8-1-2012

How to Define Mean, Variance, etc., for Heavy-Tailed Distributions: A Fractal-Motivated Approach

Vladik Kreinovich
University of Texas at El Paso, vladik@utep.edu

Olga Kosheleva
University of Texas at El Paso, olgak@utep.edu

Follow this and additional works at: http://digitalcommons.utep.edu/cs_techrep

Part of the Computer Sciences Commons, and the Mathematics Commons

Comments:
Technical Report: UTEP-CS-12-32

Recommended Citation
http://digitalcommons.utep.edu/cs_techrep/710

This Article is brought to you for free and open access by the Department of Computer Science at DigitalCommons@UTEP. It has been accepted for inclusion in Departmental Technical Reports (CS) by an authorized administrator of DigitalCommons@UTEP. For more information, please contact lweber@utep.edu.
How to Define Mean, Variance, etc., for Heavy-Tailed Distributions: A Fractal-Motivated Approach

Vladik Kreinovich and Olga Kosheleva
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
vladik@utep.edu, olgak@utep.edu

Abstract

In many practical situations, we encounter heavy-tailed distributions for which the variance — and even sometimes the mean — are infinite. We propose a fractal-based approach that enables us to gauge the mean and variance of such distributions.

1 Formulation of the Problem

There are many practical situations in which the probability distribution is drastically different from normal. In many such situations, the variance is infinite; such distributions are called heavy-tailed. These distributions surfaced in the 1960s, when Benoit Mandelbrot, the author of fractal theory, empirically studied the fluctuations and showed [11] that large-scale fluctuations follow the Pareto power-law distribution, with the probability density function $p(x) = A \cdot x^{-\alpha}$ for $x \geq x_0$, for some constants $\alpha \approx 2.7$ and $x_0$. For this distribution, variance is infinite. The above empirical result, together with similar empirical discovery of heavy-tailed laws in other application areas, has led to the formulation of fractal theory; see, e.g., [12, 13].

Since then, similar heavy-tailed distributions have been empirically found in other financial situations [2, 3, 4, 7, 14, 16, 17, 20, 21, 22, 23], and in many other application areas [1, 8, 12, 15, 19].

For heavy-tailed distributions, variance is infinite, so we cannot use variance to describe the deviation from the “average”. Thus, we need to come up with other characteristics for describing this deviation.

This situation is typical in financial and economic applications, where this deviation is known as volatility. At first, economists followed a natural idea to use standard deviation as a quantitative measure of volatility. However, since the empirical distribution is heavy-tailed, its standard deviation is infinite, so
other characteristics of volatility are needed. Such characteristics are proposed in this paper.

2 Fractal-Motivated Approach: Main Idea

Problem: reminder. In general, the mean value of a random variable with the probability density $\rho(x)$ is defined as an integral

$$E[x] \overset{\text{def}}{=} \int_{-\infty}^{\infty} x \cdot \rho(x) \, dx.$$  

Similarly, the mean value of a function $f(x)$ (e.g., of the function $f(x) = x^2$ corresponding to the second central moment) is defined as an integral

$$E[f(x)] \overset{\text{def}}{=} \int_{-\infty}^{\infty} f(x) \cdot \rho(x) \, dx.$$  

The integral over an infinite range $(-\infty, \infty)$ is, in its turn, defined as a limit of integrals over the increasing finite ranges $[-\Delta, \Delta]$:  

$$\int_{-\infty}^{\infty} f(x) \cdot \rho(x) \, dx \overset{\text{def}}{=} \lim_{\Delta \to \infty} \int_{-\Delta}^{\Delta} f(x) \cdot \rho(x) \, dx.$$  

For usual distributions, e.g., for the normal distribution, this limit is finite, so we get well-defined notions of mean, second moment, etc. However, for heavy-tailed distributions, the corresponding finite-range integrals $\int_{-\Delta}^{\Delta} f(x) \cdot \rho(x) \, dx$ tend to infinity, so in the limit, we get meaningless infinities as the values of mean, second moment, etc.

