The profound implications of the central limit theorem applied to rarefied sediment transport

David Jon Furbish

Emeritus, Vanderbilt University
May 2024

1 Preamble

The *central limit theorem* is among the crown jewels of mathematics and science. The proof of this theorem is elegant. It is a central element of numerous concepts and methods in probability and statistics. And it provides the basis for clarifying the physical behavior of certain random processes, notably the effects of adding increments represented by a random variable, as in the random-walk displacements of particles or more generally stochastic changes in the state of a system. Herein we will see that the central limit theorem points to why the Gaussian distribution is ubiquitous in many science problems, notably sediment transport.

Appendix A contains a proof of the classical central limit theorem. Here we present this theorem in practical terms. We start with the meaning of independent and identically distributed random variables, and a description of the law of large numbers, focusing on continuous random variables.

2 The Meaning of Independent and Identically Distributed

Let $x_1, x_2, ..., x_i, ..., x_n$ denote a set of random variables. In the simplest and perhaps most familiar case these might represent the values of a sample of size n. We then often assume (or assert) that the values x_i are independent and identically distributed, perhaps shortening this by saying that the x_i are i.i.d. or iid. There are two ways to think about what i.i.d. means, although the practical outcome is essentially the same.

Suppose that a stochastic process produces a measurable quantity, a random variable x, that is distributed as $f_x(x)$. By this we mean the following. First, consider a simple static system consisting of a nominally homogeneous granular material so that x represents the local porosity of the material measured with a small fixed sampling volume V_s. We then envision a Gibbs great number of such measurements systematically taken over the entirety of the granular material, and we discover that the values of x are distributed as $f_x(x)$. Or, consider a dynamical system where the state x varies as a time series. We systematically measure this state at a great number of instants over an arbitrarily long interval of time, and we discover in a frequentist sense that the values are distributed as $f_x(x)$. Alternatively, whether the system is static or varying with time, we can imagine that the distribution $f_x(x)$ is precisely defined by the sample space of x. Thus, in all cases we are envisioning that $f_x(x)$ is fixed (stationary) and well defined.

Now consider a sample-centric view of i.i.d. In this case the set $x_1, x_2, ..., x_i, ..., x_n$ represents a sample of size n drawn from the distribution $f_x(x)$, where the subscript denotes the individual measurement. Because each x_i is drawn from precisely the same distribution, by definition the values x_i are identically distributed. The idea of independence, however, is potentially more involved. Consider the example where x denotes the porosity of a homogeneous granular material. If we design an algorithm to randomly select locations for measuring x, then in this sense the values are independent. But if by chance some of these locations are close together such that the sampling volumes V_{s} overlap, then the measured values of x are not necessarily independent in a physical (or mathematical) sense given that the values x are likely to be spatially correlated. Likewise, we might randomly select values of x from a time series. But if the measured values are close together in the series or occur with a frequency close to a rational multiple of the frequency of a periodic series, then the values of x are not necessarily independent in a physical (or mathematical) sense in the presence of temporal correlation. These simple examples offer a hint of the fact that sampling is a grand challenge. For our purposes here we should be aware that the idea of random sampling to achieve i.i.d. is not necessarily straightforward.[1] And in practice, absent a careful experimental design, the i.i.d. assumption often is not satisfied. Nonetheless, we can readily envision the idealization of i.i.d.

Consider, then, an ensemble-centric view of i.i.d. We imagine a Gibbs ensemble — a great number of independent but nominally (statistically) identical systems. In the example of a granular material, no two systems are identical at the particle scale, but all are indistinguishable in a statistical sense when viewed at a larger scale. In the example of a dynamical system, no two time series of x are identical, but the statistical structure of the time series of each system is the same. In this manner each x_i is drawn from a randomly selected member of the ensemble, thus ensuring independence. That is, x_1 is drawn from the distribution $f_{x_1}(x_1)$, x_2 is drawn from the distribution $f_{x_2}(x_2)$, and so on. These distributions have means $\mu_{x_1}, \mu_{x_2}, ..., \mu_{x_n}$ and variances $\sigma_{x_1}^2, \sigma_{x_2}^2, ..., \sigma_{x_n}^2$. Now the subscript refers to the randomly selected system of the ensemble. Because the systems are nominally identical, $f_{x_1}(x_1) = f_{x_2}(x_2) = ... = f_{x_n}(x_n) = f_x(x)$ such that $\mu_{x_1} = \mu_{x_2} = ... = \mu_{x_n} = \mu_x$ and $\sigma_{x_1}^2 = \sigma_{x_2}^2 = ... = \sigma_{x_n}^2 = \sigma_x^2$, thus ensuring that the x_i are identically distributed. Notice that this outcome is identical to that of the sample-centric view above. This ensemble-centric view if i.i.d. serves us well in developing the mathematics of the law of large numbers and the central limit theorem (Appendix A), although it does not address practical problems associated with sampling. In what follows we assume the i.i.d. assumption is satisfied as a canonical starting point.

3 Law of Large Numbers

Here we focus on the weak law of large numbers (Appendix A). Let $x_1, x_2, ..., x_n$ denote a set of independent and identically distributed random variables, each with mean μ_x and variance σ_x^2. We then define a new random variable as the sum

$$S_n = \sum_{i=1}^{n} x_i \, . \tag{1}$$

[1]The problem of random sampling to achieve i.i.d. is a large topic. Sampling in the social sciences and medicine is particularly challenging, and there are numerous recognized types of sampling bias. Sampling in the physical sciences is usually tuned to the nature of the data and its sources, with a great variety of approaches across fields. In the field of sediment transport we must contend with such things as data censorship, effects of serial correlation and non-stationarity, instrument bias and filtering, resolution bias, and so on.

