PART II: MATHEMATICAL PRESENTATIONS

This part, written specially for this book, incorporates the substance of reports and memoranda written over the years. Deliberately, the chapters do not follow each other in strict logical order, and their contents overlap; therefore, they can, to a large extent, be read independently of each other. The topics of Chapters E5 and E6 are largely new but concern themes that long influenced my work. The topic of Chapter E9 is important in practical statistics. The topics of Chapters E7 and E8 have long traditions plagued with casual, questionable, or erroneous writings.

&&&&&&&&&&&&&&&&&&&&&&&&&&&&&

E5

States of randomness from mild to wild, and concentration from the short to the long run

✦ **Abstract.** An innovative useful metaphor is put forward in this chapter, and described in several increasingly technical stages. Section 1 is informal, but Sections 4 and 5 are specialized beyond the concerns of most readers; in fact, the mathematical results they use are new.

At the core is a careful examination of three well-known distributions: the Gaussian, the lognormal and the scaling with infinite variance ($\alpha < 2$). They differ deeply from one another from the viewpoint of the addition of independent addends in small or large numbers, and this chapter proposes to view them as "prototypes," respectively, of three distinct "states of randomness:" *mild, slow* and *wild*. Slow randomness is a complex inter-

mediate state between two states of greater simplicity. It too splits more finely, and there are probability distributions beyond the wild.

Given N addends, *portioning* concerns the relative contribution of the addends U_n to their sum $\Sigma_1^N U_n$. Mildness and wildness are defined by criteria that distinguish between *even* portioning, meaning that the addends are roughly equal, ex-post, and *concentrated* portioning, meaning that one or a "few" of the addends predominate, ex-post. This issue is especially important in the case of *dependent* random variables (Chapter E6), but this chapter makes a start by tackling the simplest circumstances: it deals with independent and identically distributed addends.

Classical mathematical arguments concerning the long-run $(N \to \infty)$ will suffice to distinguish between the "wild" state of randomness and the remaining states, jointly called "preGaussian."

Novel mathematical arguments will be needed to tackle the short-run ($N = 2$ or "a few"). The resulting criterion will be used to distinguish between a "mild" or "tail-mixing" state of randomness, and the remaining states, jointly called "long-tailed" or "tail-preserving." This discussion of long-tailedness may be of interest even to readers reluctant to follow me in describing the levels of randomness as "states."

In short-run partition, short-run concentration will be defined in two ways. The criterion needed for "concentration in mode" will involve the convexity of $\log p(u)$, where $p(u)$ is the probability density of the addends. The concept of "concentration in probability" is more meaningful but more delicate, and will involve a limit theorem of a new kind. Long-tailed distributions will be defined by the very important "tail-preservation criterion" under addition; it is written in shorthand as $P_N \sim NP$.

Randomness that is "preGaussian" but "tail-preserving" will be called "slow." Its study depends heavily on middle-run arguments ($N =$ "many") that involve delicate transients. ✦

THE NOTION OF CONCENTRATION being central to the study of firm sizes and price changes, this chapter is of direct relevance to economics. It shows that the economics concepts of short, middle and long-run have unsuspected parallels in probability theory: they suggest a distinction between different "states of randomness" that should prove useful in many fields of science, and also involves new mathematical results that may have enough intrinsic interest to be worth developing.

Section 1 is an informal introduction, close in style to Part I of this book. The middle part of the chapter is more technical, yet should interest many readers. Section 2 is devoted to long-run portioning and Section 3 to short-run portioning. The more specialized Section 4 proposes finer states of randomness. Section 5 is even more mathematical: it includes a proof and tackles some problems raised by the moments, and refers to "the moment problem" of classical mathematical analysis. The economic implications of short-run and long-run concentration are explored throughout the book, and serious flaws of the lognormal, in Chapter E9.

Terminology and notation. Once again, a convention is often used in this book. When there is no loss of intelligibility and the context allows, words like "Gaussian," "lognormal," "Bernoulli," "Poisson," and "scaling" will be used as common names, to avoid endless and tiresome repetition of the terms "random variable," "probability distribution," "probability density," or "density." In addition, the tail probabilities and densities will be denoted, respectively, by $P(u) = Pr\{U > u\}$ and its derivative $-P'(u)$, and $P_N(u) = Pr\{\sum_{n=1}^{N} U_{n'} > u\}$ and its derivative $-P'_N(u)$.

1. BACKGROUND AND INFORMAL PRESENTATION

The Gaussian distribution is often called "normal," because of the widespread opinion that it sets a universally applicable "norm." In the case of the phenomena studied throughout my life and described in this book, this opinion is unwarranted. In their case, randomness is highly non-Gaussian, but it is no longer possible to describe it as "pathological," "improper," "anomalous," or "abnormal." Therefore, any occurrence of *normal* in this book, as synonym of *Gaussian*, is the result of oversight, and I try not to think about the second and third syllables of *lognormal*.

1.1. The ageless competition between scaling and lognormal fit, and a motivation for introducing the notion of "states of randomness"

The innovation this chapter puts forward has many roots. One responds to a situation that plagues statistics and is a common reason for its unpopularity and ineffectiveness. All too often, reliable and competent statisticians split into camps that approach the same body of practically relevant data, and sharply disagree in their analysis. An example that concerns random variables is the very old disagreement about the distribution of income. Pareto claimed that it is scaling, and Gibrat, that it is lognormal (see Section 3 of Chapter E4, Chapter E9, and other chapters of this book). Current replays of those disagreements bring in random processes and they concern the records of price changes (Chapter E1 and Parts IV and V of this book.)

Could it be that both camps attempt to prove more than their data allow? Instead of seeking immediately to specify a distribution or a process by an analytic formula with its panoply of parameters, one should perhaps first sort out the possible outcomes into a smaller number of distinct discrete categories. The basic thought behind this classification is that, while the notion of randomness is unified from the viewpoint of mathematical axiomatics, it is of great diversity from the viewpoint of scientific modeling and related statistical tools.

Following this line of thinking, fractals led (first in finance and later in many other fields) to rather bold conclusions. To implement them, it is useful to inject a familiar metaphor and the terminology that comes with it. While a unique theory of physical interactions applies to every form of matter, the detailed consequences of those unique general laws differ sharply, for example, according to temperature and to whether the interactions are short-range or long-range. This is why physics has to distinguish between several *states of matter*, whose traditional number is three.

I propose in this chapter to argue that a similar distinction should be useful in probability theory. In the not-too-distant past, every book of statistics, as well as nearly every scientist engaged in statistical modeling in economics or elsewhere, used to deal with a special form of randomness, which will be characterized as *mild*. It will also be argued that entirely different states of randomness must be distinguished and faced. There is *wild* randomness exemplified by distributions with infinite variance. There is also an intermediate possibility exemplified by the lognormal: it is *slow randomness* – a term deliberately selected to imply what it says.

When faced with a new phenomenon or fresh dataset, the first task is to identify its state of randomness.

The implication is that, instead of ranging continuously, random variables are usefully sorted out in discrete categories exemplified by the Gaussian, the lognormal and the scaling with $\alpha < 2$. When the random variables U are defined by $\Pr(U > u) = P(u)$, the state of randomness differs sharply according to how fast the generalized inverse function P^{-1} decreases as its argument tends to 0, that is, according to how fast the moments U^q increase as $q \to \infty$. (To define P^{-1} when $P(u)$ is discontinuous, one fills each discontinuity by a vertical interval before the coordinate axes are exchanged.)

The words *mild, slow,* and *wild* were chosen to be short and without competing technical connotations (discounting that everyday usage tends to view all randomness as wild). The word "state" is also carefully chosen. Its existing technical connotations denote gases, solids and liquids; they are strong, but do not compete with the new usage; even some of its ambiguities are helpful, as I propose to argue now.

To begin with mildness, it is characterized by an *absence of structure* and in the case of random processes by a local level of statistical dependence. That is, diverse parts can be modified without much affecting the whole. Remarkably, the same properties also characterize a *gas*. Their importance will be seen in Section 2 of Chapter E8, when discussing the legitimacy of random-walk models of scaling.

