Chapter 1

HEAVY TAILS IN FINANCE FOR INDEPENDENT OR MULTIFRACTAL PRICE INCREMENTS

BENOIT B. MANDELBROT

Sterling Professor of Mathematical Sciences, Yale University, New Haven, CT 065020-8283, USA

Contents

Handbook of Heavy Tailed Distributions in Finance, Edited by S.T. Rachev

Abstract

This chapter has two goals. Section 1 sketches the history of heavy tails in finance through the author's three successive models of the variation of a financial price: mesofractal, unifractal and multifractal. The heavy tails occur, respectively, in the marginal distribution only (Mandelbrot, 1963), in the dependence only (Mandelbrot, 1965), or in both (Mandelbrot, 1997). These models increase in the scope of the "principle of scaling invariance", which the author has used since 1957.

The mesofractal model is founded on the stable processes that date to Cauchy and Lévy. The unifractal model uses the fractional Brownian motions introduced by the author. By now, both are well-understood.

To the contrary, one of the key features of the multifractals (Mandelbrot, 1974a, b) remains little known. Using the author's recent work, introduced for the first time in this chapter, the exposition can be unusually brief and mathematically elementary, yet covering all the key features of multifractality. It is restricted to very special but powerful cases: (a) the Bernoulli binomial measure, which is classical but presented in a little-known fashion, and (b) a new two-valued "canonical" measure. The latter generalizes Bernoulli and provides an especially short path to negative dimensions, divergent moments, and divergent (i.e., long range) dependence. All those features are now obtained as separately tunable aspects of the same set of simple construction rules.

My work in finance is well-documented in easily accessible sources, many of them reproduced in Mandelbrot (1997 and also in 2001a, b, c, d). That work having expanded and been commented upon by many authors, a survey of the literature is desirable, but this is a task I cannot undertake now. However, it was a pleasure to yield to the entreaties of this Handbook's editors by a text in which a new technical contribution is preceded by an introductory sketch followed by a simple new presentation of an old feature that used to be dismissed as "technical", but now moves to center stage.

The history of heavy tails in finance began in 1963. While acknowledging that the successive increments of a financial price are interdependent, I assumed independence as a first approximation and combined it with the principle of scaling invariance. This led to (Lévy) stable distributions for the price changes. The tails are very heavy, in fact, power-law distributed with an exponent $\alpha < 2$.

The multifractal model advanced in Mandelbrot (1997) extends scale invariance to allow for dependence. Readily controllable parameters generate tails that are as heavy as desired and can be made to follow a power-law with an exponent in the range $1 < \alpha < \infty$. This last result, an essential one, involves a property of multifractals that was described in Mandelbrot (1974a, b) but remains little known among users. The goal of the example described after the introduction is to illustrate this property in a very simple form.

1. Introduction: A path that led to model price by Brownian motion (Wiener or fractional) of a multifractal trading time

Given a financial price record $P(t)$ and a time lag $\mathrm{d}t$, define $L(t, \mathrm{d}t) = \log P(t + \mathrm{d}t) - \log P(t)$. The 1900 dissertation of Louis Bachelier introduced Brownian motion as a model of $P(t)$. In later publications, however, Bachelier acknowledged that this is a very rough first approximation: he recognized the presence of heavy tails and did not rule out dependence. But until 1963, no one had proposed a model of the heavy tails' distribution.

1.1. From the law of Pareto to infinite moment "anomalies" that contradict the Gaussian "norm"

All along, search for a model was inspired by a finding rooted in economics outside of finance. Indeed, the distribution of personal incomes proposed in 1896 by Pareto involved tails that are heavy in the sense of following a power-law distribution $Pr\{U > u\} = u^{-\alpha}$.

However, almost nobody took this income distribution seriously. The strongest "conventional wisdom" argument against Pareto was that the value $\alpha = 1.7$ that he claimed leads to the variance of U being infinite.

Infinite moments have been a perennial issue both before my work and (unfortunately) ever since. Partly to avoid them, Pareto volunteered an exponential multiplier, resulting in

$$Pr\{U > u\} = u^{-\alpha} \exp(-\beta u).$$

Also, Herbert A. Simon expressed a universally held view when he asserted in 1953 that infinite moments are (somehow) "improper". But in fact, the exponential multipliers are not needed and infinite moments are perfectly proper and have important consequences. In multifractal models, depending on specific features, variance can be either finite or infinite. In fact, all moments can be finite, or they can be finite only up to a critical power q_{crit} that may be 3, 4, or any other value needed to represent the data.

Beginning in the late 1950s, a general theme of my work has been that the uses of statistics must be recognized as falling into at least two broad categories. In the "normal" category, one can use the Gaussian distribution as a good approximation, so that the common replacement of the term, "Gaussian", by "normal" is fully justified. To the contrary, in the category one can call "abnormal" or "anomalous", the Gaussian is very misleading, even as an approximation.

To underline this distinction, I have long suggested – to little effect up to now – that the substance of the so-called ordinary *central limit* theorem would be better understood if it is relabeled as the *center* limit theorem. Indeed, that theorem concerns the *center* of the distribution, while the anomalies concern the *tails.* Following up on this vocabulary, the *generalized central* limit theorem that yields Lévy stable limits would be better understood if called a *tail* limit theorem. This distinction becomes essential in Section 8.5.

Be that as it may, I came to believe in the 1950s that the power-law distribution and the associated infinite moments are key elements that distinguish economics from classical physics. This distinction grew by being extended from independent to highly dependent random variables. In 1997, it became ready to be phrased in terms of randomness and variability falling in one of several distinct "states". The "mild" state prevails for classical errors of observation and for sequences of near-Gaussian and near-independent quantities. To the contrary, phenomena that present deep inequality necessarily belong to the "wild" state of randomness.

1.2. A scientific principle: scaling invariance in finance

A second general theme of my work is the "principle" that financial records are invariant by dilating or reducing the scales of time and price in ways suitably related to each other. There is no need to believe that this principle is exactly valid, nor that its exact validity could ever be tested empirically. However, a proper application of this principle has provided the basis of models or scenarios that can be called good because they satisfy all the following properties:

(a) they closely model reality,

(b) they are exceptionally *parsimonious*, being based on very few very general a priori assumptions, and

(c) they are *creative* in the following sense: extensive and correct predictions arise as consequences of a few assumptions; when those assumptions are changed the consequences also change. By contrast, all too many financial models start with Brownian motion, then build upon it by including in the input every one of the properties that one wishes to see present in the output.

1.3. Analysis alone versus statistical analysis followed by synthesis and graphic output

The topic of multifractal functions has grown into a well-developed analytic theory, making it easy to apply the multifractal formalism blindly. But it is far harder to understand it and draw consequences from its output. In particular, statistical techniques for handling multifractals are conspicuous by their near-total absence. After they become actually available, their applicability will have to be investigated carefully.

A chastening example is provided by the much simpler question of whether or not financial series exhibit global (long range) dependence. My claim that they do was largely based on *R/S* analysis which at this point relies heavily on graphical evidence. Lo (1991) criticized this conclusion very severely as being subjective. Also, a certain alternative test Lo described as "objective" led to a mixed pattern of "they do" and "they do not". This pattern being practically impossible to interpret, Lo took the position that the simpler outcome has not been shown wrong, hence one can assume that long range dependence is absent.

Unfortunately, the "objective test" in question assumed the margins to be Gaussian. Hence, Lo's experiment did not invalidate my conclusion, only showed that the test is not robust and had repeatedly failed to recognize long range dependence.

The proper conclusion is that careful graphic evidence has not yet been superseded. The first step is to attach special importance to models for which sample functions can be generated.

1.4. Actual implementation of scaling invariance by multifractal functions: it requires additional assumptions that are convenient but not a matter of principle, for example, separability and compounding

By and large, an increase in the number and specificity in the assumptions leads to an increase in the specificity of the results. It follows that generality may be an ideal unto itself in mathematics, but in the sciences it competes with specificity, hence typically with simplicity, familiarity, and intuition.

In the case of multifractal functions, two additional considerations should be heeded. The so-called multifractal formalism (to be described below) is extremely important. But it does not by itself specify a random function closely enough to allow analysis to be followed by synthesis. Furthermore, multifractal functions are so new that it is best, in a first stage, to be able to rely on existing knowledge while pursuing a concrete application. For these and related reasons, my study of multifractals in finance has relied heavily on two special cases.

One is implemented by the recursive "cartoons" investigated in Mandelbrot (1997) and in much greater detail in Mandelbrot (2001c).

The other uses compounding. This process begins with a random function $F(\theta)$ in which the variable θ is called an "intrinsic time". In the key context of financial prices, θ is called "trading time". The possible functions $F(\theta)$ include all the functions that have been previously used to model price variation. Foremost is the Wiener Brownian motion $B(t)$

postulated by Bachelier. The next simplest are the fractional Brownian motion $B_H(t)$ and the Lévy stable "flight" $L(t)$.

A separate step selects for the intrinsic trading time a scale invariant random functions of the physical "clock time" t. Mandelbrot (1972) recommended for the function $\theta(t)$ the integral of a multifractal measure. This choice was developed in Mandelbrot (1997) and Mandelbrot, Calvet and Fisher (1997).

In summary, one begins with two statistically independent random functions $F(\theta)$ and $\theta(t)$, where $\theta(t)$ is non-decreasing. Then one creates the "compound" function $F[\theta(t)] = \varphi(t)$. Choosing $F(\theta)$ and $\theta(t)$ to be scale-invariant insures that $\varphi(t)$ will be scale-invariant as well. A limitation of compounding as defined thus far is that it demands independence of F and θ, therefore restricts the scope of the compound function.

In a well-known special case called Bochner subordination, the increments of $\theta(t)$ are independent. As shown in Mandelbrot and Taylor (1967), it follows that $B[\theta(t)]$ is a Lévy stable process, i.e., the mesofractal model. This approach has become well-known. The tails it creates are heavy and do follow a power law distribution but there are at least two drawbacks. The exponent α is at most 2, a clearly unacceptable restriction in many cases, and the increments are independent.

Compounding beyond subordination was introduced because it allows α to take any value > 1 and the increments to exhibit long term dependence. All this is discussed elsewhere (Mandelbrot, 1997 and more recent papers).

The goal of the remainder of this chapter is to use a specially designed simple case to explain how multifractal measure suffices to create a power-law distribution. The idea is that $L(t, \mathrm{d}t) = \mathrm{d}\varphi(t)$ where $\varphi = B_H[\theta(t)]$. Roughly, $\mathrm{d}\mu(t)$ is $|\mathrm{d}B_H|^{1/H}$. In the Wiener Brownian case, $H = 1/2$ and $\mathrm{d}\mu$ is the "local variance". This is how a price that fluctuates up and down is reduced to a positive measure.

2. Background: the Bernoulli binomial measure and two random variants: shuffled and canonical

The prototype of all multifractals is nonrandom: it is a Bernoulli binomial measure. Its well-known properties are recalled in this section, then Section 3 introduces a random "canonical" version. Also, all Bernoulli binomial measures being powers of one another, a broader viewpoint considers them as forming a single "class of equivalence".

