Technical Appendices

A
Sampling probability distributions

There are two rough classes of methods for sampling probability distributions: those based on a *transformation method*, and those based on a *rejection method*. The transformation method is very efficient, but it requires that we can perform a definite integral over the probability distribution and then invert the results. This is not always that easy. Sometimes we can perform these step analytically, some times we need to perform the integral numerically, and then form a table which is "invertible". The rejection method will always work, but it can be inefficient, especially then the probability distribution is changing by orders of magnitude. There are also cases which are build on a hybrid of these two methods.

A.1 The transformation method

The transformation method is based on the fundamental *transformation law of probabilities*, which is simply,

$$|p(y)dy| = |p(x)dx| \quad \text{or} \quad p(y) = p(x)\left|\frac{dx}{dy}\right|$$

in order to generate random values of some arbitrary desired distribution from random values with other distributions. When the source distribution is a *uniform distribution* the transformation method is known as the *inversion method*.

First, to understand the transformation method we must first remember the definition of an *inverse function*. Two functions g and f are inverse one of the other if

$$f(g(x)) = x \quad \text{and} \quad g(f(x)) = x$$

We denote g as f^{-1} the inverse of f. Not every function f has an inverse function f^{-1}. The function f must be monotonic in a given interval to have an inverse in that interval. We must think in f^{-1} as a function that unmakes what has been made with function f. For example, a subtraction will be used to unmake an addition, a division can be used to unmake a multiplication, and a derivative can be used to unmake an integration. Provided that it exists, the inverse of a given function $y = f(x)$ can be obtained by "solving for x in terms of y": $x = g(y) = f^{-1}$. As an example,

$$\begin{aligned} f(x) &= \sqrt{(}2x - 3), \\ y = f(x) &= \sqrt{(}2x - 3), \\ x &= \frac{y^2 + 3}{2}, \\ f^{-1}(y) &= \frac{y^2 + 3}{2}. \end{aligned}$$

Taking x just as an independent variable we have

$$f^{-1}(x) = \frac{x^2 + 3}{2}. \tag{A.1}$$

Now we can proceed to explain the transformation (inversion) method. The idea behind the inversion method is that we would like to start out with a uniform probability distribution:

$$p(x)dx = \begin{cases} dx & 0 < x < 1 \\ 0 & otherwise \end{cases} \tag{A.2}$$

$$\int_{\infty}^{\infty} p(x)dx = 1 \tag{A.3}$$

and transform it into a non-uniform probability distribution p(y). (The proof presented here is from Numerical Recipes). Starting with the Fundamental Transformation Law of Probabilities,

$$p(y) = p(x)\left|\frac{dx}{dy}\right| \tag{A.4}$$

Now if we want a probability $p(y)$ to follow a function $f(y)$ we can write,

$$f(y) = p(y) = p(x)\left|\frac{dx}{dy}\right| \tag{A.5}$$

This allow us to relate a random variable x from the distribution function $p(x)$ with the random variable y from the distribution function $p(y)$. If x is from a uniform distribution in $[0,1]$, the $p(x)$ is constant so we have,

$$f(y) = \frac{dx}{dy} \tag{A.6}$$

The solution to this is the indefinite integral of $f(y)$,

$$x = F(y) = \int_{-\infty}^{y} f(z)dz \tag{A.7}$$

The relationship give us the source random variable x, given the target random variable y. So we need to be able to invert the relationship. The desired transformation which takes a uniform deviate into one distributed as $f(y)$ is therefore,

$$y(x) = F^{-1}(x) \tag{A.8}$$

where F^{-1} is the inverse function to F. Whether Eq. A.8 is feasible to implement depends on whether the inverse function of the integral of f(y) is itself feasible to compute, either analytically or numerically. Sometimes it is, and sometimes it isn't. In short, one can obtain a transformation function from $p(x) \in (0,1)$ to another function if:

* we can perform the indefinite integral, and

* we can invert the result of that indefinite integral.

Thus, the inversion method to generate random variates from a given probability distribution $p(x)$ (i.e., exponential, potential, etc.) is based in the following steps:

(i) To calculate the distribution function (normalized) of the function $p(x)$.

$$F(x) = \int_{-\infty}^{+\infty} f(x)dx = 1 \tag{A.9}$$

(ii) To calculate the cumulative function to a certain value x

$$u(x) = \int_{x_{\min}}^{x} p(y)dy, \tag{A.10}$$

where x is within the interval $x_{\min} \leq y \leq x$. Note that $0 \leq u(x) \leq 1$.

(iii) Then, to invert the function $u(x)$, obtaining $x = F^{-1}(u)$.

By generating random values of u we obtain the values x from the original distribution $p(x)$ in terms of the function $F(u)$.

