Limiting process for Einstein-Wiener random walk simulation.

Computer assignment #2, Math 504, CSUF, spring 2008

by Nasser Abbasi

February 26, 2008

Contents

1 Purpose and design of project

1.1 Nature of the project

We are solving problem #2 as described in the following screen shot (taken from the class handout)

1. Derive the Einstein-Wiener process by noting that the position of the particle is $x = j\Delta x$, where $j = X_1 + X_2 + \cdots X_n$ with $t = n\Delta t$, and the X_i are independent and identically distributed random variables which have value $+1$ with probability p, and value -1 with probability $q = 1 - p$. Take $p = q = 1/2$.

2. (a) Use the formulation in the previous exercise to simulate the random walk for $p = q = 1/2$, and a specified diffusion coefficient D. Restrict Δx and Δt so that $D = (\Delta x)^2/2\Delta t$. (b) Use the simulation model to test that in the limit as $\Delta x \to 0$ and $\Delta t \to 0$, subject to $D = (\Delta x)^2/2\Delta t$, the distribution of position, for fixed time t and given D, is normal with mean 0 and variance $\sigma^2 = 2Dt$.

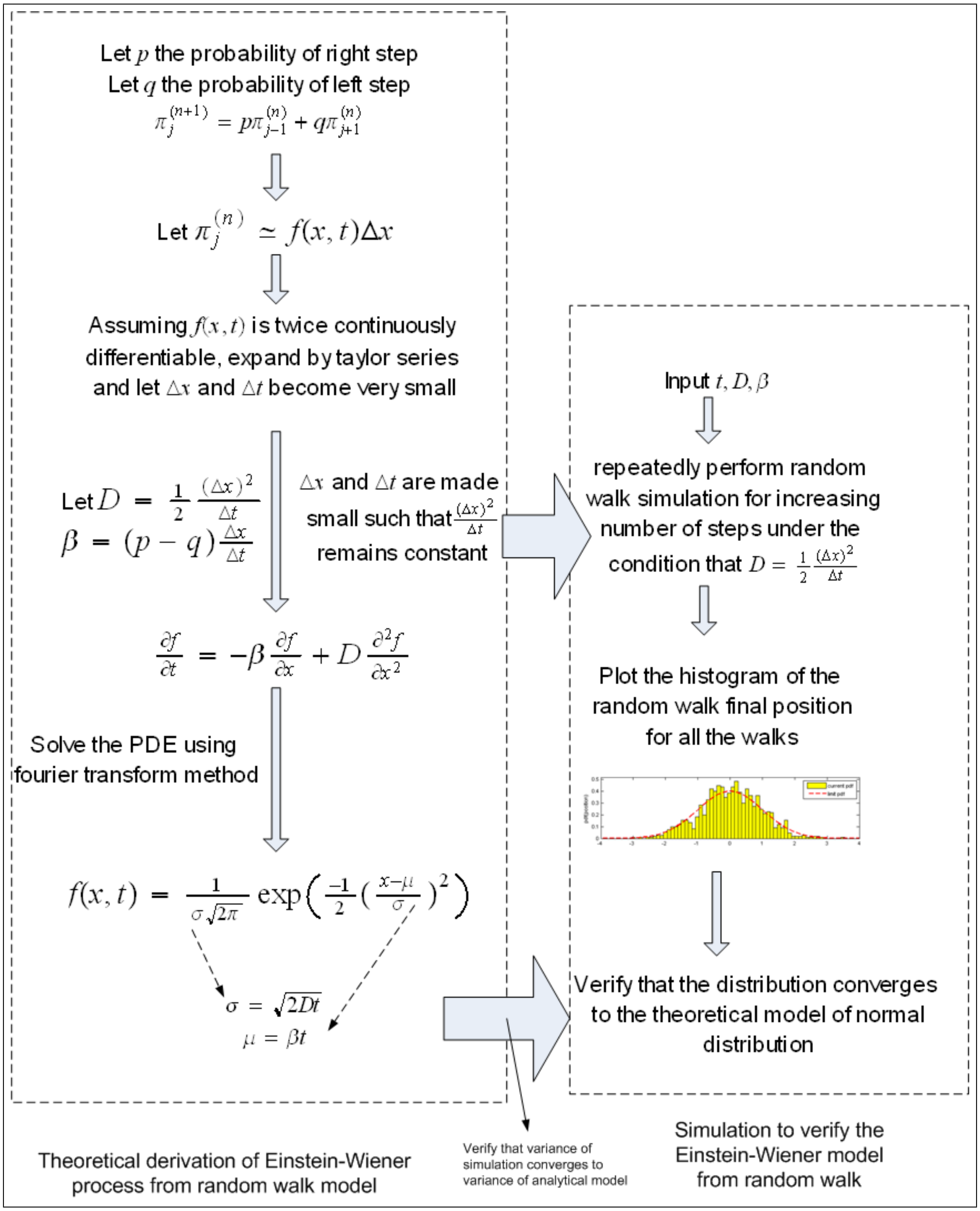
Short background on the problem: In this project we are asked to verify an analytical result derived in a handout given in the class called 'Continuos approximation to random walk'.

