

# PHYS235: Lecture Notes

Lecture 12 ~14

Lectures by Patrick H. Diamond

Notes by Chang-Chun Chen

University of California San Diego

## Contents

<b>1</b>	<b>Fractal Brownian Motion (FBM)</b>	<b>2</b>
1.1	Hurst Exponent . . . . .	2
1.2	Examples: Noah and Joseph effect (Mandelbrot & Wallis) . . . . .	3
1.3	How to calculate Hurst exponent from a time series? . . . . .	4
1.4	Gini Coefficient . . . . .	5
1.5	$1/f$ noise and Zipf's Law . . . . .	6
<b>2</b>	<b>Levy-Stable Distributions</b>	<b>8</b>
2.1	Generalized Central Limit Theorem . . . . .	8
2.2	Gaussian & Cauchy Distribution . . . . .	12
2.3	Direct evidence for Lévy flight and anomalous process . . . . .	12
<b>3</b>	<b>Continuous Time Random Walk (CTRW)</b>	<b>15</b>
3.1	Review of the Fokker-Plank Equation . . . . .	15
3.2	Why we need a CTRW? . . . . .	15
3.3	Waiting Time Model . . . . .	15
3.4	Velocity Model . . . . .	16
3.5	CTRW Summary . . . . .	17

# 1 Fractal Brownian Motion (FBM)

From previous lectures, we've discussed the fractals and the  $\beta$ -model. Since space intermittency has been discussed, now we consider Brownian time series  $\mathbf{B}(\mathbf{t})$ . We have discuss the multi-fractal structure in space with a given structure function  $\delta\mathbf{V} = \mathbf{v}(\mathbf{r} + \mathbf{l}) - \mathbf{v}(\mathbf{r})$ . Now, we are interested in the fractal in time with a given structure function in time.

## 1.1 Hurst Exponent

Harold Edwin Hurst, a British hydrologist, is the first one to investigate in the long-term storage capacity of reservoirs. His research on the fluctuations of the water level in the Nile river gave a **Hurst exponent** which has been used in finance, cardiology, and other fields. He proposed a generalization of Brownian time series by considering its expectation value of first and second order of time structure function

$$\mathbf{E}\{(B(t + \tau) - B(\tau))^2\} = \tau^1, \quad (1)$$

where the  $\mathbf{E}\{F\}$  is the expectation value of  $F$ , an arbitrary function. He generalized this equation by release the exponent from unity:

$$\mathbf{E}\{\delta B^2\} = \tau^{2H}, \quad 0 < H < 1, \quad (2)$$

where the  $H$  is the Hurst exponent (or Holder exponent). The reason that  $0 < H < 1$  is related to **Levy-stable distribution**. When  $H = 1$ , the system is governed by a Wiener process. We'll discuss Levy-stable distribution in section 2. This Hurst exponent is used as a measure of several properties–

- long-term memory of time series,
- dispersion, concentration (roughness) of data,
- and the anomalous diffusion.

The Hurst parameter can be generalized as

$$\mathbf{E}\{\delta B^q\} = \tau^{qH(q)}, \quad 0 < H < 1, \quad (3)$$

where  $q$  is the multi-scale,  $H(q)$  is the Hurst exponent with a continuous spectrum. This is for multi-fractal cases, where the expectation value of the  $q$ -th order  $\mathbf{E}\{\delta B^q\}$  is  $q$  dependent. The Hurst exponent has a relation to the box-counting dimension (or **fractal dimension  $D$** ), and can exhibit the multi-fractality:

$$\text{compare: } \begin{cases} H \sim \frac{\ln|\Delta B|}{\ln|\Delta t|} \\ D = \frac{\ln N(\epsilon)}{\ln(\frac{1}{\epsilon})}, \end{cases} \quad (4)$$

where  $\epsilon$  is the length of grids. Hence, the Hurst exponent is a counterpart of the fractal dimension for intermittency time series and it relates to **Fractal Brownian Motion (FBM)**. The relation between the Hurst exponent and the fractal dimension is

$$H + D = n + 1, \quad 0 < H < 1, \quad (5)$$

with  $n < D < n + 1$  in a  $n$  dimension space. For  $n = 1$ , we have  $H + D = 2$  with  $1 < D < 2$ . For more details about the Hurst exponent and fractal dimension please see Mandelbrot et al. (1984).

When  $H = \frac{1}{2}$ , the system dominated by Wiener Process. The expectation value of time series reduces back to the Brownian time series.

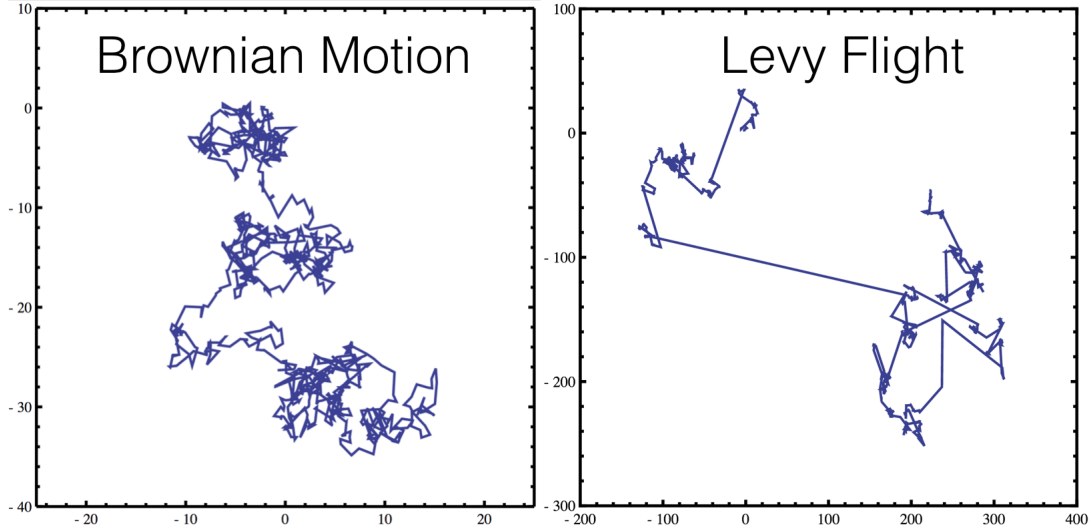


Figure 1: Brownian motion ( $H = 0.5$ ) and Levy flight ( $H > \frac{1}{2}$ ).