A.1.1 Examples

The negative exponential distribution

Imagine we want to obtain random variate values x from a given exponential function defined as

$$f(x) = e^{-\alpha x} \tag{A.11}$$

where $x \geq 0$. First we must consider the distribution function:

$$\begin{aligned} F(x) &= \int_0^\infty \beta f(x) dx = 1 \\ \int_0^\infty \beta e^{-\alpha x} dx &= 1, \quad \beta \left. \frac{e^{-\alpha x}}{-\alpha} \right|_0^\infty = 1 \\ \beta \left[0 - \frac{e^0}{-\alpha} \right] &= 1, \quad \beta \frac{1}{\alpha} = 1, \quad \beta = \alpha. \end{aligned}$$

Thus, we have the following normalized cumulative function,

$$F(x) = \int_0^x \alpha e^{-\alpha x} dx \tag{A.12}$$

Following,

$$\begin{aligned} F(x) &= \alpha \int_0^x e^{-\alpha x} dx = \alpha \left. \frac{e^{-\alpha x}}{-\alpha} \right|_0^x \\ &= \frac{\alpha}{-\alpha} \left[e^{-\alpha x} - e^0 \right] \\ &= -e^{-\alpha x} + 1 = 1 - e^{-\alpha x}. \end{aligned} \tag{A.13}$$

Finally, we have

$$u(x) = F(x) = P(\xi \leq x) = 1 - e^{-\alpha x},$$

$$e^{-\alpha x} = 1 - u(x),$$

$$-\alpha x = \ln(1 - u(x)).$$

Noting that $1 - u(x))$ is statistically identical to $u(x)$ we get,

$$x = -\frac{\ln(u(x))}{\alpha} \tag{A.14}$$

The potential distribution

To obtain a random variate from a probability function decaying potentially with exponent μ, $p(x) = x^{-\mu}$ we first must note that such probability function has a natural inferior cut-off $x = 1$ (values $x < 1$ will lead to a growing potential). Thus we have,

$$\int_1^\infty \beta x^{-\mu} = 1 \tag{A.15}$$

The distribution function is

$$\begin{aligned} \int_1^\infty \beta x^{-\mu} dx &= 1, \\ \beta \left. \frac{x^{-\mu+1}}{-\mu+1} \right|_1^\infty &= 1, \\ \beta \left[\lim_{x\to\infty} \left(\frac{x^{-\mu+1}}{-\mu+1} - \frac{1}{-\mu+1} \right) \right] &= 1, \\ \frac{\beta}{-\mu+1} \left[\lim_{x\to\infty} x^{-\mu+1} - 1 \right] &= 1. \end{aligned} \tag{A.16}$$

The limit above can be solved for $\mu > 1$. Then, if $\mu > 1$,

$$\begin{aligned} \frac{\beta}{-\mu+1} \left[\lim_{x\to\infty} x^{-\mu+1} - 1 \right] &= 1, \\ \frac{\beta}{-\mu+1} [0 - 1] &= 1, \\ -\beta &= -\mu - 1. \end{aligned} \tag{A.17}$$

Thus, we get the distribution function $F(x)$,

$$\begin{aligned} F(x) &= \int_1^x (\mu-1)x^{-\mu} dx = (\mu-1) \int_1^\infty x^{-\mu} \\ &= (\mu-1) \left[\frac{x^{1-\mu}}{1-\mu} \right]_1^x = \frac{-(1-\mu)}{1-\mu} \left[x^{1-\mu} - 1 \right] = 1 - x^{1-\mu} \end{aligned} \tag{A.18}$$

Finally, we have

$$\begin{gathered} u(x) = F(x) = P(\xi \leq x) = 1 - x^{1-\mu}, \\ x^{1-\mu} = u(x), \\ x = u(x)^{1/(1-\mu)}, \text{ for } \mu > 1. \end{gathered}$$

A.2 The acception/rejection method

The rejection method is based on a simple geometrical argument:

Draw a graph of the probability distribution $p(x)$ that we wish to generate, so that the area under the curve in any range of x corresponds to the desired probability of generating an x in that range. If we had some way of choosing a random point *in two dimensions*, with uniform probability in the *area* under the curve, then the x value of that random point would have the desired distribution.

Now on the same graph, draw any other curve $h(x)$ which has finite (not infinite) area and lies everywhere *above* your original probability distribution. This is always possible, because the original curve encloses only unit area, by definition of probability. We will call this $h(x)$ the *comparison function*. Imagine now that we have some way of choosing a random point in two dimensions that is uniform in the area under the comparison function. Whenever that point lies outside the area under the original probability distribution, we will *reject* it and choose another random point. Whenever it lies inside the area under the original probability distribution, we will *accept* it. It should be obvious that the accepted points are uniform in the accepted area, so that their x values have desired distribution. It should also be obvious that the fraction of points rejected just depends on the ratio of the area of the comparison function to the area of the probability distribution function, not on the details of shape of either function.

Now the key question is how to choose a uniform random point in two dimensions under the comparison function $h(x)$. A variant of the transformation method can be used: We must be sure that the comparison or hat function $h(x)$ has a an indefinite integral that can be known analytically, and is also analytically invertible to give x as a function of "area under the comparison function to the left of x". Then we pick a uniform deviate between 0 and A, where A is the total area under $h(x)$, and we use it to get a corresponding x. Then pick a uniform deviate between 0 and $h(x)$ as the y value for the two dimensional point. The point (x, y) is uniformly distributed in the area under the comparison function $h(x)$.

An equivalent procedure is to pick the second uniform deviate between zero and one, and accept or reject according to whether it is respectively less than or greater than the ratio $p(x)/h(x)$.

In brief, the principle of the rejection method is that the frequency distribution $f(x)$ is approximated by another distribution function $h(x)$ which is easier to calculate, and then a correction is made by randomly accepting x values with a probability $p(x) = \frac{f(x)}{kh(x)}$, and rejecting x values with probability $1 - p(x)$. The constant k is chosen so that $kh(x) \geq f(x)$ for all values

of x. Whenever an x value is rejected, the procedure starts over again with a new x. The accepted x values have the distribution $f(x)$ because the acceptance probability $p(x)$ is proportional to $f(x)/h(x)$.

