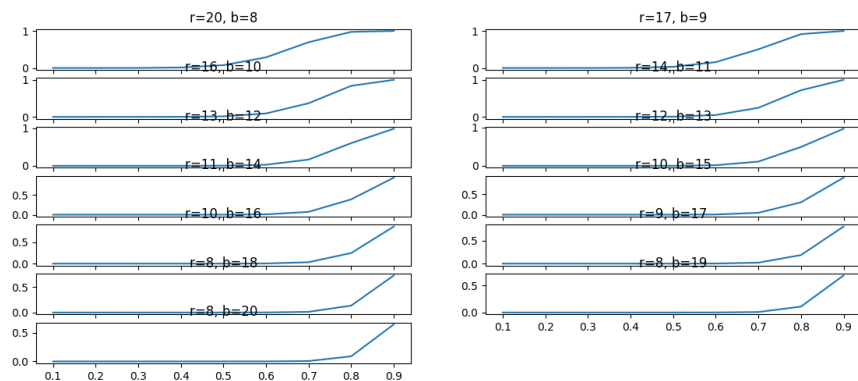


Figure 1: The following image is tau vs  $f(s)$  for various values of  $r$  and  $b$

## B

Let us take

$$f_{r,b}(s) = 1 - (1 - s^b)^r,$$

$$b = 8$$

$$r = 20$$

We know,

$$s_{A,B} = 0.75$$

$$s_{A,C} = 0.31$$

$$s_{A,D} = 0.25$$

$$s_{B,C} = 0.46$$

$$s_{B,D} = 0.63$$

$$s_{C,D} = 0.18$$

$\therefore$  similarity probability:

$f_{r,b}(s)$	$A$	$B$	$C$	$D$
$A$	$X$	0.8787	0.0017	0.0003
$B$	$X$	$X$	0.039	0.39
$C$	$X$	$X$	$X$	0.0000
$D$	$X$	$X$	$X$	$X$