A random walk is formulated, by proposing that $\pi_j^{(n)}$ which is the probability that the position of a particle at $x = j\Delta x$ and at time $n\Delta t$ can be expressed as $f(x,t)\,\Delta x$, where $f(x,t)$ represents a density per unit length, which gives a measure of the particle being at that position x at time t.

Starting with this and applying a limiting argument lead to a partial differential equation whose solution is the normal distribution function with certain mean and variance. However, the condition for arriving at the PDE was that as we make Δt and Δx small, we needed to keep the ratio $\frac{(\Delta x)^2}{\Delta t}$ constant.

In this assignment, we simulate a random walk as Δt and Δx are made smaller and smaller subject to this same condition to verify if the distribution of the final position of the random walk converges to the solution of the PDE which is normal distribution and if the converged distribution will have the same variance of $2Dt$ and same mean of βt as does the solution of the PDE.

The details of the theoretical derivation is shown in the above mentioned handout. A diagram below is made to help illustrate the overall purpose of this assignment. In this assignment, we are working on the flow shown on the right side below.



Random walk simulation to verify the Einstein-Wiener analytical derivation

1.2 Questions we are investigating

These are the questions we are trying to answer in this project

1. Does the distribution of the random walk final position generated by increasing the number of steps for fixed t (total time of the random walk) while keeping the ratio $\frac{(\Delta x)^2}{\Delta t}$constant (equal to $2D$), converges to a normal distribution (which is the solution of the Einstein-Wiener process model)?

2. Does the variance of the above distribution converges, as $\Delta t \to 0$ and $\Delta x \to 0$ under

the above mentioned condition of keeping $\frac{(\Delta x)^2}{\Delta t}$, to the analytical variance of $2Dt$ and the theoretical mean of βt?

1.3 Few words on the program

The input to the program is t, D, β where t is the total random walk time and D, β represents the terms as shown in the diagram above.

A distribution of the final random walk position is generated by running the random walk simulation a number of times (called the sample size). In each such run, we use a specific number of steps. The number of steps is increased, and we generate another distribution. We keep doing this and plot each distribution as the number of steps is increased.

At the end of the simulation, to verify that the distribution in the limit is normal. A quantile-quantile plot is made to compare the generated histogram with the theoretical standard normal distribution to see if the result is close to a straight line or not. Also a plot is made showing the convergence of the variance of the current distribution as number of steps is increased by keeping track of the relative error in the variance. In addition, the RMS error between the standard normal and the current distribution is calculated and plotted as a function of delta(T) as delta(T) is made smaller and smaller. The program is written in Matlab version 2007a and uses the statistics toolbox.

1.3.1 The following is a description of the algorithm of the program

We simulate a random walk, where each step made is either to the left or to the right with probability q and p respectively.