This method does not require that the cumulative distribution function (indefinite integral of $p(x)$) be readily computable, much less the inverse of that function -which was required for the transformation method. In the case of discrete distributions, sometimes it is used as an alternative to the chop down search. The inversion by chop down search from 0 can be used when $p(x)$ is a discrete function (i.e., Poisson, binomial, etc.). Then, the inverse function can be calculated by successively adding $p(0) + p(1) + ... + p(x)$ until the sum exceeds a random uniform variable u. The advantage of the rejection method is that the calculation time does not grow with the mean of x as in chop down searches. It is therefore used when the variance is so large that the chop down search would be more time consuming.

The disadvantage of the rejection method is that is difficult to find a good hat function $h(x)$ which is easy to calculate and at the same time approximates $f(x)$ so good that the rejection rate will be low. A bad hat function will lead to high rejection rate and hence a long execution time. Various improvements of the rejection method by quick acceptance and quick rejection schemes can be applied (i.e., the patchwork rejection method, the ratio-of-uniforms rejection method, etc.).

B

Random walks and diffusion

A classic problem in science concerns the statistical properties of random walks (Brown, 1828; Einstein, 1905). One of particular relevance is the type of diffusion generated by the underlying dynamics, which can be characterized by a parameter α. For large enough times t, the root mean squared displacement (which characterizes the spreading rate of random walkers from a source point) of the random walk scales as t^{α}. For the case of Brownian motion, whose steps increments have finite variance and no correlations, $\alpha = 1/2$, leading to normal diffusion. Anomalous diffusion is obtained when $\alpha > 1/2$ (super-diffusion) or $\alpha < 1/2$ (sub-diffusion).

B.1 Normal diffusion: Brownian motion

The statistical behavior of a random walk can be obtained by mapping the discrete rules of the walker movement into a continuous diffusion equation. Consider a particle jumping to left or right on a one-dimensional lattice. Let us assume that each jump occurs within a time interval $(t, t+\delta t)$. The probability that the particle performs a jump will be

$$P(t, t+\delta t) = \beta\, \delta t + O(\delta t) \tag{B.1}$$

where $O(\delta t)$ indicates higher order terms. If the length of a jump is given by a fixed number δx, then the transition probabilities will be:

$$P(x \to x+\delta x) = \frac{1}{2} \tag{B.2}$$

$$P(x \to x - \delta x) = \frac{1}{2} \tag{B.3}$$

Let us indicate $P_i(t) \equiv P[x(t) = i\ \delta x]$. From the previous set of hypotheses, we have:

$$P_i(t + \delta t) = (1 - \beta\ \delta t)P_i(t) + \frac{1}{2}\beta\ \delta t\Big[P_{i-1}(t) + P_{i+1}(t)\Big] + O(\delta t) \tag{B.4}$$

We can thus write:

$$P_i(t + \delta t) = -P_i(t) = -\beta\ \delta t P_i(t) + \frac{1}{2}\beta\ \delta t\Big[P_{i-1}(t) + P_{i+1}(t)\Big] + O(\delta t) \tag{B.5}$$

dividing by δt and taking the limit $\delta t \to 0$, we obtain:

$$\frac{dP_i(t)}{dt} = -\beta P_i(t) + \frac{1}{2}\beta\Big[P_{i-1}(t) + P_{i+1}(t)\Big] \tag{B.6}$$

Now we move into the continuum by removing the dependency in x to the arbitrary discretization δx. Using the notation $P_i(t) = f(x,t)$, we can write:

$$\frac{\partial f(x,t)}{\partial t} = -\beta f(x,t) + \frac{1}{2}\beta\Big[f(x - \delta x, t) + f(x + \delta x, t)\Big] \tag{B.7}$$

and performing a Taylor expansion we obtain:

$$\begin{aligned}\frac{\partial f(x,t)}{\partial t} &= -\frac{\beta}{2}\partial_x f(x,t)\delta x + \\ &+\frac{\beta}{2}\partial_x^2 f(x,t)\frac{(\delta x)^2}{2} + \\ &+\frac{\beta}{2}\partial_x f(x,t)\delta x + \frac{\beta}{2}\partial_x^2 f(x,t)\frac{(\delta x)^2}{2} + ...\end{aligned} \tag{B.8}$$

or, after simplification,

$$\frac{\partial f(x,t)}{\partial t} = \frac{\beta}{2}\frac{\partial^2 f(x,t)}{\partial x^2} + ... \tag{B.9}$$

By defining a new constant $D \equiv \beta(\delta x)^2/2$, we finally obtain the *diffusion equation* for the random walk:

$$\frac{\partial f(x,t)}{\partial t} = D\frac{\partial^2 f(x,t)}{\partial x^2} \tag{B.10}$$

This equation can be solved and gives a general expression:

$$f(x,t) = \frac{1}{\sqrt{4\pi Dt}} \exp\left[-\frac{x^2}{4Dt}\right] \tag{B.11}$$

which has been obtained by assuming an initial condition $f(x,0) = \delta(0)$ (i. e. the initial condition involves all walkers located at a single point, here the origin of coordinates).