Wildness, to the contrary, will be shown throughout this book to be characterized by the opposite qualities: *presence of structure* and long dependence. Remarkably, the same properties characterize a *solid*.

Among the long-recognized states of matter, the third and least-well understood and explained is *liquid*. Characteristic of both physical liquids and slow randomness is a surprising degree of uncertainty in the definition and many technical imperfections. Consider a glass: it behaves from many viewpoints as a solid, but physicists know that in "reality" it is a very viscous liquid. This unresolved problem of physical characterization has a surprising probabilistic counterpart in the distribution of personal income, as seen in several chapters of this book.

Nobody would suggest that income distribution is soft and akin to a gas: it is clearly hard. What remains to be established is whether the better metaphor is a "real solid,"or a "very viscous liquid." Pareto's law presses the claim that income distribution is scaling, therefore like a solid. Gibrat's writings press the counter-claim that it is lognormal, therefore like a very viscous liquid. Chapter E9 will set up a case against the lognormal, and argue that the above disagreement may be of a kind that cannot be settled by inventing better statistical methods.

1.2 The fallacy of transformations that involve "grading on the curve"

Before describing the criteria that distinguish the different states of randomness, it is necessary to dispose of a view that amounts to considering all forms of randomness as effectively equivalent. Indeed, scientists faced with clearly non-Gaussian data are often advised by statisticians to move on to a transformed scale in which everything nicely falls on the Gaussian "bell curve." In schools, the procedure is called "grading on the curve."

When pushed to its logical extreme, the underlying procedure leads to "grading by percentages." This transforms any U into a uniform random variable on [0, 1] defined by $\Pr\{I < x\} = x$. Indeed, a random variable U defined by $\Pr\{U > u\} = P(u)$ is simply the non-decreasing transform of I defined as $P^{-1}(I)$, where P^{-1} is defined in Section 1.1.

Unfortunately, transformation ceases to look attractive as soon as one faces reality. A first complication, beyond the scope of this chapter, concerns sequences of *dependent* variables: when each variable is made uniform, the rules of dependence need not transform into anything simple.

A second complication is this: money is additive, but a transform such as log (money) is not; firm sizes add up to the size of an industry, but a transform like log (firm size) is not additive. In pedantic terms, concrete economics deals with *numerical* variables that can be added, not with *ordinal* variables that can only be ordered.

A third and most important complication is that real-world distributions are not known exactly, but approximately. That is, a random variable does not come up alone, but as part of a natural "neighborhood" that also contains other variables viewed as "nearly identical" to it.

Of enormous significance are the neighborhoods that are automatically implied in every limit theorem of probability theory. For example, to say that a random variable tends to a limit, is to say that it eventually enters a suitably defined neighborhood of the limit. In the usual central limit theorem, the limit is Gaussian, and the neighborhood is defined solely on the basis of the central bell, disregarding the tails. Cramer's large deviations theory splits the neighborhood of the Gaussian in a finer way that does not concern the bell, but the tails. The concrete usefulness of a limit theorem depends initially on whether or not this neighborhood it implies is a "natural" one from the viewpoint of a specific concrete situation.

Now we can describe the major failing of the transformation of U into I : it fails to transform the natural neighborhood of U into the natural neighborhood of I.

Once again, the example of greatest relevance to this book is the notion that for some data the best methods of statistics conclude that $\log X$ is practically Gaussian. This means that the observed deviations from Gaussianity only concern the largest values of X that contribute a few percent of the whole. Faith in the significance of the Gaussian fitted to $\log X$ leads to the recommendation that these exceptional values be neglected or treated as "outliers." The trouble is that in many cases the most interesting data are those in the tail! It follows that differences between alternative notions of neighborhood are *not* matters of mathematical nit-picking.

In the light of these three "complications," the suggestion that any variable can simply be made uniform or Gaussian by transformation is ill-inspired and must be disregarded.

1.3 Portioning on the short or the long-run, and three states of randomness

The proceeding motivation gave one example of each state of randomness. It is now time to define those states. Before we do so, recall that gases, liquids and solids are distinguished through two criteria: flowing versus non-flowing, and having a fixed or a variable volume. Two criteria might define four possibilities, but "non-flowing" is incompatible with "variable volume." Adding in uncanny fashion to the value of our physical metaphor, our three states of randomness are also defined by two mathematical criteria, both deeply rooted in economic thinking. Given a sum of N independent and identically distributed random variables, those criteria hinge on two notions.

Portioning concerns the relative contribution of the addends U_n to the sum $\sum_1^N U_n$.

The *concentration ratio* of the largest addend to the sum. Loosely speaking, *concentration* is the situation that prevails when this ratio is high. This idea will, later in this chapter, be implemented in at least two distinct ways. The opposite situation, prevailing when no addend predominates, will be called *evenness*.

The issue must be raised separately on the short- and the long-run, and it will be seen that concentration in the long-run implies concentration in the short-run, but not the other way around. Hence, the contrast between concentration and evenness leads to three principal categories.

- *Mild randomness* corresponds to short- and long-run evenness.
- *Slow randomness* corresponds to short-run concentration and long-run evenness.
- *Wild randomness* corresponds to short- and long-run concentration.

In mild and wild randomness, the short- and long-run behavior are *concordant*; in slow randomness, they are *discordant*.

Here is another bit of natural and useful terminology.

• Taken together, the two non-wild states will be said to define *preGaussian* randomness, the counterpart of *flowing* for the states of matter. An alternative term is "tail-mixing."

• Taken together, the two non-mild states will be said to define *long-tailed* randomness, the counterpart of *fixed-volume* for the states of matter. An alternative term is *tail-preserving*.

Let us now dig deeper, in terms of finance and economics.

Long-run portioning and the distinction between wild and preGaussian randomness. This distinction concerns asymptotics and the long-run. Examples are the relative size of the largest firm in a large industry, the largest city in a large country, or the largest daily price increase over a significantly long period of time. PreGaussian randomness yields approximate equality in the limit, as expressed by the fact that even the largest addend is negligible in relative value. By contrast, wild randomness yields undiminishing concentration, expressed by the property that the largest relative sizes remains non-negligible even in very large aggregates.

The mathematical detail of long-run portioning is delicate and found in standard references, therefore it must and can be summarized. This will be done in Section 2. Additional information is found in Chapter E7.

Short-run portioning, and the distinction between mild and long-tailed randomness. The cleanest contrast to the long-run is the very short-run represented by two items. Given two independent and identically distributed random variables, U' and U'', and knowing the value taken by the sum $U = U' + U''$, what do we know of the distributions of U' and U''? We shall describe U as being "short-run portioned between U' and U'', " and wonder whether those parts are more or less equal, or wildly dissimilar.

As a prelude, consider two homely examples. Suppose you find out that the annual incomes of two strangers on the street add to $2,000,000. It is natural and legitimate to infer that the portioning is concentrated, that is, there is a high probability that the bulk belongs to one or the other stranger. The $2,000,000 total restricts the other person's income to be less than $2,000,000, which says close to nothing. The possibility of each unrelated stranger having an income of about $1,000,000 strikes everyone as extraordinarily unlikely, though perhaps less unlikely that if the total were not known to be $2,000,000.

To the contrary, the total energy of two sub-systems of a gas reservoir is evenly portioned: each molecule has one-half of the energy of the two together, plus a tiny fluctuation. A situation in which most of the energy concentrates in one subsystem can safely be neglected.

Rigorous mathematical argument supports the "intuition" that even portioning is very unlikely in one case, and very likely in the other. Indeed, the above two stories exemplify opposed rules of short-run portioning. Even short-run portioning will define mild randomness, and concentrated short-run portioning will define long-tailed randomness.