Let Y_i be either 1 or -1 depending if we make a right or a left step. Hence

$$Y_i = \begin{cases} 1 & probability\ p \\ -1 & probability\ q \end{cases}$$

and now if we let $X_n = Y_1 + Y_2 + \cdots + Y_n$ then the final position of the random walk can be written as

$$X_n = \Delta x \sum_{j=1}^{n} Y_j$$

where Δx is the step size. The step size is found by solving $\Delta x = \sqrt{2D\Delta t}$ where D is the diffusion parameter which is an input, and Δt is the current time step found by dividing the total simulation fixed time t, which is an input, by the current number of steps n.

$$\Delta t = \frac{t}{n}$$

This program handles a general value for β other than zero. To be able to accomplish this, we need to determine the correct starting step size n to avoid the problem with coming up with a value for the probability p being larger than 1. So, this was done in the initialization stage using this formula

$$starting\ n = round\left(\frac{t\beta^2}{2D}\right) + 1$$

And the simulation was started from the above n and not from 1.

1.3.2 A note about the quantile-quantile plot

To answer the first question of this simulation, which is to determine if the final position distribution converges to normal distribution with mean βt and variance $2Dt$, a quantile plot was used. In this plot, the quantile for the standard normal distribution was plotted against the quantile of the distribution of the final position.

The $x - axis$ of the quantile-quantile plot was found as follows

$$\begin{aligned}
n &= sample_size \\
x &= F^{-1}(([1:n] - 0.5)/n
\end{aligned}$$

Where F^{-1} is the inverse of the CDF for the standard normal distribution (the matlab function norminv() was used for this). While the $y - axis$ is the quantile of the actual data (the sample data of the final distribution of the random walk position). This was found by sorting the data from small to large and then using the resulting sorted vector as the y values. Notice that the distribution was already standardized using

$$z(i) = \frac{y(i) - \mu}{\sigma}$$

Where $\mu = \beta t$ and $\sigma = \sqrt{2Dt}$,

2 Summary of numerical results

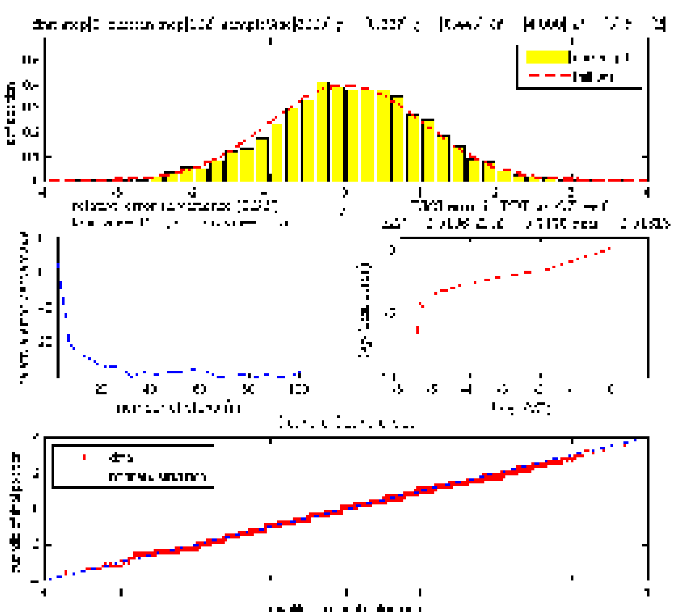
A number of experiments were performed for different input parameters. The table below lists the variance of the distribution of the final position as the number of steps is increased. The run parameters are also shown

2.1 Experiment #1 $\beta = 2, t = 2, D = 3, n = 100$

starting step number= 2, $\beta = 2, t = 2, D = 3, final\ p = 0.557, final\ q = 0.443$