2.1. *Definition and construction of the Bernoulli binomial measure*

A multiplicative nonrandom cascade. A recursive construction of the Bernoulli binomial measures involves an "initiator" and a "generator". The initiator is the interval $[0, 1]$ on which a unit of mass is uniformly spread. This interval will recursively split into halves, yielding dyadic intervals of length 2^{-k}. The generator consists in a single parameter u, variously called *multiplier* or *mass*. The first stage spreads mass over the halves of every dyadic interval, with unequal proportions. Applied to $[0, 1]$, it leaves the mass u in $[0, 1/2]$

and the mass v in $[1/2, 1]$. The $(k+1)$-th stage begins with dyadic intervals of length 2^{-k}, each split in two subintervals of length 2^{-k-1}. A proportion equal to u goes to the left subinterval and the proportion v, to the right.

After k stages, let φ_0 and $\varphi_1 = 1 - \varphi_0$ denote the relative frequencies of 0's and 1's in the finite binary development $t = 0.\beta_1\beta_2 \ldots \beta_k$. The "pre-binomial" measures in the dyadic interval $[\mathrm{d}t] = [t, t + 2^{-k}]$ takes the value

$$\mu_k(\mathrm{d}t) = u^{k\varphi_0} v^{k\varphi_1},$$

which will be called "pre-multifractal". This measure is distributed uniformly over the interval. For $k \to \infty$, this sequence of measures $\mu_k(\mathrm{d}t)$ has a limit $\mu(\mathrm{d}t)$, which is the Bernoulli binomial multifractal.

Shuffled binomial measure. The proportion equal to u now goes to either the left or the right subinterval, with equal probabilities, and the remaining proportion v goes to the remaining subinterval. This variant must be mentioned but is not interesting.

2.2. The concept of canonical random cascade and the definition of the canonical binomial measure

Mandelbrot (1974a, b) took a major step beyond the preceding constructions.

The random multiplier M. In this generalization every recursive construction can be described as follows. Given the mass m in a dyadic interval of length 2^{-k}, the two subintervals of length 2^{-k-1} are assigned the masses $M_1 m$ and $M_2 m$, where M_1 and M_2 are *independent* realizations of a random variable M called multiplier. This M is equal to u or v with probabilities $p = 1/2$ and $1 - p = 1/2$.

The Bernoulli and shuffled binomials both impose the constraint that $M_1 + M_2 = 1$. The canonical binomial does not. It follows that the canonical mass in each interval of duration 2^{-k} is multiplied in the next stage by the sum $M_1 + M_2$ of two independent realizations of M. That sum is either $2u$ (with probability p^2), or 1 (with probability $2(1-p)p$), or $2v$ (with probability $1 - p^2$).

Writing p instead of $1/2$ in the Bernoulli case and its variants complicates the notation now, but will soon prove advantageous: the step to the TVCM will simply consist in allowing $0 < p < 1$.

2.3. Two forms of conservation: strict and on the average

Both the Bernoulli and shuffled binomials repeatedly redistribute mass, but within a dyadic interval of duration 2^{-k}, the mass remains exactly conserved in all stages beyond the k-th. That is, the limit mass $\mu(t)$ in a dyadic interval satisfies $\mu_k(\mathrm{d}t) = \mu(\mathrm{d}t)$.

In a canonical binomial, to the contrary, the sum $M_1 + M_2$ is not identically 1, only its expectation is 1. Therefore, canonical binomial construction preserve mass on the average, but not exactly.

The random variable Ω. In particular, the mass $\mu([0, 1])$ is no longer equal to 1. It is a basic random variable denoted by Ω and discussed in Section 4.

Within a dyadic interval dt of length 2^{-k}, the cascade is simply a reduced-scale version of the overall cascade. It transforms the mass $\mu_k(\mathrm{d}t)$ into a product of the form $\mu(\mathrm{d}t) = \mu_k(\mathrm{d}t)\Omega(\mathrm{d}t)$ where all the $\Omega(\mathrm{d}t)$ are independent realizations of the same variable Ω.

2.4. *The term "canonical" is motivated by statistical thermodynamics*

As is well known, statistical thermodynamics finds it valuable to approximate large systems as juxtapositions of parts, the "canonical ensembles", whose energy only depends on a common temperature and not on the energies of the other parts. Microcanonical ensembles' energies are constrained to add to a prescribed total energy. In the study of multifractals, the use of this metaphor should not obscure the fact that the multiplication of canonical factors introduces strong dependence among $\mu(\mathrm{d}t)$ for different intervals dt.

2.5. *In every variant of the binomial measure one can view all finite (positive or negative) powers together, as forming a single "class of equivalence"*

To any given real exponent $g \neq 1$ and multipliers u and v corresponds a multiplier M_g that can take either of two values $u_g = \psi u^g$ with probability p, and $v_g = \psi v^g$ with probability $1 - p$. The factor ψ is meant to insure $pu_g + (1 - p)v_g = 1/2$. Therefore, $\psi[pu^g + (1 - p)v^g] = 1/2$, that is, $\psi = 1/[2EM^g]$. The expression $2EM^g$ will be generalized and encountered repeatedly especially through the expression

$$\tau(q) = -\log_2\bigl[pu^q + (1 - p)v^q\bigr] - 1 = -\log_2\bigl(2EM^q\bigr).$$

This is simply a notation at this point but will be justified in Section 5. It follows that $\psi = 2^{-\tau(g)}$, hence

$$u_g = u^g 2^{\tau(g)} \quad \text{and} \quad v_g = v^g 2^{\tau(g)}.$$

Assume $u > v$. As g ranges from 0 to ∞, u_g ranges from $1/2$ to 1 and v_g ranges from $1/2$ to 0; the inequality $u_g > v_g$ is preserved. To the contrary, as g ranges from 0 to ∞, $v_g < u_g$. For example, $g = -1$ yields

$$u_g = \frac{1/u}{1/u + 1/v} = v \quad \text{and} \quad v_g = \frac{1/v}{1/v + 1/v} = u.$$

Thus, inversion leaves both the shuffled and the canonical binomial measures unchanged. For the Bernoulli binomial, it only changes the direction of the time axis.

Altogether, every Bernoulli binomial measure can be obtained from any other as a reduced positive or negative power. If one agrees to consider a measure and its reduced powers as equivalent, *there is only one Bernoulli binomial measure.*

In concrete terms relative to non-infinitesimal dyadic intervals, the sequences representing $\log \mu$ for different values of g are mutually affine. Each is obtained from the special case $g = 1$ by a multiplication by g followed by a vertical translation.

2.6. The full and folded forms of the address plane

In anticipation of TVCM, the point of coordinates u and v will be called the *address* of a binomial measure in a *full address space*. In that plane, the locus of the Bernoulli measures is the interval defined by $0 < v,\ 0 < u$, and $u + v = 1$.

The *folded address space* will be obtained by identifying the measures (u, v) and (v, u), and representing both by one point. The locus of the Bernoulli measures becomes the interval defined by the inequalities $0 < v < u$ and $u + v = 1$.

2.7. Alternative parameters

In its role as parameter added to $p = 1/2$, one can replace u by the ("information-theoretical") fractal dimension $D = -u \log_2 u - v \log_2 v$ which can be chosen at will in this open interval $]0, 1[$. The value of D characterizes the "set that supports" the measure. It received a new application in the new notion of multifractal concentration described in Mandelbrot (2001c). More generally, the study of all multifractals, including the Bernoulli binomial, is filled with fractal dimensions of many other sets. All are unquestionably positive. One of the newest features of the TVCM will prove to be that they also allow negative dimensions.

3. Definition of the two-valued canonical multifractals

3.1. Construction of the two-valued canonical multifractal in the interval [0, 1]

The TVCM are called two-valued because, as with the Bernoulli binomial, the multiplier M can only take 2 possible values u and v. The novelties are that p need not be $1/2$, the multipliers u and v are not bounded by 1, and the inequality $u + v \neq 1$ is acceptable.

For $u + v \neq 1$, the total mass cannot be preserved exactly. Preservation on the average requires

$$EM = pu + (1 - p)v = \frac{1}{2},$$

hence $0 < p = (1/2 - v)/(u - v) < 1$.

The construction of TVCM is based upon a recursive subdivision of the interval $[0, 1]$ into equal intervals. The point of departure is, once again, a uniformly spread unit mass. The first stage splits $[0, 1]$ into two parts of equal lengths. On each, mass is poured uniformly, with the respective densities M_1 and M_2 that are independent copies of M. The second stage continues similarly with the interval $[0, 1/2]$ and $[1/2, 1]$.

3.2. *A second special two-valued canonical multifractal: the unifractal measure on the canonical Cantor dust*

The identity $EM = 1/2$ is also satisfied by $u = 1/2p$ and $v = 0$. In this case, let the lengths and number of non-empty dyadic cells after k stages be denoted by $\Delta t = 2^{-k}$ and N_k. The random variable N_k follows a simple birth and death process leading to the following alternative.

When $p > 1/2$, $EN_k = (EN_1)^k = (2p)^k = (\mathrm{d}t)^{\log(2p)}$. To be able to write $EN_k = (\mathrm{d}t)^{-D}$, it suffices to introduce the exponent $D = -\log(2p)$. It satisfies $D > 0$ and defines a fractal dimension.

When $p < 1/2$, to the contrary, the number of non-empty cells almost surely vanishes asymptotically. At the same time, the formal fractal dimension $D = -\log(2p)$ satisfies $D < 0$.

3.3. *Generalization of a useful new viewpoint: when considered together with their powers from $-\infty$ to ∞, all the TVCM parametrized by either p or $1 - p$ form a single class of equivalence*

To take the key case, the multiplier M^{-1} takes the values

$$u_{-1} = \frac{1/u}{2(p/u + (1-p)/v)} = \frac{v}{2(v+u) - 1} \quad \text{and} \quad v_{-1} = \frac{u}{2(v+u) - 1}.$$

It follows that $pu_{-1} + (1-p)v_{-1} = 1/2$ and $u_{-1}/v_{-1} = v/u$. In the full address plane, the relations imply the following: (a) the point (u_{-1}, v_{-1}) lies on the extension beyond $(1/2, 1/2)$ of the interval from (u, v) to $(1/2, 1/2)$ and (b) the slopes of the intervals from 0 to (u, v) and from 0 to (u_{-1}, v_{-1}) are inverse of one another. It suffices to fold the full phase diagram along the diagonal to achieve $v > u$. The point (u_{-1}, v_{-1}) will be the intersection of the interval corresponding to the probability $1 - p$ and of the interval joining 0 to (u, v).

3.4. *The full and folded address planes*

In the full address plane, the locus of all the points (u, v) with fixed p has the equation $pu + (1-p)v = 1/2$. This is the negatively sloped interval joining the points $(0, 1/2p)$ and $([1/2(1-p)], 0)$. When (u, v) and (v, u) are identified, the locus becomes the same interval plus the negatively sloped interval from $[0, 1/2(1-p)]$ to $(1/2p, 0)$.

In the folded address plane, the locus is made of two shorter intervals from $(1, 1)$ to both $(1/2p, 0)$ and $([1/2(1-p)], 0)$. In the special case $u + v = 1$ corresponding to $p = 1/2$, the two shorter intervals coincide.