Fractals reminder. A similar situation was one of the motivations behind the fractals; see, e.g., [12, 13]. Indeed, traditionally in mathematics, to compute the length of a smooth curve segment starting at a point $a$ and ending at a point $b$, we can pick some small number $\varepsilon > 0$, and then:

- we start with one of the curve’s endpoints $x_0 = a$,
- we find the first point $x_1$ on this curve for which the distance $d(x_0, x_1)$ between the point $x_0$ and this point $x_1$ is equal to $\varepsilon$;
- past $x_1$, we find the first point $x_2$ on the curve for which $d(x_1, x_2) = \varepsilon$;
- ...  
- past $x_k$, we find the first point $x_{k+1}$ on the curve for which $d(x_k, x_{k+1}) = \varepsilon$,
- ...  
- we stop when for some $n = n(\varepsilon)$, we reach a point $x_n$ which is $\varepsilon$-close to the other endpoint $b$: $d(x_n, b) \leq \varepsilon$.  

2
We can then define the $\varepsilon$-approximate length $L(\varepsilon)$ as the sum of the lengths of all the corresponding segments:

$$L(\varepsilon) = \sum_{i=0}^{n(\varepsilon)-1} d(x_i, x_{i+1}).$$

Since each of the segments is of length $\varepsilon$, the $\varepsilon$-length can be simply computed as value

$$L(\varepsilon) = n(\varepsilon) \cdot \varepsilon.$$

For a smooth curve, the smaller $\varepsilon$, the better the approximation, so we can define the length $L$ of the curve as the limit $L \overset{\text{def}}{=} \lim_{\varepsilon \to 0} L(\varepsilon)$.

The problem with this definition is that for some curves, when the accuracy $\varepsilon$ tends to 0, instead of a meaningful finite limit $L$, we get a meaningless infinity. This happens, e.g., when we try to measure the shoreline of Britain: the smaller $\varepsilon$, the larger the value. Similarly, we get an infinity when we try to measure the length of a path of a particle following Brownian motion.

In both cases – and in many other practical situations – when $\varepsilon$ tends to 0, the number $n(\varepsilon)$ of the corresponding points grows as $\frac{C}{\varepsilon^{f}}$ for some real numbers $C > 0$ and $f > 0$, and thus, the approximate length $L(\varepsilon) = n(\varepsilon) \cdot \varepsilon$ asymptotically tends to infinity as $\frac{C}{\varepsilon^{f}}$, where $f \overset{\text{def}}{=} f - 1$. In this case, to describe the asymptotic behavior of $L(\varepsilon)$ when $\varepsilon \to 0$, instead of a single limit value $L$, we now need two different values: $C$ and $f$. The traditional case of smooth curves corresponds to $f = 1$ and $C = L$.

The value $f$ is called a dimension of the curve. The reason for this name is as follows: $\frac{C}{\varepsilon^{f}}$ describe the largest number of points on this curve which are at least $\varepsilon$-far away from each other. On a smooth 1-D curve, this number is reached when we place points on a grid with distance $\varepsilon$ between the two neighboring points, so we get $\approx\frac{L}{\varepsilon}$ points. On a smooth 2-D surface, we have $\approx\frac{A}{\varepsilon^{2}}$ points on the corresponding grid, where $A$ is the area. In a 3-D area of volume $V$, we need $\approx\frac{V}{\varepsilon^{3}}$ grid points, etc. In all these examples, the corresponding number of points is asymptotically equal to $\approx\frac{C}{\varepsilon^{D}}$, where $D$ is the dimension of the corresponding set:

- $D = 1$ for a 1-D curve;
- $D = 2$ for a 2-D surface;
- $D = 3$ for a 3-D area, etc.

Thus, for non-smooth areas, it is natural to define dimension as the real number $f$ for which the corresponding number of points is asymptotically equal to $\frac{C}{\varepsilon^{f}}$.

In most real-life examples, in particular, for the shoreline of Britain and for the
trajectory of a Brownian particle, thus defined dimension is not an integer, it is equal to 1.5 or to some other fraction. Because of this “fractal” dimension value, the corresponding sets are called fractals.

The parameter $C$ (“measure”) serves as an analog of length.

