

关于随机行走和扩散过程的研究和模拟

陈亦林

(武汉大学 物理科学与技术学院 15 级弘毅班, 2015301020152)

摘 要: 本次实验中主要是通过 Python 语言编程对随机行走过程和扩散过程进行了研究和模拟。考虑和模拟了一维和二维的随机行走的具体过程和粒子与原点距离的平均值, 以及距离的平方的平均值与步长的关系。并且讨论了相同和不同的方向概率对其的影响。对于扩散过程, 通过数学推导, 得出了多维扩散方程的形式, 以及各维度中粒子距离原点的距离与时间(步数)的关系满足高斯分布的结论, 并且对其中的具体影响因素进行了讨论。最后还根据书上的公式, 简单的讨论了一下二维扩散体系的熵随时间的变化。

关键词: 随机行走; 扩散过程; 高斯分布; 熵; Python

Study and Simulation of Random Walk and Diffusion

Yilin Chen

(Hongyi Class, School of Physics and Technology, Wuhan University, 2015301020152)

Abstract: The random walk and diffusion are studied and simulated mainly through Python in this experiment. I have considered and simulated the specific process of one dimensional and two-dimensional random walk, the mean distance between particles and the origin, and the relationship between the mean value of distance and the step numbers. For the diffusion process, the form of multidimensional diffusion equation is obtained by mathematical deduction, and the relationship between the distance from the origin and the time (step number) in each dimension satisfies the Gauss Distribution, and the specific influencing factors are discussed. Finally, according to the formula in the book, we simply discuss the change of entropy with time in the two-dimensional diffusion system.

Key words: Random Walk; Diffusion; Gaussian Distribution; Entropy; Python

一. 引言

英国植物学家布朗(Robert Brown)发现浸没在液体中的花粉颗粒做无规则的运动, 此现象后被命名为布朗运动, 一个基于随机行走概念的数学理论即将出现。爱因斯坦(Albert Einstein)于 1905 年解释了布朗运动的原因, 认为花粉粒子受到周围介质分子撞击的不均匀性造成了布朗运动。

“随机行走”最初由 Karl Pearson 在 1905 年提出。从数学角度来看, 布朗运动是一个随机行走过程(Random Walk)。具体的说, 是连续时间、连续状态空间的马尔科夫过程。而随机行走过程在其它许多领域有着十分重要的应用。而经典的扩散过程(Diffusion), 包括扩散方程在内, 也是可以基于随机行走理论导出的。

二. 随机行走

2.1 基本模型

以一维情况为例，假设一个人在 x 轴上作一维的随机行走，初始位置在原点。他只有两个运动方向：向左或向右。并且设其每次移动步长为 1， s_i 是一个随机变量，并且令：

$$s_i = \begin{cases} 1 & \text{第 } i \text{ 步向右行走} \\ -1 & \text{第 } i \text{ 步向左行走} \end{cases} \quad (i = 1, 2, 3 \dots) \quad (1)$$

且第 i 此运动向左和向右的概率分别为：

$$P(s_i = 1) = p, P(s_i = -1) = q \quad (2)$$

显然 $p + q = 1$ ，该人第 n 次移动后所在的位置设为 x_n ，则有：

$$x_n = \sum_{i=1}^n s_i \quad (3)$$

该人经历 n 次移动后，其距离原点的平均距离为 $\langle x \rangle$ ，则有：

$$\langle x \rangle = E(x_n) = E\left(\sum_{i=1}^n s_i\right) = \sum_{i=1}^n E(s_i) = (p - q)n = (2p - 1)n \quad (4)$$

即 $\langle x \rangle$ 与 n 成正比。

该人经历 n 次移动后，其距离原点的平均距离为 $\langle x^2 \rangle$ ，则有：

$$\langle x^2 \rangle = E(x_n^2) = (2p - 1)^2 n^2 + 4p(1 - p)n \quad (5)$$

可以由上式看出：当 $p=q=0.5$ 时， $\langle x^2 \rangle = n$ ；当 p 与 q 不相等时， $\langle x^2 \rangle$ 是 n 的二次函数。