Those two intervals correspond to TVCM in the same class of equivalence. Starting from an arbitrary point on either interval, positive moments correspond to points to the same interval and negative moments, to points of the other. Moments for $g > 1$ correspond to points to the left on the same interval; moments for $0 < g < 1$, to points to the right on the same interval; negative moments to points on the other interval.

For $p \neq 1/2$, the class of equivalence of p includes a measure that corresponds to $u = 1$ and $v = [1/2 - \min(p, 1-p)]/[\max(p, 1-p)]$. This novel and convenient universal point of reference requires $p \neq 1/2$. In terms to be explained below, it corresponds to $\alpha_{\min} = -\log u = 0$.

3.5. Background of the two-valued canonical measures in the historical development of multifractals

The construction of TVCM is new but takes a well-defined place among the three main approaches to the development of a theory of multifractals.

General mathematical theories came late and have the drawback that they are accessible to few non-mathematicians and many are less general than they seem.

The heuristic presentation in Frisch and Parisi (1985) and Halsey et al. (1986) came after Mandelbrot (1974a, b) but before most of the mathematics. Most importantly for this paper's purpose, those presentations fail to include significantly random constructions, hence cannot yield measures following the power law distribution.

Both the mathematical and the heuristic approaches seek generality and only later consider the special cases. To the contrary, a third approach, the first historically, began in Mandelbrot (1974a, b) with the careful investigation of a variety of special random multiplicative measures. I believe that each feature of the general theory continues to be best understood when introduced through a special case that is as general as needed, but no more. The general theory is understood very easily when it comes last.

In pedagogical terms, the "third way" associates with each distinct feature of multifractals a special construction, often one that consists of generalizing the binomial multifractal in a new direction. TVCM is part of a continuation of that effective approach; it could have been investigated much earlier if a clear need had been perceived.

4. The limit random variable $\Omega = \mu([0, 1])$, its distribution and the star functional equation

4.1. The identity $EM = 1$ implies that the limit measure has the "martingale" property, hence the cascade defines a limit random variable $\Omega = \mu([0, 1])$

We cannot deal with martingales here, but positive martingales are mathematically attractive because they converge (almost surely) to a limit. But the situation is complicated because the limit depends on the sign of $D = 2[-pu \log_2 u - (1-p)v \log_2 v]$.

Under the condition $D > 0$, which is discussed in Section 9, what seemed obvious is confirmed: $Pr\{\Omega > 0\} > 0$, conservation on the average continues to hold as $k \to \infty$, and Ω is either non-random, or is random and satisfies the identity $E\Omega = 1$.

But if $D < 0$, one finds that $\Omega = 0$ almost surely and conservation on the average holds for finite k but fails as $k \to \infty$. The possibility that $\Omega = 0$ arose in mathematical esoterica and seemed bizarre, but is unavoidably introduced into concrete science.

4.2. Questions

(A) Which feature of the generating process dominates the tail distribution of Ω? It is shown in Section 6 to be the sign of $\max(u, v) - 1$.
(B) Which feature of the generating process allows Ω to have a high probability of being either very large or very small? Section 6 will show that the criterion is that the function $\tau(q)$ becomes negative for large enough q.
(C) Divide [0, 1] into 2^k intervals of length 2^{-k}. Which feature of the generating process determines the relative distribution of the overall Ω among those small intervals? This relative distribution motivated the introduction of the functions $f(\alpha)$ and $\rho(\alpha)$, and is discussed in Section 8.
(D) Are the features discussed under (B) and (C) interdependent? Section 10 will address this issue and show that, even when Ω has a high probability of being large, its value does not affect the distribution under (C).

4.3. Exact stochastic renormalizability and the "star functional equation" for Ω

Once again, the masses in $[0, 1/2]$ and $[1/2, 1]$ take, respectively, the forms $M_1\Omega_1$ and $M_2\Omega_2$, where M_1 and M_2 are two independent realizations of the random variable M and Ω_1, and Ω_2 are two independent realizations of the random variable Ω. Adding the two parts yields

$$\Omega \equiv \Omega_1 M_1 + \Omega_2 M_2.$$

This identity in distribution, now called the "star equation", combines with $E\Omega = 1$ to determine Ω. It was introduced in Mandelbrot (1974a, b) and has since then been investigated by several authors, for example by Durrett and Liggett (1983). A large bibliography is found in Liu (2002).

In the special case where M is non-random, the star equation reduces to the equation due to Cauchy whose solutions have become well-known: they are the Cauchy–Lévy stable distributions.

4.4. Metaphor for the probability of large values of Ω, arising in the theory of discrete time branching processes

A growth process begins at $t = 0$ with a single cell. Then, at every integer instant of time, every cell splits into a random non-negative number of N_1 cells. At time k, one deals with a clone of N_k cells. All those random splittings are statistically independent and identically distributed. The normalized clone size, defined as N_k/EN_1^k has an expectation equal to 1. The sequence of normalized sizes is a positive martingale, hence (as already mentioned) converges to a limit random variable.

When $EN > 1$, that limit does not reduce to 0 and is random for a very intuitive reason. As long as clone size is small, its growth very much depends on chance, therefore the normalized clone size is very variable. However, after a small number of splittings, a law of large numbers comes into force, the effects of chances become negligible, and the clone grows near-exponentially. That is, the randomness in the relative number of family members can be very large but acts very early.

4.5. *To a large extent, the asymptotic measure Ω of a TVCM is large if, and only if, the pre-fractal measure $\mu_k([0,1])$ has become large during the very first few stages of the generating cascade*

Such behavior is suggested by the analogy to a branching process, and analysis shows that such is indeed the case. After the first stage, the measures $\mu_1([0,1/2])$ and $\mu_1([1/2,1])$ are both equal to u^2 with probability p^2, uv with probability $2p(1-p)$, and v^2 with probability $(1-p)^2$. Extensive simulations were carried out for large k in "batches", and the largest, medium, and smallest measure was recorded for each batch. Invariably, the largest (resp., smallest) Ω started from a high (resp., low) overall level.

5. The function $\tau(q)$: motivation and form of the graph

So far $\tau(q)$ was nothing but a notation. It is important as it is the special form taken for TVCM by a function that was first defined for an arbitrary multiplier in Mandelbrot (1974a, b). (Actually, the little appreciated Figure 1 of that original paper did not include $q<0$ and worked with $-\tau(q)$, but the opposite sign came to be generally adopted.)

5.1. *Motivation of $\tau(q)$*

After k cascade stages, consider an arbitrary dyadic interval of duration $\mathrm{d}t = 2^{-k}$. For the k-approximant TVCM measure $\mu_k(\mathrm{d}t)$ the q-th power has an expected value equal to $[pu^q + (1-p)v^q]^k = \{EM^q\}^k$. Its logarithm of base 2 is

$$\log_2\{[pu^q + (1-p)v^q]^k\} = k\log_2\{pu^q + (1-p)v^q\}$$
$$= \log_2(\mathrm{d}t)[\tau(q)+1].$$

Hence

$$E\mu_k^q(\mathrm{d}t) = (\mathrm{d}t)^{\tau(q)+1}.$$

5.2. *A generalization of the role of Ω: middle- and high-frequency contributions to microrandomness*

Exactly the same cascade transforms the measure in $\mathrm{d}t$ from $\mu_k(\mathrm{d}t)$ to $\mu(\mathrm{d}t)$ and the measure in $[0,1]$ from 1 to Ω. Hence, one can write

$$\mu(\mathrm{d}t) = \mu_k(\mathrm{d}t)\Omega(\mathrm{d}t).$$

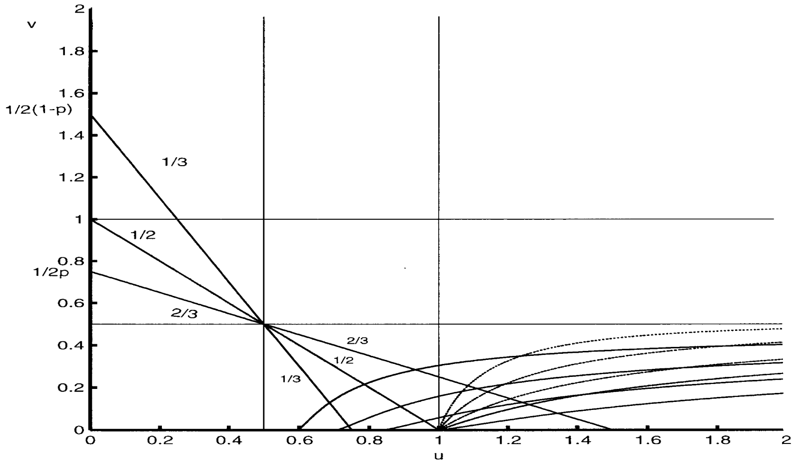

Fig. 1. The full phase diagram of TVCM with coordinates u and v. The isolines of the quantity p are straight intervals from $(1/\{2(1-p)\}, 0)$ to $(0, 1/\{2p\})$. The values p and $1-p$ are equivalent and the corresponding isolines are symmetric with respect to the main bisector $u = v$. The acceptable part of the plane excludes the points (u, v) such that either $\max(u, v) < 1/2$ or $\min(u, v) > 1/2$. Hence, the relevant part of this diagram is made of two infinite halfstrips reducible to one another by folding along the bisector. The folded phase diagram of TVCM corresponds to $v < 0.5 < u$. It shows the following curves. The isolines of $1-p$ and p are straight intervals that start at the point $(1, 1)$ and end at the points $(1/\{2p\}, 0)$ and $(1/\{2(1-p)\}, 0)$. The isolines of D start on the interval $1/2 < u < 1$ of the u-axis and continue to the point $(\infty, 0)$. The isolines of q_{crit} start at the point $(1, 0)$ and continue to the point $(\infty, 0)$. The Bernoulli binomial measure corresponds to $p = 1/2$ and the canonical Cantor measure corresponds to the half line $v = 0,\ u > 1/2$.

In this product, frequencies of wavelength $> \mathrm{d}t$, to be described as "low", contribute $\mu_k([0, 1])$, and frequencies of wavelength $< \mathrm{d}t$, to be described as "high", contribute Ω.

5.3. *The expected "partition function"* $\sum E\mu^q(\mathrm{d}_i t)$

Section 6 will show that $E\Omega^q$ need not be finite. But if it is, the limit measure $\mu(\mathrm{d}t) = \mu_k(\mathrm{d}t)\Omega(\mathrm{d}t)$ satisfies

$$E\mu^q(\mathrm{d}t) = (\mathrm{d}t)^{\tau(q)+1} E\Omega^q.$$

The interval [0, 1] subdivides into $1/\mathrm{d}t$ intervals $\mathrm{d}_i t$ of common length $\mathrm{d}t$. The sum of the q-th moments over those intervals takes the form

$$E\chi(\mathrm{d}t) = \sum E\mu^q(\mathrm{d}_i t) = (\mathrm{d}t)^{\tau(q)} E\Omega^q.$$

Estimation of $\tau(q)$ *from a sample.* It is affected by the prefactor Ω insofar as one must estimate both $\tau(q)$ and $\log E\Omega^q$.

5.4. *Form of the $\tau(q)$ graph*

Due to conservation on the average, $EM = pu + (1-p)v = 1/2$, hence $\tau(1) = -\log_2[1/2] - 1 = 0$. An additional universal value is $\tau(0) = -\log_2(1) - 1 = -1$. For other values of q, $\tau(q)$ is a cap-convex continuous function satisfying $\tau(q) < -1$ for $q < 0$.