Because the x_i are independent and identically distributed, the expected value of the sum S_n is $\mu_{x_1} + \mu_{x_2} + ... + \mu_{x_n} = n\mu_x$. In turn, because the variances of x_i are additive with zero covariance, the variance of the sum S_n is $\sigma_{x_1}^2 + \sigma_{x_2}^2 + ... + \sigma_{x_n}^2 = n\sigma_x^2$. To be clear, this is the variance of S_n, not the variance of the values x_i. That is, we can imagine creating a large number N of sums S_n. Then $n\sigma_x^2$ is the variance of these sums. This becomes important when we examine the central limit theorem below.

We now define a new random variable as

$$\overline{x}_n = \frac{S_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \, , \tag{2}$$

which is just the arithmetic average of the set of x_i, where the subscript n associates the average with the size of the set. The (weak) law of large numbers (Appendix A) then tells us that

$$\lim_{n\to\infty} \overline{x}_n = \lim_{n\to\infty} \frac{S_n}{n} = \mu_x \, . \tag{3}$$

A simple way to visualize this is as follows. Envision a random sample of x_i with relatively small n giving a value $\overline{x}_n$. By chance the value x_n does not coincide with μ_x. Repeating this N times, no two values of x_n are the same and each is similar but not identical to the mean μ_x. As n increases, individual values of x_n are on average closer to μ_x. As n approaches infinity, each random sample of x_i now closely represents all possible values of x in proportion to the occurrence of x represented by the distribution $f_x(x)$. Indeed, a histogram of the values x_i is virtually indistinguishable from the smooth distribution $f_x(x)$. As a consequence the value $\overline{x}_n$ (and indeed all N values of $\overline{x}_n$) converge to the mean μ_x.

Here we have focused on the convergence of $\overline{x}_n$ to the mean $\mathrm{E}(x) = \mu_x$. However, an expectation is not limited to the mean. The law of large numbers also applies to other expectations, for example the variance, such that the sample estimate $s_n^2 \to \sigma_x^2$ in the limit of $n \to \infty$. Moreover, whereas we assume that the values x_i are identically distributed, the law of large numbers is at work even if successive values are not independent, so long as the set x_i represents all possible values of x in the proportions represented by $f_x(x)$. Thus, systematic successive measurements of a stationary time series give a value $\overline{x}_n$, essentially a time average, that converges to μ_x in the limit of $n \to \infty$ and thus as $t \to \infty$. Also note that the law of large numbers applies in certain problems where the distribution of the random variable is itself a function of the averaging time, if the mean is fixed. In fact this is implied by the convergence of realizations of the particle flux $\hat{q}_{nx}$ to the expected value $\langle q_{nx} \rangle$ (Figure 1).

4 Classical Central Limit Theorem

Reconsider the sum S_n given by (1). The expected value of this sum is $n\mu_x$ and the variance is $n\sigma_x^2$. This implies that

$$n\sigma_x^2 = \frac{1}{N}\sum_{j=1}^{N} \left(S_{nj} - n\mu_x\right)^2 \, , \tag{4}$$

where we are appealing to the law of large numbers in letting N be arbitrarily large. Dividing (4) by n^2,

$$\frac{\sigma_x^2}{n} = \frac{1}{N}\sum_{j=1}^{N} \left(\frac{S_{nj}}{n} - \mu_x\right)^2 = \frac{1}{N}\sum_{j=1}^{N} (\overline{x}_{nj} - \mu_x)^2 = \sigma_{\overline{x}_n}^2 \, . \tag{5}$$

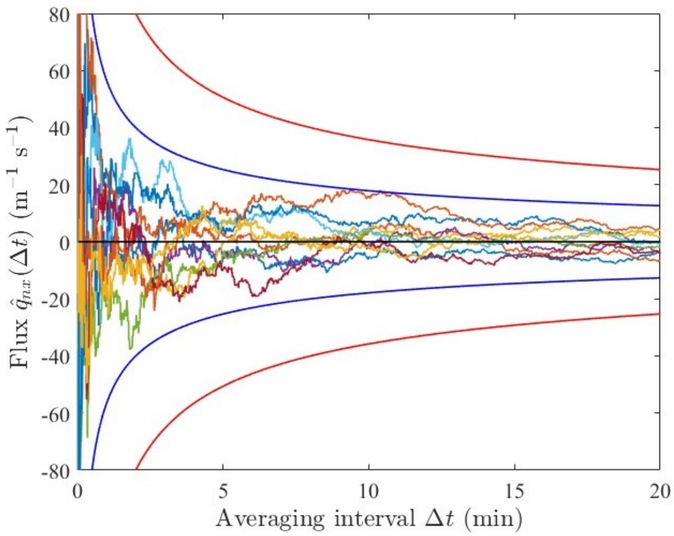

Figure 1: Plot of high-fidelity numerical simulations of 10 realizations of the particle number flux $\hat{q}_{nx}(\Delta t)$ [L^{-1} T^{-1}] associated with rain splash on a horizontal surface during steady rainfall, showing (black line) expected flux $\langle q_{nx} \rangle = 0$ together with (blue lines) ± 1 and (red lines) ± 2 standard deviations in the values of $\hat{q}_{nx}(\Delta t)$ about the expected value. Each realization is calculated as the net number $N(\Delta t)$ of particles crossing a position x per length Δy normal to x, per averaging interval Δt, namely, $\hat{q}_{nx}(\Delta t) = N(\Delta t)/\Delta y \Delta t$. Moreover, each realization arises from *precisely the same* controlling factors: the rainfall intensity, surface slope, particle size and so on. Indeed, these are examples of an infinite set of possible realizations for the same controlling factors. Similar results are obtained when the expected flux $\langle q_{nx} \rangle$ is finite with nonzero surface slope.

