On Distribution of Absolute Values

Martin Hlusek[1]

September 2011

Abstract

This note shows two important properties of absolute values of symmetrically distributed random variables. First, the characteristic function of absolute value |X| adds imaginary part which is equal to the Hilbert transform of the characteristic function of the original random variable X. Second, a sum of absolute values on any symmetric underlying distribution will maintain the property that its imaginary part will be equal to the Hilbert transform of its real part. It is shown how the first moment of E|X| can be used for estimation of dispersion of the underlying distribution and how the characteristic function of E|X| transforms to characteristic function of a sum $\sum|X_i|$. To demonstrate the results, the note shows a trivial example of a symmetric discrete value distribution and normal distribution as a special case of Levy-stable distributions. The main theorems are known facts used in signal processing. The contribution of this note is to demonstrate their applications in statistics.

The Main Theorem

This paragraph introduces the main theorem showing the equivalence between characteristic function of a symmetrically distributed random variable and characteristic function of its absolute value. It also demonstrates how absolute value moments are related to the characteristic function of the underlying random variable. The paragraph concludes with two examples, absolute value of a discrete symmetrically distributed random variable and a symmetrically distributed normal variable.

Definition 1: Let f(x) be a real valued function, we denote Hf its Hilbert transform

$$Hf = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{f(x)}{t-x} dx \tag{1}$$

where P denotes principal value in the Cauchy sense[2]. The main properties of the Hilbert transform are summarized in the appendix, which we will refer to in the subsequent text.

Definition 2: Let us define the Fourier transform of f(x) as

$$Ff \equiv \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x) e^{-i\omega x} dx \tag{2}$$

[1] hlusekm@gmail.com for comments

[2] For the purpose of this note, Cauchy's principal value can be defined as

$$P \int_{-\infty}^{+\infty} \frac{f(x)}{t-x} dx = \lim_{\varepsilon \to 0+} \left[\int_{-\infty}^{t-\varepsilon} \frac{f(x)}{t-x} dx + \int_{t+\varepsilon}^{+\infty} \frac{f(x)}{t-x} dx \right]$$

Given the definition, we have the relationship between probability distribution function (p.d.f.) f(x) and characteristic function[3] (c.f.) $\phi(t)$, $f=F\phi$.

Theorem 1: Let a random variable X has a symmetric (even) p.d.f. f(x). Then |X| has c.f. equal to $\phi(t)+iH\,\phi(t)$ if and only if $\phi(t)$ is c.f. of x.

Proof: Both implications will be proved separately.

(i) Let us assume that X has symmetric p.d.f. f(x) and c.f. $\phi(t)$, so $f=F\phi$. This implies that |X| has p.d.f. f(x)+sgn(x)f(x), so c.f. of |X| u(t)+iv(t) must satisfy,

$$F(u+iv)=f+\mathrm{sgn}(x)f$$

Taking F^{-1} of both sides and using (A3), we obtain equality

$$u+iv=\phi+iH\phi$$

Because X has symmetric p.d.f., ϕ must be real-valued and we have $u=\phi$ and $v=H\phi$.

(ii) Let us now assume that |X| has c.f. equal to $\phi+iH\phi$. Since it is a c.f., we can define p.d.f. $f_+(x)=F\,\phi+iFH(\phi)=F\,\phi+\mathrm{sgn}(x)F\,\phi$ and even function $f(x)=(f_+(x)+f_+(-x))/2$ for all real x. This implies

$$f(x)=\frac{1}{2}(F\phi+F\phi+F\phi-F\phi)=F\phi$$

which completes the proof.

The relationship between f+sgn(x)f and $F\phi+iHF\phi$ is trivial and well known. It is used in signal processing. A signal of the form $F\phi+iHF\phi$ is called an analytic signal which is known to be formed by (only) positive frequencies.

By definition, a symmetrically distributed X will have even p.d.f. f. It is easy to see that $\phi=F^{-1}f$ inherits this property (just apply transformation x→-x in (2)), so let's assume that $\phi(t)$ is even. Further, ϕ will be real valued, $|\phi|\le 1$, $\phi(0)=1$, and, assuming that EX exists, also $\phi'(0)=0$. It turns out that if EX exists, $E|X|\ge 0$ exists as well and can be calculated directly from ϕ.