Table 1: Nonlinear Model Results

Case	$0 < H < \frac{1}{2}$	$\frac{1}{2} < H < 1$
memory	short-term persistence	long-term persistence
correlation	temporally anti-correlation	positive correlation
diffusion	sub-diffusion	super-diffusion
motion	sticking process	Levy flight, ballistic
example		traffic flow
variation	mild variation	wild variation

## 1.2 Examples: Noah and Joseph effect (Mandelbrot & Wallis)

The Hurst exponent describes persistence, correlation, and the diffusion of events. A system with a high Hurst exponent ( $H > \frac{1}{2}$ ) indicates that the events are temporally correlated, with higher persistence, long memory. This suggests that a high-level events persist for a long period of time, and then followed by the low events for another long period. Thus, the time series in this system is ‘erratic’. In a longer period, however, the series of event looks stationary. Hydrological data presents two forms of erratic behavior– the Joseph-erratic and the Noah-erratic Effect. The “Joseph Effect” describes this long period of precipitation:

*Seven years of great abundance are coming throughout the land of Egypt, but the seven years of famine will follow them.*  
(Genesis, 41, 29-30)

The “Noah Effect” describe that a extreme, concentrated precipitation is possible with infinite variance.

*On that day, all the springs of the greatest deep burst forth, and the floodgate of the heavens were opened. And the rain fell in the earth forty days and forty nights.*  
(Genesis, 7, 11-12)

These Joseph- and Noah-erratic effects can simultaneously exist in a process.

By contrast, a low Hurst exponent ( $H \lesssim \frac{1}{2}$ ) describes the system has a shorter memory so that the aftereffects die out in geometric progression and decrease rapidly, suggesting a *anti-correlated process*. A low Hurst exponent system undergoes a **sticky process**– the movements of

particles in space have a tendency to return back to a long-term mean. In terms of time series, the ‘leaps’ of the intensity in a time step won’t be too far away from the mean intensity. A classical short memory mechanism can be depicted by the Gauss-Markov process ( $H = 0.5$ ). The Gauss-Markov models, however, do not match the finding by H. E. Hurst– typically the Hurst parameter of precipitation  $0.7 \sim 0.85$ . More details can be found in Mandelbrot & Wallis (1968).

The figure 2 exhibits the variation for different Hurst exponents. Figure 2 (a) shows the anti-correlated property for low Hurst exponent  $H = 0.043$ . Here, the system barely has a persistent memory. While for high Hurst exponent  $H = 0.95$  system in figure 2 (c), the long-term memory and positive correlation has been shown. The variation is “wild”. Figure 2 (b) shows when the system is dominated by the Wiener process ( $H = 0.53$ ). The variance of time series reduce back to the Brownian time series which has the same properties with the “white noise”. The variation is “mild”.

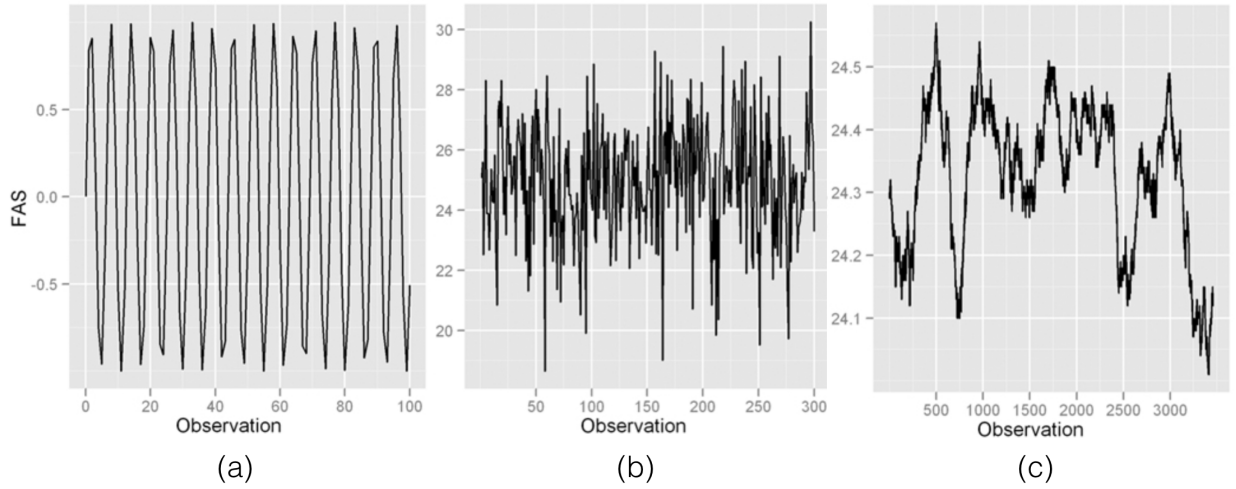


Figure 2: Variation of the time series with different Hurst parameters. (a)  $H = 0.043$ , (b)  $H = 0.53$ , (c)  $H = 0.95$ .

### 1.3 How to calculate Hurst exponent from a time series?

On the list of a recipe for the Hurst exponent, we have

- a standard deviation  $S$ ,
- and a range  $R$  of the cumulative deviation from mean

We first calculate the mean:

$$m = \frac{1}{n} \sum_{i=1}^n X_i, \quad (6)$$

where  $X_i = X_1, X_2, \dots, X_n$  is the data set in time series ( $t = t_1, t_2, \dots, t_n$ ),  $n$  is the time span of the observation (number of data points in a time series). Then, calculate the deviation from the mean for every time step:

$$Y_t = X_t - m, \quad t = 1, 2, \dots, n. \quad (7)$$

Then calculate the cumulative deviation from the mean:

$$Z_n = \sum_{i=1}^{t=n} Y_i, \quad (8)$$

where  $Z_n$  is accumulated from the beginning  $t = t_1$  to  $t = t_n$ . The **range of the cumulative deviation** could be found by calculate the  $R$ :

$$R(n) \equiv \max(Z_1, \dots, Z_n) - \min(Z_1, \dots, Z_n) \quad (9)$$

Also, the standard deviation  $S(n)$  is defined as

$$S(n) = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - m)^2}. \quad (10)$$

The Hurst exponent then can be expressed as:

$$H = \frac{\ln \mathbf{E} \left\{ \frac{R(n)}{S(n)} \right\}}{\ln(n)}, \quad (11)$$

where  $n$  indicates the time series ( $t = t_1, t_2, \dots, t_n$ ). Or alternatively,

$$Cn^H = \mathbf{E} \left\{ \frac{R(n)}{S(n)} \right\}, \quad (12)$$

where  $C$  is an arbitrary constant.