2.2 一维随机行走

2.2-1 一维随机行走，向左向右等概率时

对于一维随机行走，当 $p=q=0.5$ 时，由于设置步长为 1，因此用步数代指时间。

首先做出三次随机行走的 x （人与原点距离）与 t （时间/步数）的关系图，如下：

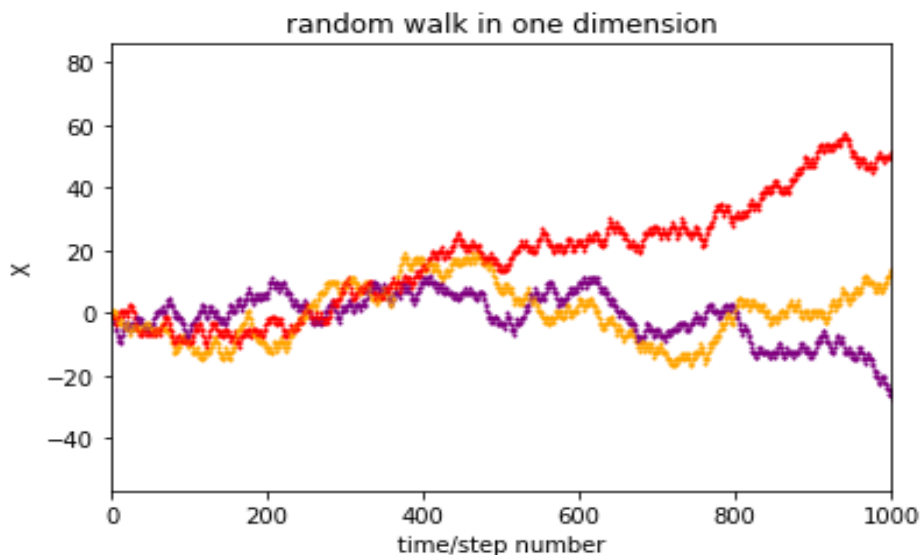


图 1 一维随机行走 $x-t$ 图

由上图可以看出，这三次完全独立随机行走，人距离原点的距离与步数的关系是完全杂乱无章，看似没有规律的。

接下来考察多个人的情况，增加行走的人相当于对于一个人增加其随机行走的次数。其与原点的距离的平均值在不同实验次数下与步数的关系图：

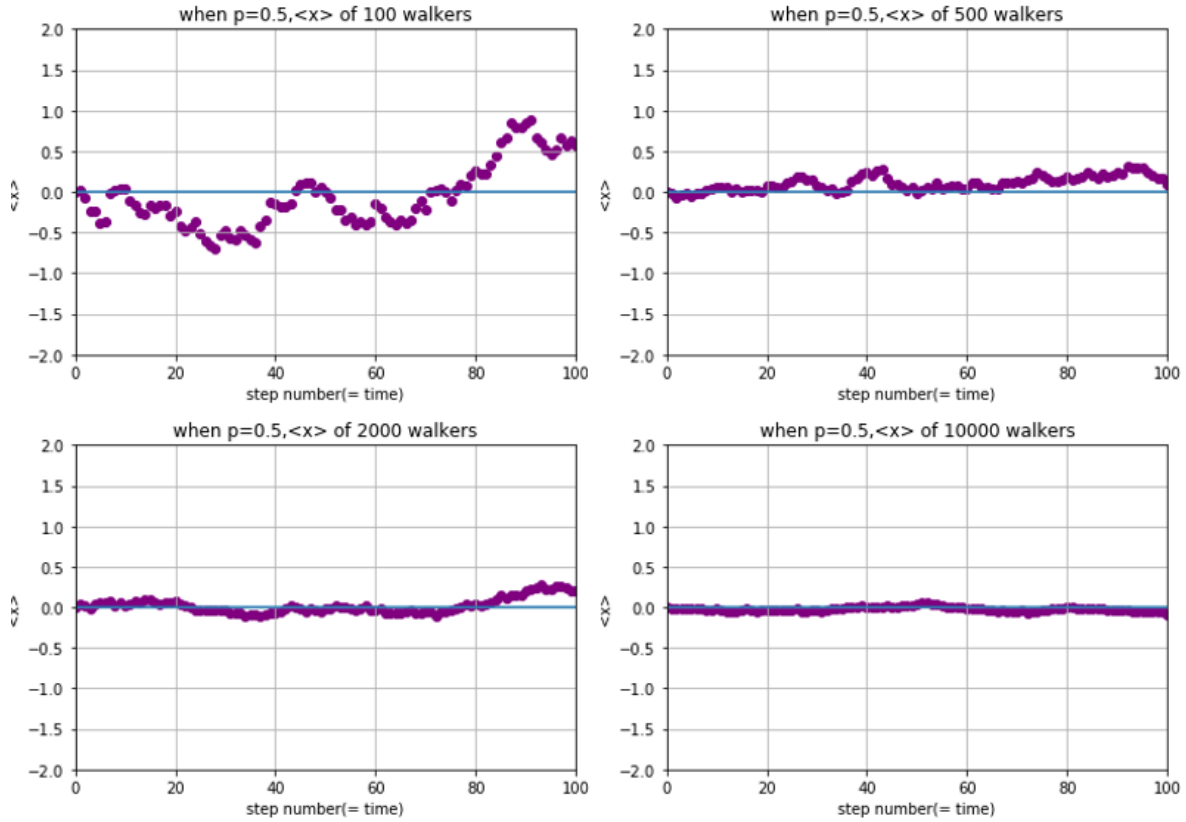


图2 一维随机行走 $\langle x \rangle$ -t 图

上图中分别考察了 100 个，500 个，2000 个，10000 个人进行相互独立的一维随机行走，其与原点的平均距离与步数的关系。根据之前的理论推导，其理论值为：

$$\langle x \rangle = (2p - 1)n$$

当 $p=0.5$ 时

$$\langle x \rangle = 0$$

即其与原点的平均距离在多次实验下，应该在 0 附近小幅度变动，无论步数是多少。

上面四条曲线表面，随着试验次数的不断增加， $\langle x \rangle$ 的值逐渐趋于 0。其从 100 次的偏离很大，到 10000 次时基本在 0 附近做很小的偏离。很好的符合了理论预测。

接下来继续考虑多个人的情况，其距离原点的距离的平方的平均值在不同步数下的关系图：

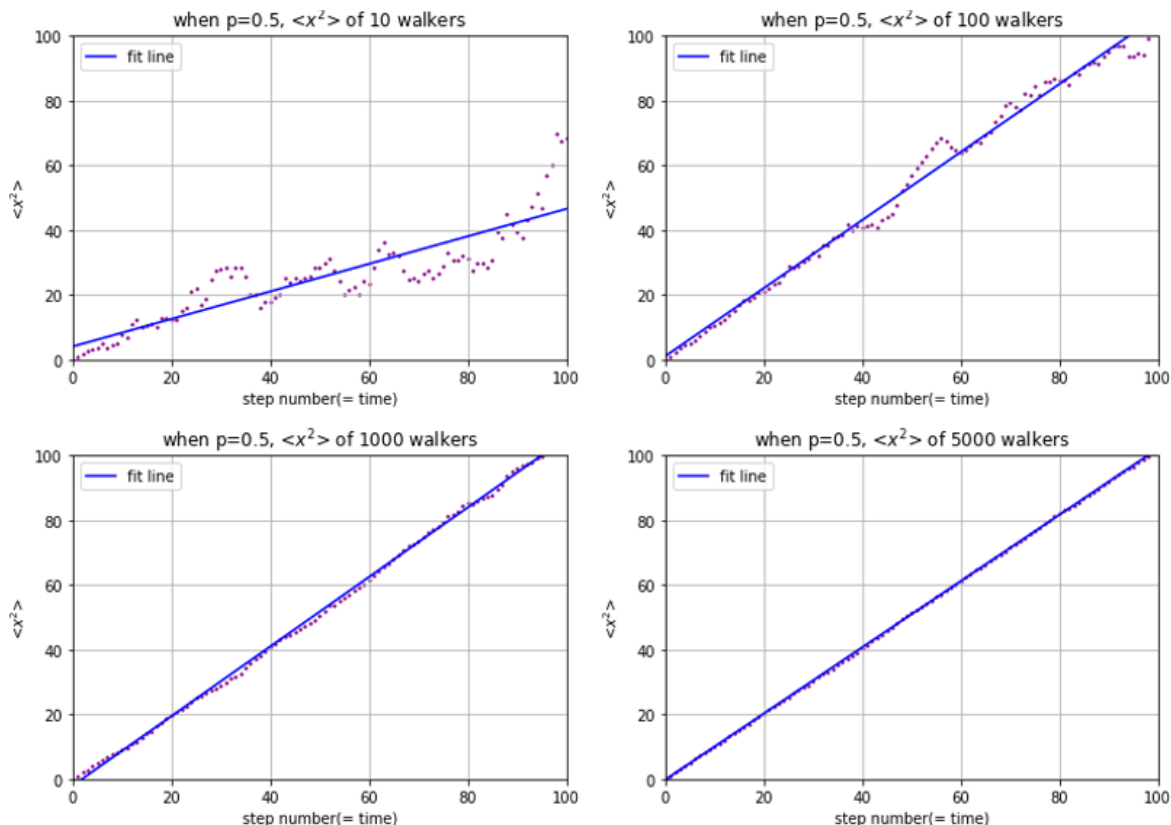


图 3 一维随机行走 $\langle x^2 \rangle$ -t 图

同理，上图中分别考察了 10 个，100 个，1000 个，5000 个人进行相互独立的一维随机行走，其与原点的距离的平均值与步数的关系。

根据之前的理论推导，其理论值为：

$$\langle x^2 \rangle = (2p - 1)^2 n^2 + 4p(1 - p)n$$

当 $p=0.5$ 时

$$\langle x^2 \rangle = n$$

即其与原点的距离的平方的平均与步数在多次实验下，应该成线性关系。

上面四条曲线表面，随着试验次数的不断增加， $\langle x \rangle$ 的指逐渐趋于 n 。其从 10 次的偏离很大，到 5000 次时基本在 n 附近做很小的偏离，也很好的符合了理论预测。