Unfortunately, the details of this distinction are not simple. In addition, short-run portioning is not a standard mathematical topic. The question was first raised and discussed heuristically in Section 2.5 of M 1960i{E10} and again in Section V.A of M 1963b{E14}, but, to my knowledge, nowhere else. The first full mathematical treatment, which is new, will be presented in Sections 3 and 5.

The middle-run. Short- and long-run considerations are familiar in economics. They are essential, but, to quote John Maynard Keynes, "in the long-run we shall all be dead." Economic long-run matters only when it approximates the middle-run reasonably, or at least provides a convenient basis for corrective terms leading to a good middle-run description.

Probability theory also favors small and large samples. Samples of a few items are handled by explicit formulas often involving combinatorics. Large samples are handled by limit theorems. Exact distributions for middle-size samples tend to involve complicated and unattractive series or other formulas that can only be handled numerically on the computer. In a way, this chapter proposes to bracket the interesting but untractable probabilistic middle-run between an already known and tractable long-run and a very different tractable short-run.

Digression concerning physics. The model for all sciences, physics, was able for a long time to limit itself to two-body or many-body problems, that is, small or large aggregates. Intermediate ("mesoscopic") phenomena were perceived as hard and only recently did they impose themselves and physics became strong and bold enough to tackle a few of them. In a few examples (some of which occur in my recent work), it is useful and possible to distinguish and describe a distinct *pre-asymptotic* regime of large but finite assemblies.

2. WILD VERSUS PRE-GAUSSIAN RANDOMNESS: CLASSICAL LIMIT THEOREMS DEFINE CONCENTRATION IN THE LONG RUN

This Section is somewhat informal, the technical aspects being available in the literature, and/or taken up in Chapter E7.

2.1 Introduction to long run portioning

Probability theory solved long ago the problems of the typical size of the largest of N addends, relative to their sum, and the problem of distribution around the typical size. The most basic distinction is based is the boundedness of the second moment. Of the many possibilities that are open, the following are the most important.

At one extreme, the addends are bounded, and the concentration is $\sim 1/N$. As $N \to \infty$, concentration converges to 0. This last conclusion also holds when $EU^2 < \infty$. Since the inequality $EU^2 < \infty$ is generally taken for granted, most scientists view the notion of concentration for large samples as completely solved by probability theory. In particular, one of the justifications of the role of the Gaussian in science is closely patterned after its role in the "theory of errors," as practiced around 1800 by Legendre and Gauss. It is taken for granted that each chance event is the observable outcome of a large number of separate additive contributions. It is also taken for granted that each contribution, even the largest, is negligible compared to the sum, both ex-ante (in terms of distributions) and ex-post (in terms of sample values). The economists' technical term for this premise is "absence of concentration in the long-run," and here it will be called "evenness in the long-run." This premise is widely believed to hold for *all* independent and identically distributed addends. In other words, identity of ex-ante distributions of the parts is believed to lead to evenness of ex-post sample values.

This common wisdom claims to solve one of the problems raised in this chapter. Observed occurrences of concentration are viewed as transients, or possibly the result of strong statistical dependence between addends.

In fact, and this is the main theme of this chapter and of the whole book, the common wisdom is simply *mathematically incorrect.* As sketched in Section 1.3 and discussed in this section, portioning in the long-run can take two distinct forms: *even*, with concentration converging to 0 as $N \to \infty$ and *concentrated*, with the largest addend remaining of the order of magnitude of the sum.

This distinction largely relies on standard results of probability theory. This book discusses its impact in economics in many places, including in the reprints on income distribution and price variation, and Chapter E13 concerned with firm sizes.

2.2 Alternative criteria of preGaussian behavior

The prototype of mild randomness is provided by the "thermal noise" that marks the difference between the statistical predictions of the theory of gases and the non-statistical predictions of the older thermodynamics. Thermal noise consists in small fluctuations around an equilibrium value. For the "astronomically" large systems that are the (successful) physical analogs of the economic long-run, those fluctuations average out into relative insignificance. If such a system is divided into many equal parts, the energy of the part with the highest energy is negligible compared to the energy of the whole.

Informal statement. More generally, the form of randomness this book calls preGaussian is defined by limit asymptotic properties that are best stated as follows.

• *The fluctuation is averaging*, or *ergodic*. The law of large numbers (LLN) shows that sample averages converge asymptotically to population expectations.

• *The fluctuation is Gaussian.* The central limit theorem (CLT) shows that the fluctuations are asymptotically Gaussian.

• *The fluctuation is Fickian.* The central limit theorem also shows that the fluctuations are proportional to $\sqrt{N}$, when N being the number of addends. For random processes, an alternative, but equivalent statement (less well-known but essential) is that events sufficiently distant in time are asymptotically independent.

More formal questions and answering statements.

Question: Take the sequential sum $\sum_{n=1}^{N} U_n$ for a sequence of independent and identically distributed random variables U_n, $1 \leq n < \infty$. Is it possible to choose the sequences A_N and B_N and define the notion of "converges to", so that $A_N\{\sum_{n=1}^{N} U_n - B_N\}$ converges to a limit?

An answer that defines preGaussian behavior: Under certain conditions described in numerable textbooks, a choice of A_N and B_N is possible in two distinct ways:

The choice of $A_N = 1/N$ and $B_N = 0$ yields the law of large numbers, in which the limit is EU, that is, non-random.

The choice of the Fickian factor $A_N \sim 1/\sqrt{N}$ and $B_N = NEU$ yields the central limit theorem, in which the limit is Gaussian, that is, random. It follows that $\sum_{n=1}^{N} U_n$ is asymptotically of the order of $\sqrt{N}$.

The notions of attraction and universality. In many contexts, physicists have no confidence in the details of their models, therefore distrust the models' consequences. An important exception is where the same consequences are shared by a "class of universality," that also includes alternatives that differ (not always slightly) from the original model. Although the word universality is rarely used by probabilists, the basic idea is very familiar to them. For example, few scientists worry about the precise applicability of a Gaussian process, because the Gaussian's domain of attraction is very broad, and slight changes in the assumptions provoke slight deviations in the consequences drawn from the model. The domain of universality of attraction to the Gaussian includes all U satisfying $EU^2 < \infty$, but also some cases when EU^2 diverges slowly enough.

2.3 Exceptions to preGaussian behavior

The preGaussian domain of universality is broad, but *bounded*. The properties of being averaging, Gaussian, or Fickian may fail. The failure of any of these properties defines the *wild* state of randomness.

Failure occurs when the population variance, or even the expectation, is infinite, when the dependence in a random process is not "short-range" or local (contrary to the locality of the Markov process) but "long-range" or global, or when total probability must be taken as infinite. Most notably, the scaling variables with $\alpha < 2$ satisfy $EU^2 < \infty$, and are *not* preGaussian.

Wild randomness and practical statistics. The preface quoted the editor's comments on M 1963b{E14} found in Cootner 1964. They, and countless other quotes by practically-minded investigators, some of them scattered throughout this book, show that non-averaging, non-Gaussian, and/or non-Fickian fluctuations were long resisted and viewed as "improper" or even "pathological." But I realized that many aspects of nature are ruled by this so-called "pathology." Those aspects are not "mental illnesses" that should or could be "healed." To the contrary, they offer science a valuable new instrument. In addition, a few specific tools available in "pure mathematics" were almost ready to handle the new needs. The new developments in science that revealed the need for those tools implied that science was moving on to a *qualitatively different* stage of indeterminism.

The editor's comments in Cootner 1964 also noted that, if it is confirmed that economic randomness is wild, some tools of statistics will be endangered. Indeed, tools developed with mild randomness in mind become questionable in the case of slow randomness. As a rule with many exceptions, they are not even close to being valid for wildly random phenomena, such as those covered by my models of price variation.

Sketch of the domains of universality of attraction to nonGaussian limits. To be outside of the Gaussian domain of attraction or universality is a great complication. In particular, each value $\alpha < 2$ defines its own domain of universality. In addition, in sharp contrast to the width of the domain of the Gaussian, each of those domains is extremely narrow and reduces to the variables for which $Pr\{U > u\} \sim u^{-\alpha}L(u)$, where $L(u)$ is logarithmic or at most slowly varying in the sense that for all h, $\lim_{u \to \infty} L(hu)/L(u) = 1$. The term $L(u)$ is largely a nuisance, and we shall not invoke it unless necessary.