Let us apply this idea to heavy-tailed distributions. The main idea of a fractal approach is that instead of simply stating that the length approximation $L(\varepsilon)$ tend to infinity, we analyze how exactly it tends to infinity – i.e., what is the asymptotic behavior of $L(\varepsilon)$ when $\varepsilon \to 0$ – and use the parameters of this asymptotic behavior as a natural generalization of the usual length. To apply this approach to our problem, it is therefore necessary not just to state that the limit of the corresponding integrals over a finite range tend to infinity, it is also necessary to analyze the asymptotic behavior of these integrals.

Asymptotic behavior of expressions corresponding to mean and variance: case of heavy-tailed distribution. For many practical heavy-tailed distributions, for large $x$, the probability density $\rho(x)$ asymptotically follows a power law (= Pareto distribution) $\rho(x) \approx \rho_0 \cdot x^{-\alpha}$. For such distributions, when $\Delta \to \infty$, we get

$$\int_{-\Delta}^{\Delta} x \cdot \rho(x) \, dx \approx \int_{-\Delta}^{\Delta} x \cdot \rho_0 \cdot x^{-\alpha} \, dx = \int_{-\Delta}^{\Delta} \rho_0 \cdot x^{1-\alpha} \, dx \approx \frac{\rho_0}{2-\alpha} \cdot \Delta^{2-\alpha}.$$  

Similarly, the expression corresponding to the second moment has the form

$$\int_{-\Delta}^{\Delta} x^2 \cdot \rho(x) \, dx \approx \int_{-\Delta}^{\Delta} x^2 \cdot \rho_0 \cdot x^{-\alpha} \, dx = \int_{-\Delta}^{\Delta} \rho_0 \cdot x^{2-\alpha} \, dx \approx \frac{\rho_0}{3-\alpha} \cdot \Delta^{3-\alpha},$$

and the expression corresponding to the $k$-th order moment, $k = 3, 4, \ldots$, has the form

$$\int_{-\Delta}^{\Delta} x^k \cdot \rho(x) \, dx \approx \int_{-\Delta}^{\Delta} x^k \cdot \rho_0 \cdot x^{-\alpha} \, dx = \int_{-\Delta}^{\Delta} \rho_0 \cdot x^{k-\alpha} \, dx \approx \frac{\rho_0}{k+1-\alpha} \cdot \Delta^{k+1-\alpha}.$$  

In all these cases, asymptotically, we have

$$E_{\Delta}[f(x)] \overset{\text{def}}{=} \int_{-\Delta}^{\Delta} f(x) \cdot \rho(x) \, dx \sim C \cdot \Delta^a$$

for some real numbers $C$ and $a \geq 0$. Thus, it is natural to arrive at the following definition.

**Definition 1.** We say that two functions $F(\Delta)$ and $F'(\Delta)$ are asymptotically equivalent (and denote it by $F(\Delta) \sim F'(\Delta)$) if the ratio $\frac{F(\Delta)}{F'(\Delta)}$ tends to 1 when $\Delta \to \infty$. 

4
Definition 2. Let $f(x)$ be a function, and let $\rho(x)$ be a probability density function. If asymptotically, when $\Delta \to \infty$, we have
\[ E_\Delta[f(x)] = \int_{-\Delta}^{\Delta} f(x) \cdot \rho(x) \, dx \sim C \cdot \Delta^a \]
for some real numbers $C$ and $a \geq 0$, then the pair $(C, a)$ is called the fractal-motivated mean of $f(x)$ under the distribution $\rho(x)$. This pair is denoted by $E_\rho[f(x)]$.

Comment. When $a = 0$, we get the usual definition of the expected value ($=$ mean).

Let us now describe how to deal with these fractal-motivated characteristics.

3 How to Compare Different Values of Fractal-Motivated Mean

In many applications, it is important to be able compare the means. For example, a standard way to describe preferences of a decision maker is to use the notion of utility $u$; see, e.g., [5, 6, 9, 10, 18]. According to decision theory, a user prefers an alternative for which the expected value $E[u] = \int \rho(x) \cdot u(x) \, dx$ of the utility is the largest possible.