2.2-2 一维随机行走，向左向右不等概率时

对于一维随机漫走，当向左和向右的概率 p 与 q 不相等时，我们考察一下 $\langle x \rangle$ 与 $\langle x^2 \rangle$ 的变化情况，应该与之前 $p=q=0.5$ 有所不同，本质上来说，后者只是前者的一个特殊情况。

首先，我们做出一维随机漫走时，不同 p 情况下，人与原点的平均距离与步数的关系图，为了明显的贴合理论曲线，我们选择 1000 次实验的综合结果：

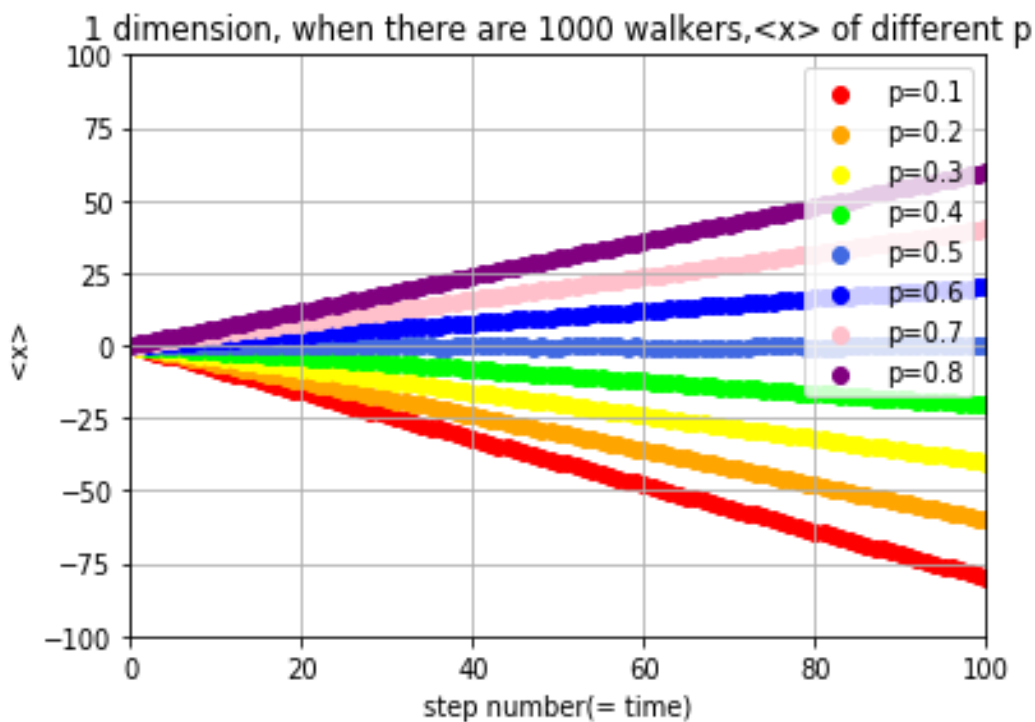


图4 一维随机行走,不同 p 情况下, $\langle x \rangle$ - t 图(1000 walkers)

根据与之前相同的理论推导, 当 p 不同时

$$\langle x \rangle = (2p - 1)n$$

可以看出, $\langle x \rangle$ 与 n 依然满足线性关系, 而在 1000 次实验综合下, 图中 $\langle x \rangle$ 的确只在 n 附近做很微小的震荡, 线性关系符合较好。

且对于和为 1 的 p 的两种情况, $\langle x \rangle$ 与 n 的系数互为相反数, 这一点在曲线上体现为两条斜率相反的过原点的直线, 从图中 $p=0.2$ 与 $p=0.8$ 等曲线看出, 这一点也和理论符合的很好。

接下来考虑一维随机漫走时, 不同 p 情况下, 人与原点的距离的平方与步数的关系图, 为了看得更加明显, 我们选择 5000 次实验的综合结果:

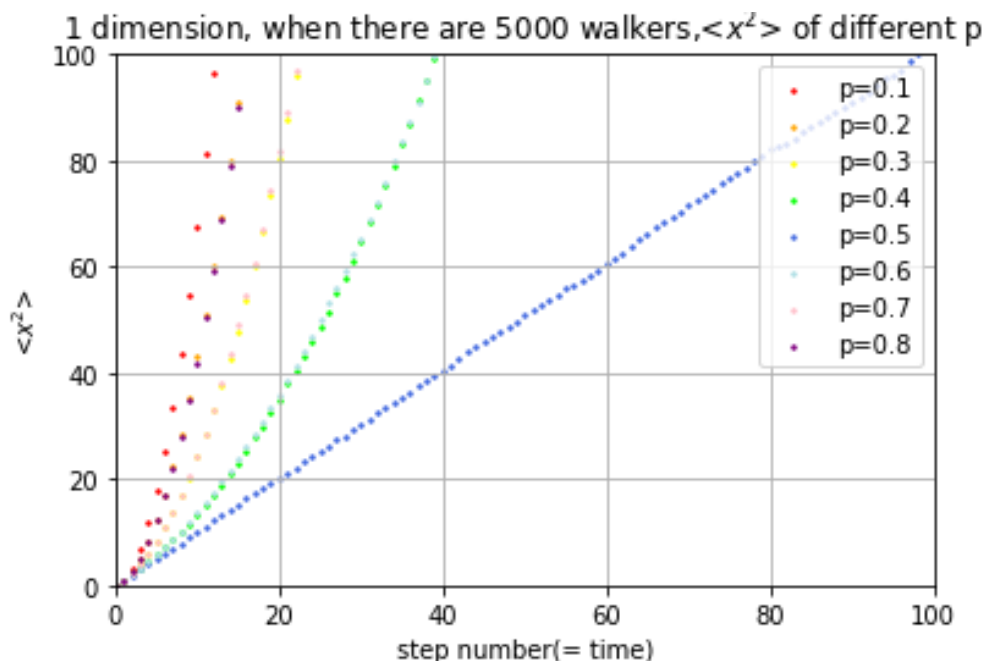


图5 一维随机行走,不同 p 情况下, $\langle x^2 \rangle$ - t 图 (5000 walkers)