If U_n is in the domain of universality of $\alpha < 2$, the limit is a random variable called *L-stable*, which is widely discussed and used in the papers reprinted in this book.

In the absence of slowly varying term $L(u)$, the choice of A_N is $A_N = N^{1/\alpha}$ for all α, therefore U_N^{Σ} is asymptotically of the order of $N^{1/\alpha}$. The choice of B_N is NEU when $1 < \alpha < 2$, and 0 when $0 < \alpha < 1$.

2.4 Comments on the middle-run and slow randomness

Adding new evidence that the world is not a simple place and science is more difficult than mathematics, the limit theorems of probability do not really matter, unless they also help describe the middle-run. Unfortunately, the middle-run is complicated, hence the existence of a third "middle" state did not fully impress itself on my work until a recent careful look at its foremost example, the lognormal (Chapter E9). Since industrial concentration is incontrovertible (see Chapter E13), the very fact that the lognormal is continually proposed to model industrial concentration means that it cannot really be counted as mild. What is it?

On the long-run, it is indeed averaging, Gaussian and Fickian, therefore, preGaussian. In the middle-run, however, its "nice" asymptotic properties are irrelevant and give no hint of the fact that the strict lognormal yields a "very erratic" sample averages. The statistician who is invited to examine those averages, and not the distribution itself, should conclude that those averages behave "as if" the addends were wild. In other words, a more correct interpolate of the middle-run behavior is obtained if one does not start with the lognormal, but a wild approximation to the lognormal. For actual data that are neither exactly lognormal nor exactly wild, my long-term goal has been to develop viewpoints and techniques that illuminate the middle-run and can be used as the starting point for improvements.

3. MILD VERSUS LONG-TAILED RANDOMNESS: CONCENTRATION IN THE SHORT RUN, CONVEXITY OF $\log p(u)$ AND THE TAIL PRESERVATION RELATION $P_N(u) \sim NP(u)$

Section 2 divides all forms of randomness into wild – defined by concentrated long-run portioning, and preGaussian – defined by even long-run portioning. Our next goal is to subdivide this second category into two categories to be called mild and long-tailed. This will be done in stages.

- A first criterion will be based on *concentration in mode*; it is very simple, but has many flaws.

- A more intrinsic second criterion of wider applicability will be based on asymptotic *concentration in probability*. It will lead to the "tail-preservation" relation $P_N(u) \sim NP(u)$.

The term *slow*, is justified by arguments that cast doubt on the acceptability of slow random models in scientific work. It is best to phrase those arguments in the specific context of the lognormal distribution. This will be done in Chapter E9.

The tail-preservation relation is not, in itself, new to probability theory, since it occurs in the classical "extreme value problem." Indeed, let the random variables $U_j (1 \le j \le N)$ be independent and identically distributed, with the tail probability $P(u)$, and let $\tilde{P}_N(u)$ be the tail probability of $\tilde{U}_N = \max(U_j)$. It is well-known that $1 - \tilde{P}_N(u) = \{1 - P(u)\}^N$. In the tail where $P(u) \ll 1$ and $\tilde{P}_N(u) \ll 1$, we find in *all cases* that $\tilde{P}_N \sim NP$.

However, the material that follows *does not concern* the extreme value problem, it merely injects some considerations relative to extreme values into the classical study of *sums*. A striking consequence is that, in this new context, the tail preservation relation for sums holds for *some*, but *not all*, probability distributions. By ceasing to hold universally, it ceases to be a trivial property; instead, it takes up a central role in a fundamental distinction between one state of randomness (mild) and the other states taken together (long-tailed.)

3.1 The doubling convolution and the short-run portioning ratio

As agreed, we denote the common probability density of U' and U'' by $p(u)$. The probability density of $U = U' + U''$, denoted $p_2(u)$, is given by the doubling convolution $p_2(x) = \int p(u)p(x-u)du$. When u is known, the conditional probability density of u' is given by the following expression, to be called "portioning ratio"

$$\frac{p(u')p(u-u')}{p_2(u)}.$$

The denominator is a constant and it remains to study the numerator.

Min (U', U'') and max (U', U'') can be compared in many different ways. The conditional expectation of U', knowing $U = u$, is of no help, since it is always $u/2$, and the conditional expectation of $\min(u', u'')$ is not given by any manageable expansion.

To the contrary, it is often easy to study the location of the most probable values of $\min(u', u'')$ and $\max(u', u'')$, which statisticians call "modes." Those locations lead to a criterion based on the convexity of $\log p(u)$, which will serve to define "concentration versus evenness in mode."

The mode is of little use in probability, but in this instance turns out to be surprisingly close to being satisfactory. Indeed, a more searching stage of this study shows that, under suitable additional assumptions, the integral $\int p(u')p(u-u')du'$ is dominated by values the conditional density $p(u')p(u-u')$ takes in intervals near the modes, while the remaining intervals have a negligible contribution. The underlying mathematical theorem concerns concentration "in probability," but in some cases also holds in the "almost sure" sense.

The proof of this basic theorem also yields the fundamental "tail-preservation criterion" written in shorthand as $P_N \sim NP$. In due time, the assumption of the basic theorem are bound to be improved. Therefore, I propose to define "long-tailedness" as meaning "tail-preserving."

3.2 Sufficient criterion of evenness or concentration "in mode": the graph of $\log p(u)$ is cap- or cup-convex for sufficiently large values of u

In many important cases, the maximum of the product $p(u')p(u-u')$ occurs either near $u' = u/2$, or near $u' = 0$ and $u' = u$. Take logarithms and write

$$\Delta(u) = 2 \log p\left(\frac{u}{2}\right) - [\log p(0) + \log p(u)].$$

When the convexity of $\log p(u)$ is uniform for all u, the sign of $\Delta(u)$ is independent of u.

• The case when the graph of $\log p(u)$ is cap-convex, like the typographical sign $\cap$. In that case, the portioning ratio is *maximum* for $u' = u/2$, and portioning is even in terms of the mode.

• The boundary case when the graph of $\log p(u)$ is straight. In that case, the addends are exponential, and the portioning ratio is a constant.

• The case when the graph of $\log p(u)$ is cup-convex, like the typographical sign $\cup$. In that case, the portioning ratio is *minimum* for $u' = u/2$ and portioning is concentrated in terms of the mode.

Distributions with uniform convexity of $\log p(u)$ suffice to show that the distinction between mild and long-tailed randomness cannot be identified with the distinction between even and concentrated short-run portioning.

3.3 Simple examples of uniform convexity

Every Poisson always yields even *short-run portioning in mode.* When $p(u) = e^{-\gamma}\gamma^u/u!$, the convexity of $\log p(u)$ is that of $\log u!$, which is cap-convex all $u > 0$. The portioning ratio is

$$\frac{p(u')p(u-u')}{p_2(u)} = \frac{e^{-\gamma}e^{-\gamma}}{e^{-2\gamma}} \frac{\gamma^{u'}\gamma^{u-u'}}{(2\gamma)^u} \frac{u!}{u'!(u-u')!}.$$

The non-constant third term is a binomial coefficient that peaks at $u = u/2$ if u is even, and at $(u \pm 1)/2$ if u is odd. At those points, the portioning ratio $p(u')p(u-u')/p(u)$ has a maximum. Even portioning was to be expected: the Poisson distribution rules the number of points of a Poisson process that fall in an interval of given length.