#### 1.4 Gini Coefficient

The range  $R(n)$  is actually related to Gini coefficient ( $G$ ), or **Gini index**, developed by the Italian statistician and sociologist Corrado Gini. This Gini index is a measurement of statistical dispersion of the income or wealth distribution of a nation's residents, and is the most commonly used measurement of inequality. The form of Gini index is

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n |x_i - x_j|}{2n \sum_{i=1}^n x_i}, \quad (13)$$

where  $n$  is the total population, and  $x_i$  is the wealth or income of person  $i$ . Notice that the denominator is a the total wealth of the population times  $2n$ , which has area  $A + B$ , and the numerator has area equal to  $A$  (see figure 3). Therefore, the Gini coefficient can be rewritten as

$$G = \frac{\text{area}(A)}{\text{area}(A + B)}.$$

Comparing the elements of the Gini coefficient and the Hurst exponent, one can realize that the "accumulated mean" in the Hurst calculation resembles the "perfect distribution line" where every one has a "mean" wealth. If we, in Hurst exponent calculation, sort  $X_i$  from low to high and assume the  $X_i$  are all positive, then  $Y_t$  can be resembled to the difference between the perfect distribution (the accumulated mean) and the real distribution curve (the accumulated distribution of  $X_i$ , see figure 3). Thus, the accumulated deviation for all population  $Z_{n=all}$  will be the area  $A$ . Since we sorted the "deviation"  $Y_t$  from low to high, we have  $Z_1 \leq Z_2 \leq \dots \leq Z_n$ , which gives  $\max(Z_1, \dots, Z_n) = Z_n = A$  and  $\min(Z_1, \dots, Z_n) = Z_1 = Y_1 = X_1 - m$ . If we have large  $n$ , then we'll have  $A \gg (X_1 - m)$ . As a result, the range  $R(n)$  in Hurst exponent calculation can be approximated as  $A - X_1 + m \sim A$ :

$$\lim_{n \rightarrow \infty} R(n) \sim \text{area}(A) \propto \text{Gini coefficient}. \quad (14)$$

Thus, the range  $R(n)$  is related to the Gini coefficient. Cynical questions for Gini coefficient is what're the diffusion in population  $D(x)$  and time  $D(t)$ .

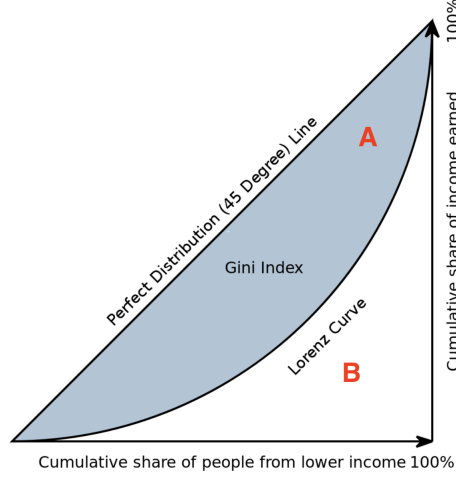


Figure 3: Gini index. Let shaded region has a area  $A$ , and the area under the distribution curve (here is the Lorenz Curve) is  $B$ .

### 1.5 $1/f$ noise and Zipf's Law

Consider the frequency spectrum for the generalized time series.

$$\langle B^2(\omega) \rangle = \int_t^{t+\tau} d\tau \langle B(t)B(t+\tau) \rangle. \quad (15)$$

We have:

$$\langle B^2(\omega) \rangle = \omega^{-\gamma} \quad (16)$$

where  $\gamma$  is defined as:

$$\gamma \equiv 2H - 1 = \begin{cases} 0, & H = \frac{1}{2}, \text{ white noise} \\ 1, & H = 1, \frac{1}{f} \text{ noise (or pink noise), "Joseph" effect.} \end{cases} \quad (17)$$

White noise is the generalized mean-square derivative of the Wiener process or Brownian motion. Notice that the "noise" here is nothing to do with the "thermal noise"; instead, the "noise" here describes the profile for the time series spectra. Mandelbrot and Van Ness proposed that the exponent of the power spectrum could take non-integer values and be related to FBM.

The  $\frac{1}{f}$  noise is ubiquitous in physical system, and is the drive of the Self Organized Criticality (SOC) correlated to avalanches.

Notice that the form of spectrum density  $\langle B_{(\omega)}^2 \rangle \propto \frac{1}{\omega}$  is similar to the **Zipf's Law**:

$$f_{(k)} \propto \frac{1}{k^s}, \quad (k = \text{integer}), \quad (18)$$

where  $f_{(k)}$  is the appearance frequency of a word with a rank  $k$  in the words' frequency of table, and  $s$  is the value of the exponent characterizing the distribution. When  $s = 1$ , we have

$$\begin{aligned} f_{(k)} &\propto \frac{1}{k}, \quad (k = \text{integer}) \\ f_{(\Delta x)} &\propto \frac{1}{\Delta x} \end{aligned} \quad (19)$$

and this relates to the  $\frac{1}{f}$  noise, and the distribution with higher events (large  $\Delta x$ ) has a power-law decay. Zipf's law is proposed by an American linguist George Kingsley Zipf, delineating the

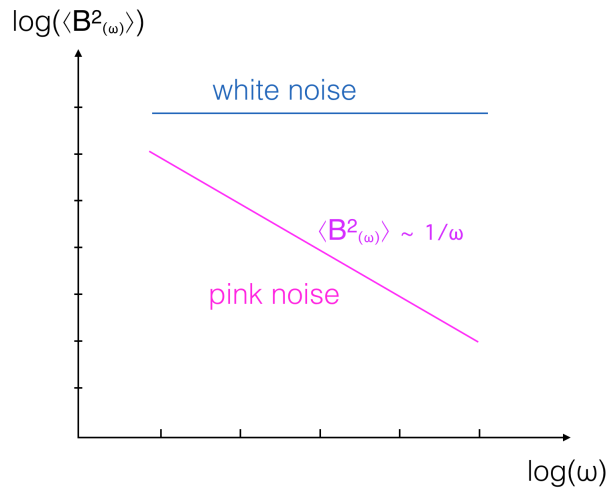


Figure 4: Power spectral density of the white noise and pink noise.

fact that the frequency of any word is inversely proportional to its rank in the frequency table. In human languages, word frequencies have a very heavy-tailed distribution, and can therefore be modeled reasonably well by a Zipf distribution with an  $s$  close to 1. The  $n^{\text{th}}$  most common frequency of a word occurs  $\frac{1}{n}$  as often as the most common frequency. Notice that the rank  $k$  is an integer. But with a wide range and good approximation, the Zipf distribution is related to the  $\frac{1}{f}$ -noise power spectrum distribution.