More formally (Appendix A), the central limit theorem says that the variance $\sigma^2_{\overline{x}_n}$ of the average $\overline{x}_n$ is

$$\sigma^2_{\overline{x}_n} = \frac{\sigma^2_x}{n} \, . \tag{6}$$

Moreover, this theorem gives the remarkable result that the average $\overline{x}_n \sim \mathcal{N}(\mu_x, \sigma^2_x/n)$, regardless of the form of the underlying distribution $f_x(x)$, so long as its mean μ_x and variance σ^2_x are defined. This result also holds for discrete distributions $p_x(x)$. We must note, however, that the sampling distribution $f_{\overline{x}_n}(\overline{x}_n)$ of $\overline{x}_n$ is only approximately Gaussian. In a strict sense this sampling distribution converges to a normal distribution only in the limit of $n \to \infty$, where in general the central part of the distribution in the vicinity of the mean converges faster than the tails.

A simple way to visualize this outcome of the central limit theorem is to reconsider the sampling described above in relation to the law of large numbers. Envision a large number N of random samples of x_i, each with modest size n. The histogram of each sample crudely mimics the smooth distribution $f_x(x)$. But the average $\overline{x}_n$ of each sample, a single number, reveals nothing about the appearance of the histogram from which it is obtained. The N averages $\overline{x}_n$ are centered about the mean μ_x, and a histogram of these averages roughly mimics a Gaussian distribution. As the sample size n increases, this histogram formed from the N averages x_n becomes virtually indistinguishable from a smooth Gaussian distribution and is increasingly centered on μ_x with deceasing variance.

5 Implications

5.1 A Statistics Aside

Those familiar with parametric statistics will recognize that (6) is the basis for defining the so-called standard error of the estimate $\overline{x}_n$ of the mean μ_x defined as $SE = \left(\sigma^2_{\overline{x}_n}\right)^{1/2} = \sigma_{\overline{x}_n}$. Thus,

$$SE = \frac{\sigma_x}{\sqrt{n}} \, . \tag{7}$$

Because we rarely know the precise values of the mean and variance of the underlying distribution $f_x(x)$, the standard deviation σ_x is in practice replaced with the sample estimate s_x of the standard deviation, and we then form "error bars" and a "confidence interval" using the sample average, for example, $\overline{x}_n \pm SE$ or $\overline{x}_n \pm 2SE$ with $SE = s_x/\sqrt{n}$. Because of the popularity of the standard error in statistical analyses, let us briefly note several items that should not go unsaid before returning to our main objectives.

First, the word error is a misnomer. The quantity $\sigma_{\overline{x}_n}$ is just the standard deviation of the random variable $\overline{x}_n$. Unless it specifically refers to, say, measurement error, this standard deviation has nothing to do with errors. Rather, it reflects the natural variability in the values of x, characterized by the distribution $f_x(x)$, leading to variability in the average $\overline{x}_n$ calculated from n values of x. To view the quantity $\sigma_{\overline{x}_n}$ as error reflects a 20th century style of frequentist thinking preoccupied with the perceived importance of expected values in attempting to codify our understanding of things based on what can be concluded from dichotomous hypothesis testing involving arbitrary thresholds of "significance" — versus engaging with variability as an inherent feature of a system, including its physical basis. Second, to claim that a confidence interval as defined above gives a probability (e.g. a 95% "confidence") that the "true" mean μ_x falls within the interval is wrong. Yet hundreds of scientists make this mistaken claim every year. In fact, with probability equal to one the mean μ_x is either within or outside the specified interval, and one can never know this unless the mean μ_x is known *a priori*. Such intervals pertain to assessments of sampling based on the properties of the sampling distribution $f_{\overline{x}_n}(\overline{x}_n)$, not to assessments of the location of the mean μ_x. Third, a rule of thumb is often offered to the effect that if $n \geq 30$ then the sampling distribution $f_{\overline{x}_n}(\overline{x}_n)$ is sufficiently well approximated by a Gaussian distribution that calculations of probability based on the Gaussian can be trusted in (statistical) inferential matters. In fact, this rule of thumb has no justification and can lead to flawed analyses. These items are among the many "don'ts" of statistics (Furbish and Schmeeckle, 2020). On the other hand, when combined with other techniques, and depending on availability of information regarding the distribution $f_x(x)$, the so-called standard error can be valuable in sampling design to anticipate sample sizes needed to achieve a desired convergence of the estimate $\overline{x}_n$ to the mean μ_x. We now set this topic aside to focus on probabilistic implications of the central limit theorem.