In the traditional statistical approach, when the mean (expected value) of a variable is a finite number, we can simply compare these numbers. In the fractal-motivated approach, we can compare them asymptotically:

Definition 3. We say that the pair $(C, a)$ is smaller than the pair $(C', a')$ (and denote it by $(C, a) < (C', a')$) if for every two functions $F(\Delta)$ and $F'(\Delta)$,
- if $F(\Delta) \sim C \cdot \Delta^a$ and $F'(\Delta) \sim C' \cdot \Delta^{a'}$,
- then $F(\Delta) < F'(\Delta)$ for all sufficiently large $\Delta$, i.e., there exists a $\Delta_0$ such that for all $\Delta \geq \Delta_0$, we have
  \[ F(\Delta) < F'(\Delta). \]

One can easily show that this seemingly abstract definition can be easily checked by comparing pairs $(C, a)$ and $(C', a')$:

Proposition 1. $(C, a) < (C', a')$ if and only if:
- either $a < a'$
- or $(a = a'$ and $C < C')$.
In other words, this comparison is simply a lexicographic comparison of the pairs \((C, a)\) and \((C', a')\). First, we compare the second components \(a\) and \(a'\):

- if \(a < a'\), we conclude that \((C, a) < (C', a')\);
- if \(a' < a\), we conclude that \((C', a') < (C, a)\);
- if \(a = a'\), then we have to compare \(C\) and \(C'\).

After the second comparison:

- if \(C < C'\), we conclude that \((C, a) < (C', a')\);
- if \(C' < C\), we conclude that \((C', a') < (C, a)\);
- if \(a = a'\) and \(C = C'\), then the means \((C, a)\) and \((C', a')\) simply coincide, so we cannot make a meaningful comparison.

4 How to Compute the Mean of the Linear Combination

How the mean of linear combination is computed in traditional statistics: reminder. In the traditional statistics, if we have a function \(f(x)\) with a known mean \(E[f(x)]\), then for each constant \(c \neq 0\), the can compute the mean of the function \(c \cdot f(x)\) as \(E[c \cdot f(x)] = c \cdot E[f(x)]\).

Similarly, if have two functions \(f_1(x)\) and \(f_2(x)\) with known means \(E[f_1(x)]\) and \(E[f_2(x)]\), then we can compute the mean of the sum \(f(x) = f_1(x) + f_2(x)\) as the sum of the means:

\[
E[f_1(x) + f_2(x)] = E[f_1(x)] + E[f_2(x)].
\]

As a result, if we have several functions \(f_1(x), \ldots, f_m(x)\) with known means \(E[f_1(x)], \ldots, E[f_m(x)]\), then the mean of their linear combination

\[
f(x) = c_1 \cdot f_1(x) + \ldots + c_m \cdot f_m(x)
\]

can be computed as a linear combination of the means

\[
E[f(x)] = c_1 \cdot E[f_1(x)] + \ldots + c_m \cdot E[f_1(x)].
\]

Fractal-motivated case: multiplication by a constant.

**Definition 4.** Let \((C, a)\) be a pair and let \(c \neq 0\) be a real number. We say that \((C', a')\) is the result of multiplying \((C, a)\) by \(c\) (and denote it by \(c \cdot (C, a)\)) if for every function \(F(\Delta)\):

- if \(F(\Delta) \sim C \cdot \Delta^a\),
- then \(c \cdot F(\Delta) \sim C' \cdot \Delta^{a'}\).
One can easily check that this operation reduces to multiplying $c$ and $C$:

**Proposition 2.** For every $C$, $a$, and $c$, we have $c \cdot (C, a) = (c \cdot C, a)$.

**Corollary 1.** For every function $f(x)$ for which the fractal-motivated mean $E_{fa}[f(x)]$ is defined, and for every constant $c \neq 0$, the fractal-motivated mean of $c \cdot f(x)$ is also defined and is equal to the result of multiplying $c$ by the corresponding mean:

$$E_{fa}[c \cdot f(x)] = c \cdot E_{fa}[f(x)].$$

**Fractal-motivated case: sum.**

**Definition 5.** Let $(C, a)$ and $(C', a')$ be pairs. We say that a pair $(C'', a'')$ is the sum of the pairs $(C, a)$ and $(C', a')$ (and denote it by $(C'', a'') = (C, a) + (C', a')$) if for every two functions $F(\Delta)$ and $F'(\Delta)$:

- if $F(\Delta) \sim C \cdot \Delta^a$ and $F'(\Delta) \sim C' \cdot \Delta^{a'}$,
- then for $F''(\Delta) \overset{\text{def}}{=} F(\Delta) + F'(\Delta)$, we have

$$F''(\Delta) = F(\Delta) + F'(\Delta) \sim C'' \cdot \Delta^{a''}.$$  

One can easily check how we can perform this operation:

**Proposition 3.** The operation $(C, a) + (C', a')$ if defined for all pairs except for the case when $a = a'$ and $C + C' = 0$, and has the following form:

- if $a < a'$, then $(C, a) + (C', a') = (C', a')$;
- if $a' < a$, then $(C, a) + (C', a') = (C, a)$;
- if $a = a'$ and $C + C' \neq 0$, then $(C, a) + (C', a') = (C + C', a)$.

**Corollary 2.** For every two functions $f(x)$ and $f'(x)$ for which the fractal-motivated means $E_{fa}[f(x)]$ and $E_{fa}[f'(x)]$ are defined and for which the sum $E_{fa}[f(x)] + E_{fa}[f'(x)]$ is defined, the fractal-motivated mean of the sum $f(x) + f'(x)$ is also defined and is equal to the sum of the corresponding means:

$$E_{fa}[f(x) + f'(x)] = E_{fa}[f(x)] + E_{fa}[f'(x)].$$
Fractal-motivated case: general linear combination.

**Definition 6.** Let \((C_1, a_1), \ldots, (C_m, a_m)\) be pairs, and let \(c_1, \ldots, c_m\) be real numbers. We say that a pair \((C, a)\) is the linear combinations of the pairs \((C_i, a_i)\) with coefficients \(c_i\) (and denote it by \((C, a) = \sum_{i=1}^{m} c_i \cdot (C_i, a_i)\)) if for every \(m\) functions \(F_1(\Delta), \ldots, F_m(\Delta)\):

- if \(F_i(\Delta) \sim C_i \cdot \Delta^{a_i}\) for all \(i\),

- then for \(F(\Delta) \triangleq \sum_{i=1}^{m} c_i \cdot F_i(\Delta)\), we have

\[
F(\Delta) = \sum_{i=1}^{m} c_i \cdot F_i(\Delta) \sim C \cdot \Delta^{a}.
\]

One can easily check how we can perform this operation:

**Proposition 4.** The operation \(\sum_{i=1}^{m} c_i \cdot (C_i, a_i)\) if defined for all the case except for the case when \(\sum\{C_i : a_i = \max(a_1, \ldots, a_m)\} = 0\), and has the following form:

\[
\sum_{i=1}^{m} c_i \cdot (C_i, a_i) = (C, a),
\]

where \(a \triangleq \max(a_1, \ldots, a_m)\) and \(C = \sum\{C_i : a_i = a\}\).

**Corollary 3.** For every \(m\) functions \(f_1(x), \ldots, f_m(x)\) for which the fractal-motivated means \(E_{fr}[f_i(x)]\) are defined and for every numbers \(c_1, \ldots, c_m\) for which the linear combination \(\sum_{i=1}^{m} c_i \cdot E_{fr}[f_i(x)]\) is defined, the fractal-motivated mean of the linear combination \(\sum_{i=1}^{m} c_i \cdot f_i(x)\) is also defined and is equal to the linear combination of the corresponding means:

\[
E_{fr}\left[\sum_{i=1}^{m} c_i \cdot f_i(x)\right] = \sum_{i=1}^{m} c_i \cdot E_{fr}[f_i(x)].
\]

5 How to Generalize this Approach to Multi-D Case

**Definition.** What if we have a random vector \(x = (x_1, \ldots, x_n)\) and we want to describe the mean and the variance of a function \(f(x) = f(x_1, \ldots, x_n)\)? In
In this case, we can similarly limit all the variables \( x_i \) to values for which \( |x_i| \leq \Delta \) for all \( i \), and then consider the asymptotics of such means

\[
\int_{-\Delta}^{\Delta} \cdots \int_{-\Delta}^{\Delta} f(x_1, \ldots, x_n) \cdot \rho(x_1, \ldots, x_n) \, dx_1 \cdots dx_n \sim C \cdot \Delta^a.
\]