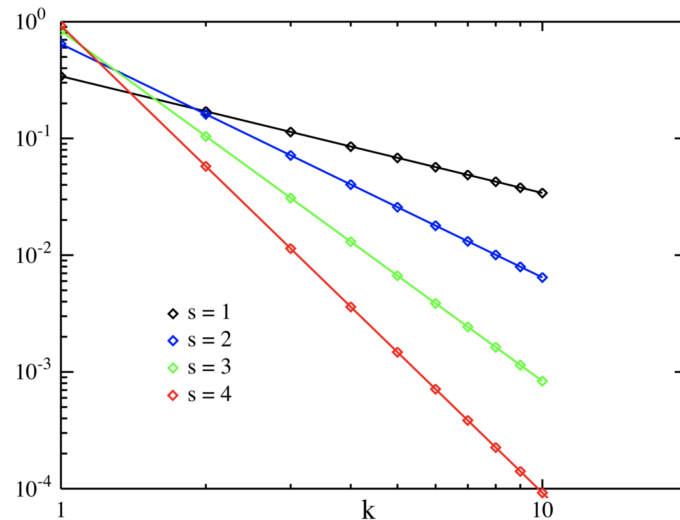


Figure 5: Zipfian distribution. The frequency of human language is described by Zipf's Law with  $s = 1$ .

## 2 Lévy-Stable Distributions

Since we are interested in the process where  $H \neq \frac{1}{2}$ , we need to broaden our view of random processes– real life is outside Central Limit Theorem (CLT). The CLT points out that when the variance of the independent and identically distributed (IID) <sup>1</sup> variables are finite, the attractor distribution is the normal distribution. We need to find out the attractor for the processes beyond CLT in a stable <sup>2</sup> distribution family, which is also referred to as the **Lévy alpha-stable distribution**.

### 2.1 Generalized Central Limit Theorem

A simple way to derive the form of Lévy alpha-stable distribution is to start with the **Chapman-Kolmogorov equation** and assume this stochastic process is Markovian. So we have:

$$P_{i(\mathbf{x})} = \int dy P_{i-1}(\mathbf{y}) P_i(\mathbf{x}|\mathbf{x} - \mathbf{y}), \quad (20)$$

$P_{i(\mathbf{x})}$  is the probability to find a particle at position  $\mathbf{x}$  at step  $i$ ,  $P_i(\mathbf{x}|\mathbf{x} - \mathbf{y})$  is the probability for a particle moving from position  $\mathbf{x} - \mathbf{y}$  to position  $\mathbf{x}$  through a path  $\mathbf{y}$ , and  $P_{i-1}(\mathbf{x} - \mathbf{y})$  is the probability to find particles at position  $\mathbf{x} - \mathbf{y}$  at step  $i - 1$ .

This equation can be extended by convoluting several points in-between positions  $\mathbf{x}$  and  $\mathbf{x} - \mathbf{y}$  (see figure 6). We define  $\Delta t = t_{i+1} - t_i$  and  $N\Delta t = t_N - t_0$ , and assume  $N \gg 1$ . Assuming the Markovian process that the time and space interval is stationary, we have

$$P_{(x_{i+1}, t_{i+1}; x_i, t_i)} = P_{(x_{i+1} - x_i; t_{i+1} - t_i)} = P_{(x_{i+1} - x_i; \Delta t)}, \quad (21)$$

where  $P_{(x_{i+1}, t_{i+1}; x_i, t_i)}$  is probability for a particle move from  $x_i$  to  $x_{i+1}$  in time interval  $t_i$  to  $t_{i+1}$ . Therefore, the probability of transition from  $(x_0, t_0) \rightarrow (x_N, t_N)$  is

$$P_{N(x_0, t_0; x_N, t_N)} = \int dx_0 \dots \int dx_{N-1} \cdot P_{(x_0, t_0; x_1, t_1)} P_{(x_1, t_1; x_2, t_2)} \dots P_{(x_{N-1}, t_{N-1}; x_N, t_N)} \quad (22)$$

$$= \int dy_1 \dots \int dy_N \cdot P_{(y_1, \Delta t)} P_{(y_2, \Delta t)} \dots P_{(y_N, \Delta t)}, \quad (23)$$

where  $y_i = x_{i+1} - x_i$ . And we introduce generating functions

$$\begin{cases} \hat{P}_{(k)} = \int e^{iky_j} P_{(y_j, \Delta t)} dy_j, \\ \hat{P}_{N(k)} = \int e^{iky^N} P_{(y^N, N\Delta t)} dy^N, \end{cases} \quad (24)$$

where  $y^N = \sum_{i=1}^N y_i = x_N - x_1$ .

By playing the Fourier transform, we have:

$$P_{N(\mathbf{x})} = \int e^{ikx} \prod_{i=1}^N \hat{P}_{N(k)} dk. \quad (25)$$

And we assume that all the distribution is IID, so that

$$\begin{aligned} \hat{P}_{1(k)} = \hat{P}_{2(k)} = \hat{P}_{3(k)} = \dots = \hat{P}_{N(k)} \equiv \hat{P}_{(k)} \\ \lim_{N \rightarrow \infty} \prod_{i=1}^N \hat{P}_{(k)} = \hat{P}_{N(k)}. \end{aligned} \quad (26)$$

<sup>1</sup>In probability theory and statistics, a collection of random variables is independent and identically distributed (IID) if each random variable has the same probability distribution as the others and all are mutually independent.

<sup>2</sup>In probability theory, if a linear combination of two independent random variables has the same distribution, location, and scale parameters, then we could say that this distribution is stable.



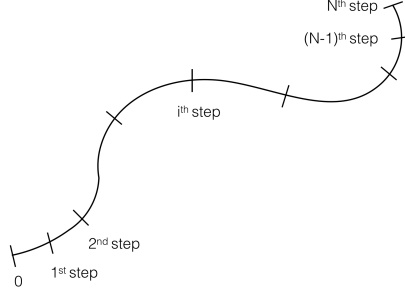


Figure 6: Lévy alpha-stable PDF.

This indicates that the infinite product of a sequence of functions becomes a defined function, i.e. the generating function of the Lévy distribution (RHS in eq. 26). This is the key point in the Lévy stable distribution.