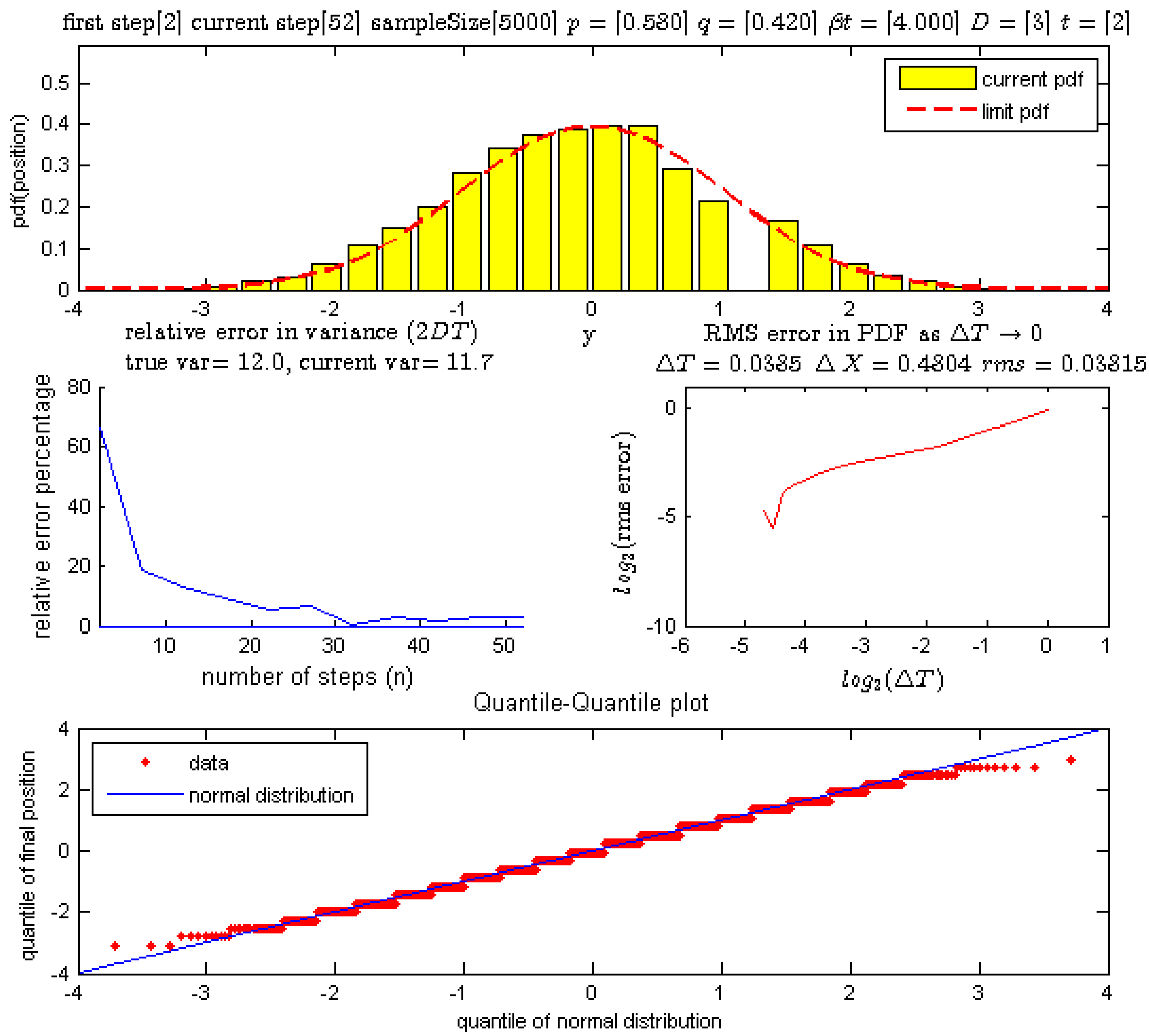
sample size 5000, number of bins 40, seed= 123456

n (number of steps)	Variance	True variance (2Dt)	Δt
2	3.92	12	1
7	9.73	12	0.2857
12	10.43	12	0.1667
17	10.9	12	0.1176
22	11.37	12	0.0909
27	11.19	12	0.0741
32	12.02	12	0.0625
...	...	...	...
67	12.05	12	0.0299
72	11.89	12	0.0278
77	12.16	12	0.0260
82	11.99	12	0.0244
87	11.78	12	0.0230
92	12.03	12	0.0217
97	11.88	12	0.0206
102	11.47	12	0.0196