For TVCM, a more special property is that $\tau(q)$ is asymptotically linear: assuming $u > v$, and letting $q \to \infty$:

$$\tau(q) \sim -\log_2 p - 1 - q\log u \quad \text{and} \quad \tau(-q) \sim -\log_2(1-p) - 1 + q\log v.$$

The sign of $u - 1$ affects the sign of $\log u$, a fact that will be very important in Section 6.

Moving as little as possible beyond these properties. The very special tau function of the TVCM is simple but Figure 2 suffices to bring out every one of the delicate possibilities first reported in Mandelbrot (1974a), where $-\tau(q)$ is plotted in that little appreciated Figure 1.

Other features of τ that deserve to be mentioned. Direct proofs are tedious and the short proofs require the multifractal formalism that will only be described in Section 11.

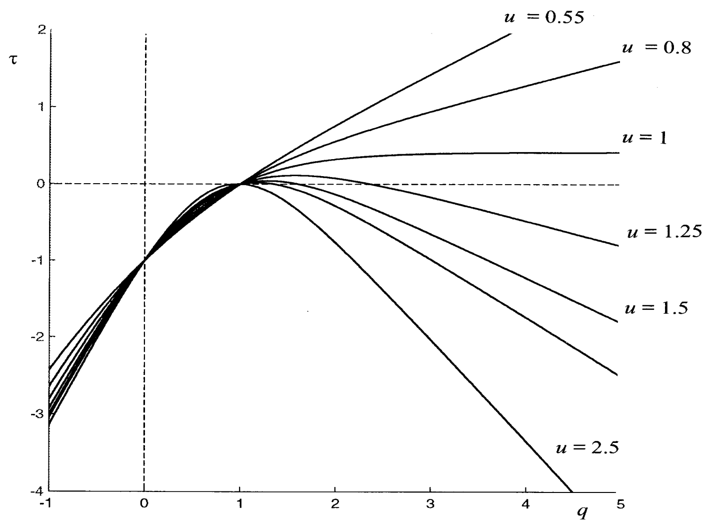

Fig. 2. The function $\tau(q)$ for $p = 3/4$ and varying g. By arbitrary choice, the value $g = 1$ is assigned $u = 1$, from which follows that $g = -1$ is assigned to the case $v = 1$. Behavior of $\tau(q)$ for the value $g > 0$: as $q \to -\infty$, the graph of $\tau(q)$ is asymptotically tangent to $\tau = -q\log_2 v$, as $q \to \infty$, the graph of $\tau(q)$ is asymptotically tangent to $\tau = -q\log_2 u$. Those properties are widely believed to describe the main facts about $\tau(q)$. But for TVCM they do not. Thus, $\tau(q)$ is also tangent to $\tau = q\alpha^*_{\max}$ and $\tau = q\alpha^*_{\min}$. Beyond those points of tangency, f becomes < 0. For $g > 1$, that is, for $u > 1$, $\tau(q)$ has a maximum. Values of q beyond this maximum correspond to $\alpha_{\min} < 0$. Because of the capconvexity of $\tau(q)$, the equation $\tau(q) = 0$ may, in addition to the "universal" value $q = 1$, have a root $q_{\text{crit}} > 1$. For $u > 2.5$, one deals with a very different phenomenon also first described in Mandelbrot (1974a, b). One finds that the construction of TVCM leads to a measure that degenerates to 0.

The quantity $D(q) = \tau(q)/(q-1)$. This popular expression is often called a "generalized dimension", a term too vague to mean anything. $D(q)$ is obtained by extending the line from (q, τ) to $(1, 0)$ to its intercept with the line $q = 0$. It plays the role of a critical embedding codimension for the existence of a finite q-th moment. This topic cannot be discussed here but is treated in Mandelbrot (2003).

The ratio $\tau(q)/q$ *and the "accessible" values of* q. Increase q from $-\infty$ to 0 then to $+\infty$. In the Bernoulli case, $\tau(q)/q$ increases from $\alpha_{\max}$ to ∞, jumps down to $-\infty$ for $q = 0$, then increases again from $-\infty$ to $\alpha_{\min}$. For TVCM with $p \neq 1/2$, the behavior is very different. For example, let $p < 1/2$. As q increases from 1 to ∞, $\tau(q)$ increases from 0 to a maximum $\alpha^*_{\max}$, then decreases. In a way explored in Section 10, the values of $\alpha > \alpha^*_{\max}$ are not "accessible".

5.5. *Reducible and irreducible canonical multifractals*

Once again, being "canonical" implies conservation on the average. When there exists a microcanonical (conservative) variant having the same function $f(\alpha)$, a canonical measure can be called "reducible". The canonical binomial is reducible because its $f(\alpha)$ is shared by the Bernoulli binomial. Another example introduced in Mandelbrot (1989b) is the "Erice" measure, in which the multiplier M is uniformly distributed on [0, 1]. But the TVCM with $p \neq 1/2$ is not reducible.

In the interval [0, 1] subdivided in the base $b = 2$, reducibility demands a multiplier M whose distribution is symmetric with respect to $M = 1/2$. Since $u > 0$, this implies $u < 1$.

6. When $u > 1$, the moment $E\Omega^q$ diverges if q exceeds a critical exponent q_{crit} satisfying $\tau(q) = 0$; Ω follows a power-law distribution of exponent q_{crit}

6.1. *Divergent moments, power-law distributions and limits to the ability of moments to determine a distribution*

This section injects a concern that might have been voiced in Sections 4 and 5. The canonical binomial and many other examples satisfy the following properties, which everyone takes for granted and no one seems to think about: (a) $\Omega = 1$, $E\Omega^q < \infty$, (b) $\tau(q) > 0$ for all $q > 0$, and (c) $\tau(q)/q$ increases monotonically as $q \to \pm\infty$.

Many presentations of fractals take those properties for granted in all cases. In fact, as this section will show, the TVCM with $u > 1$ lead to the "anomalous" divergence $E\Omega^q = \infty$ and the "inconceivable" inequality $\tau(q) < 0$ for $q_{\text{crit}} < q < \infty$. Also, the monotonicity of $\tau(q)/q$ fails for all TVCM with $p \neq 1/2$.

Since Pareto in 1897, infinite moments have been known to characterize the power-law distributions of the form $Pr\{X > x\} = x^{-q_{\text{crit}}}$. But in the case of TVCM and other canonical multifractals, the complicating factor $L(x)$ is absent. One finds that when $u > 1$, the overall measure Ω follows a power law of exponent q_{crit} determined by $\tau(q)$.

6.2. Discussion

The power-law "anomalies" have very concrete consequences deduced in Mandelbrot (1997) and discussed, for example, in Mandelbrot (2001c).

But does all this make sense? After all, $\tau(q)$ and $E\Omega^q$ are given by simple formulas and are finite for all parameters. The fact that those values cannot actually be observed raises a question. Are high moments lost by being unobservable? In fact, they are "latent" but can be made "actual" by a process is indeed provided by the process of "embedding" studied elsewhere.

An additional comment is useful. The fact that high moments are non-observable does not express a deficiency of TVCM but a limitation of the notion of moment. Features ordinarily expressed by moments must be expressed by other means.

6.3. An important apparent "anomaly": in a TVCM, the q-th moment of Ω may diverge

Let us elaborate. From long past experience, physicists' and statisticians' natural impulse is to define and manipulate moments without envisioning or voicing the possibility of their being infinite. This lack of concern cannot extend to multifractals. The distribution of the TVCM within a dyadic interval introduces an additional critical exponent q_{crit} that satisfies $q_{\text{crit}} > 1$. When $1 < q_{\text{crit}} < \infty$, which is a stronger requirement that $D > 0$, the q-th moment of $\mu(\mathrm{d}t)$ diverges for $q > q_{\text{crit}}$.

A stronger result holds: the TVCM cascade generates a measure whose distribution follows the power law of exponent q_{crit}.

Comment. The heuristic approach to non-random multifractals fails to extend to random ones, in particular, it fails to allow $q_{\text{crit}} < \infty$. This makes it incomplete from the viewpoint of finance and several other important applications.

The finite q_{crit} has been around since Mandelbrot (1974a, b) (where it is denoted by α) and triggered a substantial literature in mathematics. But it is linked with events so extraordinarily unlikely as to appear incapable of having any perceptible effect on the generated measure. The applications continue to neglect it, perhaps because it is ill-understood. A central goal of TVCM is to make this concept well-understood and widely adopted.

6.4. An important role of $\tau(q)$: if $q > 1$ the q-th moment of Ω is finite if, and only if, $\tau(q) > 0$; the same holds for $\mu(\mathrm{d}t)$ whenever $\mathrm{d}t$ is a dyadic interval

By definition, after k levels of iteration, the following symbolic equality relates independent realizations of M and μ. That is, it does not link random variables but distributions

$$\mu_k\big([0,1]\big) = M\mu_{k-1}\big([0,1]\big) + M\mu_{k-1}\big([0,1]\big).$$

Conservation on the average is expressed by the identity $E\mu_{k-1}([0,1]) = 1$. In addition, we have the following recursion relative to the second moment.

$$E\mu^2\big([0,1]\big) = 2EM^2\big[E\mu_{k-1}^2\big([0,1]\big)\big] + 2EM^2\big[E\mu_{k-1}\big([0,1]\big)\big]^2.$$

The second term to the right reduces to 1/2. Now let $k \to \infty$. The necessary and sufficient condition for the variance of $\mu_k([0, 1])$ to converge to a finite limit is

$$2\big(EM^2\big) < 1 \quad \text{in other words} \quad \tau(2) = -\log_2\big(EM^2\big) - 1 > 0.$$

When such is the case, Kahane and Peyrière (1976) gave a mathematically rigorous proof that there exists a limit measure $\mu([0, 1])$ satisfying the formal expression

$$E\mu^2\big([0, 1]\big) = \frac{1}{2(1 - 2^{\tau(2)})}.$$

Higher integer moments satisfy analogous recursion relations. That is, knowing that all moments of order up to $q - 1$ are finite, the moment of order q is finite if and only if $\tau(q) > 0$.

The moments of non-integer order q are more delicate to handle, but they too are finite if, and only if, $\tau(q) > 0$.

6.5. Definition of q_{crit}; proof that in the case of TVCM q_{crit} is finite if, and only if, $u > 1$

Section 5.4 noted that the graph of $\tau(q)$ is always cap-convex and for large $q > 0$,

$$\tau(q) \sim -\log_2\big(pu^q\big) + -1 \sim -\log_2 p - 1 - q \log_2 u.$$

The dependence of $\tau(q)$ on q is ruled by the sign of $u - 1$, as follows.

- *The case when $u < 1$, hence $\alpha_{\min} > 0$.* In this case, $\tau(q)$ is monotone increasing and $\tau(q) > 0$ for $q > 1$. This behavior is exemplified by the Bernoulli binomial.
- *The case when $u > 1$, hence $\alpha_{\min} < 0$.* In this case, one has $\tau(q) < 0$ for large q. In addition to the root $q = 1$, the equation $\tau(q) = 1$ has a second root that is denoted by q_{crit}.