From this distribution, we can compute the moments and their dynamics, using the integral

$$\langle x^\alpha(t)\rangle = \int_{-\infty}^{+\infty} x^\alpha P(x,t)dt$$

It can be shown, for example, that the average $\langle x\rangle$ remains constant over time, i.e.

$$\frac{d\langle x\rangle}{dt} = 0$$

consistently with the symmetry of the RW rules. The dispersal, as measured by the variance, can also be computed. Now we have

$$\frac{d\langle x^2(t)\rangle}{dt} = 2D$$

which can be easily solved, leading to a linear relation

$$\langle x^2(t)\rangle = 2Dt$$

Since the variance $\sigma^2(x)$ is defined as

$$\sigma^2(x) = \langle x^2(t)\rangle - \langle x(t)\rangle^2$$

and $\langle x(t)\rangle^2 = 0$, we have a standard deviation for the moment of the walkers, $\sigma(t)$, given by

$$\sigma(t) = \sqrt{2Dt}$$

For a population of random walkers, the previous result indicates that their spatial spreading is limited by a general, simple scaling law

$$\sigma(t) \approx (Dt)^{1/2}$$

which can actually be used as a test for diffusive motion in a given population of individuals.

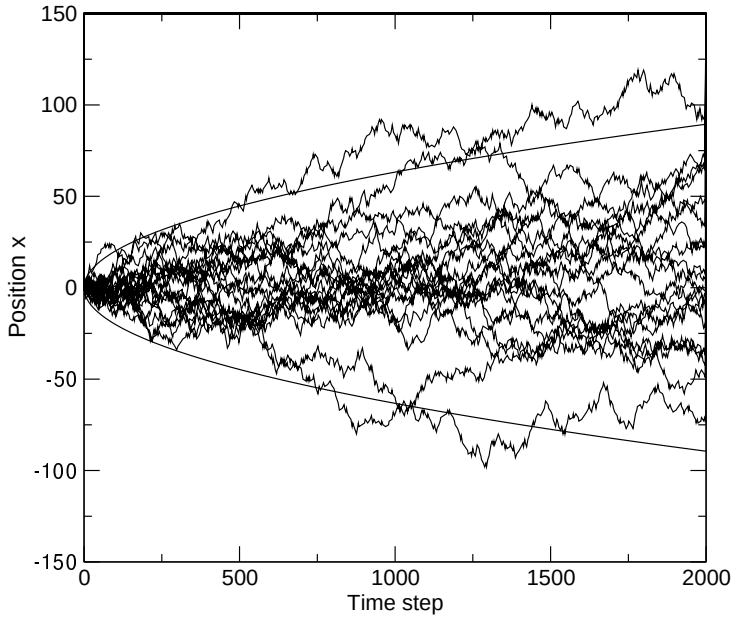

Figure B.1: Paths followed with time by random walkers in one dimension from a small population. The parabola describes the standard deviation of the dispersal.

B.2 Anomalous diffusion: fractional Brownian motion and Lévy statistics

Anomalous diffusion, that is, root mean squared (rms) displacement departures from the scaling relationship with time $rms \approx t^{1/2}$ can emerge by either a time dependence of a process' variance (Mandelbrot's *Joseph's effect*), or by the statistics of the random variable (Mandelbrot's *Noah's effect*) (Mandelbrot, 1977, 1982).

Let us now consider the first case. Imagine a process having a Gaussian statistics but with variance that is not linear in time. To accomplish this, consider a random walk process $X(t)$ with the correlation function

$$\langle X(t)X(t')\rangle = 2D|t-t'|^{2H} \tag{B.12}$$

This process is simple a Brownian motion for $H = 1/2$, but is called anomalous diffusion for other values of H. In terms of the random walk model for $H > 1/2$, the walker tends to continue in the direction he or she has previously been walking. The higher the H value the greater the tendency to persist in one's direction of motion. For $H < 1/2$ the walkers actively choose to avoid continuing in a given direction, and changes in direction more frequently than in Brownian motion. The behavior is called anti-persistence. The statistics of the random walk process are not changed by the correlation function B.12; they remain Gaussian but their variance changes, yielding

$$P(\Delta x, \Delta t) = (4\pi D\Delta t^{2H})^{-1/2} exp[-(\Delta x)^2/4D\Delta t^{2H}] \tag{B.13}$$

where Δx is the change in the displacement of the walker during the time interval Δt. Eq. B.13 is the probability density for *fractional Brownian motion* (*fBm*). The variance of fBm, denoted by σ^2, can be expressed as $\sigma \propto t^{2H}$ where t is the time over which the process has evolved and H is the Hurst exponent ($H \in 0,1$). It is worth to pointing out that these anomalous processes have inverse power law spectra, $1/f^{2H+1}$, where f is the frequency. Note the relationship between the exponents in the variance and the spectrum. Such kind of processes are described by *fractional Brownian motion* (*fBm*).

It is interesting to scale the displacement by λ and the time interval by β so that from Eq. B.13 we obtain

$$P(\lambda\Delta x, \beta\Delta t) = \beta^{-H} P(\Delta x, \Delta t) \tag{B.14}$$

as long as $\lambda = \beta^H$. Fo $H = 1/2$ the scaling relation is exactly the same for the probability density of Brownian motion. Thus, we see that the *fBm*

is self-affine, with the random variables $\beta^H X(\beta t)$ and $X(t)$ having the same distribution.

