Derivations For Random Walker

usaid.malik

May 2024

1 Problem Statement

Given a random walker and d dimensions after T time steps what is the expected distance (L2 squared norm) that the walker will have traveled? The walker can take a step forward or a step backward in any direction with equal probability for taking a step backward and forward.

2 Expectation

the vector $\vec{x}$ is a vector with d dimensions.

L2 squared norm for $\vec{x}$ is $||\vec{x}||_2^2 = \sum_{i=1}^{d} x_i^2$

$E(||\vec{x}||_2^2) = \sum_{i=1}^{d} E(x_i^2)$ By definition and linearity of expectation

x_i^2 is a random variable representing an entry in the vector $\vec{x}$. the random variable is the sum of each step taken in that direction after T time steps. The probability of incrementing a step in that direction i is represented by the random variable s_i where

$$s_i = \left\{ \begin{array}{lr} 0, & \text{with probability } 1 - 1/d \\ -1, & 1/2d \\ 1, & 1/2d \end{array} \right\}$$

where the $1/d$ term comes from the probability of the walker taking a step in that direction.

As such then:

$$x_i^2 = (\sum_{j=1}^{T} s_i)^2 = \sum_{j=1}^{T} s_i \sum_{k=1}^{T} s_k = s_1^2 + ... + s_T^2 + s_1 s_2 + ... + s_j s_k + s_{n-1} s_n$$

Then by the linearity of expectation :

$$E(x_i^2) = E(s_1^2) + ... + E(s_T^2) + E(s_1 s_2) + ... + E(s_j s_k) + E(s_{n-1} s_n)$$

$E(s_1^2) = 0^2 * 1 - 1/d + -1^2 * 1/2d + 1^2 * 1/2d = 1/d$ $E(s_j s_k) = E(s_j) * E(s_k) = 0$ because of independence and the expectation of the random variable s_i

Therefore $E(x_i^2) = T/d$

and for the entire vector L2 norm then

$$E(||\vec{x}||_2^2) = \sum_{i=1}^{d} E(x_i^2) = T$$

which doesnst tell us where a walker is only that the L2 squared norm from the origin is expected to be T

3 Variance

It is also helpful to know the variance of the entries of the vector.

$$VAR(x_i^2) = E((x_i^2)^2) - E(x_i^2)^2$$
$$E((x_i^2)^2) - (T/d)^2$$

$E((x_i^2)^2) = E((s_1^2 + ... + s_T^2 + s_1 s_2 + ... + s_j s_k + s_{T-1} s_T)^2)$

$(s_1^2 + ... + s_T^2 + s_1 s_2 + ... + s_j s_k + s_{T-1} s_T) * (s_1^2 + ... + s_T^2 + s_1 s_2 + ... + s_j s_k + s_{T-1} s_T)$
$s_1^4 + ... + s_T^4 + s_1^2 s_2^2 + ... + s_j^2 s_k^2 + ... s_{T-1}^2 s_T^2 + ... s_1^3 s_2 + ... s_i^3 s_j + ... + s_i^2 s_j s_k ... + s_i s_j s_k s_l$ where $j \neq i \neq k \neq l$ but it shouldn't matter for the expectation because if any items have a single independent s then it will be a 0 expectation because of independence so in reality the expectation will only be over these terms (or at least that's what i hope is right)

$s_1^4 + ... + s_T^4 + s_1^2 s_2^2 + ... + s_j^2 s_k^2 + ... s_{T-1}^2 s_T^2$ as such $E(s^4) = 1/d$ and $E(s_1^2 s_2^2) = 1/d^2$ because of independence and the values of the squared expectations

Therefore since there are T of the quadratic terms the expectation sum is now

T/d

and since there are $T(T-1)/2 - T$ squared pairs where the $-T$ comes from taking out the quadratic pairs the expectation sum for the squared terms is $(T(T-1)/2 - T)/d^2$

Adding everything togethere the variance for the x_i^2 term is

$VAR(x_i^2) = T/d + (T(T-1)/2 - T)/d^2 - (T/d)^2$

As the variance of each entry is known it can then be found the variance of the L2 norm squared for this situation as such

$$VAR(||\vec{x}||_2^2) = E((||\vec{x}||_2^2)^2) - E(||\vec{x}||_2^2)^2$$

$$VAR(||\vec{x}||_2^2) = E((||\vec{x}||_2^2)^2) - T^2$$

$E((||\vec{x}||_2^2)^2) = E((\sum_{i=1}^{d} x_i^2)^2)$ where the inner portion can be expanded to be