根据与之前相同的理论推导，当 p 不同时

$$\langle x^2 \rangle = (2p - 1)^2 n^2 + 4p(1 - p)n$$

从上面图中明显看出，除了 $p=0.5$ 以外，其余的 $\langle x^2 \rangle$ 与 n 都不再是线性关系，而是二次函数关系，这点图中曲线符合的很好。

同时根据上式还可以看出，且对于和为 1 的 p 的两种情况， $\langle x^2 \rangle$ 与 n 的关系应该是相同的，而图中 $p=0.2$ 与 $p=0.8$ ， $p=0.3$ 与 $p=0.7$ ， $p=0.4$ 与 $p=0.6$ 的模拟的点所在曲线也基本重合。这也和理论符合的很好。

2.3 二维随机行走，以 $p=0.5$ 为例

根据理论易知， $\langle r \rangle$, $\langle r^2 \rangle$ 与 n 的关系，二维随机行走相较于一维随机漫走，并没有本质上的差别。做出三次随机行走的 r （人与原点距离）与 t （时间/步数）的关系图，如下：

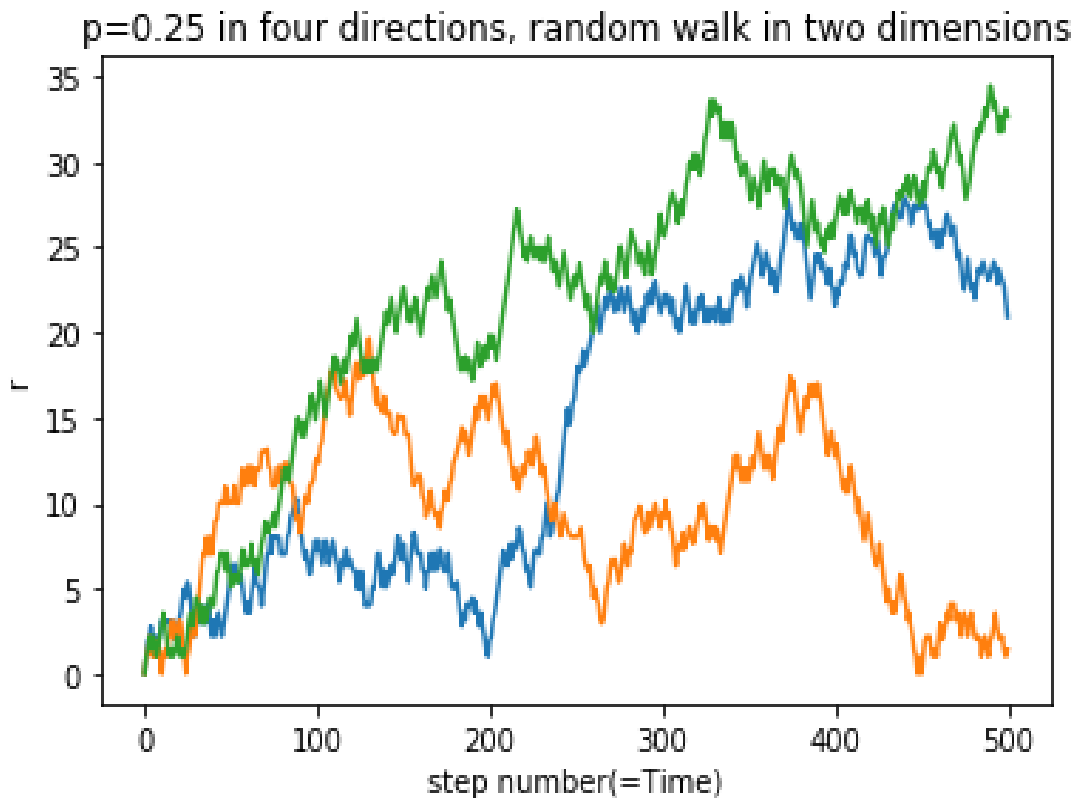


图 6 二维随机行走 x - t 图

同理考察多个人的情况，其距离原点的距离的平均值在不同实验次数下与步数的关系图：

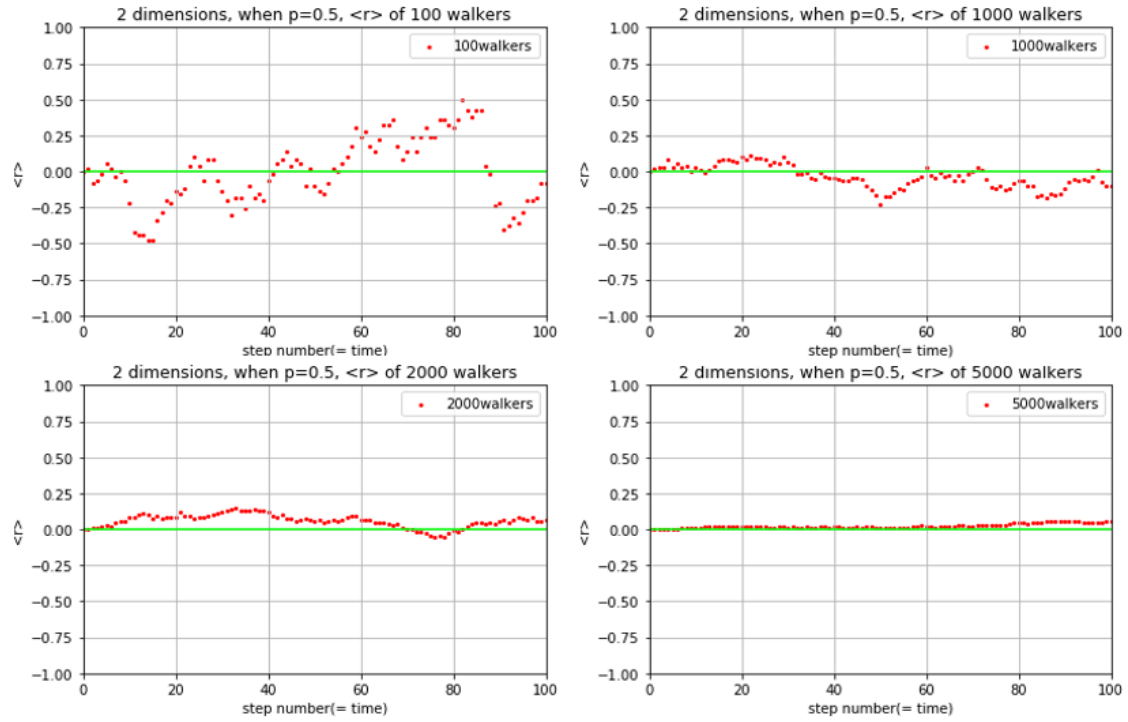


图 7 二维随机行走 $\langle x \rangle$ -t 图

同理，下面考察了不同试验次数下，人与原点的距离的平均值与步数的关系：

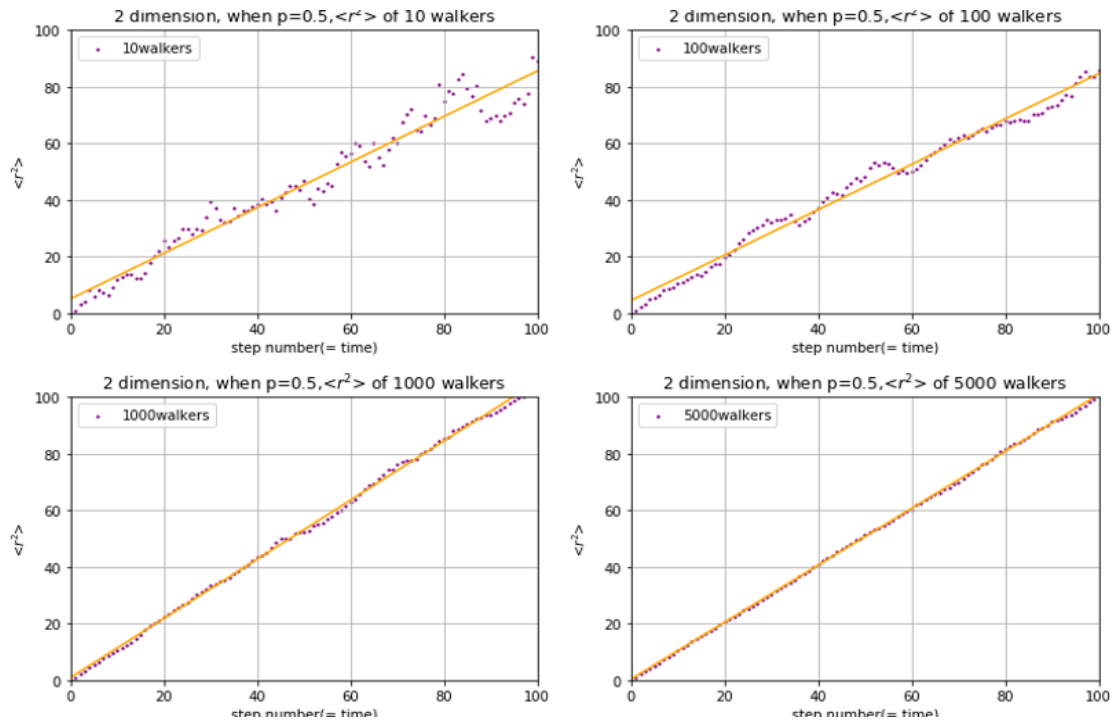


图 8 二维随机行走 $\langle x^2 \rangle$ -t 图

这些关系与一维没有本质差异，就不赘述了，二维的主要意义是引出下文要讨论的扩散。

三. 扩散过程

3.1 多维扩散过程的基本原理

扩散过程本质上可以由随机漫走模型推演过来,下面以多维空间的普适情况来研究这两者的联系,并且导出扩散方程。

考虑一个随机行走粒子,初始点位于 s 维空间的原点,设每一步粒子移动量为 Δx_N ,该量由概率分布 $p_N(\mathbf{r})$ 决定,这些步都是相互独立且等概率的,因此有

$$p_N(\mathbf{r}) = p(\mathbf{r}) \quad (6)$$

且假设这些步的同方向性,所以 $p(\mathbf{r})$ 仅仅是径向距离 r 的函数,这种条件通常也消除偏移,则:

$$\langle \Delta x_N \rangle = 0 \quad (7)$$

令 x_N 表示 N 步后粒子的位置, $P_N(\mathbf{R})$ 为联合概率密度函数,对于满足相互独立等概率的分布,有以下递归式:

$$P_{N+1}(\mathbf{R}) = \int p(\mathbf{r}) P_N(\mathbf{R} - \mathbf{r}) d^s \mathbf{r} \quad (8)$$

上式中,一维时 $s=1$,二维时 $s=2$,以此类推,这就是 Bachelier 方程。

下面为了求解连续解,当 $N \rightarrow \infty$ 时,在上式积分中应用 Taylor 展开式,得到:

$$\begin{aligned} P_{N+1}(\mathbf{R}) &= \int p(\mathbf{r}) \left[P_N(\mathbf{R}) - \mathbf{r} \cdot \nabla P_N(\mathbf{R}) + \frac{1}{2} \mathbf{r} \cdot \nabla \nabla P_N \cdot \mathbf{r} + \dots \right] d^s \mathbf{r} \\ &= P_N(\mathbf{R}) - 0 + \frac{1}{2} \sum_i \sum_j \langle r_i r_j \rangle \frac{\partial^2 P_N}{\partial R_i \partial R_j} + \dots \\ &= P_N(\mathbf{R}) - 0 + \frac{\langle \mathbf{r} \cdot \mathbf{r} \rangle}{2} \nabla^2 P_N(\mathbf{R}) + \dots \end{aligned} \quad (9)$$

假设时间步长为 Δt ,并且定义时间为 $t=N\Delta t$,当 $N \rightarrow \infty$ 时,极限分布 $\rho(\mathbf{R}, t)$ 的定义为:

$$P_N(\mathbf{R}) = \rho(\mathbf{R}, t) \quad (10)$$

上式满足:

$$\frac{\partial \rho}{\partial t} = D \nabla^2 \rho \quad (11)$$

其中, D 为扩散函数:

$$D = \frac{\langle r^2 \rangle}{2d\Delta t} \quad (12)$$

注意初始条件:粒子开始于原点,即:

$$\rho(\mathbf{R}, t) = \delta(\mathbf{R}) \quad (13)$$

接着利用傅里叶变换:

$$\rho(\mathbf{R}, t) = \frac{1}{(2\pi)^s} \int e^{-i\mathbf{k} \cdot \mathbf{R}} \hat{\rho}(\mathbf{k}, t) d^s \mathbf{k} \quad (14)$$

得到如下微分方程：

$$\frac{\partial \hat{\rho}}{\partial t} = -Dk^2 \hat{\rho} \quad (15)$$

上式解为：

$$\hat{\rho}(\mathbf{k}, t) = e^{-Dk^2 t} \hat{\rho}(\mathbf{k}, 0) = e^{-Dk^2 t} \quad (16)$$

作傅里叶逆变换，得到：

$$\rho(\mathbf{R}, t) = \frac{e^{-\frac{R^2}{4Dt}}}{(4\pi Dt)^{\frac{s}{2}}} \quad (17)$$

在对应的离散问题中，有：

$$P_N(\mathbf{R}, t) \sim \frac{\rho^{-\frac{sR^2}{2\langle r^2 \rangle N}}}{\left(\frac{2\pi\langle r^2 \rangle N}{st}\right)^{\frac{s}{2}}} \quad (18)$$