Comment. In terms of the function $f(\alpha)$ graphed on Figure 3, the values 1 and q_{crit} are the slopes of the two tangents drawn to $f(\alpha)$ from the origin $(0, 0)$.

Within the class of equivalence of any p and $1 - p$; the parameter g can be "tuned" so that q_{crit} begins by being > 1 then converges to 1; if so, it is seen that D converges to 0.

- Therefore, the conditions $q_{\text{crit}} = 1$ and $D = 0$ describe the same "anomaly".

In Figure 1, isolines of q_{crit} are drawn for $q_{\text{crit}} = 1, 2, 3$, and 4. When $q = 1$ is the only root, it is convenient to say that $q_{\text{crit}} = \infty$. This isoset $q_{\text{crit}} = \infty$ is made of the half-line $\{v = 1/2$ and $u > 1/2\}$ and of the square $\{0 < v < 1/2, 1/2 < u < 1\}$.

6.6. The exponent q_{crit} can be considered as a macroscopic variable of the generating process

Any set of two parameters that fully describes a TVCM can be called "microscopic". All the quantities that are directly observable and can be called macroscopic are functions of those two parameters.

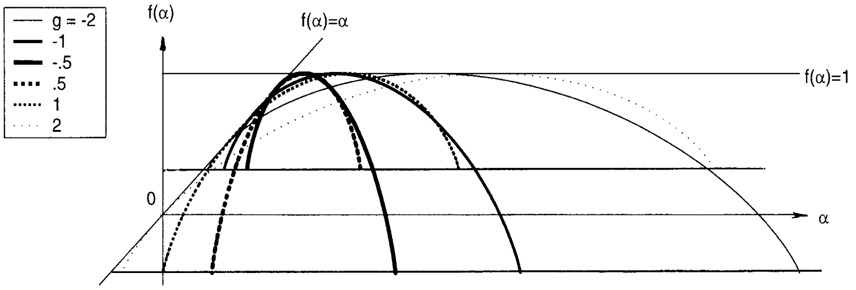

Fig. 3. The functions $f(\alpha)$ for $p = 3/4$ and varying g. All those graphs are linked by horizontal reductions or dilations followed by translation and further self-affinity. It is widely anticipated that $f(\alpha) > 0$ holds in all cases, but for the TVCM this anticipation fails, as shown in this figure. For $g > 0$ (resp., $g < 0$) the left endpoint of $f(\alpha)$ (resp., the right endpoint) satisfies $f(\alpha) < 0$ and the other endpoint, $f(\alpha) > 0$.

For the general canonical multifractal, a full specification requires a far larger number of microscopic quantities but the same number of macroscopic ones. Some of the latter characterize each sample, but others, for example q_{crit}, characterize the population.

7. The quantity α: the original Hölder exponent and beyond

The multiplicative cascades – common to the Bernoulli and canonical binomials and TVCM – involve successive multiplications. An immediate consequence is that both the basic $\mu(\mathrm{d}t)$ and its probability are most intrinsically viewed through their logarithms. A less obvious fact is that a normalizing factor $1/\log(\mathrm{d}t)$ is appropriate in each case. An even less obvious fact is that the normalizations $\log\mu/\log\mathrm{d}t$ and $\log P/\log\mathrm{d}t$ are of far broader usefulness in the study of multifractals. The exact extend of their domain of usefulness is beyond the goal of this chapter, but we keep some special cases that can be treated fully by elementary arguments.

7.1. The Bernoulli binomial case and two forms of the Hölder exponent: coarse-grained (or coarse) and fine-grained

Recall that due to conservation, the measure in an interval of length $\mathrm{d}t = 2^{-k}$ is the same after k stages and in the limit, namely, $\mu(\mathrm{d}t) = \mu_k(\mathrm{d}t)$. As a result, the coarse-grained Hölder exponent can be defined in either of two ways,

$$\alpha(\mathrm{d}t) = \frac{\log\mu(\mathrm{d}t)}{\log(\mathrm{d}t)} \quad \text{and}$$

$$\tilde{\alpha}(\mathrm{d}t) = \frac{\log\mu_k(\mathrm{d}t)}{\log(\mathrm{d}t)}.$$

The distinction is empty in the Bernoulli case but prove prove essential for the TVCM. In terms of the relative frequencies φ_0 and φ_1 defined in Section 2.1,

$$\alpha(\mathrm{d}t) = \tilde{\alpha}(\mathrm{d}t) = \alpha(\varphi_0, \varphi_1) = -\varphi_0 \log_2 u - \varphi_1 \log_2 v$$
$$= -\varphi_0(\log_2 u - \log_2 v) - \log v.$$

Since $u > v$, one has $0 < \alpha_{\min} = -\log_2 u \leqslant \alpha = \tilde{\alpha} \leqslant \alpha_{\max} = -\log_2 v < \infty$. In particular, $\alpha > 0$, hence $\tilde{\alpha} > 0$. As $\mathrm{d}t \to 0$, so does $\mu(\mathrm{d}t)$, and a formal inversion of the definition of α yields

$$\mu(\mathrm{d}t) = (\mathrm{d}t)^{\alpha}.$$

This inversion reveals an old mathematical pedigree. Redefine φ_0 and φ_1 from denoting the finite frequencies of 0 and 1 in an interval, into denoting the limit frequencies at an instant t. The instant t is the limit of an infinite sequence of approximating intervals of duration 2^{-k}. The function $\mu([0, t])$ is non-differentiable because $\lim_{\mathrm{d}t \to 0} \mu(\mathrm{d}t)/\mathrm{d}t$ is not defined and cannot serve to define the local density of μ at the instant $\mathrm{d}t$.

The need for alternative measures of roughness of a singularity expression first arose around 1870 in mathematical esoterica due to L. Hölder. In fractal/multifractal geometry this expression merged with a very concrete exponent due to H.E. Hurst and is continually being generalized. It follows that for the Bernoulli binomial measure, it is legitimate to interpret the coarse αs as finite-difference surrogates of the local (infinitesimal) Hölder exponents.

7.2. *In the general TVCM measure, $\alpha \neq \tilde{\alpha}$, and the link between "$\alpha$" and the Hölder exponent breaks down; one consequence is that the "doubly anomalous" inequalities $\alpha_{\min} < 0$, hence $\tilde{\alpha} < 0$, are not excluded*

A Hölder (Hurst) exponent is necessarily positive. Hence negative $\tilde{\alpha}$s cannot be interpreted as Hölder exponents. Let us describe the heuristic argument that leads to this paradox and then show that $\tilde{\alpha} < 0$ is a serious "anomaly": it shows that the link between "some kind of α" and the Hölder exponent requires a searching look. The resolution of the paradox is very subtle and is associated with the finite q_{crit} introduced in Section 6.5.

Once again, except in the Bernoulli case, $\Omega \neq 1$ and $\mu(\mathrm{d}t) = \mu_k(\mathrm{d}t)\Omega(\mathrm{d}t)$, hence

$$\alpha(\mathrm{d}t) = \tilde{\alpha}(\mathrm{d}t) + \frac{\log \Omega(\mathrm{d}t)}{\log \mathrm{d}t.}$$

In the limit $\mathrm{d}t \to 0$ the factor $\log = \Omega / \log(\mathrm{d}t)$ tends to 0, hence it seems that $\alpha = \tilde{\alpha}$. Assume $u > 1$, hence $\alpha_{\min} < 0$ and consider an interval where $\tilde{\alpha}(\mathrm{d}t) < 0$. The formal equality

$$\text{“}\mu_k(\mathrm{d}t) = (\mathrm{d}t)^{\tilde{\alpha}}\text{”}$$

seems to hold and to imply that "the" mass in an interval increases as the interval length $\to 0$. On casual inspection, this is absurd. On careful inspection, it is not – simply because the variable $\mathrm{d}t = 2^{-k}$ and the function $\mu_k(\mathrm{d}t)$ both depend on k. For example, consider the point t for which $\varphi_0 = 1$. Around this point, one has $\mu_k = u\mu_{k-1} > \mu_{k-1}$. This inequality is not paradoxical.

Furthermore, Section 8 shows that the theory of the multiplicative measures introduces $\tilde{\alpha}$ intrinsically and inevitably and allows $\tilde{\alpha} < 0$.

Those seemingly contradictory properties will be reexamined in Section 9. Values of $\mu(\mathrm{d}t)$ will be seen to have a positive probability but one so minute that they can never be observed in the way $\alpha > 0$ are observed. But they affect the distribution of the variable Ω examined in Section 4, therefore are observed indirectly.

8. The full function $f(\alpha)$ and the function $\rho(\alpha)$

8.1. The Bernoulli binomial measure: definition and derivation of the box dimension function $f(\alpha)$

The number of intervals of denumerator 2^{-k} leading to φ_0 and φ_1 is $N(k, \varphi_0, \varphi_1) = k!/(k\varphi_0)!(k\varphi_1)!$, and $\mathrm{d}t$ is the reduction ratio r from $[0, 1]$ to an interval of duration $\mathrm{d}t$. Therefore, the expression

$$f(k, \varphi_0, \varphi_1) = -\frac{\log N(k, \varphi_0, \varphi_1)}{\log(\mathrm{d}t)} = -\frac{\log[k!/(k\varphi_0)!(k\varphi_1)!]}{\log(\mathrm{d}t)}$$

is of the form $f(k, \varphi_0, \varphi_1) = -\log N/\log r$. Fractal geometry calls this the "box similarity dimension" of a set. This is one of several forms taken by *fractal dimension*. More precisely, since the boxes belong to a grid, it is a *grid fractal dimension.*

The dimension function $f(\alpha)$. For large k, the leading term in the Stirling approximation of the factorial yields

$$\lim_{k\to\infty} f(k, \varphi_0, \varphi_1) = f(\varphi_0, \varphi_1) = -\varphi_0 \log_2 \varphi_0 - \varphi_1 \log_2 \varphi_1.$$

8.2. The "entropy ogive" function $f(\alpha)$; the role of statistical thermodynamics in multifractals and the contrast between equipartition and concentration

Eliminate φ_0 and φ_1 between the functions f and $\alpha = -\varphi_0 \log u - \varphi_1 \log v$. This yields in parametric form a function, $f(\alpha)$. Note that $0 \leqslant f(\alpha) \leqslant \min\{\alpha, 1\}$. Equality to the right is achieved when $\varphi_0 = u$. The value α where $f = \alpha$ is very important and will be discussed in Section 9. In terms of the reduced variable $\varphi_0 = (\alpha - \alpha_{\min})/(\alpha_{\max} - \alpha_{\min})$, the function $f(\alpha)$ becomes the "ogive"

$$\tilde{f}(\varphi_0) = -\varphi_0 \log_2 \varphi_0 - (1 - \varphi_0)\log_2(1 - \varphi_0).$$

This $\tilde{f}(\varphi_0)$ can be called a universal function. The $f(\alpha)$ corresponding to fixed p and varying g are affine transforms of $\tilde{f}(\varphi_0)$, therefore of one another. The ogive function $\tilde{f}$ first arose in thermodynamics as an entropy and in 1948 (with Shannon) entered communication theory as an information. Its occurrence here is the first of several roles the formalism of thermodynamics plays in the theory of multifractals.