The existence is clear from the fact that the distribution function of |X| is, up to a constant, the same as for X on the positive semi-axis. Therefore, if EX exists, E|X| must exist as well (the tail of the distribution vanishes sufficiently fast). Now, knowing that E|X| exists, it must be equal to

$$E|X|=\frac{\partial}{\partial t}H\phi(0)=\frac{\partial}{\partial t}\frac{1}{\pi}P\int_{-\infty}^{+\infty}\frac{\phi(x)}{t-x}dx=-\frac{1}{\pi}P\int_{-\infty}^{+\infty}\frac{\phi(x)}{x^2}dx \tag{3}$$

We can go on, assume existence of EX^2 and find out that

$$EX^2=E|X|^2=\frac{\partial^2}{\partial t^2}\phi(0)\ \text{ or }\ \frac{\partial^2}{\partial t^2}H\phi(0)=0 \tag{4}$$

[3] For a definition of characteristic functions and their properties, see any basic statistical book, for example [1].

Or, in general, that the even moments are identical to those of X and that the odd moments 2k+1 (if they exist) are equal to

$$E|X|^{2k+1} = \frac{(-1)^{k+1}(2k+1)!}{\pi} P\int_{-\infty}^{+\infty} \frac{\phi(x)}{x^{2k}}dx \tag{5}$$

Notes:

- We expect that E|X| will carry some information about how 'wide' the underlying distribution of X is without explicit request of the existence of the second moment. Intuitively, we expect that the more weight is put 'away from zero' the higher E|X| will be.
- For ϕ even, we have

$$P\int_{-\infty}^{+\infty} \frac{\phi(x)}{x^{2k}}dx = \frac{2}{\pi} P\int_{0}^{+\infty} \frac{\phi(x)}{x^{2k}}dx \tag{6}$$

- It turns out that

$$-P\int_{-\infty}^{+\infty} \frac{\phi(x)}{x^{2}}dx = P\int_{-\infty}^{+\infty} \frac{1-\phi(x)}{x^{2}}dx \tag{7}$$

  because $P\int 1/x^2 dx = 0$. This follows from integration of $1/z^2$ over upper semi-circle in the complex plain and from the fact that $1/z^2$ has a simple pole of second order at zero, where Res $1/z^2 = 0$. Also, because of (7) and the fact that $\phi(0)=1$ and $\phi'(0)=0$, the integrand of (3) can be continuously extended to zero.

- If $1-\phi$ has a holomorphic extension in upper convex plain $\Psi(z)$, then, thanks to the fact that $\phi(0)=1$ and $\phi'(0)=0$, and $|\phi|\leq 1$, $\Phi(z)$ has a simple pole of second order in $z=0$ and

$$\int_{-\infty}^{+\infty} \frac{1-\phi(x)}{x^{2}}dx = -i\pi\Psi'(0) \tag{8}$$

Example 1: Let us demonstrate the derived results by a trivial example of a symmetrically distributed discrete random variable X. Let us write the probability distribution function $f(x)=K\delta(g(x))$, for an even function g having 2n roots at $\pm x_1,\ldots, \pm x_n$ $x_k>0$. K is a normalization constant such that $K\int\delta(g(x))dx=1$.

Characteristic function ϕ of X is equal to[4]

$$\phi(t) = K\int_{-\infty}^{+\infty} e^{ixt}\delta(g(x))dx = 2K\sum_{k=1}^{n}\frac{e^{ix_k t}+e^{-ix_k t}}{2|g'(x_k)|} = 2K\sum_{k=1}^{n}\frac{\cos(x_k t)}{|g'(x_k)|} \tag{9}$$

and its Hilbert transform

$$H\phi(t) = 2K\sum_{k=1}^{n}\frac{\sin(x_k t)}{|g'(x_k)|} \tag{10}$$

[4] We use identity $\delta(g(x)) = \sum_{g(x_k)=0}\frac{\delta(x-x_k)}{|g'(x_k)|}$

We can derive directly the expected result that

$$E|X| = \frac{\partial}{\partial t} H\phi(0) = 2K\sum_{k=1}^{n} \frac{x_k}{|g'(x_k)|} \tag{11}$$

i.e. that the expected value E|X| is the weighted average of the positive roots of g(x). Or, we can use (3) and (7) and write

$$E|X| = 2K\sum_{k=1}^{n} \frac{1}{|g'(x_k)|} \frac{1}{\pi} P\int_{-\infty}^{+\infty} \frac{1-\cos(x_k t)}{x^2} dx$$

Since 1-cos(xt) = Re(1-e^{ixz}), using (8) we obtain

$$\frac{1}{\pi}\int_{-\infty}^{+\infty} \frac{1-\cos(x_k t)}{x^2} dx = -i(ix_k)$$

and, consequently, the same result as (11).