Often in anomalous diffusive processes there is a change in the statistics of the variable (in comparison to the statistics of the variable given rise to normal diffusion) rather than a change in the variance, and a Lévy stable distribution is required to describe the process. Thus, anomalies in the diffusion emerge because of non-Gaussian statistics. For example, the random walker steps may involve power-law distributions of the form Eq. C.7, and thus we get Lévy-stable distribution statistics (Eq. C.6.

The scaling property of the Lévy distribution is obtained by scaling the displacement with λ and the time with β to obtain

$$P_{\mathrm{L}}(\lambda x, \beta t) = \beta^{-1/\mu} P_{\mathrm{L}}(x, t) \tag{B.15}$$

as long as $\lambda = \beta^{1/\mu}$. Note that Eq. B.15 has the same scaling form as *fBm* Eq. B.14 if the Lévy index μ is the same as $1/H$. Thus the self-affine scaling results from Lévy statistics with $0 < \mu < 2$ or from Gaussian statistics with a power-spectrum with $0 < H \leq 1$. The scaling relation is the same in both cases and therefore cannot be used to distinguish between the two. Note that Eq. B.15 and Eq. B.14 are equivalent only for $2 \geq \mu \geq 1$ for which $1/2 \leq \mu \leq 1$, indicating that the Lévy processes also give rise to anomalous diffusion (t^{α},$\alpha > 1$). Is for this range of anomaly, that we cannot distinguish between Gaussian statistics *fBm* or Lévy statistics underlying the process, unless we have some knowledge of the causing mechanism, that is, a well defined variable key on the process to be measured. When the Lévy index is in the interval $0 < \mu \leq 1$, the parameter H is greater than unity, and as unity is the greatest value for H in a diffusive process since it corresponds to ballistic motion, such scaling indicates a Lévy process rather than a *fBm*.

The statistics of intermittence, and more concretely *the time between reorientations in an animal's movement*, is a key variable mastering anomalous diffusion phenomena. Assuming large-scale animal movement as an intrinsic discrete process, one should try to identify re-orientation mechanisms, and compute the statistics of such variable in order to know if the underlying statistics giving rise to anomalous diffusion is due to a Lévy statistics or not.

C
The Lévy-stable distribution

C.1 The Lévy stable distribution

A collection of measurements or quantities, often called ensemble, can be described by a distribution function. This function, also called the *probability density*, associates a probability with the occurrence of an event (or quantity). Usually the average value of the quantity of interest has the greatest probability, and deviations from the average are less probable; the greater the deviation from the average, the less the probability. This regularity is observed in even the most random processes, for example, in the error associated with making a measurement. This regularity in the form of frequency distribution of errors refers to the "Law of Frequency of Error" or Gauss' Law of error and the normal (Gaussian) distribution. When distributions have finite *moments* to all orders (average, variance, kurtosis, etc.), these moments can be used to determine the probability density. The moments of a distribution $f(x)$ are defined as $\sum x^r f(x)dx$ where r is the moment order.

In some cases, such as for the bell-shaped (normal) distribution, the second moment (variance) is sufficient to determine the complete behavior of the ensemble. However, when distributions have sufficiently long-tails, the first few moments will not characterize the distribution because they diverge (they never stabilize, always increasing or decreasing). Distributions that with infinite moments characterize processes describe by non-integer exponents and contain surprises that run counter to our intuition.

An example of such distributions are inverse power-law distributions. Important empirical examples of inverse power-law distributions are the

Zipf's Law, where $f(n)$ is the relative frequency of the nth word in the order of the frequency of usage of that in a given language $f(n) \approx 1/n$; the distribution of the frequency of distribution of incomes in western societies, the Pareto's Law where $N(x)$ is the number of people with income x or larger and ν is a positive number (Pareto's exponent) then $N(x) \approx 1/n^{\nu}$, or the Latka's Law, inverse squared distribution of the percentage of authors publishing exactly n papers as a function of n ($N(n) \approx 1/n^2$)

The geometric notion of a fractal carries over into the domain of statistics in that it is the details of the random fluctuations in the process of interest that are revealed at smaller and smaller scales as the interval of observation is magnified. Paul Pierre Lévy (1886-1971) studied the general properties of such processes in the 1920s and 1930s and generalize the *Central Limit Theorem* to include those phenomena for which the second moment diverges. Among the varied and interesting properties of the *Lévy stable distribution* (process) is the fact that it satisfies a scaling law, indicating that the statistical fluctuations in the underlying process are self-similar.

A *stable* distribution is a distribution obtained by summing up (or averaging) variables with given distributions. The *Central Limit Theorem*, demonstrated by Carl F. Gauss, states that the sum (properly normalized) of many independent aleatory variables with distributions with finite first (mean) and second (variance) moments converge to the Gaussian distribution (which is a stable distribution). Indeed, in the *Generalized Central Limit Theorem*, Paul Lévy demonstrated that the Gaussian distribution is a particular case of a more general law. In the Generalized Central Limit Theorem it is stated that the sum of (properly normalized) of many independent variables (properly normalized) with any given distribution of either converging or diverging moments, converge to a family of distributions named Lévy stable distribution. For the case of variables with distributions with finite moments the tail of the Lévy stable distribution (the asymptotic behavior) converges to the Gaussian distribution. For the case of infinite moment distributions (such as inverse power-laws with exponent α) the asymptotic behavior of the Lévy stable distribution converges to an inverse power-law with exponent $\beta = \alpha - 1$. The whole probabilistic theory developed by P. Lévy can be studied from his main books (Lévy, 1925, 1937, 1948).