An essential but paradoxical feature. Equilibrium thermodynamics is a study of various forms of *near-equality*, for example postulates the equipartition of states on a surface in phase space or of energy among modes. In sharp contrast, multifractals are characterized by extreme *inequality* between the measures in different intervals of common duration $\mathrm{d}t$. Upon more careful examination, the paradox dissolves by being turned around: the main tools of thermodynamics can handle phenomena well beyond their original scope.

8.3. The Bernoulli binomial measure, continued: definition and derivation of a function $\rho(\alpha) = f(\alpha) - 1$ that originates as a rescaled logarithm of a probability

The function $f(\alpha)$ never fully specifies the measure. For example, it does not distinguish between the Bernoulli, shuffled and canonical binomials. The function $f(\alpha)$ can be generalized by being deduced from a function $\rho(\alpha) = f(\alpha) - 1$ that will now be defined. Instead of dimensions, that deduction relies on probabilities. In the Bernoulli case, the derivation of ρ is a minute variant of the argument in Section 8.1, but, contrary to the definition of f, the definition of ρ easily extends to TVCM and other random multifractals.

In the Bernoulli binomial case, the probability of hitting an interval leading to φ_0 and φ_1 is simply $P(k, \varphi_0, \varphi) = N(k, \varphi_0, \varphi_1)2^{-k} = k!/(k\varphi_0)!(k\varphi_1)!2^{-k}$. Consider the expression

$$\rho(k, \varphi_0, \varphi_1) = -\frac{\log[P(k, \varphi_0, \varphi_1)]}{\log(\mathrm{d}t)},$$

which is a rescaled but not averaged form of entropy. For large k, Stirling yields

$$\begin{aligned}\lim_{k\to\infty} \rho(k, \varphi_0, \varphi_1) = \rho(\varphi_0, \varphi_1) &= -\varphi_0 \log_2 \varphi_0 - \varphi_1 \log_2 \varphi_1 - 1 \\ &= f(\alpha) - 1.\end{aligned}$$

8.4. Generalization of $\rho(\alpha)$ to the case of TVCM; the definition of $f(\alpha)$ as $\rho(\alpha) + 1$ is indirect but significant because it allows the generalized f to be negative

Comparing the arguments in Sections 8.1 and 8.2 link the concepts of fractal dimension and of minus log (probability). However, when $f(\alpha)$ is reported through $f(\alpha) = \rho(\alpha) + 1$, the latter is not a mysterious "spectrum of singularities". It is simply the peculiar but proper way a probability distribution must be handled in the case of multifractal measures. Moreover, there is a major a priori difference exploited in Section 10. Minus log (probability) is not subjected to any bound. To the contrary, every one of the traditional definitions of fractal dimension (including Hausdorff–Besicovitch or Minkowski–Bouligand) necessarily yields a positive value.

The point is that the dimension argument in Section 8.1 does not carry over to TVCM, but the probability argument does carry over as follows. The probability of hitting an interval leading to φ_0 and φ_1 now changes to $P(k,\varphi_0,\varphi_1) = p(\varphi_0 k)!/(k\varphi_0)!(k\varphi_1)!$ One can now form the expression

$$\rho(k,\varphi_0,\varphi_1) = -\frac{\log[P(k,\varphi_0,\varphi_1)]}{\log(\mathrm{d}t)}.$$

Stirling now yields

$$\begin{aligned}\rho(\varphi_0,\varphi_1) &= \lim_{k\to\infty} \rho(k,\varphi_0,\varphi_1)\\ &= \{-\varphi_0\log_2\varphi_0 - \varphi_1\log_2\varphi_1\} + \{\varphi_0\log_2 p + \varphi_1\log_2(1-p)\}.\end{aligned}$$

In this sum of two terms marked by braces, we know that the first one transforms (by horizontal stretching and translation) into the entropy ogive. The second is a linear function of φ, namely $\varphi_0[\log_2 p - \log_2(1-p)] + \log_2(1-p)$. It transforms the entropy ogive by an affinity in which the line joining the two support endpoints changes from horizontal to inclined. The overall affinity solely depends on p, but φ_0 depends explicitly on u and v.

This affinity extends to all values of p. Another property familiar from the binomial extends to all values of p. For all u and v, the graphs of $\rho(\alpha)$, hence of $f(\alpha)$ have a vertical slope for $q = \pm\infty$.

Alternatively, $\rho(\varphi_0,\varphi_1) = -\varphi_0\log_2[\varphi_0/p] - \varphi_1\log_2[\varphi_1/(1-p)]$.

8.5. Comments in terms of probability theory

Roughly speaking, the measure μ is a *product* of random variables, while the limit theorems of probability theory are concerned with *sums*. The definition of α as $\log\mu(\mathrm{d}t)/\log(\mathrm{d}t)$ replaces a product of random variables M by a weighted sum of random variables of the form $\log M$. Let us now go through this argument step by step in greater rigor and generality. One needs a cumbersome restatement of $\alpha_k(\mathrm{d}t)$.

The low frequency factor of $\mu_k(\mathrm{d}t)$ and the random variable H_{low}. Consider once again a dyadic cell of length 2^{-k} that starts at $t = 0.\beta_1\beta_2\ldots\beta_k$. The first k stages of the cascade can be called of *low frequency* because they involve multipliers that are constant over dyadic intervals of length $\mathrm{d}t = 2^{-}k$ or longer. These stages yield

$$\mu_k(\mathrm{d}t) = M(\beta_1)M(\beta_1,\beta_2)\cdots M(\beta_1,\ldots,\beta_k) = \prod M.$$

We transform $\mu_k(\mathrm{d}t)$ into the *low frequency* random variable

$$\begin{aligned}H_{\mathrm{low}} &= \frac{\log[\mu_k(\mathrm{d}t)]}{\log(\mathrm{d}t)}\\ &= \frac{1}{k}\big[-\log_2 M(\beta_1) - \log_2 M(\beta_1,\beta_2) - \cdots\big].\end{aligned}$$

We saw in Section 4.5 that the first few values of M largely determine the distribution of Ω. But the last expression involves an operation of averaging in which the first terms contributing to $\mu(\mathrm{d}t)$ are asymptotically washed out.

8.6. *Distinction between "center" and "tail" theorems in probability*

The quantity $\tilde{\alpha}_k(\mathrm{d}t) = \varphi_0 \log_2 u - \varphi_1 \log_2 v$ is the average of a sum of variables $-\log M$; but why is its distribution is not Gaussian and the graph of $\rho(\alpha)$ is an entropy ogive rather than a parabola? Why is this so? The law of large numbers tells us that $\tilde{\alpha}_k(\mathrm{d}t)$ almost surely converges to its expectation which tells us very little. A tempting heuristic argument continues as follows. The central limit theorem is believed to ensure that for small $\mathrm{d}t$, $H_{\mathrm{low}}(\mathrm{d}t)$ becomes Gaussian, therefore the graph of $\log p(\mathrm{d}t)$ should be expected to be a parabola. This being granted, why is it that the Stirling approximation yields an entropy ogive – not a parabola?

In fact, there is no paradox of any kind. While the central limit theorem is indeed central to probability theory, all it asserts in this context is that, asymptotically, the Gaussian rules the *center* of the distribution, its "bell". Renormalizations reduce this center to the immediate neighborhood of the top of the $\rho(\alpha)$ graph and the central limit theorem is correct in asserting that the top of the entropy ogive is locally parabolic. But in the present context this information is of little significance. We need instead an alternative that is only concerned with the tail behavior which it ought to blow up. For this and many other reasons, it would be an excellent idea to speak of *center,* not *central* limit theorem. The tail limit theorem is due to H. Cramer and asserts that the tail consisting in the bulk of the graph is not a parabola but an entropy ogive.

8.7. *The reason for the anomalous inequalities $f(\alpha) < 0$ and $\alpha < 0$ is that, by the definition of a random variable $\mu(\mathrm{d}t)$, the sample size is bounded and is prescribed intrinsically; the notion of supersampling*

The inequality $\rho(\alpha) < -1$ characterizes events whose probability is extraordinarily small. The finding that this inequality plays a significant role was not anticipated, remains difficult to understand and appreciate, and demands comment.

The common response is that even extremely low probability events are captured if one simply takes a sufficiently long sample of independent values. But this is impossible, even if one forgets that, in the present uncommon context, the values are extremely far from being statistically independent. Indeed, the choice the duration $\mathrm{d}t = 2^{-k}$ has two effects. Not only does it fix the distribution of $\mu(\mathrm{d}t)$, but it also sets the sample size at the value $N = 1/\mathrm{d}t = 2^k$. Roughly speaking, a sample of size N can only reveal values having a probability greater than $1/N$, which means $\rho(\alpha) > -1$.

In summary, it is true that decreasing $\mathrm{d}t$ to 2^{-k-1} increases the sample size. But it also changes the distribution and does so in such a way that the bound $\rho = -1$ remains untouched.

This bound excludes ∂u items of information that correspond to $f(\alpha) < 0$ (for example, the value of q_{crit} when finite). Those items remain hidden and latent in the sense that they cannot be inferred from one sample of values of $\mu(\mathrm{d}t)$. Ways of revealing those values, supersampling and embedding, are examined in Mandelbrot (1989b, 1995) and forthcoming Mandelbrot (2003).

Figure 3 shows, for $p = 3/4$, how the graph of $f(\alpha)$ depends on g.

8.8. *Excluding the Bernoulli case $p = 1/2$, TVCM faces either one of two major "anomalies": for $p > -1/2$, one has $f(\alpha_{\min}) = 1 + \log_2 p > 0$ and $f(\alpha_{\max}) = 1 + \log_2(1 - p) < 0$; for $p < 1/2$, the opposite signs hold*

The fact that the values of $\rho(\alpha_{\min}) = f(\alpha_{\min}) - 1$ and $\rho(\alpha_{\max}) = f(\alpha_{\max}) - 1$ are logarithms of probabilities confirms and extends the definition of $p(\alpha) = f(\alpha) - 1$ as a limit rescaled probability. Here, those endpoint values of $f(\alpha)$ are independent of g and the affinity that deduces them from the entropy ogive (with ends on the horizontal axis) characterizes the class of equivalence of p and $1-p$. If, and only if, $p = 1/2$ and $u+v = 1$, that is, in the familiar Bernoulli binomial case, one has $\rho(\alpha_{\min}) = \rho(\alpha_{\max}) = \log_2(1/2) = -1$ hence $f(\alpha_{\min}) = f(\alpha_{\max}) = 0$. When $u + v \neq 1$, one of the endpoints satisfies $f > 0$ and the other satisfies $f < 0$. Sections 8.9 and 10 shall examine the sharply differing consequences of those inequalities.

8.9. *The "minor anomalies" $f(\alpha_{\max}) > 0$ or $f(\alpha_{\min}) > 0$ lead to sample function with a clear "ceiling" or "floor"*

Suppose that $f(\alpha_{\min}) = 0$ and $f(\alpha_{\max}) = 0$, as is the case for $p = 1/2$. Then, using terms often applied to the printed page – but after it has been turned 90° to the side – the sample functions are "non-justified" or "ragged" for both high and low values. That is, the values tend to be unequal; one is clearly larger than all others, a second is clearly the second largest, etc.