Example 2: Let us apply the results to a less trivial case of symmetric Levy-stable distributions whose c.f. $\phi(t) = \exp(-c|t|^{\alpha})$, where $c>0$ is a scale constant measuring how 'wide' the bell curve of p.d.f. is and α is a constant related to its tail behavior. For $\alpha = 1$ and $\alpha = 2$, the c.f. describes well-known Cauchy and Gaussian distributions ($c = \sigma^2/2$). While Cauchy distribution does not have the first (and the second) moment, Gaussian distribution has all moments. It can be shown that for $1<\alpha<2$, Levy-stable distributions have the first moment but do not have the second - a property often attributed to so called leptokurtic property.

Theoretically, we should try to derive $H\exp(-c|t|^{\alpha})$ and calculate its first derivative at zero (which exists for $1<\alpha$). Practically, $H\exp(-c|t|^{\alpha})$ is not an analytical function (except for special cases described below), so applying formula (3) is an easier way to see that

$$E|X| = -\frac{2}{\pi} P\int_{0}^{+\infty} \frac{e^{-cx^{\alpha}}}{x^2} dx = -\frac{2c^{1/\alpha}}{\pi\alpha}\Gamma(-\frac{1}{\alpha}) = \frac{2c^{1/\alpha}}{\pi}\Gamma(1-\frac{1}{\alpha}) \tag{12}$$

which makes sense for all $1<\alpha$ and, as expected, approaches infinity for $\alpha \to 1+$. Formula (12) is known (see [2], for example), but direct calculation of E|X| is more cumbersome. Note, that for all $1<\alpha$, the first moment of E|X| carries information about the dispersion scale c. For the Gaussian case, the formula reduces to

$$E|X| = \frac{2c^{1/2}}{\pi}\sqrt{\pi} = \sigma\sqrt{\frac{2}{\pi}} \tag{13}$$

so the first moment, up to a constant, is the dispersion of the underlying distribution. Similarly, the third moment can be found as

$$E|X|^3 = \sigma^3\sqrt{\frac{8}{\pi}} \tag{14}$$

Let us come back to the problem of deriving $H\exp(-c|t|^{\alpha})$ to demonstrate how the Hilbert transform of a relatively simple function leads to surprisingly complicated results.

From (A8), we have for even ϕ

$$H(\phi) = tH(t^{-1}\phi) \tag{15}$$

and thus using (A2) and (A4)

$$\frac{\partial}{\partial t} H\phi = H(t^{-1}\phi) - tH(t^{-2}\phi) + tH(t^{-1}\frac{\partial}{\partial t}\phi) \tag{16}$$

Since ϕ is even, $\phi(x)/x$ is odd, so if we can apply (A8) again to $H(t^{-1}(\phi/t))$ to obtain identity

$$tH(t^{-2}\phi) = H(t^{-1}\phi) + \frac{2}{\pi} P\int_0^{+\infty} \frac{\phi(x)}{x^2} dx$$

and substituting into (16) we derive a first order differential equation for $H\phi$

$$\frac{\partial}{\partial t} H\phi = tH(t^{-1}\frac{\partial}{\partial t}\phi) - \frac{2}{\pi} P\int_0^{+\infty} \frac{\phi(x)}{x^2} dx \tag{17}$$

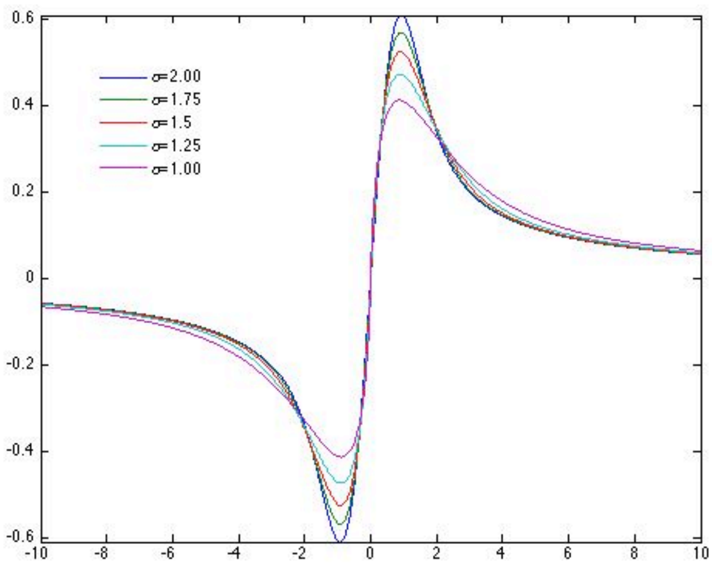
Note that the last term in (17) is $\partial/\partial\, tH\phi(0)$ (if it exists).