Every Gaussian yields even *short-run portioning in mode*. Here, $\log p(u)$ is essentially $-u^2$, which is cap-convex uniformly for all u. The portioning ratio is

$$\frac{\dfrac{1}{\sqrt{2\pi}} \exp\left(-\dfrac{u^2}{2}\right) \dfrac{1}{\sqrt{2\pi}} \exp\left\{-\dfrac{(u-u')^2}{2}\right\}}{\dfrac{1}{\sqrt{2\pi}} \exp\left\{-\dfrac{u^2}{2}\right\}} = \frac{1}{\sqrt{\pi}} \exp\left\{-\left(u' - \frac{u}{2}\right)^2\right\}.$$

Thus, a Gaussian is evenly partitioned with a Gaussian "error-term" for which variance is $1/2$, that is, does not depend on u.

Every scaling yields concentrated *short-run portioning in mode for all* α. When $p(u)=\alpha u^{-\alpha-1}$ for $u>1$, $\log p(u)$ is cup-convex uniformly for all $u>1$. The same is true of $\log p(u)+\log p(x-u)$, and the portioning ratio is

$$\frac{1}{2}\frac{u'^{-\alpha-1}(u-u')^{-\alpha-1}}{u^{-\alpha-1}}.$$

It is cup-convex and largest for $u=1$ and $u=x-1$.

The family $p(u)=\exp(-u^{w})$, *is split, the nature of short-run portioning being dependent on the sign of* $w-1$. The convexity of $\log p(u)$ is, again, uniform for all $u>0$, but here it depends on the sign of $w-1$. Portioning is even for $w>1$ and concentrated for $w<1$. The family $\exp(-u^{w})$ is often praised in the literature for the ability of one analytic expression to account for very different behaviors, according to whether $w>1$ or $w<1$. This versatility can also be interpreted in a negative light, as a form of insensitivity to profound differences.

3.4 The lognormal and other examples of non-constant convexity of log $p(u)$; mixed rules of short-run portioning

For many usual distributions, the graph of $\log p(u)$ is cap-convex for all u. But a bell where the graph of $\log p(u)$ is cap-convex is often flanked by one or two tails where the graph is cup-convex. In those mixed cases, portioning depends on u: it is *even for* u *near the mode (i.e., where* $p(u)$ *is largest) and concentrated for large* u. Let us examine a few examples.

The Cauchy. Here,

$$p(u)=\frac{1}{\pi(1+u^{2})}\text{ and }p_2(u)=\frac{1}{2\pi(1+u^{2}/4)}.$$

Here the convexity of $\log p(u)$ changes for $u=\pm1$. Hence, portioning is in mode even for $|u|<2$, and concentrated for $|u|>2$.

The "Cournot" (positive L-stable density with $\alpha=1/2$*).* Here,

$$p(u)=\frac{1}{\sqrt{2}\,\pi}e^{-1/2u}u^{-3/2},\text{ and }p_2(u)=\frac{2}{\sqrt{2}\,\pi}e^{-4/2u}u^{-3/2}.$$

The lognormal. If $EU=1$, there is a single parameter $\sigma^2/2$, and

$$\log p(u) = -\log(\sigma\sqrt{2\pi}) - \log u - \frac{(\log u + \sigma^2/2)^2}{2\sigma^2}.$$

Here, the convexity of $\log p(u)$ changes when $\log u_0 = 1 - 3\sigma^2/2$. Hence, portioning is even when $u < 2u_0$, and concentrated when $u > 2u_0$. When σ^2 is large, so that the lognormal is very skew, the bell lies almost entirely to the left of $EU = 1$ and its total probability is small. Portioning is then most likely to be concentrated and the non-mild character of the lognormal is obvious. When σ^2 is small, so that the lognormal is near-Gaussian with a small tail added, the bell includes $EU = 1$ and its total probability is near 1. Portioning is then most likely to be even, and the lognormal may seem mild.

Note. This example raises an issue of wider applicability. Ostensibly, portioning in the case $N = 2$ is a short-run notion. But in the case of near-Gaussian lognormals, concentration only occurs in very large assemblies.

The log Bernoulli e^B. This is the exponential of a Bernoulli; it has a finite upper bound $\exp(\max B)$, therefore the limit arguments concerning $u \to \infty$ have no meaning for it. As the sum of two addends approaches 2 $\exp(\max B)$, the portioning between the addends necessarily becomes even.

3.5 The problematic gamma family $p(u) = u^{\gamma-1}e^{-u}/\Gamma(\gamma)$; portioning in mode is even for $\gamma > 1$ and concentrated for $\gamma < 1$

The concentration in mode based on the convexity of $\log p(u)$ proves unreasonable in the case of the gamma distribution. In that case, $U' + U''$ is a gamma of parameter 2γ, hence

$$\frac{p(u')p(u-u')}{p_2(u)} = \frac{\Gamma(2\gamma)}{[\Gamma(\gamma)]^2}\,\frac{u'^{\gamma-1}(u-u')^{\gamma-1}}{u^{2\gamma-1}}.$$

The exponential special case $\gamma = 1$ marks the boundary between two opposite rules of portioning in mode.

When $\gamma > 1$, portioning in mode is even and the maximum at $u/2$ becomes increasingly more accentuated as $\gamma \to \infty$. For integer values of γ, this result was to be expected, because the resulting gamma is the sum of γ independent exponential variables, therefore becomes increasingly close to Gaussian.

When $\gamma < 1$, to the contrary, portioning in mode is concentrated. However, this behavior is not *due* to the tail behavior of the gamma, rather to its behavior near $u = 0$. The tail is *shorter* in the cap-convex case $\gamma < 1$ than in the cup-convex case $\gamma > 1$.

In summary, the gamma shows the need of a criterion of "mildness" that goes beyond the convexity of log $p(u)$.

The multiplicative character of concentration in mode is the gamma case. For the gamma, $W = U'/U$ is independent of u, and has the beta density

$$\Gamma(2\gamma)[\Gamma(\gamma)]^{-2} w^{\gamma-1}(1-w)^{\gamma-1}.$$

Therefore, the fluctuating term can be described as multiplicative. Now apply the same argument formally to the asymptotically scaling. The concentration ratio converges to $w^{-\alpha-1}(1-w)^{-\alpha-1}$. This limit is non-integrable near $w > 0$ and $w = 1$, implying that for the scaling, $w \to 0$ or 1 as $u \to \infty$. The underlying reason is that in the scaling case, the distribution of $\min(u', u - u')$ is independent of u for large u, hence the fluctuating term is *not* multiplicative but *additive*.

3.6 Evenness and concentration "in probability," and the criterion $p_2(u) \sim 2p(u)$ of tail preservation under addition as defining long-tailedness

The study of concentration in mode has the virtue of extreme simplicity. The results are surprisingly adequate, but exceptions must be avoided without artificiality. The smallness of the number of exceptions is largely serendipitous, because the criterion based solely on the maxima of $p(u')p(u - u')$ is an extraordinarily crude one. The real question is more searching: is the relative value U'/u likely to lie in a suitably narrow neighborhood of the maximum or maxima?

Definition of short-run concentration in probability. This definition is geared to the case when there is concentration in mode, that is, $p(u')p(u - u')$ is maximum for u' near 0 and u' near u. In that case, given a value of $\tilde{u}$ that satisfies $\tilde{u} < u/2$ and may depend on u, one can split the doubling convolution in three parts, as follows

$$p_2(x) = \int_0^x p(u)p(x-u)du = \left\{ \int_0^{\tilde{x}} + \int_{\tilde{x}}^{x-\tilde{x}} + \int_{x-\tilde{x}}^{x} \right\} p(u)p(x-u)du$$
$$= I_L + I_0 + I_R.$$

I propose to describe $p(u)$ as *short-run concentrated in probability* if it is possible to select $\tilde{u}(u)$ so that the middle interval $(\tilde{u}, u - \tilde{u})$ has the following two properties as $u \rightarrow \infty$.

- The relative probability in the middle interval $I_0/p_2(u)$, *tends* to 0.
- The relative length of the middle interval $(u - 2\tilde{u})u$ *does not tend* to 0.

This second requirement opens two sub-possibilities.