Now we have:

$$P_{N(x)} = \int e^{ikx} [\hat{P}(k)]^N \frac{dk}{2\pi}, \quad (27)$$

where  $P(k) = \sum_{n=0}^{\infty} (-i)^n \frac{k^n m_n}{n!}$ , where  $m_n = i^n \frac{\partial^n P}{\partial k^n}$  and thus  $\hat{P}(k) = 1 - im_1 k - \frac{1}{2} m_2 k^2$ . The  $m_n$  here is the moments  $m_n = \langle x^n \rangle$ . If we let  $\hat{P}(k) \equiv e^{\psi(k)}$ , we have  $\psi(k) = -iC_1 K_1 - \frac{1}{2} C_s k_2^2 + \dots$ . Here  $C_n$  are the **cumulants**–  $C_1 = m_1$ , and  $C_2 = m_2 - m_1^2 = \sigma^2$  (variance). Notice that the property of  $\psi$  that  $k \frac{\partial \psi}{\partial k} = \psi$  implies the scalar invariance and leads us to the self-similarity of the Lévy alpha-stable distribution. After some calculations with an assumption of a zero mean  $C_1 = 0$ , we have the distribution of the generalized CLT:

$$P_{N(x)} = \int e^{ikx} e^{N\psi(k)} \frac{dk}{2\pi} = \frac{1}{2\pi} \frac{1}{(NC_2)^{1/2}} e^{-\frac{x^2}{NC_2}}. \quad (28)$$

The CLT is still hold, but we have larger deviation on the tail. For Lévy alpha-stable distribution, the higher moments are divergent, and thus the cumulants become dubious– indicating the fat-tail effect. This indicates that we have anomalies in the higher order moments.

When the number  $N$  is large enough, the function  $[\hat{P}_{N(k)}]^n$  will converge to a defined function  $\hat{P}_N(C_n k)$ . The mapping from  $[\hat{P}_{N(k)}]^n$  to  $\hat{P}_N(C_n k)$  indicates  $\hat{P}_N(C_n k)$  similar to a fixed point<sup>3</sup>, but in the function space. This fixed point acts as an *attractor*<sup>4</sup> in Lévy alpha-stable distribution that attracts the system to a stable equilibrium (see figure 7). Notice that the importance of Lévy stable probability distributions is that they are “attractors” for properly normed sums of IID variables. That is, the garden-variety Gaussian, one of the cases of Lévy distribution, is an attractor.

This equation  $[\hat{P}_{N(k)}]^n = \hat{P}_N(C_n k)$  has the solution:

$$\hat{P}_{a,\alpha(k)} = e^{-a|k|^\alpha}, \quad (29)$$

<sup>3</sup>A fixed point will be mapped to itself by a function or a transformation. The final state that a dynamical system evolves towards corresponds to an attracting fixed point of the evolution function for that system, such as the center bottom position of a damped pendulum, the level and flat water line of sloshing water in a glass, or the bottom center of a bowl contain a rolling marble.

<sup>4</sup>In the mathematical field of dynamical systems, an attractor is a set of numerical values toward which a system tends to evolve, for a wide variety of starting conditions of the system. An attractor can be a point, a finite set of points, a curve, a manifold, or even a complicated set with a fractal structure known as a strange attractor.

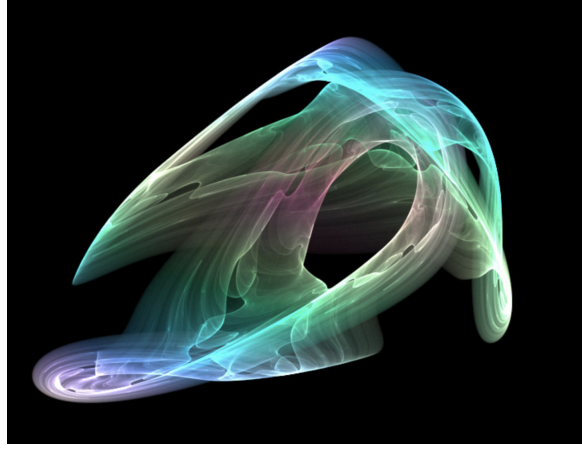


Figure 7: Visual representation of a strange attractor.

where  $a = a(t)$  is a function of  $t$ . And we have Lévy alpha-stable distribution  $\mathcal{L}_\alpha(a, x)$ :

$$\begin{aligned} P_{(x_0, t_0; x_n, t_N)} = P_{a, \alpha(x)} = \mathcal{L}_\alpha(a, x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx} \hat{P}_{a, \alpha(k)} dk \\ &= \frac{1}{|x|^{1+\alpha}} \left[ c^\alpha (1 + \text{sgn}(x)\beta) \sin\left(\frac{\pi\alpha}{2}\right) \frac{\Gamma(\alpha+1)}{\pi} \right], \end{aligned} \quad (30)$$

where  $\alpha$  is the **stability parameter** that characterizes the distribution. This is a generalization of CLT. It states that the sum of a number of random variables with symmetric distributions having power-law tails (Paretian tails), decreasing as  $|x|^{-\alpha-1}$  where  $0 < \alpha \leq 2$ , and therefore have infinite variance. The divergence of variance means a standard Fokker-Plank approach to solving for the time evolution of the distribution function is not applicable.

The width of distribution for different  $\alpha$  can be derived by discretizing the  $\mathcal{L}_\alpha(a, x)$  into  $N$  time steps:

$$P_{N(x)} = \frac{1}{N^{\frac{1}{\alpha}}} \mathcal{L}_\alpha\left(a, \frac{x}{N^{\frac{1}{\alpha}}}\right). \quad (31)$$

Thus, the width is proportional to  $\propto N^{\frac{1}{\alpha}}$ . When  $\alpha < 2$ , the distribution has a width that is broader than  $N^{\frac{1}{\alpha}}$ . This indicates a super-diffusive flow. If  $\alpha = 1$ , the width is proportional to  $N$ , indicating that the width is spreading ballistically. Notice that an asymptotic behavior for  $|x| \rightarrow \infty$  has a propagator:

$$p_\alpha(x, t) \sim \frac{t}{|x|^{\alpha+1}}, \quad 0 < \alpha \leq 2. \quad (32)$$

At any time  $t$ , it doesn't have finite second moment since  $\alpha < 2$ . To maintain the equal probability at every time and space, as the clock runs ( $t$  increases), the particle should go to a larger  $x$  ( $x$  increases). This indicates that the tail of the distribution function is expanding outward with time. This is an **expanding tail solution**. This relates to the **Black Swan Theory** popularized by a statistician and risk analyst N. N. Taleb. This theory is built to explain the disproportionate role of high-profile, hard-to-predict, and rare events that are beyond the realm of normal expectations in history, science, finance, and technology.