For special choice of $\phi = \exp(-ct^2)$, $t^{-1}\phi' = -2c\phi$ and (17) becomes

$$\frac{\partial}{c^{1/2}\partial t} H\phi + 2c^{1/2} tH(\phi) = \frac{2}{\sqrt{\pi}} \tag{18}$$

Figure 1: $H\exp(-|t|^\alpha)$ for different parameters α

Dawson integral for $\alpha = 2$



Knowing that $H\phi(0)=0$ (imaginary part of c.f.), equation (18) is the differential equation for Dawson integral[5] multiplied by $2/\sqrt{\pi}$ and argument $c^{1/2}t$, so the solution is $H\phi(t)= 2/\sqrt{\pi}$ DawsonF($c^{1/2}t$). Figure 1 shows $H\exp(-|t|^{\alpha})$ for different values of α (and c=1).

Sum of Absolute Values

For practical purposes, we need to know how the distribution changes when we add absolute values of independent random variables, so we can form statistics suitable for estimation of the dispersion parameter c. It turns out that no matter what the underlying distributions are, the c.f. of the sum of absolute values will maintain the property that its imaginary part is Hilbert transform of its real part. Intuitively, it should not be surprising, as $|x_1|+|x_2|$ can be viewed as a single random variable $|y|$ and Theorem 1 states that its characteristic function must have the form $\phi+H\phi$. However, it is not entirely clear how ϕ is related to the c.f. of x_1 and x_2. The paragraph concludes with two examples.

Theorem 2: Let X_1 and X_2 be symmetrically distributed independent random variables whose c.f. are ϕ_1 and ϕ_2, then $|X_1| + |X_2|$ has characteristic function

$$c.f.(|X_1|+|X_2|)=\phi_1\phi_2-H\phi_1 H\phi_2+i(\phi_1 H\phi_2+\phi_2 H\phi_1) \qquad (19)$$

where $H(\phi_1\phi_2-H\phi_1 H\phi_2)= \phi_1 H\phi_2+ \phi_2 H\phi_1$.

Proof: First, let us show that independence of X_1 and X_2 implies independence of $|X_1|$ and $|X_2|$. Probability of one conditional on another for any non-negative x_1 and x_2 is

$$P(|X_1|=x_1/|X_2|=x_2)=P(|X_1|=x_1/X_2=-x_2)P(X_2=-x_2)+P(|X_1|=x_1/X_2=x_2)P(X_2=x_2)$$

and from independence of X_1 and X_2 we have

$$P(|X_1|=x_1/X_2=-x_2)=\frac{P(X_1=x_1)P(X_2=-x_2)+P(X_1=-x_1)P(X_2=-x_2)}{P(X_2=-x_2)}=P(X_1=x_1)+P(X_1=-x_1)$$

and therefore

$$P(|X_1|=x_1/|X_2|=x_2)=P(|X_1|=x_1)(P(X_2=-x_2)+P(X_2=x_2))=P(|X_1|=x_1)P(|X_2|=x_2)$$

Given the independence of $|X_1|$ and $|X_2|$, we have immediately that the c.f. of $|X_1| + |X_2|$ has the form of (19). What remains to show is the imaginary part is the Hilbert transform of the real part. This is a well-known fact from signal processing, namely, that a multiple of two analytic signals remains an analytic signal.

We can restrict our proof to ϕ_1 and ϕ_2 even functions and demonstrate that the multiple $(\phi_1+iH\phi_1)(\phi_2+iH\phi_2)$ can be written using inverse Fourier transform as $F^{-1}h+iHF^{-1}h$ for a symmetric p.d.f. h. We will construct function h.