- When $p(u)$ is only moderately long-tailed, the relative length of the middle interval tends to 1. The density $p(u)p(\alpha - u)$ concentrated arbitrarily tightly around its mode. Concentration in probability is replaced by a stronger property: almost sure concentration.

- When $p(u)$ is extremely long-tailed, the relative length of the middle interval tends to a limit, or may have a lower bound > 0 and an upper bound < 1.

The "tail-preservation criterion." Section 5.1 will insure that short-run concentration in probability prevails when log $p(u)$ is smoothly varying, decreasing and cup-convex and its derivative $p'(u)/p(u)$ tends rapidly enough to 0 as $u \rightarrow \infty$.

In addition to concentration in probability, the same proof yields a very perspicuous criterion, namely $p_2(u) \sim 2p(u)$. In terms of the tail probabilities $P(u)$ and $P_2(u)$ of U and $U' + U''$, this criterion reads

$$P_2(u) \sim 2P(u).$$

More generally, writing $P(u)$ and $P_N(u)$ for the tail probabilities of U and of a sum of N variables with the same distribution, one obtains

$$P_N(u) = NP(u).$$

This criterion expresses that the tail behavior of U is preserved under finite addition. The notion of tail preservation, first introduced in M 1960i{E10}, recently turned out to be related to classical and seemingly unrelated considerations in classical "fine mathematical analysis," to be described in Section 5.

Rescaling of tails and a property of scaling distribution. When the tail is asymptotically scaling as in the case of the L-stable distributions, the tail conservation relation acquires a special meaning. It shows that

$$\text{"scale of } U\text{"} = \text{"scale of } U' \times 2^{1/\alpha}\text{."}$$

This result also holds when $p(u) = u^{-\alpha}L(u)$, where $L(u)$ is slowly varying, that is, it satisfies $L(hu)/L(u) \to 1$ for all $h > 0$) as $u \to \infty$.

Tail conservation holds for lognormals, but fails to have this special meaning. Lognormality is *not* preserved by addition.

3.7 Mild randomness and mixing behavior when log $p(u)$ is cap-convex

To appreciate the meaning of the criterion $P_N \sim NP$, let us examine cases where it does not hold.

The borderline exponential case. Here, $\sum_{n=1}^{N} U$ is a gamma variable; therefore as $u \to \infty$, P_N/P does not tend to the constant N, but increases like u^{N-1}.

The case of evenness in mode. When the convolution integrand $p(u)p(x-u)$ has a maximum at $u = \alpha/2$, the tail of $p_2(u)$ is little affected by the behavior of $p(u)$ in the tail. But it is greatly affected by its behavior part-way through the tail. The result is that P_N/P can increase very fast. In the Gaussian case with $EU = 0$, when $N \gg 1, P_N/P \sim 1/[\sqrt{N}\,P]$, which grows very fast as $u \to \infty$. Instead of tail preservation, one encounters an interesting "mixing" behavior whose intensity can be measured by the rate of growth of P_N/P.

3.8 Portioning and the tail-preservation relation $P_N \sim NP$, when N is a small integer above 2

In an equilateral triangle of height u, the distances from a point P to the three sides add up to u, therefore can represent u', u'' and u^0 in the portioning of the sum $u = u' + u'' + u^0$ into its contributing addends u', u' and u^0. When the U are exponential, the conditional distribution of P is uniform within this triangle. When U is mild, the conditional distribution concentrates near the center. When U is short-run concentrated in probability, the conditional distribution concentrates near the corners.

The same distinction holds for $N = 4, 5$ etc.... It is of help in gaining a better understanding of the problematic gamma family. The γ exponent of a sum of N gammas is $N\gamma$, which exceeds 1 as soon as $N > 1/\gamma$. Therefore, U can conceivably be called mild if $\sum_{n=1}^{N} U$ has a cap-convex log $p_N(u)$ for all N above some threshold. Starting with $\gamma = 2^k$, where k is a large integer, evenness decreases until $\gamma = 1$. There, a boundary is crossed and portioning becomes increasingly concentrated.

However, as $N \to \infty$, an altogether different classification takes over, as seen in Section 3.

Next, consider portioning of a sum of four addends $U_1 + U_2 + U_3 + U_4$ into two sums of two addends $U_1 + U_2$ and $U_3 + U_4$. Indeed, even when $\log p(u)$ is cup-convex for all u, one part of the graph of $\log p_2(u)$ is bell shaped. In the scaling case, the bell continues by a cup-convex tail. In the gamma case with $1/2 < \gamma < 1$, the tail is cap-convex.

A seeming paradox of immediate practical importance: when $P_N(u) \sim NP(u)$, *the cup-convexity of* $\log P_N(u)$ *for large* u *is preserved for all* N*; this is true both when* U *is wild and when it is preGaussian.* Taking the word "addends" as model, "limitands" is a self-explanatory term for "items" that are made to tend to a limit. The items may be sets, graphs of functions, or analytic expressions. Let $P(L_n)$ and $P(L)$ be properties of each limitand L_n and the limit $L = \lim_{n \to \infty}$, respectively. The "intuition" that $P(L) = \lim_{n \to \infty} P(L_n)$ is often wrong. It used to be that it only failed for artificial mathematical counter-examples, but no longer. Define L_N as the graph of the function $\log p_N(u)$ relative to the sum of N long-tailed random variables U and the property $P(L_N)$ as asserting that, for all N, the graph L_N is cup-convex for large u. Two possibilities are open: When U is wild, this convexity property is indeed preserved in the limit; However, when U is slow, this property fails in the limit, since the limit is the graph of $\log p(u)$ for the Gaussian.

4. A MORE REFINED TENTATIVE SUBDIVISION, YIELDING SEVEN STATES OF RANDOMNESS

The criteria stated in Section 1 and elaborated in Sections 2 and 3 leave open many conceptual and practical "details."

4.1 The boundary between mild and long-tailed and "borderline mild" randomness

Sections 3.3 to 3.6 and Section 4 imply that the exact separation between the mild and the long-tailed states of randomness is not unique, and depends upon the definition selected for the notion of concentration. Within the problematic gamma family (Section 3.6), the convexity of $\log p(u)$ defines a different boundary for each value of N. Granted this fuzziness, one may as well accept the existence of a transitional state between proper mildness and proper long-tailedness.

4.2 Extreme randomness

Wild randomness was characterized by the fact that the largest of many addends is of the same order of magnitude as their sum. But it is possible for concentration to be even more extreme. In the example of the tail probability $P(u) = 1/\log u$, concentration converges to 1 as $N \to \infty$; asymptotically, it becomes absolute. The same is true whenever $P(u)$ is a "slowly varying function," in the sense that, for all $h > 0, \lim_{u \to \infty} P(hu)/P(u) \to 1$. Those $P(u)$ define a state of randomness beyond the wild. I never encountered it in practice.

4.3 The contrast between localized and delocalized moments

Take a hard look at the formula $EU^q = \int_0^\infty u^q p(u) du$. For the scaling, the

integrand is maximum at the trivial values 0 or ∞. But in non-trivial cases, the integrand *may* have a sharp global maximum for some value $\tilde{u}_q$ defined by the equation

$$0 = \frac{d}{du}(q \log u + \log p(u)) = \frac{q}{u} - \left| \frac{d \log p(u)}{du} \right|.$$

The dependence of $\tilde{u}_q$ on q is ruled, once again, by the convexity of $\log p(u)$.

- When $\log p(u)$ is rectilinear, the $\tilde{u}_q$ are uniformly spaced.
- When $\log p(u)$ is cap-convex, $\tilde{u}_q/q$ is decreasing; that is, the $\tilde{u}_q$ are increasingly tightly spaced.
- When $\log p(u)$ is cup-convex, $\tilde{u}_q/q$ is increasing; that is, the $\tilde{u}_q$ are increasingly loosely spaced.