Here it follows a derivation of the Lévy stable distribution form (West, 1996) (see also (Montroll & Lebowitz, 2002):

Imagine a random walk whose steps become arbitrarily large, resulting in the divergence of the second moment. We can handle such an eventuality by changing the normalization on the random walk displacement series as

follows:

$$Y_N(t) = \frac{1}{N^{1/\mu}} \sum_{j=1}^{N} X_j \tag{C.1}$$

so that $X(t) = N^{1/\mu} Y_N(t)$ is the displacement of the random walk after N steps, $t = N\tau$, and μ is a positive parameter. Lévy asked: When is this distribution of the normalized sum $P(y)$ the same as that of the individual steps $p(x)$? Stated differently, this asks when is the character of the whole the same as that of it parts, which is a statistical form of requiring the process to be fractal.

One distribution for which the second moment diverges is the hyperbolic or inverse power-law distribution.

$$p(x) = \frac{c}{|x|^{\beta+1}}, \qquad 0 < \beta < 2 \tag{C.2}$$

where c is the normalization constant. The characteristic function, the Fourier transform of the probability density, corresponding to the fractal distribution Eq. C.2 is determined by using a Tauberian Theorem to be

$$\tilde{p}(k) \cong 1 - c|k|^{\beta} \cong e^{-c|k|^{\beta}} \tag{C.3}$$

for small k. The statistics of the random walk process can be determined by using the convolution theorem to obtain

$$\tilde{P}(k, N) = e^{-cN|k|^{\beta}} \tag{C.4}$$

which is the characteristic function for a random walk consisting of N steps carried out with the fractal distribution Eq. C.2 for the individual steps. In the continuum limit we can replace Eq. C.4 with

$$\tilde{P}(k, t) = e^{-\gamma t|k|^{\beta}} \tag{C.5}$$

so that the probability density, the inverse Fourier transform of the characteristic function, is

$$P_{\mathrm{L}}(y, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{iky} e^{-\gamma t|k|^{\beta}} dk \qquad 0 < \beta \leq 2 \tag{C.6}$$

which is the expression for the centro-symmetric Lévy stable distribution in one dimension. The Lévy distribution (Eq. C.6I is indeed, a family of distributions that have all the required properties of a probability distributions when $0 < \beta \leq 2$ with the cases $\beta = 1$ and $\beta = 2$ corresponding respectively to the Cauchy and Gauss distributions. In Montroll & Lebowitz

(2002) it is shown that for large values of the random variable, the Lévy distribution becomes an inverse power-law (Pareto) distribution

$$P_{\mathrm{L}}(y,t) \approx \frac{1}{|y|^{\mu+1}}, \qquad 0 < \mu < 2 \tag{C.7}$$

which has the same form as the distribution for the individual steps of the random walk with $\mu = \beta$.

Therefore, asymptotic behavior of Eq. C.6 for large $|y|$ is interesting in that it has the inverse power (Pareto) form for all $0 < \beta < 2$ ($P_{\mathrm{L}}(y,t) \approx 1/x^{\mu+1}$), with the Gauss case $\beta = 2$ being singular, having a quadratic exponential decay ($P_{\mathrm{L}}(y,t) \approx e^{x^2}$).

In practice, we generate Lévy type of random walks with uniform distributions for the turning angles and power-law distributions for the move lengths (i.e., $P(\ell) = \ell^{-\mu}$, exponent μ named Lévy index). The inverse power-law distributions for the move lengths represent the tail of the Lévy-stable distributions with Lévy stable index $\beta = \mu - 1$. If the power-law exponent μ lies in the interval $1 \leq \mu < 3$, then the Lévy stable distribution of the sums of such variables also has a power-law form, with exponent $\beta = \mu - 1$. For $\mu > 3$ the form of the Lévy stable distribution of the sums converges to a Gaussian distribution due to the Central Limit Theorem. Thus we recover Brownian motion for $\mu \geq 3$. The case of $\mu \leq 1$ does not correspond to normalizable distributions.

D
The stochastic time

In physical sciences, the notion of "probability" means quite specifically the *fraction of trials* that yield a particular *event* in the limit of infinitely many repeated trials. This "frequency interpretation" allows one to straightforwardly *derive*, not merely postulate, the following three *laws of probability* (Gillespie, 1992):

- *Range Law.* The probability of an event E is a real number $P(E)$ satisfying $0 \leq P(E) \leq 1$, with the circumstance $P(E) = 0$ corresponding to E *never* occurring, and the circumstance $P(E) = 1$ corresponding to E *always* occurring.

- *Addition law.* If $P(E)$ and $P(F)$ are the respective probabilities of two events E and F, and if these two events are *mutually exclusive* (i.e., they never occur together) then the probability of the event "either E or F" is $P(E \text{ or } F) = P(E) + P(F)$

- *Multiplication law.* If $P(E)$ is the probability of an event E, and $P(F/E)$ is the probability of an event F *given* that event E occurs, then the probability of the event "both E and F" is $P(EF) = P(E) \times Pr(F/E)$. In the frequency interpretation the conditional probability of F given E is the ratio of the number of trials yielding both events E and F to the number of trials yielding event E irrespective of event F, in the limit of infinitely many trials. When E and F are independent events, the probability of the event "both E and F" is $P(EF) = P(E) \times P(F)$.