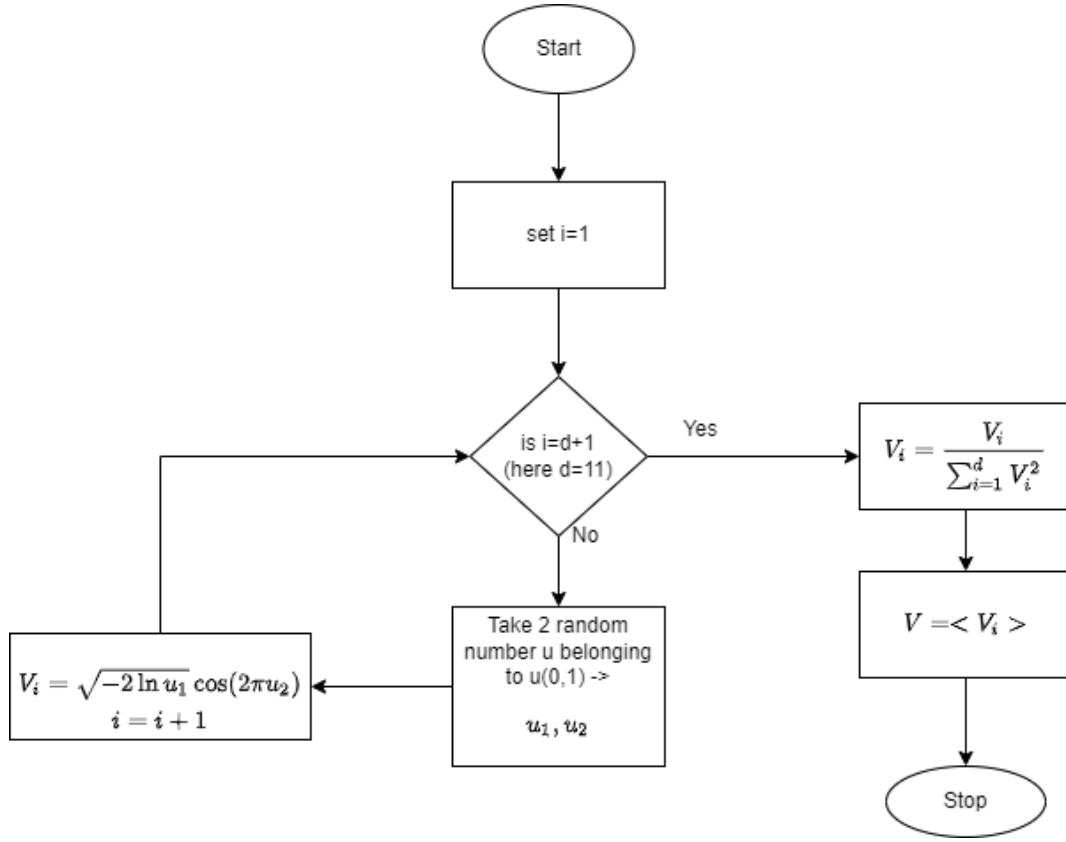
## 2 Generating Random Directions

### A

The flowchart below explains an algorithm to create a random d dimensional unit vector.

### B

Refer to Figure 2



### 3 Angular Hashed Approximation

#### A

Refer to Figure 3 There are a total of 36049 pairs of vectors of the 101025 pairs that have a similarity  $> \tau(0.70)$

#### B

Refer to Figure 4 Interestingly none of the pairs that were created had a similarity of  $> \tau(0.70)$  which is surprising considering the distribution was Gaussian

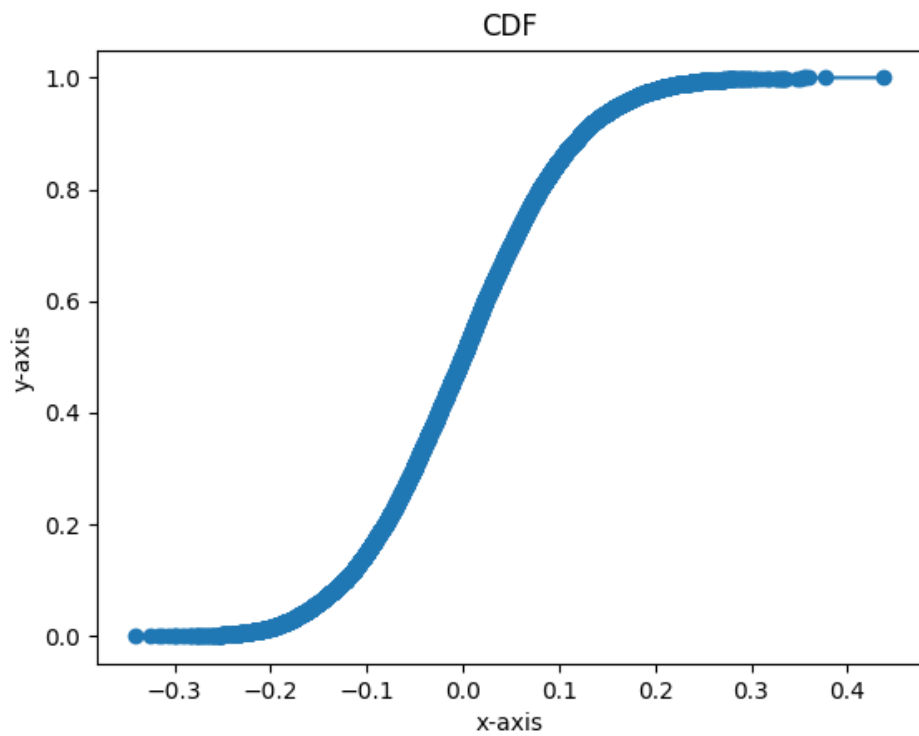


Figure 2: 'CDF of dot products of  $t$  random unit vectors ( $d=110$ ,  $t=180$ )'