However, knowing $\tilde{u}_q$ is not enough; one must also know $u^q p(u)$ in the neighborhood of $\tilde{u}_q$. The function $u^q p(u)$ often admits a "Gaussian" approximation obtained by the "steepest descents" expansion

$$\log [u^q p(u)] = \log p(u) + qu = \text{constant} - (u - \tilde{u}_q)^2 \tilde{\sigma}_q^{-2/2}.$$

When $u^q p(u)$ is well-approximated by a Gaussian density, the bulk of EU^q originates in the "q-interval" defined as $[\tilde{u}_q - \tilde{\sigma}_q, \tilde{u}_q + \tilde{\sigma}_q]$.

The usual typical examples yield the following results. The Gaussian q-intervals greatly overlap for all values of σ. The Gaussian's moments will be called *delocalized*. The lognormal's q-intervals are uniformly spaced and their width is independent of q; therefore, when the lognormal is sufficiently skew, the q-interval and the $(q+1)$-interval do not overlap. The lognormal's moments will be called *uniformly localized*. In other cases, neighboring q-intervals cease to overlap for sufficiently high q. Such moments will be called *asymptotically localized*.

The notion of localization involves an inherent difficulty. Working in the "natural" scale is essential to problems involving addition, but here it is irrelevant. That is, it suffices to show that $u^q p(u)$ has a good Gaussian approximation in terms of either u or any increasing transform $v = y(u)$.

Example of the density $\exp(-u^w)$. Here, $q = w\tilde{u}_q^w$, hence, $\Delta\tilde{u}_q = \tilde{u}_q - \tilde{u}_{q-1} \sim q^{1/w-1}$; in addition, $\tilde{\sigma}_q^{-2} = wq\tilde{u}_q^{-2}$, hence, $\tilde{\sigma}_q \sim q^{1/w-1/2}$. It follows that $\tilde{\sigma}_q / \Delta\tilde{u}_q \sim \sqrt{q}$. That is, the q-intervals overlap for all values of w. (The same result is obtained using the free variable $v = \log u$.)

Example of the density $\exp[-(\log u)^w]/u$. The expression $u^q p(u) du$, if reexpressed in the variable $V = \log U$, becomes $\exp[-(v^w - qv)]$. One finds

$$\tilde{v} \sim (q/w)^{1/(w-1)} \text{ and } \Delta\tilde{v}_q \sim w^{-1/w} q^{(2-w)/(w-1)},$$

and

$$\tilde{\sigma}_q \sim w - \frac{1}{2(w-1)} (w-1)^{-1/2} q^{(2-w)/2(w-1)}.$$

It follows that $\tilde{\sigma}_q / \Delta\tilde{v}_q \sim q^{(w-2)/2(w-1)}$. When $w > 2$, all the moments of U are delocalized. When $w \le 2$, they are localized. In the lognormal case $w = 2$, $\tilde{\sigma}_q / \Delta\tilde{v}_q$ is a constant that $\to 0$ as $w \to \infty$ and in the case beyond the lognormal, $w < 2$, $\tilde{\sigma}_q / \Delta\tilde{v}_q$ decreases as $q \to \infty$.

4.4 A tentative list of seven states of randomness

We see that the "slow" state between mild and wild splits into distinct states. Altogether, we shall face seven states of randomness, which we now list, together with examples. Alternative criteria involve the rate of increase as function of q of the moment EU^q or the scale factor $[EU^q]^{1/q}$.)

• *Proper mild randomness.* Short-run portioning is even for $N = 2$. Examples: the Gaussian, the distribution $P(u) = \exp(-u^w)$ with $w > 1$, and the gamma density $-P'(u) = u^{\gamma-1}\exp(-u)/\Gamma(\gamma)$ with $\gamma > 1$.

Mild randomness is loosely characterized, *either* by P^{-1} increasing near $x=0$ no faster than $|\log x|$, *or* by $[EU^q]^{1/q}$ increasing near $q \to \infty$ no faster than q.

• *Borderline mild randomness.* Short-run portioning is concentrated for $N=2$, but becomes even when N exceeds some finite threshold. Examples: the exponential $P(u)=e^{-u}$, which is the limit case of the preceding non-Gaussian examples for $w=\gamma=1$, and more generally the gamma for $\gamma<1$.

• *Slow randomness with finite and delocalized moments.* It is loosely characterized, *either* by P^{-1} increasing faster than $|\log x|$ but no faster than $|\log x|^{1/w}$, with $w<1$, *or* by $[EU^q]^{1/q}$ increasing faster than q but no faster than a power $q^{1/w}$. Examples: $P(u)=\exp(-u^w)$ with $w<1$, and $P(u)=\exp[-(\log u)^\lambda]$ with $\lambda>2$.

• *Slow randomness with finite and localized moments.* It is loosely characterized by *either* P^{-1} increasing faster than any power $|\log x|^{1/2}$ but less rapidly than any function of the form $\exp(|\log x|^\gamma)$ with $\gamma<1$, *or* by $[EU^q]^{1/q}$ increasing faster than any power of q, but remaining finite. Examples: the lognormal and $P(u)=\exp[-(\log u)^\lambda]$ with $\lambda \le 1$.

• *Pre-wild randomness.* It is loosely characterized *either* by P^{-1} increasing more rapidly than any functions of the form $\exp(|\log x|^\gamma)$ with $\gamma<1$ but less rapidly than $x^{-1/2}$, or by $[EU^q]^{1/q}$ being infinite when $q \ge \alpha > 2$. Examples: the scaling $P(u)=u^{-\alpha}$ with $\alpha>2$. The power U^q becomes a wild random variable if $q>\alpha/2$.

• *Wild randomness.* It is characterized by $EU^2=\infty$, but $EU^q<\infty$ for some $q>0$, however small. Examples: the scaling $P(u)=u^{-\alpha}$ with $\alpha<2$.

• *Extreme randomness.* It is characterized by $EU^q=\infty$ for all $q>0$. Example: $P(u)=1/\log u$.

4.5 Aside on the medium-run in slow randomness: problems of "sensitivity" and "erratic behavior"

In the slow state of randomness, the middle run poses many problems. The case of the lognormal is investigated in Chapter E9, to which the reader is referred. A more general discussion begins in a straightforward fashion, but is too lengthy to be included here.

5. MATHEMATICAL TREATMENT OF THE TAIL PRESERVATION CRITERION $P_N \sim NP$, AND ROLE OF LONG-TAILEDNESS IN CLASSICAL MATHEMATICAL ANALYSIS

This Section, more mathematical in tone than the rest of this chapter, begins with an important proof and then digresses on some definitions and references.

5.1 Theorem: the tail-preservation criterion $p_2 \sim 2p$ and short-run ($N = 2$) concentration both follow when the function $\log p(s)$ is decreasing and cup-convex and has a derivative that tends rapidly to 0 as $s \to \infty$

Let us repeat the definition of I_L, I_0 and I_L:

$$p_2(u) = \int_0^u p(s)p(u-s)ds = \left\{ \int_0^{\tilde{u}} + \int_{\tilde{u}}^{u-\tilde{u}} + \int_{u-\tilde{u}}^{u} \right\} p(s)p(u-s)ds$$
$$= I_L + I_0 + I_R.$$

Bounds on $I_L = I_R$. To establish concentration in probability, it suffices to prove that, as $s \to \infty$, $I_0/I \to 0$ but $1 - 2\tilde{u}/u$ does not tend to 0 . But we shall prove a far stronger result, namely that $I_L = I_R$ can be approximated by $p(s)$, in the sense that, given $\varepsilon > 0$, one can select $\tilde{u}$ so that, for large enough u,

$$(1-\varepsilon)p(u) < I_L = I_R < (1+\varepsilon)p(u).$$

The assumption that $p(u)$ is decreasing yields the following bounds valid for all $\tilde{u}$.

$$p(u)\int_0^{\tilde{u}} p(s)ds \leq I_L = I_R \leq p(u-\tilde{u})\int_0^{\tilde{u}} p(s)dx \leq p(u-\tilde{u}).$$

The desired lower bound of $I_L = I_R$ is achieved if $\int_0^{\tilde{u}} p(s)ds > 1 - \varepsilon$. This inequality will follow automatically from the fact that the upper bound will require that $\tilde{u} \to \infty$ with u.