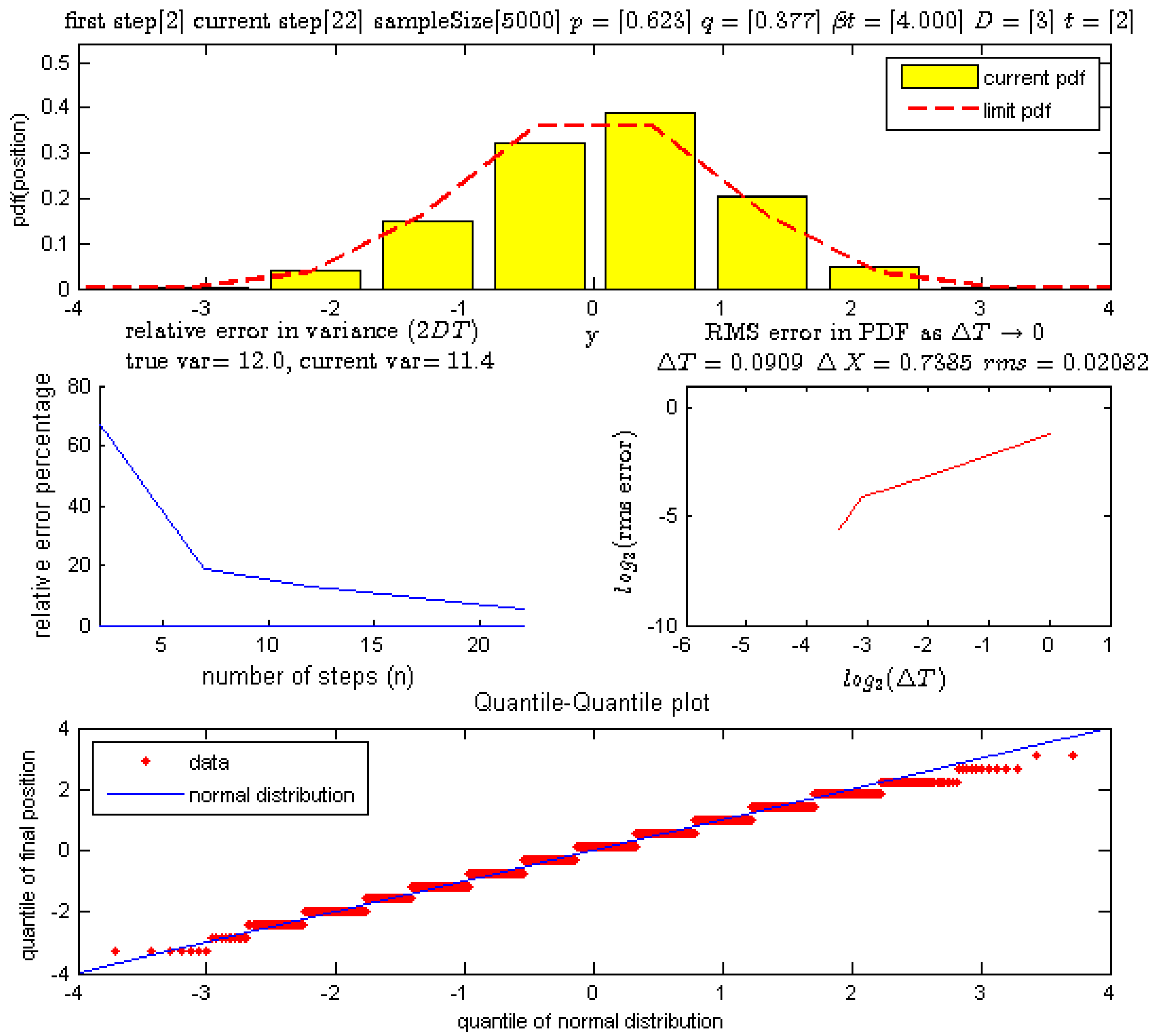
2.2 Experiment #2 $\beta = 2, t = 2, D = 3, n = 50$

Since the parameters t, D, β, then running for $n = 50$ will produce the same numerical values already contained in the first experiment when looking at the table above up to $n = 50$ (the program starts by seeding the random number generator, so nothing will change here and we will just produce a subset of the result already produced in first experiment). So I will just show the final plot, showing the convergence of the histogram and the quantile-quantile plot



2.3 Experiment #3 $\beta = 2, t = 2, D = 3, n = 20$

Again, as described at the start of experiment 2 above, this is a subset of the first experiment. We will show the final plot only to show how close to the standard normal the final position histogram is.