The equation 30 resembles the **Pareto distribution** which is named after the Italian civil engineer, economist, and sociologist Vilfredo Pareto. Its probability density function is

$$f_X(x) = \begin{cases} \frac{\alpha x_m^\alpha}{x^{\alpha+1}} & x \geq x_m, \\ 0 & x < x_m, \end{cases} \quad (33)$$

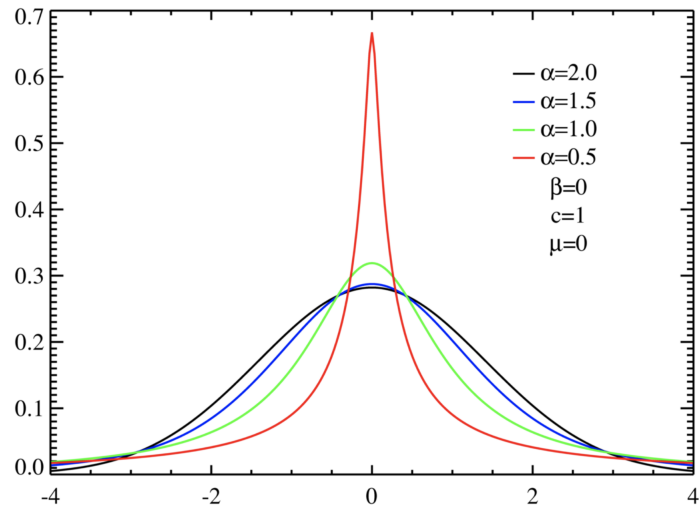


Figure 8: Lévy alpha-stable PDF. The black line ( $\alpha = 2$ ) represents a Gaussian distribution. The Green line ( $\alpha = 1$ ) represents the Cauchy distribution. If we compare all these lines, one can find that at large  $x$ , the smaller  $\alpha$  are, the fatter tail they'll have.

where  $x_m$  is the (necessarily positive) minimum possible value of  $X$ , and  $\alpha$  is a positive parameter. The Pareto distribution has colloquially become known and referred to as the Pareto principle, or “80-20 rule”. The empirical observation has found that the 80-20 distribution fits a wide range of cases, including natural phenomena and human activities.

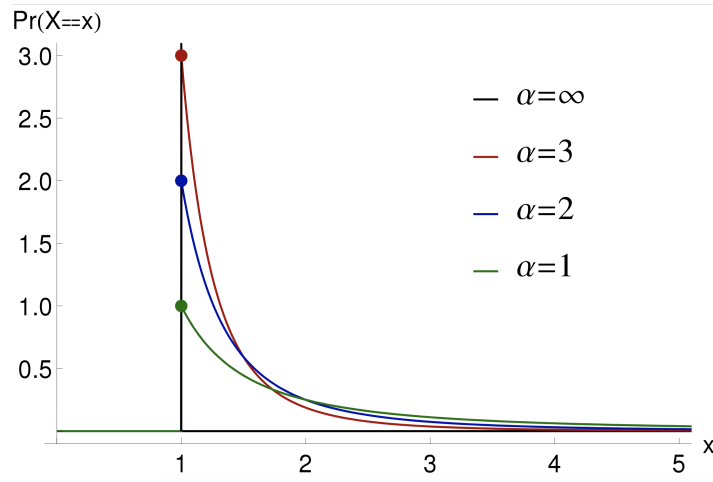


Figure 9: Pareto distribution.

Though the Zipf's law is a discrete probability distribution, the Zipf's Law can be derived from the Pareto distribution if the  $x$  values (incomes) are binned into  $N$  ranks.

## 2.2 Gaussian & Cauchy Distribution

When  $\alpha = 1$  the Lévy-stable distribution reduces to the Cauchy distribution (or **Lorentz distribution**), and when  $\alpha = 2$  it'll reduce to the Gaussian distribution:

$$\mathcal{L}_\alpha(a, x) \propto \begin{cases} \frac{1}{x^2+a^2}, & \text{Cauchy distribution } (\alpha = 1) \\ e^{-x^2/a}, & \text{Gaussian distribution } (\alpha = 2). \end{cases} \quad (34)$$

Notice that the terms  $\frac{(\dots)}{|x|^{1+\alpha}}$  and  $\frac{1}{x^2+a^2}$  imply a fat power-law tail in the distribution. In particular, Mandelbrot (1960) referred to those maximally skewed in the positive direction with  $1 < \alpha < 2$  as “**Pareto-Lévy distributions**”, which he regarded as better descriptions of stock and commodity prices than the garden-variety normal distribution. The Pareto-Lévy distribution is also known for exhibiting its “80-20 Law”.

Notice that when  $\alpha = 2$ , the parameter  $a$  can be represented as  $Dt$ , where  $D$  is diffusivity. We can recover the diffusion function by taking derivative of  $P_{\alpha=2}(x,t)$  from equation (29) and assuming  $a = Dt$ :

$$\frac{\partial P_{\alpha=2}(x,t)}{\partial t} = \int e^{ikx} (-D|k|^2) e^{-Dt|k|^2} = D \frac{\partial^2}{\partial x^2} P_{\alpha=2}(x,t). \quad (35)$$

Clearly, we have parameter  $D$  is the diffusivity. Let's generalize this the diffusivity (still assuming  $a = Dt$ ), for  $\alpha \neq 2$ . We have

$$\frac{\partial P_{\alpha}(x,t)}{\partial t} = \int e^{ikx} (-D|k|^\alpha) e^{-Dt|k|^\alpha} = D \frac{\partial^\alpha}{\partial x^\alpha} P_{\alpha}(x,t), \quad (36)$$

where  $\alpha$  can be an arbitrary rational number that satisfies  $0 < \alpha \leq 2$ , and therefore  $\frac{\partial^\alpha}{\partial x^\alpha}$  is fractional derivative. This leads us to the **fractional calculus**. A critical question thus arises:

- What is the physical meaning of the parameter  $a/t$ , when  $\alpha \neq 2$ ?

To get the physical meaning of  $a$ , we might start from the basic continuum equations, i.e. Hasegawa-Mina or gyro-kinetics, then proceed to the Lévy flight formulation.

All in all, a Lévy Walk is self-similar, and has long excursions because of the fat-tail effect—large events weight more (see figure 9). The large excursion in a Lévy flight introduces the *non-locality in the flux-gradient relation*.