5.2 Emergence of Gaussian Behavior

Consider a physical problem that illustrates the elements of the central limit theorem in a tangible manner related to particle transport. First, recall that the expected value of the sum S_n given by (1) is equal to $n\mu_x$ and the variance of this sum is $n\sigma_x^2$. We presented the result that, according to the central limit theorem, the average $\overline{x}_n = S_n/n \sim \mathcal{N}(\mu_x, \sigma_x^2/n)$. But because $\overline{x}_n$ and S_n differ only by the specified factor n as $S_n = n\overline{x}_n$, this also implies that $S_n \sim \mathcal{N}(n\mu_x, n\sigma_x^2)$. That is, the sum S_n is approximately Gaussian with sufficiently large n. We return to this point momentarily.

Consider the stochastic one-dimensional movement of a particle parallel to the y axis. We choose a fixed interval of time Δt during which the particle is displaced by an amount x_i. Note that the instantaneous velocity of the particle might fluctuate during Δt. Nonetheless we can define its velocity as $v_{\mathrm{p}i} = x_i/\Delta t$, which is an average over Δt. Thus the displacement $x_i = u_{\mathrm{p}i}\Delta t$. As above, consider the set of random variables $x_1, x_2, ..., x_n$. These now represent a set of n particle displacements such that $f_x(x; \Delta t)$ denotes the distribution of possible displacements with mean μ_x. Notice that we have added Δt after a semicolon in the functional notation to emphasize that this distribution is specific to the interval Δt. Further notice that $x_1, x_2, ..., x_n = v_{\mathrm{p}1}\Delta t, v_{\mathrm{p}2}\Delta t, ..., v_{\mathrm{p}n}\Delta t$. Now the subscripts $i = 1, 2, ...$ imply successive displacements in time.

Now reconsider the sum S_n. A particle starting at the initial position $y(0) = 0$ at time $t = 0$ must be at a position $y(t) = S_n$ after n displacements. That is,

$$S_n = y(t) = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} u_{\mathrm{p}i}\Delta t = \Delta t \sum_{i=1}^{n} u_{\mathrm{p}i}\,. \tag{8}$$

Moreover, because each displacement occurs over an interval Δt, the total time $t = n\Delta t$. This also implies that the expected value of $y(t)$ is $\mathrm{E}[y(t)] = n\mu_x$ and the variance is $\mathrm{V}[y(t)] = n\sigma_x^2$. In turn, we expect from the central limit theorem that

$$f_y(y,t) = \frac{1}{\sqrt{2\pi \mathrm{V}[y(t)]}} \exp\left[-\frac{(y - \mathrm{E}[y(t)])^2}{2\mathrm{V}[y(t)]}\right] = \frac{1}{\sqrt{2\pi n\sigma_x^2}} \exp\left[-\frac{(y - n\mu_x)^2}{2n\sigma_x^2}\right]. \tag{9}$$

We now write the expected displacement as $\mathrm{E}[y(t)] = n\mu_x = n(\mu_x/\Delta t)\Delta t$. Here, $\mu_x/\Delta t = \langle v_{\mathrm{p}}\rangle$ is the expected particle velocity so $\mathrm{E}[y(t)] = \langle v_{\mathrm{p}}\rangle t$ with $n\Delta t = t$. In turn we write the variance as $\mathrm{V}[y(t)] = n\sigma_x^2 = n(\sigma_x^2/\Delta t)\Delta t$. By convention we denote $\sigma_x^2/\Delta t = 2\kappa_y$ where κ_y denotes the particle diffusivity, so $\mathrm{V}[y(t)] = 2\kappa_y t$. Substituting these expressions into (9) then gives

$$f_y(y,t) = \frac{1}{\sqrt{4\pi\kappa_t t}} \exp\left[-\frac{(y - \langle v_{\mathrm{p}}\rangle t)^2}{4\kappa_y t}\right]. \tag{10}$$

That is, the distribution of particle positions y is a Gaussian distribution whose mean and variance increase linearly with time t.

It can be shown that (10) is a solution of a Fokker–Planck-like equation — an advection–diffusion equation — having the form,

$$\frac{\partial f_y(y,t)}{\partial t} = -\langle v_{\mathrm{p}}\rangle \frac{\partial f_y(y,t)}{\partial y} + \kappa_y \frac{\partial^2 f_y(y,t)}{\partial y^2}\,. \tag{11}$$

This expression is obtained from a master equation, a general probabilistic description of the time evolution of the distribution $f_y(y,t)$ of particle states y, taking into account the difference in the rate at which particles arrive at the position y at time t from all possible preceding positions, and the rate at which particles leave the position y. Importantly, this derivation involving a master equation makes no reference to the central limit theorem. Let us now highlight several points.

First, aside from the kinematic definitions of the average particle velocity $\langle v_{\mathrm{p}}\rangle$ and the diffusivity κ_y, the arguments leading to (10) involve no physics. The result embodied in (10) therefore is simply the *necessary* probabilistic outcome of adding random increments. Thus, we should not be surprised that (10), representing a solution of (11), is independently obtained from a master equation. Moverover, the displacements x_i are not restricted to particle displacements. These displacements more generally can refer to changes in a measure of the state of a system. For example, if the displacements x_i represent incremental changes in the local elevation $y(t)$ of a granular surface, then the average velocity represents the expected erosion or deposition rate and the diffusivity characterizes the noisiness of fluctuations about the expected rate. The key lesson is this: Owing to the central limit theorem we should anticipate the likely appearance of Gaussian behavior in diverse stochastic systems, regardless of the detailed physics involved.