To the contrary, TVCM with $p \neq 1/2$ yield either $f(\alpha_{\max}) > 0$ or $f(\alpha_{\min}) > 0$. Sample functions have a conspicuous "ceiling" (resp., a "floor"). That is, a largest (resp., smallest) value is attained repeatedly for values of t belonging to a set of positive dimension. To use the printers' vocabulary, when one side is "ragged" the other is "justified". On visual inspection of the data, the ceiling is always visible; the floor merges with the time axis, except when one plots $\log[\mu(\mathrm{d}t)]$.

9. The fractal dimension $D = \tau'(1) = 2[-pu \log_2 u - (1 - p)v \log_2 v]$ and multifractal concentration

The function $f(\alpha)$ satisfies $f(\alpha) \leqslant \alpha$, with equality $f(\alpha) = \alpha$ when $\alpha = D = \tau\prime(1)$. From the value of $\alpha = D$ follows one of the most important properties of multifractals. Mandelbrot (2001d) proposed to call it "multifractal concentration". This section will first examine its opposite, which is asymptotic negligibility.

9.1. *In the Bernoulli binomial measures weak asymptotic negligibility holds but strong asymptotic negligibility fails*

Recall that during construction, the total binomial measure of [0, 1] remains constant and equal to 1. But the first few stages of construction make its distribution become very unequal and a few values that stand out as sharp spikes. After k stages, the maximum measure is u^k, which is far larger than the minimum measure v^k. From the relations

$$2^{-k} = \mathrm{d}t, \quad 2^k = N, \quad -\log_2 u = \alpha_{\min} < 1, \quad \text{and} \quad -\log_2 v = \alpha_{\min} > 1,$$

it follows that

$$u^k = b^{(-\log_b u)(-k)} = (\mathrm{d}t)^{\alpha_{\min}} = N^{-\alpha_{\min}}.$$

In words: even the maximum u^k tends to 0. This is a *weak* form of asymptotic negligibility following a power-law.

The preceding result holds for every multifractal for which there is an $\alpha_{\min} > 0$ that plays the same role as in the binomial case. (In more general multifractals the same role is held by some $\alpha^*_{\min} > \max\{\alpha_{\min}, 0\}$.)

Similarly, the total contribution of any fixed number of largest spikes is asymptotically negligible.

9.2. *For the Bernoulli or canonical binomials, the equation $f(\alpha) = \alpha$ has one and only one solution; that solution satisfies $D > 0$ and is the fractal dimension of the "carrier" of the measure*

We now proceed to the total contribution of a number of spikes that is no longer fixed but increases with N. In the simplest of all possible worlds, many spikes would have been more or less equal to the largest, and the sum of all the other spikes would have been negligible. If so, the sum of $N^{\alpha_{\min}}$ spikes would have been of the order of $N^{\alpha_{\min}} N^{-\alpha_{\min}} = 1$.

While the world is actually more complicated there is an element of orderliness. The equality $\varphi_0 = u$ is achieved for $\alpha = f(\alpha) = -u \log u - v \log v = D$. For finite but large k, it follows that

$$\mu(k, \varphi_0, \varphi_1) \sim 2^{-k\alpha} = 2^{-kD} \quad \text{and} \quad N(k_1\varphi_0, \varphi_1) \sim 2^{kf(\alpha)} = 2^{kD}.$$

Hence,

$\mu(k_1\varphi_0,\ \varphi_1) N(k_1\varphi_0\varphi_1)$ is approximately equal to 1.

Actually, this product is necessarily $\leqslant 1$ but the difference tends to 0 as $k \to \infty$. That is, an increasingly overwhelming bulk of the measure tends to "concentrate" in the cells where $\alpha = D$. The remainder is small, but in the theory of multifractals even very small remainders are extremely significant for some purposes.

9.3. The notion of "multifractal concentration"

A key feature of multifractals is a subtle interaction between number and size that is elaborated upon in Mandelbrot (2001d). Section 9.2 showed that the contributions that are large are too few to matter. The small contributions are very numerous, but so extremely small that their total contribution is negligible as well. The bulk of the measure is found in a rather inconspicuous intermediate range one can call "mass carrying". Since $D > \alpha_{\min}$, the N^D spikes of size N^{-D} are far smaller than the largest one. Separately, each is asymptotically negligible. But their number N^D is exactly large enough to insure that their total contribution is nearly equal to the overall measure 1. When a sample is plotted, this range does not stand out but it makes a perfect match between size and frequency.

Practically, the number of visible peaks is so small compared to N^D that a combination of the peaks and the intermediate range is still of the order of N^D. The combined range has the advantage of simplicity, since it includes the N^D largest values. Note that the peaks tend to be located in the midst of stretches of values of intermediate size.

9.4. The case of TVCM with $p < 1/2$, allows D to be positive, negative, or zero

Using the alternative expression for $f(\alpha)$ given in Section 8.4, the identity $f(\alpha) = \alpha$ demands the equality of the two expressions

$$f(\alpha) = -\varphi_0 \log_2\left[\frac{\varphi_0}{p}\right] - \varphi_1 \log_2\left[\frac{\varphi_1}{1-p}\right] \quad \text{and} \quad \alpha = -\varphi_0 \log_2 u - \varphi_1 \log_2 v.$$

The solution is, obviously, $\varphi_0 = pu$ and $\varphi_1 = (1-p)v$. The sum $\varphi_1 + \varphi_1$ is 1, as it must. Hence, $D = -pu \log_2 u - (1-p)v \log_2 v$, as announced. The novelty is that TVCM allow $D > 0$, $D = 0$, and $D < 0$.

Familiar role of D under the inequality $D > 0$. Mandelbrot (1974a, b) obtained the following criterion, which has become widely known and includes the TVCM case. When positive, D is the fractal dimension of the "set that supports" the measure. Figure 1 shows isolines of D for $D = 0, 1/4, 1/2$, and $3/4$. The isoline for $D = 1$ is made of the interval $\{u = 1,\ 0 < v < 1\}$ and the half-line $\{v = 1,\ u \geqslant 1\}$. The key result is that, contrary to the Bernoulli binomial case, the half line $1 < q < \infty$ subdivides into up to three subranges of values.

Largely unfamiliar consequence of the inequality $D < 0$. For all non-random multifractals, $\tau\prime(1) > 0$. A casual acquaintance with multifractals takes for granted that this is not changed by randomness. But Mandelbrot (1974a, b) also allows for an alternative possibility, which has so far remained little known. The example of TVCM shows that, in a canonical case, the formally evaluated D can be negative. In the example of TVCM, D is negative when the point (u, v) falls in a domain to the bottom right of the folded phase diagram in Figure 1. The consequences of $D < 0$ are drastic: the multifractal reduces to 0 almost surely and is called degenerate.

A classical "pathological limit" as metaphor. This limit behavior of the distribution of μ seems incompatible with the fact that $E\mu = 1$ by definition. But in fact, no contradiction is observed. A convincing idea of the distribution is provided for each p, by the behavior of the $g \to \infty$ limit of the weights $u^g 2^{\tau(g)}$ and $v^g 2^{\tau(g)}$. This recalls a classical counterexample of analysis, namely, the behavior for $k \to \infty$ of the variable P_k defined as follows: $P_k = k$ with the probability $1/k$ and $P_k = 0$ with the probability $1 - 1/k$. For finite k, one has $EP_k = 1$. But in the limit $k \to \infty$, $P_\infty = 0$, hence $EP_\infty = 0$, so that in the limit the expectation drops discontinuously from 1 to 0. In practice, the preasymptotic measure is extremely small with a high probability and huge with a tiny probability.

The condition $D = 0$. It defines the threshold of degeneracy.

10. A noteworthy and unexpected separation of roles, between the "dimension spectrum" and the total mass Ω; the former is ruled by the accessible α for which $f(\alpha) > 0$, the latter, by the inaccessible α for which $f(\alpha) < 0$

Brought together, Sections 4, 7, 8, and 9 imply, in plain words, that what you do not necessarily see may affect you significantly. This section serves to underline that the notion of canonical multifractal is very subtle and deserves to be well-understood and further discussed.

10.1. Definitions of the "accessible ranges" of the variables: qs from $q^*_{\min}$ to $q^*_{\max}$ and αs from $\alpha^*_{\min}$ to $\alpha^*_{\max}$; the accessible functions $\tau^*(q)$ and $f^*(\alpha)$

Mandelbrot (1995) worked to introduce to the function $f^*(\alpha) = \max\{0, f(\alpha)\}$. That is,

- In the interval $[\alpha^*_{\min}, \alpha^*_{\max}]$ where $f(\alpha) > 0$, $f^*(\alpha) = f(\alpha)$;
- When $f(\alpha) \leqslant 0$, $f^*(\alpha) = 0$.

The graph of $f^*(\alpha)$ is identical to that of $f(\alpha)$ except that the "tails" with $f < 0$ are truncated so that $f^* > 0$. In terms of $\tau(q)$, the equality $f(\alpha) = 0$ corresponds to lines that are tangent to the graph of $\tau(q)$ and also go through $(0, 0)$. In the most general case, those lines' slopes are $\alpha^*_{\min}$ and $\alpha^*_{\max}$ and the points of contact are denoted by $q^*_{\max}$ (satisfying >0) and $q^*_{\min}$ (satisfying <0). Therefore, the function $f^*(\alpha)$ corresponds to the following truncated function $\tau^*(q)$.

- When $q < q^*_{\min}$, $\tau^*(q) = \alpha^*_{\max} q$;
- When $q^*_{\min} < q < q^*_{\max}$, $\tau^*(q) = \tau(q)$;
- When $q > q^*_{\max}$, $\tau^*(q) = \alpha^*_{\min} q$.

In other words, the graph of τ^* is identical to that of τ except that beyond $q^*_{\max}$ or $q^*_{\min}$ it follows the tangents that go through the origins. Therefore it is straight.

For the TVCM, one has either $\alpha^*_{\max} = \alpha_{\max}$ with $q^*_{\min} = -\infty$, or $\alpha^*_{\min} = \alpha_{\min}$ with $q^*_{\max} = \infty$.

10.2. A confrontation

Section 4 noted that the largest values of $\Omega([0, 1])$ are generated when a sample cascade begins with a few large values. Section 7 noted that the value of $\Omega([0, 1])$ – irrespective of size – ceases, for $k \to \infty$, to have any impact on α. Section 8 noted that, again for $k \to \infty$, values of α such that $f(\alpha) < 0$ have a vanishing probability of being observed. Section 9.1 followed up by defining the accessible function $f(\alpha)$. Section 9 returned to large values of $\Omega([0, 1])$ and noted their association with $q_{\text{crit}} < \infty$. The values of α they involve satisfy $\alpha < 0$, hence a fortiori $f(\alpha) < 0$. Those values do not occur in multifractal decomposition, yet they are extremely important.

10.3. The simplest cases where $f(\alpha) > 0$ for all α, as exemplified by the canonical binomial

Here, the large values of Ω are ruled by the left-most part of the graph of $f(\alpha)$. That is, the same graph controls those large values and the distribution of $\Omega([0, 1])$ among the $1/\mathrm{d}t$ intervals of length $\mathrm{d}t$.