---

5 Dawson integral (or function) has an integral representation

$$DawsonF(t)=e^{-t^2}\int_0^t e^{x^2}\,dx$$

For properties, see http://mathworld.wolfram.com/DawsonsIntegral.html, for example.

Let us assume that X_1 and X_2 have (even) p.d.f. f_1 and f_2. Then the multiple of c.f. can be written as

$$(\phi_1 + iH\phi_1)(\phi_2 + iH\phi_2) = F^{-1}(f_1 + \mathrm{sgn}(x) f_1) * (f_2 + \mathrm{sgn}(x) f_2) \tag{20}$$

where * denotes convolution. Evaluating each convolution and using the fact that f_1 and f_2 are even, we obtain

$$f_1 * f_2 + \mathrm{sgn}(x) f_1 * \mathrm{sgn}(x) f_2 = 2\int_0^{|x|} f_1(|x| - u) f_2(u)du \tag{21}$$

and

$$f_1 * \mathrm{sgn}(x) f_2 + \mathrm{sgn}(x) f_1 * f_2 = 2\,\mathrm{sgn}(x)\int_0^{|x|} f_1(|x| - u) f_2(u)du \tag{22}$$

so the candidate for distribution h(x) is

$$h(x) = 2\int_0^{|x|} f_1(|x| - u) f_2(u)du \tag{23}$$

It is easy to see that h is a proper p.d.f., i.e. non-negative, normalized to 1 and symmetric in x. Consequently, multiple of c.f. (20) can be written as

$$(\phi_1 + iH\phi_1)(\phi_2 + iH\phi_2) = F^{-1}(h + \mathrm{sgn}(x) h) = F^{-1}h + iHF^{-1}h \tag{24}$$

which completes the proof.

Notes:

- By induction, the theorem can be extended to any finite number of $|X_i|$, i = 1,...,N. In general, the real part of (19) is across even functions (even number of Hilbert transforms applied) and the imaginary part across odd functions (odd number of Hilbert transforms applied).
- When X_i are identically distributed with c.f. ϕ, (19) becomes

$$(\phi + iH\phi)^N = \sum_{k-even=0}^{N} \binom{N}{k} (-1)^{k/2} (H\phi)^k \phi^{N-k} + i \sum_{k-odd=0}^{N} \binom{N}{k} (-1)^{(k-1)/2} (H\phi)^k \phi^{N-k} \tag{25}$$

  N stands for 'degrees of freedom', similarly as for χ^2, for example. Note that (25) does not say that the c.f. for sum $|X_1| + \cdots + |X_N|$ will be the same or similar to that of individual $|X_i|$. It just shows the transformation of c.f. and guarantees that the result will have imaginary part equal to the Hilbert transform of the real part.
- Note that (25) includes ϕ^N for k = 0, which the c.f. for $X_1 + \cdots + X_N$.

Example 3: Let us take ϕ and $H\phi$ from Example 1 and apply transformation (19). The real part of c.f. of $|X_1| + |X_2|$ will be

$$\phi^2 - (H\phi)^2 = 4K^2 \sum_{k,l=1}^{n} \frac{\cos(x_k t)\cos(x_l t) - \sin(x_k t)\sin(x_l t)}{|g'(x_k)||g'(x_l)|} = 4K^2 \sum_{k,l=1}^{n} \frac{\cos((x_k + x_l)t)}{|g'(x_k)||g'(x_l)|} \tag{26}$$

and imaginary part

$$2\phi H\phi = 4K^2 \sum_{k,l=1}^{n} \frac{\sin((x_k + x_l)t)}{|g'(x_k)||g'(x_l)|} \tag{27}$$

and thus $E(|X_1|+|X_2|)$ is the normalized sum over all combinations of two positive roots of g.

Example 4: Let us conclude the note with a practical example of estimating standard deviation from absolute values of independent normally distributed random variables $X_1,\cdots,X_N \sim N(0,\sigma^2)$.