## 2.3 Direct evidence for Lévy flight and anomalous process

As we knew, the normal diffusion can be derived from the Fokker-Plank approach and can be described by Lévy alpha-stable distribution with stability parameter  $\alpha = 2$ . In this case, the relation between space and time in diffusion is

$$\langle (x - \langle x \rangle)^2 \rangle = \langle \delta x^2 \rangle \propto t^\gamma, \quad \text{where } \gamma = 1. \quad (37)$$

For anomalous diffusion, however, is denoted when  $\gamma \neq 1$ :

$$\begin{cases} 0 < \gamma < 1 & \text{sub-diffusion,} \\ \gamma > 1 & \text{super-diffusion.} \end{cases} \quad (38)$$

And there are two methods to deal with anomalous diffusion. One is the Continuous Time Random Walk (CTRW) and the other is the Fractal Kinetics (FK). Both methods are similar but beyond the Fokker-Plank approach. We'll discuss this in section 3.

Solomon et al. (1993) did an observation of anomalous diffusion and Lévy flights in a two-dimensional rotating flow. The setup is a rotating annular tank at a frequency of 1.5 Hz. The inner and outer radii are 10.8 and 43.2 cm respectively. A laminar velocity field is maintained in the flow. A sheared counter-rotating azimuthal jet is created and leads to a chain of vortices that move around the annulus. They follow the trajectories of a large number of particles for a long time and have direct evidence for Lévy flight and anomalous diffusion. A bunch of tracers are designed to follow the chaotic trajectories, sticking to vortices intermittently. They move as a “Lévy flight” for a long distance. These tracers collect the movements of trajectories and thus the variance in angle displacement  $\langle(\theta - \langle\theta\rangle)^2\rangle$  can be calculated. The azimuthal displacement as a function of time for the particles is in figure 10 and 11.

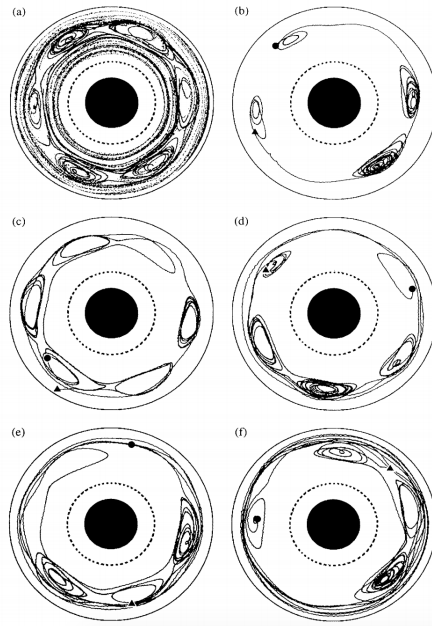


Figure 10: Vortices and time evolution of particle trajectories in a rotating tank.

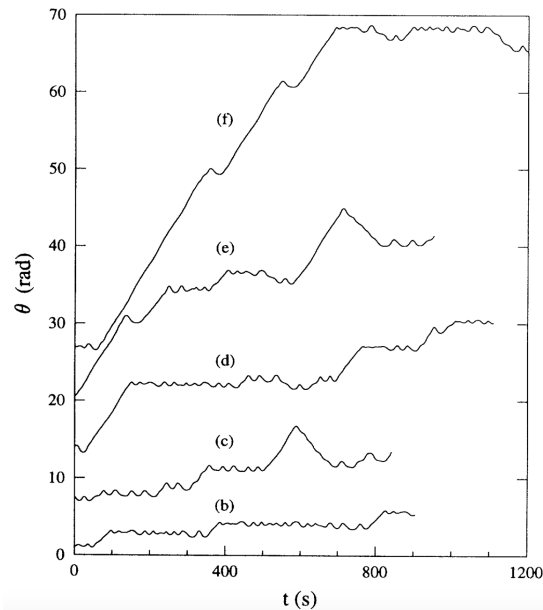


Figure 11: The azimuthal displacement as function of time for the particles. The “flat” parts of curves denotes the sticking process. The diagonal lines are the “Lévy flight” between two sticky events.

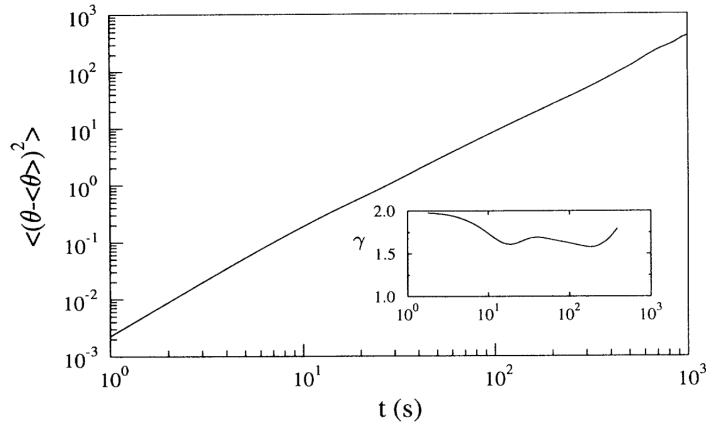


Figure 12: Variance of azimuthal displacement of a distribution of tracer particles for a time-periodic laminar flow.

In the figure 11, the “flat” parts of curves denote the sticking process. These flat parts corresponds to oscillatory movements of the tracers within one vortex ring– i.e. circulations. Here the flow is sub-diffusive that retard the motion of particles– the particles are trapped in a confined region. On the other hand, the diagonal, steeper lines represent the “Lévy flight” between two sticky events (flat oscillations). The tracer is moving from one vortex to another, thus trajectories of tracer in the experiment have long excursions which indicate the “Lévy flights” that form fractal scaling trajectories. In this region, the excursion is super-diffusive, and it’s second moment is divergent.

They also estimate how the exponent  $\gamma$  evolves with time. Observation results indicates that the transport of tracer particles in laminar flow is super-diffusive ( $\gamma > 1$ ), with an average exponent  $\gamma = 1.65$  (see figure 12). Notice that this experiment suggests a combination of “Joseph effect”– the long persistent oscillating movement, and the “Noah effect”– the sudden growths in the angle variance (see figure 11).