10.4. The extreme case where $f(\alpha) < 0$ and $\alpha < 0$ both occur, as exemplified by TVCM when $u > 1$

Due to the inequality $f(\alpha) < \alpha$, the graph of $f(\alpha)$ never intersects the quadrant where $\alpha < 0$ and $f > 0$. The key unexpected fact is that the portions of $f(\alpha)$ within other quadrants play more or less separate roles. In the TVCM case, those quadrants are parts of one (analytically simple) function. But in general they are nearly independent of each other.

The function $f(\alpha)$ was defined as having a graph that lies in the non-anomalous quadrant $\alpha > 0$ and $f > 0$. This f determines completely the multifractal decomposition of our TVCM measure, in particular, the dimension D and the exponents $q^*_{\min}$, $q^*_{\max}$, $\alpha^*_{\min}$ and $\alpha^*_{\max}$.

To the contrary, q_{crit} is entirely determined by the doubly anomalous left tail located in the quadrant characterized by $f(\alpha) < 0$ and $\alpha < 0$. A priori, it was quite unexpected that this quadrant should exist and play *any* role, least of all a central role, in the theory of multifractals. But in fact, q_{crit} has a major effect on the distribution, hence the value of the total measure in an interval.

10.5. The intermediate case where $\alpha_{\min} > 0$ but $f(\alpha) < 0$ for some values of α

When $p < 1/2$, but $u < 1$ so that $q_{\text{crit}} = \infty$ and all moments are finite, large values of μ have a much lower probability than when $u > 1$. As always, however, their probability distribution continues to be determined by the left tail of the probability graph where $f < 0$.

11. A broad form of the multifractal formalism that allows $\alpha < 0$ and $f(\alpha) < 0$

The collection of rules that relate $\tau(q)$ to $f(\alpha)$ is called "multifractal formalism". TVCM was specifically designed to understand multifractals directly, thus avoiding all formalism.

However, general random multifractals more than TVCM demand their own broad multifractal formalism. Once again, the most widely known form of the multifractal formalism does not allow randomness and yields $f(\alpha) > 0$, but the broad formalism first introduced in Mandelbrot (1974a, b) concerns a generalized function for which $f(\alpha) < 0$ is allowed.

11.1. The broad "multifractal formalism" confirms the form of $f(\alpha)$ and allows $f(\alpha) < 0$ for some α

Through a point on the graph of coordinates q and $\tau(q)$, draw the tangent to that graph. Under wide conditions, the tangent's slope is $\alpha(q)$ and its intercept by the ordinate axis is $-f(q)$. Thus

$$\alpha(q) = \frac{\mathrm{d}\tau(q)}{\mathrm{d}q} \quad \text{and} \quad -f(q) = \tau(q) - q\frac{\mathrm{d}\tau(q)}{\mathrm{d}q}.$$

Through the quantities $\alpha(q)$ and $f(q)$, a function $f(\alpha)$ is defined by using q as parameter.

The slope $f'(\alpha)$ is the inverse of the function $\alpha(q)$. The tangent of slope $f'(\alpha)$ intersects the line $\alpha = 0$ at the point of ordinate $-\tau(q)$. The $D(q)$ tangent's equation being $-\tau(q) + q\alpha$, its intersection with the bisector satisfies the condition $-\tau + q = \alpha$, hence $D = \tau(q)/(q-1)$. This is the critical embedding dimension discussed in Section 5.4.

11.2. The Legendre and inverse Legendre transforms and the thermodynamical analogy

The transforms that replace q and $\tau(q)$ by α and $f(\alpha)$, or conversely, are due to Legendre. They play a central role in thermodynamics, as does already the argument that yielded $f(\alpha)$ and $\rho(\alpha)$ in the original formalism introduced in Mandelbrot (1974a, b).

Acknowledgments

The figures in this chapter were prepared by Evgenyi Vilenkin, Yale class of 2003.

References

Chernoff, H., 1952. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. The Annals of Mathematical Statistics 23, 493–507.

Cootner, P.H. (Ed.), 1964. The Random Character of Stock Market Prices. MIT Press, Cambridge, MA.

Daniels, H.E., 1954. Saddlepoint approximations in statistics. The Annals of Mathematical Statistics 25, 631–649.

Durrett, R., Liggett, T.M., 1983. Fixed points of the smoothing transformation. Zeitschrift für Wahrscheinlichkeitstheorie 64, 275–301.

Frisch, U., Parisi, G., 1985. Fully developed turbulence and intermittency. In: Ghil, M. (Ed.), Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics, North-Holland, pp. 84–86. Excepted in Mandelbrot (1999a).

Halsey, T.C., Jensen, M.H., Kadanoff, L.P., Procaccia, I., Shraiman, B.I., 1986. Fractal measures and their singularities: the characterization of strange sets. Physical Review A 33, 1141–1151.

Hentschel, H.G.E., Procaccia, I., 1983. The infinite number of generalized dimensions of fractals and strange attractors. Physica (Utrecht) 8D, 435–444.

Kahane, J. P., Peyrière, J., 1976. Sur certaines martingales de B. Mandelbrot. Advances in Mathematics 22, 131–145. Translated in Mandelbrot (1999a) as Chapter N17.

Kolmogorov, A.N., 1962. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. Journal of Fluid Mechanics 13, 82–85.

Liu, Q.S., 2002. An extension of a fundamental equation of Poincaré and Mandelbrot. Asian Journal of Mathematics 6, 145–68.

Lo, A.W., 1991. Long-term memory in stock market prices. Econometrica 59, 1279–1313.

Mandelbrot, B.B., 1963. The variation of certain speculative prices. Journal of Business (Chicago) 36, 394–419. Reprinted in Cootner (1964), as Chapter E 14 of Mandelbrot (1997), in Telser (2000), and several other collections of papers on finance.

Mandelbrot, B.B., 1965. Une classe de processus stochastiques homothétiques à soi; application à la loi climatologique de H.E. Hurst. Comptes Rendus (Paris) 260, 3274–3277. Translated as Chapter H9 of Mandelbrot (2002).

Mandelbrot, B.B., 1967. The variation of some other speculative prices. Journal of Business (Chicago) 40, 393–413. Reprinted as Chapter E14 of Mandelbrot (1997), pp. 419–443, in Telser (2000), and several other collections of papers on finance.

Mandelbrot, B.B., 1972. Possible refinement of the lognormal hypothesis concerning the distribution of energy dissipation in intermittent turbulence. In: Rosenblatt, M., Van Atta, C. (Eds.), Statistical Models and Turbulence. Springer-Verlag, New York, pp. 333–351. Reprinted in Mandelbrot (1999a) as Chapter N14.

Mandelbrot, B.B., 1974a. Intermittent turbulence in self similar cascades; divergence of high moments and dimension of the carrier. Journal of Fluid Mechanics 62, 331–358. Reprinted in Mandelbrot (1999a) as Chapter N15.

Mandelbrot, B.B., 1974b. Multiplications aléa.atoires itérées et distributions invariantes par moyenne pondérée aléatoire, Comptes Rendus (Paris) A 278, 289–292 and 355–358. Reprinted in Mandelbrot (1999a) as Chapter N16.

Mandelbrot, B.B., 1982. The Fractal Geometry of Nature. Freeman, New York.

Mandelbrot, B.B., 1984. Fractals in physics: squid clusters, diffusions, fractal measures and the unicity of fractal dimension. Journal of Statistical Physics 34, 895–930.

Mandelbrot, B.B., 1989a. Multifractal measures, especially for the geophysicist. Pure and Applied Geophysics 131, 5–42.

Mandelbrot, B.B., 1989b. A class of multinomial multifractal measures with negative (latent) values for the "dimension" $f(\alpha)$. In: Pietronero, L. (Ed.), Fractals' Physical Origin and Properties. Plenum, New York, pp. 3–29.

Mandelbrot, B.B., 1990a. Negative fractal dimensions and multifractals. Physica A 163, 306–315.

Mandelbrot, B.B., 1990b. New "anomalous" multiplicative multifractals: left-sided $f(\alpha)$ and the modeling of DLA. Physica A 168, 95–111.

Mandelbrot, B.B., 1990c. Limit lognormal multifractal measures. In: Gotsman, E.A., Neéman, Y., Voronel, A. (Eds.), Frontiers of Physics: Landau Memorial Conference. Pergamon, New York, pp. 309–340.

Mandelbrot, B.B., 1995. Negative dimensions and Hölders, multifractals and their Hölder spectra, and the role of lateral preasymptotics in science. In: Bonami, A., Peyrière, J. (Eds.), J.P. Kahane's meeting, Paris, 1993. The Journal of Fourier Analysis and Applications, 409–432.

Mandelbrot, B.B., 1997. Fractals and Scaling in Finance: Discontinuity, Concentration, Risk (Selecta Volume E). Springer-Verlag.

Mandelbrot, B.B., 1999a. Multifractals and $1/f$ Noise: Wild Self-Affinity in Physics (Selecta Volume N). Springer-Verlag.

Mandelbrot, B.B., 1999b. A multifractal walk through Wall Street. Scientific American, February issue, 50–53.

Mandelbrot, B.B., 2001a. Scaling in financial prices, I: Tails and dependence. Quantitative Finance 1, 113–124. Reprint: Farmer, D., Geanakoplos, J. (Eds.), Beyond Efficiency and Equilibrium. Oxford University Press, UK, 2002.

Mandelbrot, B.B., 2001b. Scaling in financial prices, II: Multifractals and the star equation. Quantitative Finance 1, 124–130. Reprint: Farmer, D., Geanakoplos, J. (Eds.), Beyond Efficiency and Equilibrium. Oxford University Press, UK, 2002.

Mandelbrot, B.B., 2001c. Scaling in financial prices, III: Cartoon Brownian motions in multifractal time. Quantitative Finance 1, 427–440.

Mandelbrot, B.B., 2001d. Scaling in financial prices, IV: Multifractal concentration. Quantitative Finance 1, 558–559.

Mandelbrot, B.B., 2001e. Stochastic volatility, power-laws and long memory. Quantitative Finance 1, 427–440.

Mandelbrot, B.B., 2002. Gaussian Self-Affinity and Fractals (Selecta Volume H). Springer-Verlag.

Mandelbrot, B.B., 2003, forthcoming.

Mandelbrot, B.B., Calvet, L., Fisher, A., 1997. The multifractal model of asset returns. Large deviations and the distribution of price changes. The multifractality of the Deutschmark/US Dollar exchange rate. Discussion Papers numbers 1164, 1165, and 1166 of the Cowles Foundation for Economics at Yale University, New Haven, CT. Available on the web at the following addresses.
http://papers.ssrn.com/sol3/paper.taf? ABSTRACT ID=78588.
http://papers.ssrn.com/sol3/paper.taf? ABSTRACT ID=78606.
http://papers.ssrn.com/sol3/paper.taf? ABSTRACT ID=78628.

Mandelbrot, B.B., Taylor, H.M., 1967. On the distribution of stock price differences. Operations Research 15, 1057–1062.

Obukhov, A.M., 1962. Some specific features of atmospheric turbulence. Journal of Fluid Mechanics 13, 77–81.

Telser, L. (Ed.), 2000. Classic Futures: Lessons from the Past for the Electronic Age. Risk Books, London.