Second, the probabilistic arguments leading to (10) are scale independent. That is, as probabilistic constructs (10) and (11) equally pertain to the behavior of Brownian particles and sediment particles. The physics of these different systems are then distinguished by the specific ingredients of the expected velocity $\langle v_\mathrm{p} \rangle$ and the diffusivity κ_x.

Third, in the specific case of sediment particle states y, the descriptions of behavior provide by (10) and (11) represent a particle-centric view of things. This view is entirely agnostic to the presence or absence of continuum conditions. These equations describe the probabilistic ensemble behavior (sensu Gibbs, 1902) of an individual particle having little to do with a continuum, or they may equally apply to a great number of particles in any realization (Schumer et al., 2009; Furbish et al., 2018; Furbish and Doane, 2021).

Consider the form of the distribution $f_x(x; \Delta t)$ of displacements x within the context of remarks offered by the celebrated probabilist and physicist Edwin Jaynes (2003, p. 206):

> This is just the process by which noise is produced in Nature — by addition of many small increments, one at a time... Once a Gaussian form is obtained, it is preserved; this process can be stopped at any point, and the resulting final distribution still has the Gaussian form. What is at first surprising is that this stable form is independent of the distribution... of the small increments.

That is, the emergence of the Gaussian and its persistence through time is remarkably insensitive to the form of $f_x(x; \Delta t)$. For example, if $f_x(x; \Delta t)$ has a mean $\mu_x = 0$, then the expected value $\mathrm{E}[y(t)] = \langle v_\mathrm{p} \rangle t = 0$ with $\langle v_\mathrm{p} \rangle = 0$. Now (10) becomes

$$f_y(y, t) = \frac{1}{\sqrt{4\pi\kappa_t t}} \exp\left[-\frac{y^2}{4\kappa_y t}\right], \tag{12}$$

where the associated Fokker–Planck-like equation is a diffusion equation,

$$\frac{\partial f_y(y, t)}{\partial t} = \kappa_y \frac{\partial^2 f_y(y, t)}{\partial y^2}. \tag{13}$$

Here, $f_x(x; \Delta t)$ might be a symmetrical Gasussian distribution or a symmetrical Laplace distribution; or $f_x(x; \Delta t)$ could be asymmetrical with $\mu_x = 0$. In these examples the displacements x are both positive and negative. If instead the distribution $f_x(x; \Delta t)$ involves only positive displacements or only negative displacements, then the expected value $\mathrm{E}[y(t)] = \langle v_\mathrm{p} \rangle t$ is finite with $\langle v_\mathrm{p} \rangle \neq 0$. Now the behavior of $f_y(y, t)$ is described by (10) and (11), and we say that $\langle v_\mathrm{p} \rangle$ represents a *drift speed* or an advective speed. Here the distribution $f_x(x; \Delta t)$ might be an exponential distribution or a gamma distribution.

In this same vein, it is well known that the binomial and Poisson distributions converge to the Gaussian distribution with a large number of trials n (binomial) or large time t (Poisson). Numerous formal proofs of this are available. In the case of the binomial distribution we may think of this distribution as representing the possible set of integer-valued displacements occurring after n trials, analogous to the distribution of displacements $f_x(x; \Delta t = t/n)$. In the case of the Poisson distribution we may think of this distribution as representing the possible set of integer-valued displacements occurring during a specified interval Δt, again analogous to the distribution of displacements $f_x(x; \Delta t)$. The sum S_n in (1) or (8) thus consists of a set of binomial or Poisson distributed increments. In both case the arguments presented above lead to the conclusion that the binomial and Poisson distributions converge to the Gaussian distribution.

Because the velocity is defined as $v_\mathrm{p} = x/\Delta t$ so that $x = v_\mathrm{p}\Delta t$, the distribution of velocities $f_{v_\mathrm{p}}(v_\mathrm{p}; \Delta t)$ has the same form as the distribution $f_x(x; \Delta t)$ of displacements x. It is therefore equally correct to claim that Gaussian behavior reflects varying particle velocities. Let us also recall that the developments above assume that $f_x(x; \Delta t)$ has finite mean and variance. Things change if $f_x(x; \Delta t)$ is heavy-tailed with undefined moments. For simplicity we also assumed continuous particle motions. In the case of bed load particles the effects of rest times must be addressed in describing displacements.

Let us end with a simple but profound truth attributable to the central limit theorem. Sediment particles experience varying velocities and displacements during transport — a hallmark of their behavior. Regardless of the detailed physics involved, formulations of rarefied sediment transport that do not explicitly acknowledge the existence and effects of particle diffusion are wrong. Particle diffusion is an inherent feature of transport — a probabilistic consequence of varying particle velocities and displacements — and its effects must figure into formulations of transport.

A Proof of the Central Limit Theorem

There are several proofs of the central limit theorem. Here we appeal to proofs involving the moment generating function and the characteristic function of a random variable. The moment generating function and the characteristic function provide ways to analytically work with probability distributions and their moments as alternatives to working directly with the probability density functions. We start with proofs of the weak law of large numbers, followed by proofs of the classical central limit theorem.[2]

A.1 Moment Generating Function

Let x denote a random variable distributed as $f_x(x)$ with mean μ_x and variance σ_x^2. The moment generating function of x is defined as

$$M_x(t) = \mathrm{E}(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f_x(x)\,\mathrm{d}x\,, \tag{14}$$

for argument t (not to be confused with time). Notice that, according to the law of the unconscious statistician, $M_x(t)$ is just the expected value of e^{tx}. Also notice that (14) is like a Fourier transform,

[2]For typical problems the weak law and the strong law of large numbers give the same conclusion; the nature of the convergence differs. Similarly, various versions of the central limit theorem differ in how they are obtained, and in the assumptions regarding whether the random variables are independent and identically distributed.

but without the imaginary number i. For later reference, the moment generating function of a constant a is $M_a(t) = \mathrm{E}(e^{ta}) = e^{ta}$.