这就是 s 维时，对于一个各向同性随机行走 $P_N(\mathbf{R})$ 的长时间统计结果，趋近于高斯分布。

对于一维运动有 $s=1$ ，上式化为为：

$$P_N(\mathbf{R}, t) \sim \frac{2R^2}{N\sqrt{t}} \rho^{-\frac{R^2}{a^2 Nt}} \quad (19)$$

这就是高斯分布，二维也满足。说明当 N 足够大时，粒子与原点的距离的概率分布，应该趋近于高斯分布。

3.2 一维扩散过程

根据之前的推导，一维扩散过程中，当 N 足够大时，粒子与原点的距离的概率分布，应该趋近于高斯分布：

$$P_N(\mathbf{R}, t) \sim \frac{2R^2}{N\sqrt{t}} \rho^{-\frac{R^2}{a^2 Nt}}$$

但是这是各向同性的情况，也就是向左向右概率相等，而当其不相等时，易知粒子与原点距离的概率分布仍然趋近于高斯分布，但是应该会有些变化，如下图：

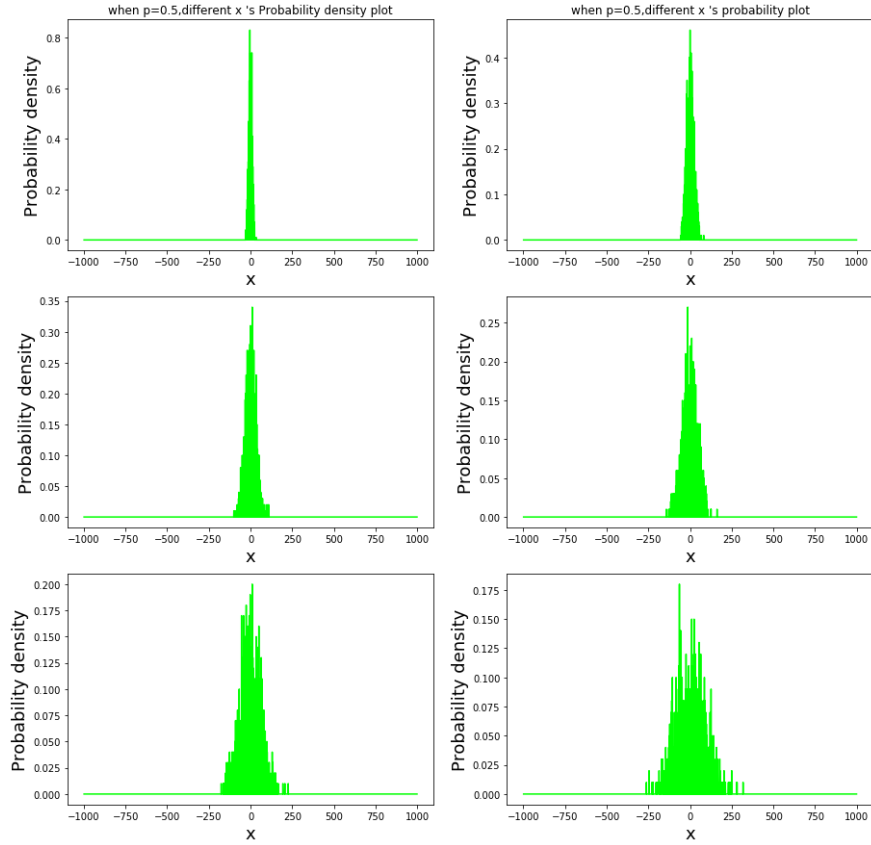


图 9 一维扩散过程， $p=0.5$ x-t 图

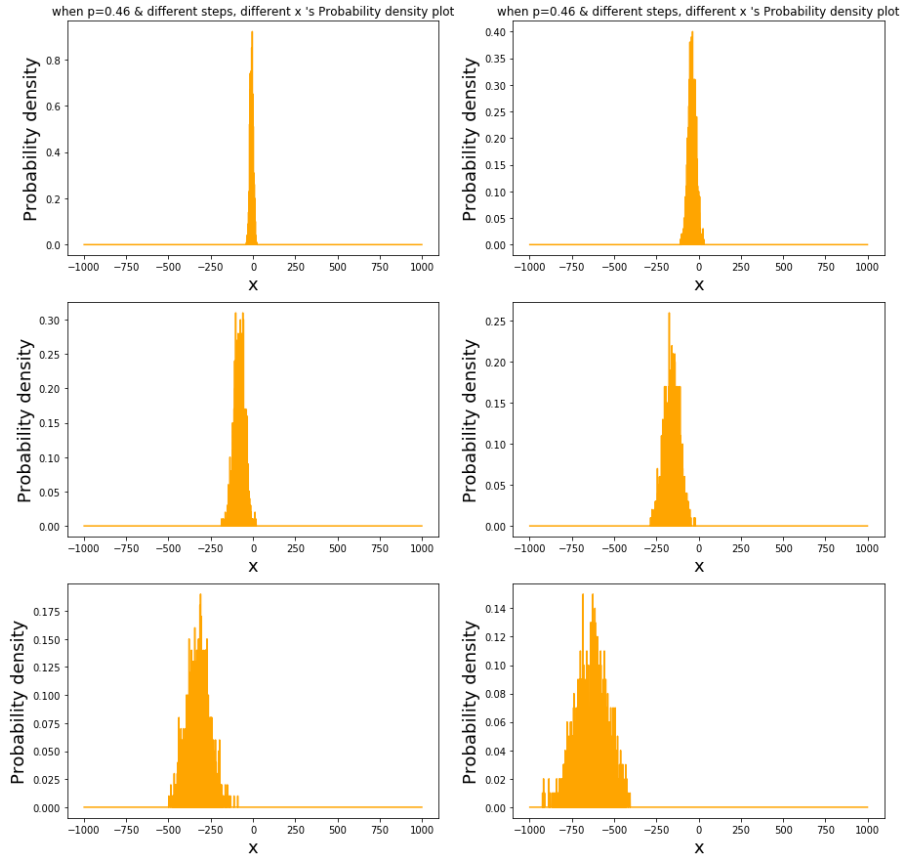


图 10 一维扩散过程， $p=0.46$ x-t 图

为了获得更直观的理解，我们把不同 steps 的拟合曲线以下面这种方式展示出来：

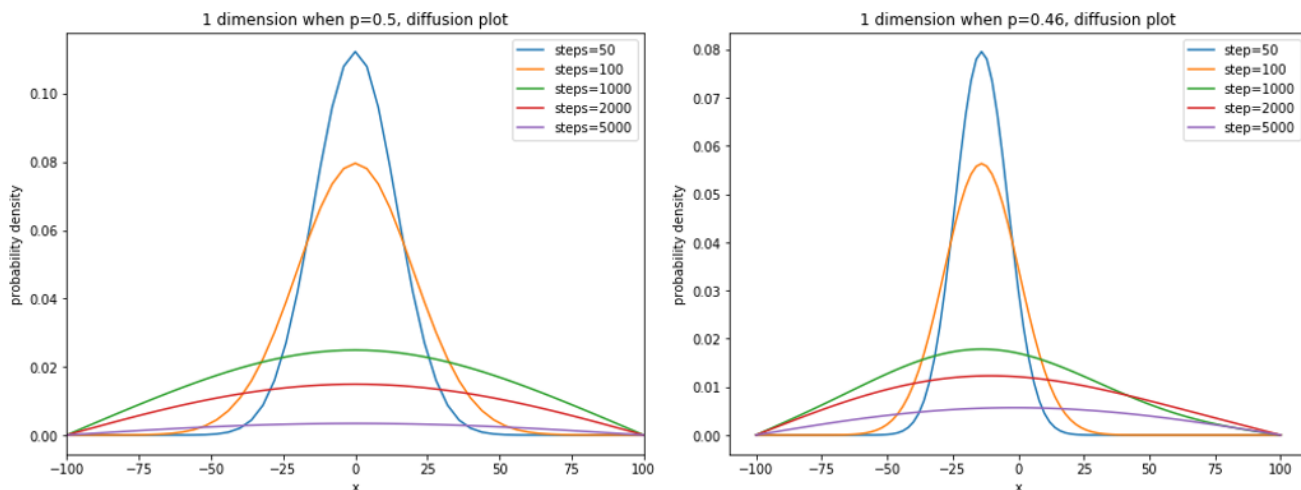


图 11 一维扩散过程， $p=0.5$ & $p=0.46$ x-t 图

从上图可以得出三个基本结论：

1. 进行足够多的实验后，粒子与原点的距离的概率分布，的确趋近于高斯分布。