$(\sum_{i=1}^{d} x_i^2)^2 = x_1^4 + ... + x_d^4 + x_1^2 x_2^2 + x_i^2 x_j^2 + x_{d-1}^2 x_d^2$ where $i \neq j$ then finding the expectation for each term is then $E(x_i^4) = T/d + (T(T-1)/2 - T)/d^2 = \frac{T}{d} + \frac{T^2 - 3T}{2d^2}$ where this was found previously and since the squared pairs are

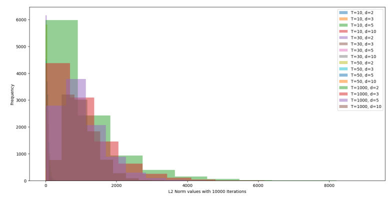

Figure 1: Graph of distribution of frequencies of $||x||_2^2$ with 10,000 iterations

independent there exepcaation is the product of each independent expectation which is $E(x_i^2 * x_j^2) = (T/d)^2$ where this was found previously.

As such since there are d quadratic terms and $\frac{d(d-1)}{2} - d$ squared terms the expected value for the L2 norm squared squared can be written as

$E((||x||_2^2)^2) = d(\frac{T}{d} + \frac{T^2-3T}{2d^2}) + (\frac{d(d-1)}{2} - d)(\frac{T}{d})^2$

Then the variance can be written as

$$VAR(||\vec{x}||_2^2) = d(\frac{T}{d} + \frac{T^2 - 3T}{2d^2}) + (\frac{d(d-1)}{2} - d)(\frac{T}{d})^2 - T^2$$

$$VAR(||\vec{x}||_2^2) = T + \frac{T^2 - 3T}{2d} + \frac{T^2(d^2 - 3d)}{2d^2} - T^2$$

4 Probability Distribution

As the random variable $||x||_2^2$ is a sum of random variables squared which are themselves the sum of uniformly distributed random variables. the variable $||x||_2^2$ follows a Chi squared distribution given enough sample points as its sums are gaussian random variables given enough sampling points, else wise the distribution remains unknown. with mean T and variance given in the previous section for the random variable

Each portion of teh vector is a gaussian random variablewith mean 0 and variance T/d which given enough time steps however initially the distribution is unknown.

5 Emperical graphing of L2 squared norm

As was expected theoretically the variable $||x||_2^2$ does follow a chi squared distribution:

And the entries of the vector similarly also follow a normal distribution

Since each squared random variable has a mean of T/d and a variance of $VAR(x_i^2) = T/d + (T(T-1)/2 - T)/d^2 - (T/d)^2$

the chi squared distribution above with d degrees of freedom (the number of entries of the vector being added in) will have a mean of $d * T/d = T$ and a

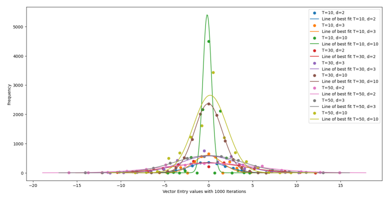

Figure 2: normal distribution for entry values of the vector with 1000 iterations only

variance of twice the sums of teh variances of the squared variables so

$$T + \frac{5T(T-1)}{d} + \frac{T(d-1)}{d} - T^2$$

this should always be positive hopefully as long as, na its always pos W variance W calculation.

its pos as long as

also T and d are always greater than or equal to 0 besides d

$T^2 < T + \frac{5T(T-1)}{d} + \frac{T(d-1)}{d}$

Dividing out by T

$T < 1 + \frac{5(T-1)}{d} + \frac{(d-1)}{d}$

$dT < 5T + 2d - 6$ multiplying by d and doing some combine

$T(d-5) < 2d - 6$ true as long as something.

its a chi squared distribution with a variance of the distribution along the varaince of the original gaussian mean ish.

6 References

Prof Christopher Musco Problem Solutions and Problem Set: https://www.chrismusco.com/amlds2023/homewo