We now expand e^{tx} as a Taylor series,

$$e^{tx} = 1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + ... + \frac{t^nx^n}{n!} + ... \tag{15}$$

Substituting this into (14),

$$M_x(t) = \int_{-\infty}^{\infty} \left(1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + ... + \frac{t^nx^n}{n!} + ...\right) f_x(x)\,\mathrm{d}x\,. \tag{16}$$

From this it is clear that

$$\begin{aligned} M_x(t) = \mathrm{E}(e^{tx}) &= 1 + t\mathrm{E}(x) + \frac{t^2\mathrm{E}(x^2)}{2!} + \frac{t^3\mathrm{E}(x^3)}{3!} + ... + \frac{t^n\mathrm{E}(x^n)}{n!} + ... \\ &= 1 + tm_1 + \frac{t^2m_2}{2!} + \frac{t^3m_3}{3!} + ... + \frac{t^nm_n}{n!} + ...\,, \end{aligned} \tag{17}$$

where $m_n = \mathrm{E}(x^n)$ denotes the nth moment about the origin. In turn,

$$m_n = \mathrm{E}(x^n) = M_x^{(n)}(0) = \left.\frac{\mathrm{d}^nM_x}{\mathrm{d}t^n}\right|_{t=0}\,. \tag{18}$$

To see that this is the case, simply take the nth derivative of (17) with respect to t then set $t = 0$. This provides an algorithm for computing the nth moment of a distribution, hence the name moment generating function.

Note that not all distributions have moment generating functions. One example is the log-normal distribution, although all moments of the log-normal distribution exist.

A.2 Characteristic Function

Let x denote a random variable distributed as $f_x(x)$ with mean μ_x and variance σ_x^2. The characteristic function of x is defined as

$$\phi_x(t) = \mathrm{E}(e^{itx}) = \int_{-\infty}^{\infty} e^{itx} f_x(x)\,\mathrm{d}x\,, \tag{19}$$

for argument t. Notice that, according to the law of the unconscious statistician, $\phi_x(t)$ is just the expected value of e^{itx}. Also notice that (19) is essentially a Fourier transform involving the imaginary number i defined by $i^2 = -1$. For later reference, the characteristic function of a constant a is $\phi_a(t) = \mathrm{E}(e^{ita}) = e^{ita}$.

We now expand e^{itx} as a Taylor series,

$$e^{itx} = 1 + itx - \frac{t^2x^2}{2!} - \frac{it^3x^3}{3!} + ... \tag{20}$$

Substituting this into (19),

$$\phi_x(t) = \int_{-\infty}^{\infty} \left(1 + itx - \frac{t^2x^2}{2!} - \frac{it^3x^3}{3!} + ...\right) f_x(x)\,\mathrm{d}x\,. \tag{21}$$

From this it is clear that

$$\phi_x(t) = \mathrm{E}(e^{itx}) = 1 + it\mathrm{E}(x) - \frac{t^2\mathrm{E}(x^2)}{2!} - \frac{it^3\mathrm{E}(x^3)}{3!} + ...$$

$$= 1 + itm_1 - \frac{t^2 m_2}{2!} - \frac{it^3 m_3}{3!} + ... \,, \tag{22}$$

where $m_n = \mathrm{E}(x^n)$ denotes the nth moment about the origin.

As a Fourier transform of a probability density function, the characteristic function completely specifies the density function. In contrast to the moment generating function, the characteristic function exists for all distributions of real-valued random variables.

A.3 Law of Large Numbers

A.3.1 Proof Involving the Moment Generating Function

Let $x_1, x_2, ..., x_n$ denote a set of independent and identically distributed random variables, each with mean μ_x and variance σ_x^2. We now define a new random variable as the sum

$$S_n = \sum_{i=1}^{n} x_i \,. \tag{23}$$

We then want to show that the moment generating function $M_{S_n}(t)$ of S_n is

$$M_{S_n}(t) = [M_x(t)]^n \,, \tag{24}$$

where $M_x(t)$ is the moment generating function of x given by (17). To do this we use the definition of the moment generating function and write

$$M_{S_n}(t) = \mathrm{E}(e^{tS_n}) = \mathrm{E}\left[e^{t(x_1+x_2+...+x_n)}\right] = \mathrm{E}(e^{tx_1}e^{tx_2}...e^{tx_n}) \,. \tag{25}$$

Because the x_i are independent,

$$M_{S_n}(t) = \mathrm{E}_{x_1}(e^{tx_1})\mathrm{E}_{x_2}(e^{tx_2})...\mathrm{E}_{x_n}(e^{tx_n}) = M_{x_1}(t)M_{x_2}(t)...M_{x_n}(t) \,. \tag{26}$$

In turn, because the x_i are identically distributed, $M_{x_1}(t) = M_{x_2}(t) = ... = M_{x_n}(t) = M_x(t)$, which leads to the result (24).