2. 随着步数的增加，也就是 N 的增大，近似的高斯分布的峰值减小，全区域内的积分总量不变，相当于扩散过程的进行， N 增大，粒子扩散越散，但是仍然近似符合高斯分布。且该关系符合之前理论推导时， N 增大， $P_N(\mathbf{R})$ 峰值减小的结论。
3. 改变 p ，也就是向左和向右的概率大小，不影响该高斯分布函数的形状，但是会使得峰值位置发生偏离。左右概率相等时，峰值处于原点；不等时，峰值偏离原点。

3.3 二维扩散过程

二维扩散过程可以看成两个一维扩散过程的综合，其 x , y 方向的就是一维扩散过程，同样粒子与原点的分量距离趋近于高斯分布，其余各结论也十分相似。

同理，我们也考虑两种情况，一种是 x, y 方向的扩散系数相等，一种是不相等。前者相当于一个团簇在各向同性的介质中扩散，一种相当于在各向异性的介质中扩散。

我们考虑书上所说的“融化的方糖”模型，我们有一块方糖，分别在一种 x, y 方向的扩散系数相等，一种是不相等的咖啡中将其融化，观察其扩散过程随时间的变化：

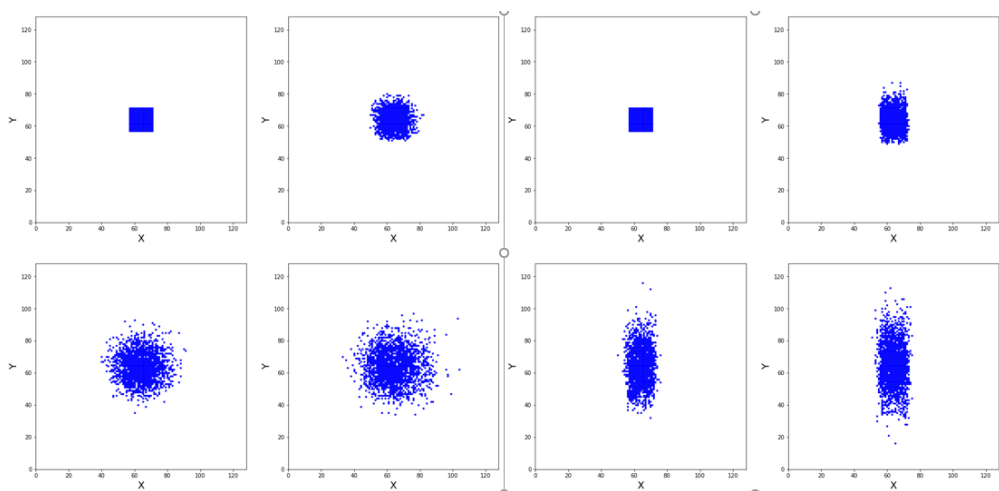


图 10 二维方糖扩散过程，各向同性与各向异性溶剂内扩散过程图

可以看出，各向同性溶剂中方糖最终扩散形状趋近于圆，而各向异性溶剂中方糖最终扩散形状趋近于椭圆，这与我们的理论预期是相符合的。

接着，我们观察其在 x, y 方向分量的扩散分布，以各向同性溶剂中情况为例：

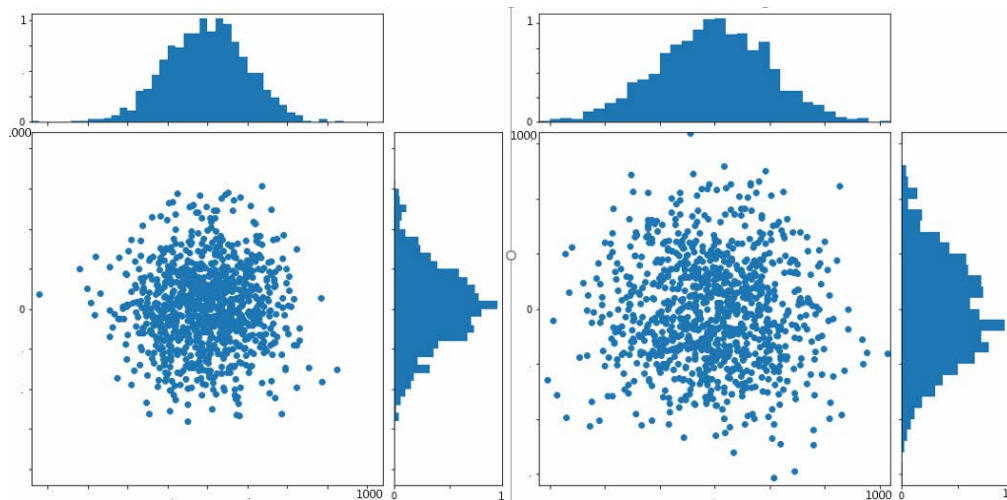


图 11 二维方糖在各向同性溶剂中的扩散过程，以及 x, y 分量图

上图展示了 steps 为 1000 时，二维扩散过程张中的两个时间的扩散情况，左图的时间在右图之前，可以很明显的看出，方糖随时间扩散变得更为松散，同时注意展示的 x, y 方向的概率分布，粒子与原点距离的概率分布，的确趋近于高斯分布，且左图的峰值比右图更高，并且更集中在远点区域，这与之前的分析相符合。