The desired upper bound is insured if $p(u-\tilde{u})/p(u) \leq 1 + \varepsilon$. Assuming $\varepsilon \ll 1$, this reads $\log p(u-\tilde{u}) - \log p(u) < \varepsilon$. Assume that $g(s) = -(d/ds) \log p(s)$ exists and $\to 0$ as $s \to \infty$. Then the desired upper bound requires $\tilde{u} < \varepsilon g(u)$. The condition that $g(s) \to 0$ insures that $\tilde{u} \to \infty$ with u, therefore insures the validity of the lower bound to $I_L = I_R$.

Examples: The scaling cases $p(u) \sim u^{-\alpha-1}$ yields $\tilde{u}/u < \varepsilon/(\alpha+1)$, a constant. The cases $p(u) \sim \exp(-u^w)$ yield $\tilde{u} < \varepsilon u^{1-w}/w$, which increases with u , while $\tilde{u}/u < \varepsilon u^{-w}/w$ decreases. Now assume that $L(u)$ is slowly varying, which means that, for every μ, we have $L(\mu u)/L(u) \to 1$ as $u \to \infty$, and consider the density $p(u) \sim \exp(-u/L(u))$; the fact that this density is cup-convex implies that $L(u) \to \infty$; the resulting densities $p(u)$ yield $\tilde{u} < \varepsilon L(u)$, which again increases while $\tilde{u}/u$ decreases.

Finally, let us check that in the problematic gamma case, the desired upper bound is not available. This case is an example of $p(u) = \exp[-u - L(u)]$; the fact that this density is cup-convex, again implies $L(u) \to \infty$. Now, $\tilde{u}$ decreases as $u \to \infty$, albeit slowly. Therefore, the lower bound fails to hold, and the approach is not effective.

Upper bound on I_0. Because of the cup-convexity of $p(s)p(u-s)$, one has

$$I_0 < (u - 2\tilde{u})p(\tilde{u})p(u - \tilde{u}).$$

The condition $u - 2\tilde{u} \leq u$, and the selection of an upper bound for $I_L = I_R$ have already insured that $p(u - \tilde{u}) \leq p(u)(1 + \varepsilon)$; hence

$$I_0 < (1 + \varepsilon)p(u)[up(\tilde{u})].$$

Return to the example of $p(s)$ considered in discussing the upper bound for $I_L = I_R$. Aside from $p(s) = \exp(-u/L(\underline{s})$, they yield $up(\tilde{u}) \to 0$, as $u \to \infty$. The example $p(s) = \exp[-uL(\underline{s})]$ is more complicated and depend on $L(\underline{s})$. Indeed, as $s \to \infty$, $\log[up(\tilde{u})] \sim \log u - \varepsilon L(u)/L[\varepsilon L(u)]$ behaves like $\log u - \varepsilon L(u)$. This expression may converge to $-\infty$, as for example when $L(u) = (\log u)^2$; in those cases $I_0 \to 0$. But this expression may also converge to $+\infty$; in those cases, it does not yield, it is a bound of I_0, and more detailed study is needed to tell whether $I_0 \to 0$. Obviously the issue is far from settled, but this is not the place to pursue the finer study of the domain of validity of the concentration in probability theory.

5.2 A digression: complications concerning the moments, the moment problem, and roles of long-tailedness in classical analysis

Thus far in this chapter, the finiteness of the moments was important, but their actual values and this behavior of EU^q as $q \to \infty$ were barely mentioned. In the slowly random case with $EU^q < \infty$, this behavior of EU^q is a genuinely hard problem. It is even a topic in what is called "fine (or hard) mathematical analysis" that repeatedly attracted the best minds. Unfortunately, the pure mathematical results are not of direct help to users: the complications that attract the mathematicians' interest prove to be a burden in concrete uses.

Convergence of the Taylor expansion of the characteristic function, and a related alternative definition of long-tailed randomness. It is widely taken for granted that the characteristic function (Fourier transform)

$$\varphi(s) = Ee^{isu} = \int_0^\infty e^{isu} p(u) du$$

always has the Taylor expansion

$$\varphi(s) = Ee^{isu} = \sum i^q s^q = \sum i^q s^q \frac{EU^q}{q!}.$$

When lim $\sqrt[q]{EU^q/q!}$ exists, this limit is the inverse of the radius of convergence of this Taylor series. (When there is no limit, limit sup $\sqrt[q]{EU^q/q!}$ always exists and is the inverse of the radius of convergence.)

For the exponential, the series expansion does indeed represent the analytic function $\varphi(s) = \Sigma i^q s^q = 1/(1 - is)$. The radius of convergence is 1, and $\gamma(s) = Ee^{isU}$.

For the Gaussian $\varphi(s) = \exp(-2\sigma^2 s^2)$. Here, $EU^q = 0$ if q is odd and $EU^q = q!/2(q/2)!$ if q is even. The formal Taylor expansion has an infinite radius of convergence, defining $\exp(-2\sigma^2 s^2)$ as an "entire function."

But the lognormal yields lim sup $\sqrt[q]{EU^q/q!} = \infty$. The function $\gamma(s)Ee^{isU}$ is well-defined, but its formal Taylor series *fails to converge for* $s \neq 0$.

There is a strong temptation to dismiss those properties of the lognormal as meaningless mathematical blips. But they could also provide yet another alternative definition of long-tailed randomness. To do so, it is useful, when $\sqrt[q]{EU^q/q\,!}$ becomes infinite for $q \geq \alpha$, to also write lim sup $\sqrt[q]{EU^q/q!} = \infty$. When this is done, the criterion

$$\text{lim sup} \sqrt[q]{EU^q/q!} < \infty \text{ versus lim sup} \sqrt[q]{EUq/q!} = \infty$$

is a criterion of mild versus long-tailed randomness.

The moment problems and additional possible definitions of long-tailed randomness. The following questions were posed by Thomas Stieltjes (1856-1894). Given a sequence M_q, does there exist a measure U (a generalized probability distribution) such that $EU^q = M_q$? If U exists, is it unique? Stieltjes 1894 gave the lognormal as one of several examples where U exists, but is not unique. (See also p. 22 of Shohat & Tamarkin 1943). This property was rediscovered in Heyde 1963, recorded in Feller 1950 (Vol 2, 2nd edition, p. 227), and mentioned in studies of turbulence, including M 1974f{N15}, without suggesting any practical consequence.

The the available partial criteria are either sufficient or necessary, and are not the same on the line and the half-line. Loosely speaking, each known criterion is a way to distinguish between short and long-tailedness the murky border region around mild randomness. The same is true of the criteria encountered in the theory of "quasi-analytic" functions. Some criteria are worth mentioning:

Krein implicitly defines long-tailedness by the convergence of $J=\int_0^\infty \log\, p(u)(1+u^2)^{-1}du$. Koosis 1988-92 is a two-volume treatise that describes many problems where the conditions $J=-\infty$ and $J>-\infty$ are, respectively, the correct ways of expressing that the density $p(u)$ is short or long-tailed. Krein's definition is far more general than the convexity of $\log p(u)$. It is also a little more restrictive, because of the difference it makes between the forms $-u/\log u$ and $-u/(\log u)^2$ for $\log p(u)$.

Carleman implicitly defines long-tailedness by the convergence of $C=\sum(EU^q)^{-1/(2q)}$. For a distribution on the positive half-line to be determined by its moments, a sufficient condition is $C=\infty$. When U is bounded, $EU^q=(u_{\max})^Q$, therefore $C=\infty$. The exponential or the Gaussian also yields $C=\infty$. But $C<\infty$ holds for the scaling and the lognormal.

To conclude, my doubling criterion $P_2=2P$ is a new addition to an already overflowing collection. Who knows, perhaps this newcomer may add fresh spice to an aging mathematical game, or conversely.