In deterministic models time evolution is related to the transition rates of the specific events governing the system (i.e., parameters of the model

with dimension $[T^{-}1]$). In stochastic models, events are described as probabilities, and therefore, transition rates describing the dynamics are "transition rates of probabilities" or "probability rates" with no dimensionality. The *fundamental hypothesis* of stochastic formulation is that, instead of transition rates r of specific events, we should consider the probability rates of specific events in a given interval of time. The only *assumption* to be made by this computational method is that there exists a dt-independent scalar r_i such that:

> $r_i dt \equiv$ probability that a randomly selected element (or combination of elements) of the system at time t will experience a transition event i in the next infinitesimal time interval $[t, t+dt)$

The above three laws of probability, together with the *fundamental stochastic hypothesis* will allow us to relate the *probability rates of events* to deterministic *rates of events*.

Consider a system consisting of N elements. Each element can be in any one of S possible states $\{S_j\} = S_1, .., S_s$ (for example two states: occupied or available $s = 2$). This elements can experience M elementary transition events $\{E_i\} = E_1, E_2, ..., E_M$ (for example, extinction and colonization, $M = 2$). The specific event depending on their states. The state of the "whole system" at time t can be characterized by a vector of states $\vec{N}(t) = N_1, N_2, ..., N_s$ where $\{N_j\}$ represent the number of elements in state S_j, with $\sum N_j = N$, *while* the set of independent events $\{E_i\}$ describes the dynamics of the system.

The principle theoretical construct upon which this numerical procedure is based is not the *temporal evolution* of the *probability distribution function of abundances or densities* of certain states of the elements of the system (i.e., occupied/available patches), nor any of its derived quantities (i.e., first and second moments of that distribution) as it is done in the "master equation approach". Instead, the approach is based on the *temporal evolution* of what should be called a *reaction probability density function*, that is, the *temporal evolution* of the probability *that* a specific transition event *occurs in* a given interval of time.

This quantity is defined as:

> $p(\tau, E_i / \vec{N}(t), t) d\tau \equiv$ probability that given *the state of the system* $\vec{N}(t)$, the next event will occur after time τ in the infinitesimal time interval $d\tau$ and will be the event E_i.

In the language of probability theory, p is the joint density function for the two random variables "time to the next event" and "index of the next

event", with the possible values of these two random variables being represented respectively by the real variable $\tau(0 \leq \tau < \infty)$ and the integer variable $i(1 \leq i \leq M)$.

This probability can be calculated as the product of "the probability at time t that *no* event will occur in the interval $(t, t+\tau)$", times "the subsequent probability that only one event E_i will occur on the next differential time interval $(t+\tau, t+\tau+d\tau)$".

First, the "probability that exactly *one* event E_i occurs in the system during the interval $(t, t+dt)$" can be derived in the following way. Imagine that every element N in the system at time t is assigned a unique label. Then, we could distinguish a specific *class within all the elements that could undergo the event* E_i, let us denoted these elements h_i. By Eq. D, each of the elements h_i has probability $r_i dt$ of undergoing a transition event in $[t, t+dt)$. And by the range and addition laws, each has probability $1 - r_i dt$ of non-undergoing a transition event in $[t, t+dt)$. The multiplication law then implies that the probability that a *particular element* h_i *does* undergo a transition event E_i during the interval $[t, t+dt)$ while the others *do not*, is

$$\begin{aligned} r_i dt(1 - r_i dt)^{h_i - 1} &= r_i dt[1 - \sum_{i=1}^{h_i - 1} r_i dt + q(dt^2)] \\ &= r_i dt + o(dt) \end{aligned} \tag{D.1}$$

where we have scooped all the terms involving higher powers of dt (i.e., $(dt^j$ being $j > 1))$ into a single expression $o(dt) = \sum_{i=1}^{h_i - 1} r_i dt + q(dt^2)$:

$$r_i dt(1 - r_i dt)^{h_i - 1} = r_i dt + o(dt) \tag{D.2}$$

We can now use the addition law to calculate the probability that *any* of the h_i distinct elements at time t will undergo a transition event *alone* in $[t, t+dt)$ as the sum of their separate probabilities (because these events are mutually exclusive). Since each of the h_i terms in this probability sum is equal to $r_i dt + o(dt)$, then the sum is equal to

$$h_i[r_i dt + o(dt)] = r_i h_i dt + o(dt) \tag{D.3}$$

The term $o(dt)$ can be interpreted as the probability of *more than one event* occurring in the interval $[t, t+dt)$. From the multiplication law and Eq. D, we see that the probability for k transition events to occur in $[t, t+dt)$ must be proportional to $(dt)^k$, and therefore will be $o(dt)$ for $k > 1$.