**Case of independent random variables: mean of the product.** It is known that in the traditional statistics, if we have two independent random variables \( x_1 \) and \( x_2 \), then for every two functions \( f_1(x_1) \) and \( f_2(x_2) \), the mean of the product \( f(x_1, x_2) \equiv f_1(x_1) \cdot f_2(x_2) \) is equal to the product of the corresponding means:

\[
E[f_1(x_1) \cdot f_2(x_2)] = E[f_1(x_1)] \cdot E[f_2(x_2)].
\]

**Definition 7.** Let \( (C, a) \) and \( (C', a') \) be pairs. We say that a pair \( (C'', a'') \) is the product of the pairs \( (C, a) \) and \( (C', a') \) (and denote it by \( (C'', a'') = (C, a) \cdot (C', a') \)) if for every two functions \( F(\Delta) \) and \( F'(\Delta) \):

- if \( F(\Delta) \sim C \cdot \Delta^a \) and \( F'(\Delta) \sim C' \cdot \Delta^{a'} \),
- then for \( F''(\Delta) \equiv F(\Delta) \cdot F'(\Delta) \), we have

\[
F''(\Delta) = F(\Delta) \cdot F'(\Delta) \sim C'' \cdot \Delta^{a''}.
\]

One can easily check how we can perform this operation:

**Proposition 5.** For all pairs \( (C, a) \) and \( (C', a') \), we have

\[
(C, a) \cdot (C', a') = (C \cdot C', a + a').
\]

**Corollary 4.** If variables \( x_1 \) and \( x_2 \) are independent, then for every two functions \( f_1(x_1) \) and \( f_2(x_2) \) for which the fractal-motivated means \( E_{\mathbb{R}}[f_i(x_i)] \) are defined, the fractal-motivated mean of the product \( f_1(x_1) \cdot f_2(x_2) \) is also defined and is equal to the product of the corresponding means:

\[
E_{\mathbb{R}}[f_1(x_1) \cdot f_2(x_2)] = E_{\mathbb{R}}[f_1(x_1)] \cdot E_{\mathbb{R}}[f_2(x_2)].
\]

These results can be easily generalized to an arbitrary number of independent variables:
Definition 8. Let \((C_1, a_1), \ldots, (C_m, a_m)\) be pairs. We say that a pair \((C, a)\)

is the product of the pairs \((C_1, a_1), \ldots, (C_m, a_m)\) (and denote it by \((C, a) = (C_1, a_1) \cdot \ldots \cdot (C_m, a_m)\)) if for every \(m\) functions \(F_1(\Delta), \ldots, F_m(\Delta)\):

- if \(F_i(\Delta) \sim C_i \cdot \Delta^{a_i}\) for all \(i\),

- then for \(F(\Delta) \overset{\text{def}}{=} F_1(\Delta) \cdot \ldots \cdot F_m(\Delta)\), we have

\[ F(\Delta) = F_1(\Delta) \cdot \ldots \cdot F_m(\Delta) \sim C \cdot \Delta^{a} \]

One can easily check how we can perform this operation:

Proposition 6. For all pairs \((C_1, a_1), \ldots, (C_m, a_m)\), we have

\[ (C_1, a_1) \cdot \ldots \cdot (C_m, a_m) = (C_1 \cdot \ldots \cdot C_m, a_1 + \ldots + a_m) \]

Corollary 5. If the variables \(x_1, \ldots, x_m\) are independent, then for every \(m\) functions \(f_1(x_1), \ldots, f_m(x_m)\) for which the fractal-motivated means \(E_{fr}[f_i(x_i)]\) are defined, the fractal-motivated mean of the product \(f_1(x_1) \cdot \ldots \cdot f_m(x_m)\) is also defined and is equal to the product of the corresponding means:

\[ E_{fr}[f_1(x_1) \cdot \ldots \cdot f_m(x_m)] = E_{fr}[f_1(x_1)] \cdot \ldots \cdot E_{fr}[f_m(x_m)] \]

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, by Grant 1 T36 GM078000-01 from the National Institutes of Health, and by a grant on F-transforms from the Office of Naval Research.

References