We now define a new random variable Y_n as

$$Y_n = \frac{S_n}{n} \,, \tag{27}$$

which is the arithmetic average $\overline{x}_n$ of the random variables $x_1, x_2, ..., x_n$. Now let $s = t/n$. By arguments identical to those above we can write the moment generating function $M_{Y_n}(t)$ of Y_n as

$$M_{Y_n}(t) = \mathrm{E}(e^{tY_n}) = \mathrm{E}(e^{sS_n}) = [M_x(s)]^n = \left[M_x\left(\frac{t}{n}\right)\right]^n \,. \tag{28}$$

We now write $M_x(s) = M_x(t/n)$ as a Taylor series to give

$$M_{Y_n}(t) = \left[1 + \frac{tm_1}{n} + o\left(\frac{t^2}{n^2}\right)\right]^n \,, \tag{29}$$

where $o(t^2/n^2)$ indicates that the second and higher-order terms ultimately vanish faster than the first-order term in the limit of $n \to \infty$. Taking this limit,

$$\lim_{n\to\infty} M_{Y_n}(t) = \lim_{n\to\infty} \left[1 + \frac{tm_1}{n} + o\left(\frac{t^2}{n^2}\right)\right]^n = e^{tm_1} = e^{t\mu_x} . \tag{30}$$

Recall that the moment generating function of a constant a is equal to e^{ta}. Thus, the result $e^{t\mu_x}$ is just the moment generating function of the mean μ_x. This implies that the average $Y_n = \overline{x}_n$ converges to the ensemble mean μ_x in the limit of $n \to \infty$, thus completing the proof.

A.3.2 Proof Involving the Characteristic Function

A proof of the law of large numbers using the characteristic function closely follows the developments above. Here we may start with the sum Y_n defined above, namely,

$$Y_n = \frac{S_n}{n} , \tag{31}$$

which is the arithmetic average $\overline{x}_n$ of the random variables $x_1, x_2, ..., x_n$. Again letting $s = t/n$, the characteristic function $\phi_{Y_n}(t)$ of Y_n is

$$\phi_{Y_n}(t) = \mathrm{E}(e^{itY_n}) = \mathrm{E}(e^{isS_n}) = [\phi_x(s)]^n = \left[\phi_x\left(\frac{t}{n}\right)\right]^n . \tag{32}$$

Writing $\phi_x(s) = \phi_x(t/n)$ as a Taylor series,

$$\phi_{Y_n}(t) = \left[1 + \frac{itm_1}{n} + o\left(\frac{t^2}{n^2}\right)\right]^n . \tag{33}$$

Taking the limit,

$$\lim_{n\to\infty} \phi_{Y_n}(t) = \lim_{n\to\infty} \left[1 + \frac{itm_1}{n} + o\left(\frac{t^2}{n^2}\right)\right]^n = e^{itm_1} = e^{it\mu_x} . \tag{34}$$

Recall that the characteristic function of a constant a is equal to e^{ita}. Thus, the result $e^{it\mu_x}$ is just the characteristic function of the mean μ_x. As concluded above, the average $Y_n = \overline{x}_n$ converges to the ensemble mean μ_x in the limit of $n \to \infty$, thus completing the proof.

A.4 Central Limit Theorem

A.4.1 Proof Involving the Moment Generating Function

We first need to show that the moment generating function of a standard normal distribution $f_z(z)$ with mean $\mu_z = 0$ and variance $\sigma_z^2 = 1$ is $M_z(t) = e^{t^2/2}$. To do this we write

$$\begin{aligned} M_z(t) = \mathrm{E}(e^{tz}) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} e^{tz} \,\mathrm{d}z \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2 + tz} \,\mathrm{d}z . \end{aligned} \tag{35}$$

Focusing on the exponent,

$$-\frac{1}{2}z^2 + tz = -\frac{1}{2}\left(z^2 - 2tz + t^2\right) + \frac{1}{2}t^2 = -\frac{1}{2}(z-t)^2 + \frac{1}{2}t^2\,,$$

so that (35) becomes

$$M_z(t) = e^{t^2/2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-t)^2}\,\mathrm{d}z\,. \tag{36}$$

We then define $u = z - t$ so that $\mathrm{d}u = \mathrm{d}z$. A change of variable then gives

$$M_z(t) = e^{t^2/2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\,\mathrm{d}u\,. \tag{37}$$

Because the integrand in (37) is the standard normal distribution the integral equals unity, leading to the result that $M_z(t) = e^{t^2/2}$.

Let $x_1, x_2, ..., x_n$ denote a set of independent and identically distributed random variables, each with mean μ_x and variance σ_x^2. We now define a new random variable as the sum

$$S_n = \sum_{i=1}^{n} x_i\,. \tag{38}$$

Because the x_i are independent and identically distributed, the expected value of the sum S_n is $\mu_{x_1} + \mu_{x_2} + ... + \mu_{x_n} = n\mu_x$ and the variance of S_n is $\sigma_{x_1}^2 + \sigma_{x_2}^2 + ... + \sigma_{x_n}^2 = n\sigma_x^2$. We now define a new random variable as

$$Z_n = \frac{S_n - n\mu_x}{\sqrt{n}\sigma_x} = \sum_{i=1}^{n} \frac{x_i - \mu_x}{\sqrt{n}\sigma_x} = \sum_{i=1}^{n} \frac{z_i}{\sqrt{n}} = \frac{S_z}{\sqrt{n}}\,, \tag{39}$$

where each $z_i = (x_i - \mu_x)/\sigma_x$ has zero mean and unit variance, and S_z denotes the sum of the z_i.