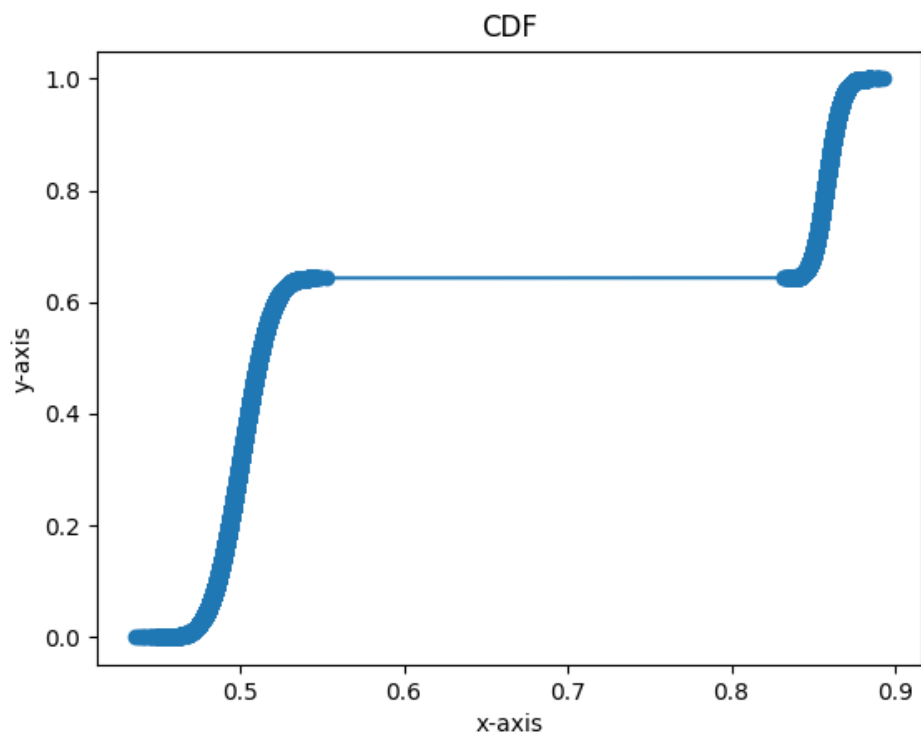


Figure 3: 'CDF of angular similarity of  $\binom{n}{2}$  values'

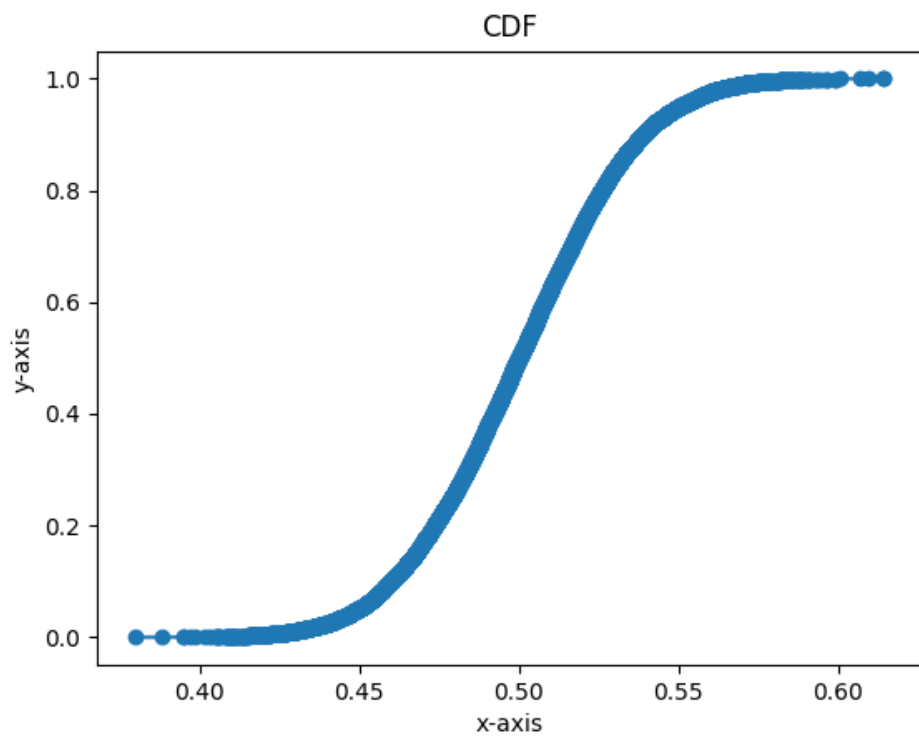


Figure 4: 'CDF of angular similarity of  $\binom{t}{2}$  values'