Second, to calculate $p_0(\tau / \vec{N}(t), t)$, the probability that no event will occur in $[t, t+\tau)$, we start by considering again the fact that each of the

elements in the set h_i has a probability of non experiencing a transition event in the interval $[t, t+dt)$ of $1 - r_i dt$. So by the multiplication law, the probability that *none of the* h_i sites will undergo a transition even E_i in $[t, t+dt)$ is

$$(1 - r_i dt)^{h_i} = 1 - r_i h_i dt + o(dt) \tag{D.4}$$

Therefore, appealing once again to the multiplication law, the probability that no E_1 transition will occur *and* no E_2 transition...*and* no E_M transition will occur in $[t, t+dt)$, is

$$\prod_{i=1}^{M}(1 - r_i h_i dt + o(dt)) = 1 - \sum_{i=1}^{M} r_i h_i dt + o(dt) \tag{D.5}$$

This is valid for infinitesimal intervals of time $[t, t+dt)$. However, $p_0(\tau/\vec{N}(t), t)$ is the probability that no event will occur in the interval $[t, t+\tau)$. To compute p_0 we have to imagine the interval $[t, t+\tau)$ to be subdivided into k subintervals of equal length $\epsilon = \tau/k$. The probability that *none* of the events $E_i, ..., E_M$ occurs in the first subinterval $(t+\epsilon)$ is,

$$\prod_{i=1}^{M}(1 - r_i h_i \epsilon + o(\epsilon)) = 1 - \sum_{i=1}^{M} r_i h_i \epsilon + o(\epsilon) \tag{D.6}$$

This is also valid for the subsequent probability that no reaction occurs in $(t+\epsilon, t+2\epsilon)$, and then in $(t+2\epsilon, t+3\epsilon)$ and so on. Since there are k such ϵ subintervals between $(t+\tau)$, then

$$p_0(\tau/\vec{N}(t), t) = [1 - \sum_{i=1}^{M} r_i h_i \epsilon + o(\epsilon)]^k \tag{D.7}$$

Finally, we have that the *reaction probability density function* can be expressed as,

$$p(\tau, E_i/\vec{N}(t), t)d\tau = [1 - R(M)\epsilon + o(\epsilon)]^k [r_i h_i d\tau + o(d\tau)], \tag{D.8}$$

where $R(M)$ is the total transition rate at time t,

$$R(M) = \sum_{i=1}^{M} r_i h_i. \tag{D.9}$$

Dividing this equation by $d\tau$ and taking the limit $d\tau \to 0$, we get

$$p(\tau, E_i/\vec{N}(t), t) = [1 - R(M)\epsilon + o(\epsilon)]^k r_i h_i. \tag{D.10}$$

We can rewrite the first factor on the right as

$$
\begin{aligned}
[1 - R(M)\epsilon + o(\epsilon)]^k &= \left(1 - \frac{R(M)k\epsilon + ko(\epsilon)}{k}\right)^k \\
&= \left(1 - \frac{R(M)\tau + \tau[o(\epsilon)/\epsilon]}{k}\right)^k. \qquad \text{(D.11)}
\end{aligned}
$$

where the last step uses the fact that $k\epsilon = \tau$. Now letting $k \to \infty$, and noting that $o(\epsilon)/\epsilon \to 0$ in that limit, we get

$$
\lim_{k\to\infty}[1 - R(M)\epsilon + o(\epsilon)]^k = \lim_{k\to\infty}\left(1 - \frac{R(M)\tau}{k}\right)^k. \qquad \text{(D.12)}
$$

The last limit, is the standard limit formula for the exponential function so we can write,

$$
\lim_{k\to\infty}\left(1 - \frac{R(M)\tau}{k}\right)^k = e^{-R(M)\tau}. \qquad \text{(D.13)}
$$

Substituting this result in Eq. D.10, and then multiplying and dividing by $R(M)$, we conclude that the function p defined in Eq. D.8 is given by

$$
p(\tau, E_i/\vec{N}(t), t) = e^{-R(M)\tau}\frac{r_i h_i}{R(M)}. \qquad \text{(D.14)}
$$

Eq. D.15 provides the mathematical basis for the stochastic simulation algorithm. This equation says that the "time to next reaction" and "what reaction will be" are statistically independent random variables, the former having the exponential density function with decay constant $R(M)$, and the latter having the integer density function $r_i h_i / R(M)$. So we can express p as,

$$
p(\tau, E_i/\vec{N}(t), t) = p(\tau) \times p(i). \qquad \text{(D.15)}
$$

To have exact stochastic simulations in "well stirred" systems (non spatial effects are considered), we have to generate only two random numbers x_1 and x_2. The first one is used to randomly sample the negative exponential distribution

$$
p(\tau) = e^{-R(M)\tau}, \qquad \text{(D.16)}
$$

in order to obtain τ, the "time to next reaction" or "inter-event time". By using the *inversion method* to sample probability distributions (see General Appendix A) we can take τ to be

$$\tau = -\frac{\ln x_1}{R(M)}. \tag{D.17}$$

The second random number is used to obtain i, that is, which specific event will occur. This can be done by taking i as the smallest integer for which

$$\sum_{i=1}^{M} \frac{r_i h_i}{R(M)} > x_2. \tag{D.18}$$

In spatially extended systems the the stochastic algorithm must also take into account the place "where the next reaction will occur" (Alonso & McKane, 2002; Alonso, 2003), and thus the *reaction probability density function* is not used as defined here. However, as time to next interaction is independent of "where" and "which" reaction will occur, the spatially extended versions of the stochastic algorithm also make use of the inter-event time negative exponential probability distribution originally obtained by employing the *reaction probability density function* concept.