Following the developments in preceding sections the moment generating function $M_{Z_n}(t)$ of Z_n is

$$M_{Z_n}(t) = \mathrm{E}(e^{tZ_n}) = \mathrm{E}(e^{sS_z}) = [M_{Z_n}(s)]^n = \left[M_z\left(\frac{t}{\sqrt{n}}\right)\right]^n\,, \tag{40}$$

with $s = t/\sqrt{n}$. Writing $M_z(s) = M_z(t/\sqrt{n})$ as a Taylor series,

$$M_{Z_n}(t) = \left[1 + \frac{tm_1}{\sqrt{n}} + \frac{t^2 m_2}{2n} + o\left(\frac{t^3}{(\sqrt{n})^3}\right)\right]^n\,. \tag{41}$$

From our results above, $m_1 = 0$ and $m_2 = 1$, so

$$M_{Z_n}(t) = \left[1 + \frac{t^2}{2n} + o\left(\frac{t^3}{(\sqrt{n})^3}\right)\right]^n\,. \tag{42}$$

Taking the limit,

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty}\left[1 + \frac{t^2}{2n} + o\left(\frac{t^3}{(\sqrt{n})^3}\right)\right]^n = e^{t^2/2}\,, \tag{43}$$

which is the moment generating function of a standard normal distribution. This implies that the distribution of Z_n approaches $\mathcal{N}(0,1)$ as $n \to \infty$. Because of the definition of z, it also implies that the average $\overline{x}_n \sim \mathcal{N}(\mu_x, \sigma_x^2/n)$ with sufficiently large n, thus completing the proof.

A.4.2 Proof Involving the Characteristic Function

A proof of the central limit theorem using the characteristic function closely follows the developments above. We first need to show that the characteristic function of a standard normal distribution $f_z(z)$ with mean $\mu_z = 0$ and variance $\sigma_z^2 = 1$ is $\phi_z(t) = e^{-t^2/2}$. To do this we write

$$\phi_z(t) = \mathrm{E}(itz) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} e^{itz} \mathrm{d}z = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2 + itz} \, \mathrm{d}z \,. \tag{44}$$

Focusing on the exponent,

$$-\frac{z^2}{2} + itz = \left(-\frac{z^2}{2} + itz + \frac{t^2}{2} \right) - \frac{t^2}{2} = -\frac{1}{2}(z - it)^2 - \frac{t^2}{2} \,, \tag{45}$$

so that (44) becomes

$$\phi_z(t) = e^{-t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z-it)^2} \, \mathrm{d}z \,. \tag{46}$$

We then define $u = z - it$ so that $\mathrm{d}u = \mathrm{d}z$. A change of variable then gives

$$\phi_z(t) = e^{-t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} \, \mathrm{d}u \,, \tag{47}$$

which leads to the result that $\phi_z(t) = e^{-t^2/2}$.

Let $x_1, x_2, ..., x_n$ denote a set of independent and identically distributed random variables, each with mean μ_x and variance σ_x^2. As in the preceding section we define the sum S_n given by (38) and the random variable Z_n given by (39). Following the developments in the preceding sections the characteristic function $M_{Z_n}(t)$ of Z_n is

$$\phi_{Z_n}(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{t^3}{(\sqrt{n})^3} \right) \right]^n \,. \tag{48}$$

Taking the limit,

$$\lim_{n \to \infty} \phi_{Z_n}(t) = \lim_{n \to \infty} \left[1 - \frac{t^2}{2n} + o\left(\frac{t^3}{(\sqrt{n})^3} \right) \right]^n = e^{-t^2/2} \,, \tag{49}$$

which is the characteristic function of a standard normal distribution. As stated above, this implies that the distribution of Z_n approaches $\mathcal{N}(0, 1)$ as $n \to \infty$. Because of the definition of z, it also implies that the average $\overline{x}_n \sim \mathcal{N}(\mu_x, \sigma_x^2/n)$ with sufficiently large n, thus completing the proof.

References

[1] Furbish, D. J. and Doane, T. H. (2021) Rarefied particle motions on hillslopes – Part 4: Philosophy. *Earth Surface Dynamics*, 9, 629–664, https://doi.org/10.5194/esurf-9-629-2021, 2021.

[2] Furbish, D. J. and Schmeeckle, M. W. (2020) Catching up to a 21st century view of statistics in the doing and reporting of research in the Earth sciences. https://my.vanderbilt.edu/davidjonfurbish/files/2013/06/21st-Century-Statistics.pdf

[3] Furbish, D. J., Schumer, R. and Keen-Zebert, A. (2018) The rarefied (non-continuum) conditions of tracer particle transport in soils, with implications for assessing the intensity and depth dependence of mixing from geochronology. *Earth Surface Dynamics*, 6, 1169–1202, https://doi.org/10.5194/esurf-6-1169-2018.

[4] Gibbs, J. W. (1902) *Elementary Principles in Statistical Mechanics*. Yale University Press, New Haven, Connecticut.

[5] Jaynes, E. T. (2003) *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge.

[6] Schumer, R., Meerschaert, M. M. and Baeumer, B. (2009) Fractional advection–dispersion equations for modeling transport at the Earth surface. *Journal of Geophysical Research – Earth Surface*, 114, F00A07, https://doi.org/10.1029/2008JF001246.