2.4 Experiment #4 $\beta = 2, t = 2, D = 3,\ n = 7000$

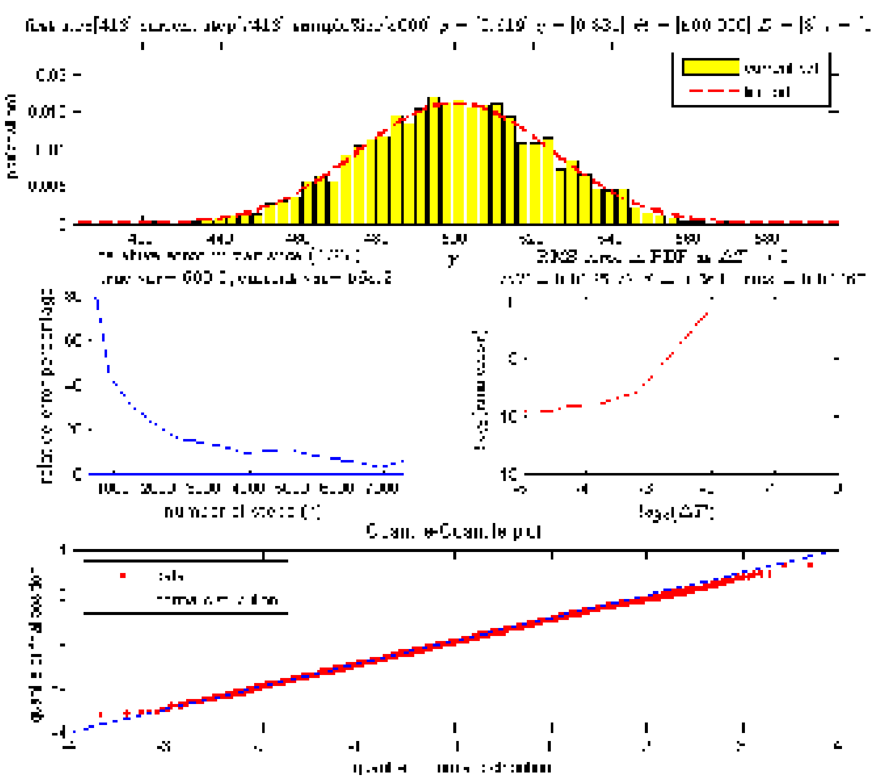
The following 2 experiments are not required to do, but they are extra experiments I already done and included here.

$$\text{starting step number} = 400,\ \beta = 5, t = 100, D = 3, final\ p = 0.623, final\ q = .377$$

$$\text{sample size 5000, number of bins 60, seed} = 123456$$

$$\text{final } \Delta x = 0.2945 \text{ final } \Delta t = 0.0145$$

Experiment number	n (number of steps)	Variance	True variance (2Dt)	Δt
1	400	1.89	600	0.2392
2	900	340	600	0.1089
3	1400	420	600	0.0705
4	1900	464	600	0.0521
5	2400	504	600	0.0414
6	2900	514	600	0.0343
7	3400	525	600	0.0293
8	3900	546	600	0.0255
9	4400	536	600	0.0226
10	4900	533	600	0.0203
11	5400	552	600	0.0185
12	5900	558	600	0.0169
13	6400	567	600	0.0156
14	6900	583	600	0.0145