### 3 Continuous Time Random Walk (CTRW)

#### 3.1 Review of the Fokker-Plank Equation

Starting from the Chapman-Kolmogorov equation, we have:

$$P_{(x_1, t_1 | x_3, t_3)} = \int dx_2 P_{(x_1, t_1 | x_2, t_2)} P_{(x_2, t_2 | x_3, t_3)}. \quad (39)$$

We expand this equation into several steps between  $(x_1, t_1)$  and  $(x_3, t_3)$  with  $\Delta t$  is stationary and  $\Delta x$  has its own distribution. Then we can derive the **Fokker-Plank equation**

$$\frac{\partial}{\partial t} P_{(x, t | x', t')} = - \frac{\partial}{\partial x} \underbrace{\left[ A_{(x, t)} P_{(x, t | x', t')} - \frac{1}{2} \frac{\partial}{\partial x} B_{(x, t)} P_{(x, t | x', t')} \right]}_{\text{probability flux}}, \quad (40)$$

where  $A_{(x, t)}$  is the drift velocity,  $B_{(x, t)}$  is the diffusivity:

$$\begin{aligned} A_{(x, t)} &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \langle \Delta x \rangle \\ B_{(x, t)} &= \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} \langle (\Delta x)^2 \rangle \end{aligned} \quad (41)$$

The relation between these two  $A_{(x, t)}$  and  $B_{(x, t)}$  is:

$$A_{(x, t)} = \frac{1}{2} \frac{\partial}{\partial x} B_{(x, t)}. \quad (42)$$

More details are in Lecture notes in 2016.

#### 3.2 Why we need a CTRW?

One of the ways to analyze the anomalous diffusion is using the CTRW method. recall that in Fokker-Plank approach, it's necessary to have  $\Delta x$  distribution— each step of a particle motion can vary. The distribution of  $\Delta t$ , however, is a delta function— the time step is fixed and acts like a clock. We release the constraint of time step from the role of a clock, let it take a statistical distribution, and let it evolve dynamically. Thus, we have two stochastic variables and we need to specify the distribution functions for both  $\Delta x$  and  $\Delta t$ . The distribution of  $\Delta t$  indicates that the system's memory can be enlarged or shortened, corresponding to Lévy flights and sticking processes respectively. This indicates that the CTRW is non-local and thus is non-Markovian in space-time.

First of all, we start with a general form of the Chapman-Kolmogorov equation:

$$Q(x, t) = \int d(\Delta x) \int_0^t d(\Delta t) \cdot Q(x - \Delta x; t - \Delta t) \cdot P(\Delta x, \Delta t), \quad (43)$$

where  $Q(x, t)$  is an orbit distribution of space and time, and  $P(\Delta x, \Delta t)$  is the joint PDF of the two stochastic variables  $\Delta x$  and  $\Delta t$ . This indicates that the weighting is fractalized. CTRW is played in two different ways— the waiting time model and the velocity model.

#### 3.3 Waiting Time Model

In the waiting time model, the stochastic distribution of  $\Delta t$  and  $\Delta x$  are independent, i.e. separable:

$$P_{(\Delta x, \Delta t)} = P_{(\Delta x)} P_{(\Delta t)}. \quad (44)$$

We expand  $Q(x - \Delta x; t - \Delta t)$  to first order in  $\Delta x$ :

$$Q_w(x, t) = \int_0^t d(\Delta t) Q(x; t - \Delta t) \Phi_w(\Delta t) \quad (45)$$

where  $Q_w(x, t)$  is a relabeling to indicate a jump PDF derived by the waiting model, and  $\Phi(\Delta t)$  is the probability to wait for at least  $\Delta t$ . Of course, this equation can be extended by the Fokker-Plank method with a well-behaved  $P(\Delta x)$ . The probability  $\Phi(\Delta t)$  is given by

$$\Phi_w(\Delta t) = \int_{\Delta t}^{\infty} dt' P(t'). \quad (46)$$

Thus, the distribution for the time step  $P(t')$  need to be specified.

This model can be used to deal with sticking processes. The reason is that to discuss the sticking process, we need to represent long-time trapping of particles in a confined region. And this can be approached by setting spatial steps small and have a Pareto-Lévy distribution in time steps which allows large  $|\Delta t|$  event highly possible. Thus, this model is especially suitable for **sub-diffusive** systems.

### 3.4 Velocity Model

In the velocity model,  $\Delta t$  is defined as a traveling time  $\frac{\Delta r}{V}$ , where  $\Delta r$  is the distance of traveling, and  $V$  is the velocity constant. With a constant velocity, we can tie the time step  $\Delta t$  to the spatial step  $\Delta x$ . the joint PDF can be expressed as

$$P_{(\Delta x, \Delta t)} = \delta\left(\Delta t - \frac{\Delta x}{V}\right) P_{(\Delta x)} \quad (47)$$

Substituting this joint PDF into the Chapman-Kolmogorov equation, we have

$$Q_v(x, t) = \int_{-Vt}^{Vt} d(\Delta x) \int_0^t d(\Delta t) Q(x - \Delta x; t - \Delta t) \Phi_v(\Delta x, \Delta t), \quad (48)$$

where  $Q(x, t)$  is an orbit distribution of space and time, and  $\Phi_v(\Delta x, \Delta t)$  is the probability to make a step of at least length  $|\Delta x|$  with duration  $\Delta t$ . The  $\Phi_v(\Delta x, \Delta t)$  can be expressed as

$$\Phi_v(\Delta x, \Delta t) = \frac{1}{2} \delta(|\Delta x| - V\Delta t) \int_{|\Delta x|}^{\infty} dx' \int_{\Delta t}^{\infty} dt' \delta(dt' - \frac{|dx'|}{V}) P(dx') \quad (49)$$

Thus, the distribution for the spatial step  $P(dx')$  need to be specified.

This velocity model is especially suitable for the movements of Lévy flights. Since the Lévy flight can have a long excursion within a small time duration, it can be approached by setting the time steps is small, but with Lévy distributed spatial steps, which the fat-tails effect allows a higher weighted large  $|\Delta x|$ . Thus, the velocity model is workable for **super-diffusive** cases.

For the cases that both time steps  $\Delta t$  and spatial steps  $|\Delta x|$  are very small, the CTRW method will reduce to the case of *normal diffusion*.



### 3.5 CTRW Summary

The key assumption for the CTRW is that the distribution depends on both of a stochastic variable of time-step and that of spatial-step. This leads to the fact that the CTRW is non-local and thus non-Markovian in time.

If we relate the Lévy flights and sticky processes to the *percolation theory*, one can suggest that above the percolation threshold, the system is super-diffusive and has Lévy flights. In this regime, the cells in phase space overlap, which allows long-distance “kicks” to be more possible. While when the percolation is below the threshold, the system is sub-diffusive and the flow manifests a sticky process.

### References

Mandelbrot, B. 1960, *International Economic Review*, 1, 79

Mandelbrot, B. B., Passoja, D. E., & Paullay, A. J. 1984, *Nature*, 308, 721

Mandelbrot, B. B., & Wallis, J. R. 1968, *Water resources research*, 4, 909

Solomon, T., Weeks, E. R., & Swinney, H. L. 1993, *Physical Review Letters*, 71, 3975