为了获得更直观的理解，我们将二位扩散过程在三维图像中表现出来， z 坐标是粒子距离原点距离的概率分布（注意具体数值，图中是为了看得清楚所以不同的图取了不同的坐标尺度）， x, y 轴是粒子的位置坐标，以各向同性的情况为例：

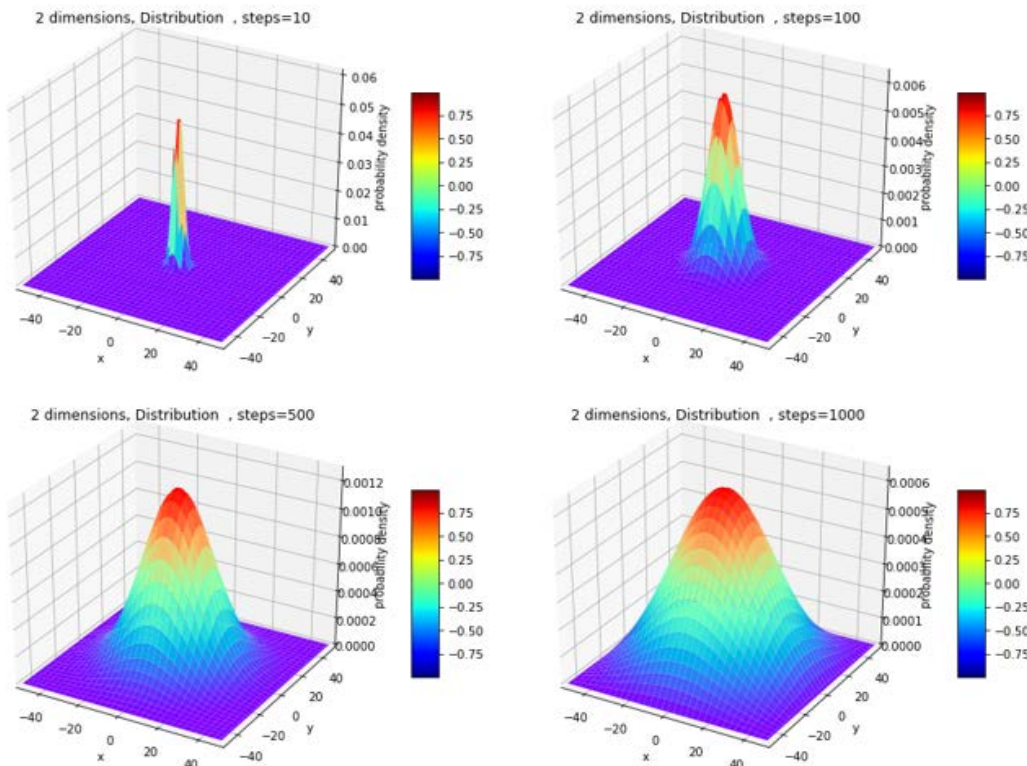


图 12 二维方糖在各向同性溶剂中的扩散过程，粒子与原点距离的概率分布三维图

从这里我们很直观的得到了和前面相同的结论，就不赘述了。

3.4 二维扩散过程中熵的变化

根据书上所给的公式，我们模拟 1000 个粒子的二维扩散过程，其熵为：

$$S = - \sum_i P_i \ln P_i \quad (20)$$

这里只如书上一样，简单的作一下体系的熵随时间（step numbers）的变化曲线图：

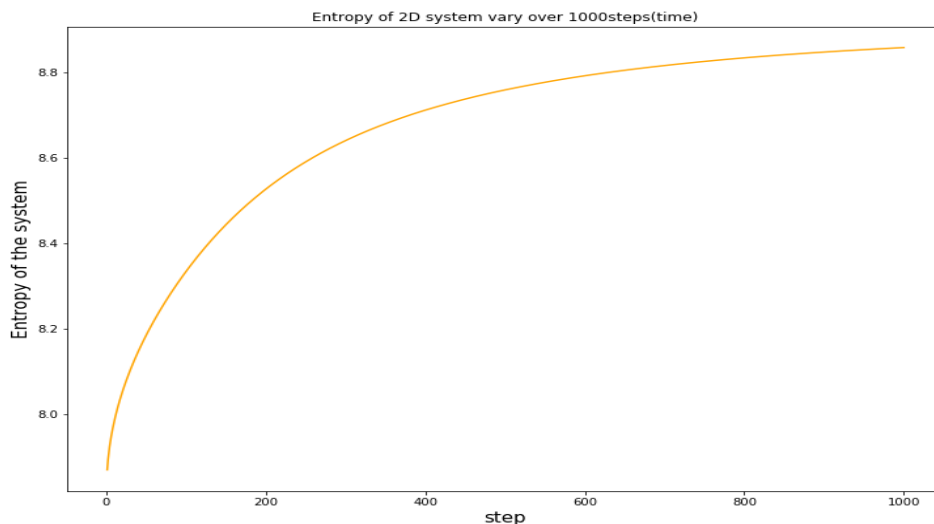


图 13 二维扩散过程体系的熵随时间的变化曲线图

可以明显的看出，体系的熵增一开始随时间的变化速率很快，之后曲线的斜率降低表明：体系的熵增一开始随时间的变化速率逐渐降低。

可以预见，随着时间的增加，体系的熵最终趋近于一个常数。体系的熵不再发生大的变化，即扩散过程基本结束，整个溶剂内几乎均匀地分散着被扩散的物质，体系达到平衡稳定的状态。

四. 总结

本次实验通过 Python 语言编程对随机行走过程和扩散过程进行了研究和模拟。我们考虑和模拟了一维和二维的随机行走的具体过程和粒子与原点的距离的平均值，以及距离的平方的平均值与步长的关系。得出了粒子与原点的距离的平均值，以及距离的平方的平均值与步长的具体表达式，模拟效果也很好的验证了理论。并且讨论了相同和不同的方向概率对其的影响。对于扩散过程，通过数学推导，得出了多维扩散方程的形式，以及各维度中粒子距离原点的距离与时间（步数）的关系满足高斯分布的结论，并且对其中的具体影响因素进行了讨论。最后还根据书上的公式，简单的讨论了一下二维扩散体系的熵随时间的变化。得出了随着时间的增加，体系的熵最终趋近于一个常数的结论。

参考文献

- [1] Nicholas J.Giordano Hisao Nakanishi, 计算物理（第二版），清华大学出版社，184-197.
- [2] 齐民友，刘禄勤，概率论与数理统计（第二版），高等教育出版社，132-135.
- [3] 李宏亮，焦潍莘，随机行走的扩散方程，陕西师范大学出版社，63-65.