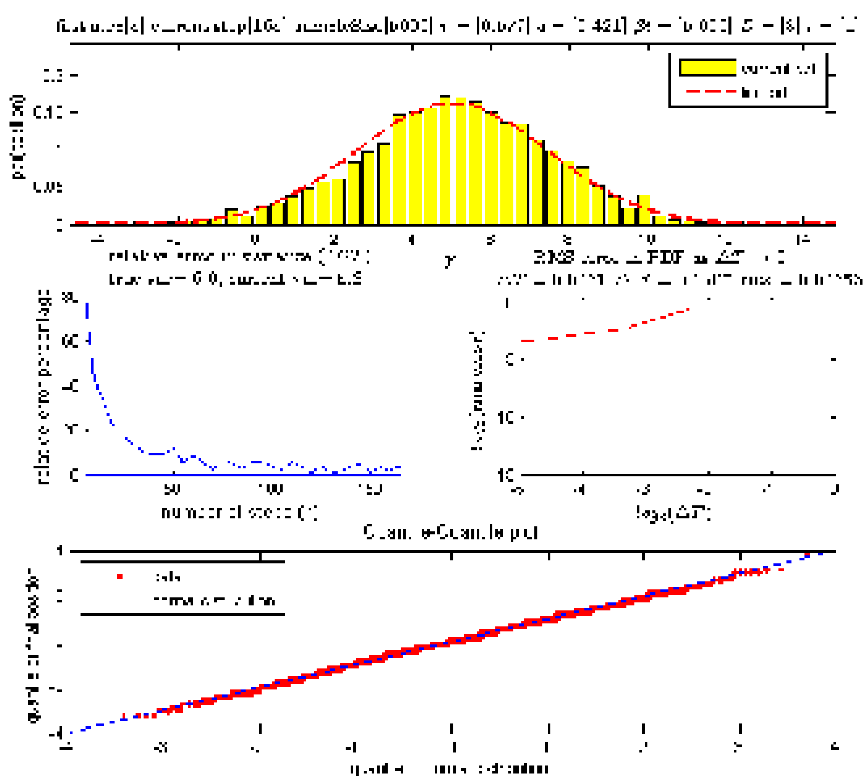
2.5 Experiment #5 $n = 160, \beta = 5, t = 1, D = 3$

starting step number$= 5, \beta = 5, t = 1, D = 3, final\ p = 0.579, final\ q = 0.421$

sample size 5000, number of bins 50, seed$= 123456$

final $\Delta x = 0.1907$, final $\Delta t = 0.0061$

Experiment number	n (number of steps)	Variance	True variance ($2Dt$)	Δt
1	5	1.019	6	0.2
2	10	3.4	6	0.1
3	15	4.09	6	0.0667
4	20	4.74	6	0.05
5	25	5	6	0.4
6	30	5.18	6	0.0333
7	35	5.43	6	0.0286
8	40	5.466	6	0.0250
9	45	5.3	6	0.0222
10	50	5.66	6	0.02
11	55	5.4	6	0.0182
12	60	5.85	6	0.0167
...	...	...	...	...
31	150	5.78	6	0.0065
32	155	5.909	6	0.0063
33	160	5.75	6	0.0061



3 Discussion of numerical results

From the above tables we observe that as Δt becomes smaller, the variance of the sample of the final position becomes closer to the variance predicted by the model which is $2Dt$.

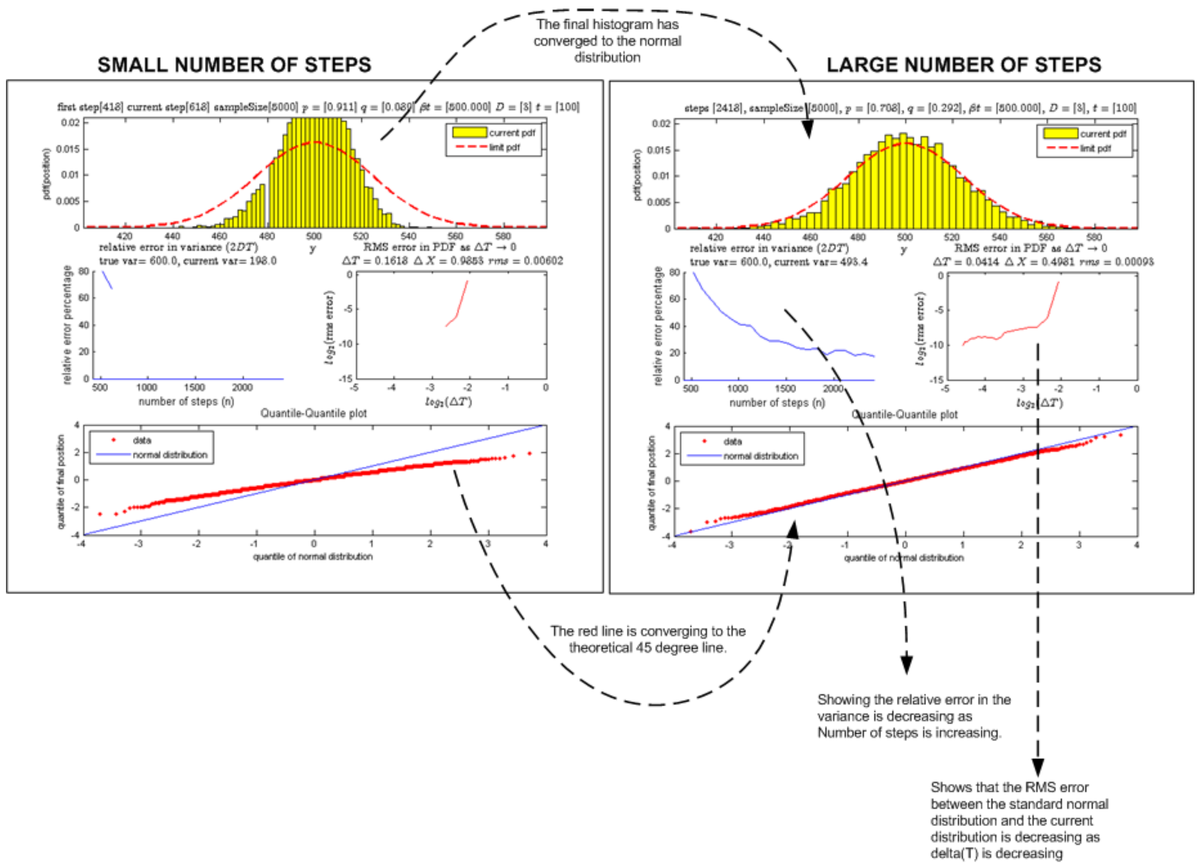
The mean remains the same which is βt.

We observe that if the total walk time is large (experiment #4) , then more steps are needed to bring Δt to be small enough so that the variance becomes close to $2Dt$.. This answers the second question we are set to solve in this project which is Does the variance of the above distribution converges, as $\Delta t \rightarrow 0$ and $\Delta x \rightarrow 0$ under the above mentioned condition of keeping $\frac{(\Delta x)^2}{\Delta t}$, to the analytical variance of $2Dt$ and the theoretical mean of βt?

Now to answer the first question of convergence of the histogram of the final position to the normal.

Looking at the quantile plots we observe that as more steps are used (hence smaller Δt and smaller Δx) then the quantile-quantile plot was tilting closer and closer to the straight line at 45^0 which would be the case when we plot the quantile of 2 data sets coming from the same distribution. This concludes that the final distribution of the random walk position converges to normal distribution with the above parameters.

The following diagram below shows a run where on the left side there is a plot showing the quantile plot when the number of steps is small. The plot on the right side shows the quantile plot at the end of the run when n was large. We see that the quantile plot line is now almost exactly over the 45^0 line, confirming that the data is coming from normal distribution.



Therefore, we have answered the 2 questions this simulation was designed to answer.

3.1 Final observation

In doing the above experiments, it was observed that the relative error in the variance of the final position as n increased does approach the true variance $2Dt$ but the convergence is not smooth. As the relative error (around 5% to 10%), then increasing n more can cause the error to sometimes increase and not decrease as one would expect. Meaning the relative error is not monotonic decreasing as n increases. However, as n becomes very large, the trend is for the relative error is to decrease. I can only contribute this behavior to some sort of statistical error. This needs to be investigated more.

4 Source code listing