Not surprisingly, the average of absolute values will converge to standard deviation multiplied by $\sqrt{2/\pi}$ (as we have seen in Example 2).

$$E\frac{1}{N}\sum_{k=1}^{N}|X_k| = \sqrt{\frac{2}{\pi}}\sigma \tag{28}$$

However, what is interesting is the variance of this statistic. From independence of $|X_i|$ we have

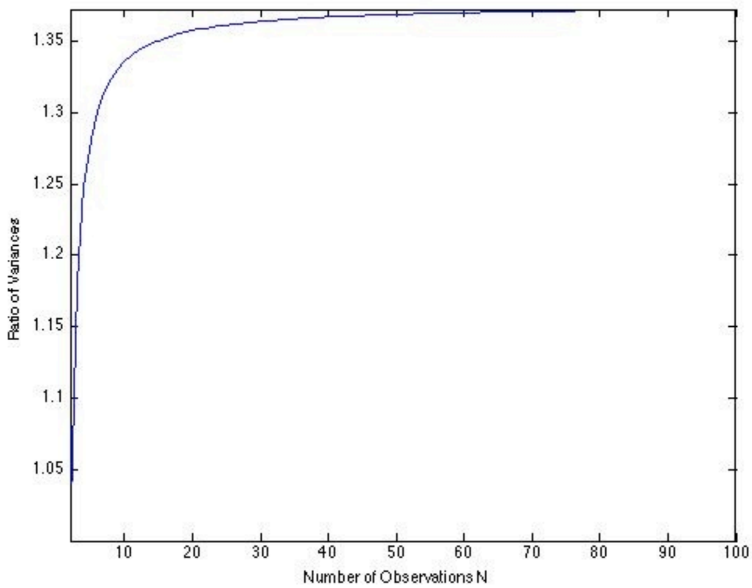
$$Var\frac{1}{N}\sum_{k=1}^{N}|X_k| = \frac{1}{N^2}\sum_{k=1}^{N}Var(|X_k|) = \frac{1}{N^2}\sum_{k=1}^{N}EX_k^2 - E^2(|X_k|) = \frac{\sigma^2}{N}(1-\frac{2}{\pi}) \tag{29}$$

For comparison, traditional estimate of standard deviation based on squares will have the first and second moments equal to[6]

$$E\sqrt{\frac{1}{N}\sum_{k=1}^{N}x_k^2} = \sqrt{\frac{2}{N}}\frac{\Gamma(N/2)}{\Gamma((N-1)/2)}\sigma \tag{30}$$

$$Var\sqrt{\frac{1}{N}\sum_{k=1}^{N}x_k^2} = \frac{\sigma^2}{N}\left(N-1-\frac{2\Gamma^2(N/2)}{\Gamma^2((N-1)/2)}\right) \tag{31}$$

Figure 2: $\left(N-1-\frac{2\Gamma^2(N/2)}{\Gamma^2((N-1)/2)}\right)/\left(1-\frac{2}{\pi}\right)$



Note that (29) and (31) are equal for N = 2 but for any N>2 expression (31) is strictly greater than (29). Of course, asymptotically, variance of both will vanish for sufficiently high N. However, the estimate

[6] Can be shown by deriving moments of $\sqrt{X^2}$ with N-degrees of freedom. See http://mathworld.wolfram.com/StandardDeviationDistribution.html for example.

based on absolute values is more efficient. Intuitively, it makes sense as high values of $|X_i|$ are magnified by taking squared values and thus estimate (31) will have a slower convergence. Figure 2 shows the ratio of (31) and (29) for $N=2,\cdots,100$.

The estimate based on absolute values (28) has additional advantage over the traditional estimate (30). The advantage is that (28) will converge even if (30) fails to converge to a constant. There is number of indications that financial time-series follow Levy-stable distributions with tail coefficient $1.5<\alpha<2$ and thus have infinite variance (see [3], for example). In this case, traditional dispersion estimates will fail as they will oscillate for any high number of observations. However, we have seen in Example 2 that $E|X|$ still exists for a range of Levy-stable distributions with tail coefficient $\alpha>1$ and thus (28) will still converge to a single number. However, misspecification of α will create a bias. To see how big the bias is, let us assume that we use the sum of absolute values to estimate dispersion $c^{1/\alpha}$ (equivalent of σ). Our correct estimation statistic should be[7]

$$\hat{c}^{1/\alpha} = \frac{\Gamma(\frac{1}{\alpha})\sin(\frac{\pi}{\alpha})}{2N} \sum_{i=1}^{n} |X_i| \tag{32}$$

but assuming $\alpha=2$, we will calculate

$$\hat{\hat{c}}^{1/\alpha} = \frac{\sqrt{\pi}}{2N} \sum_{i=1}^{n} |X_i| \tag{33}$$

and thus overestimating true dispersion by a factor of

$$\frac{\sqrt{\pi}}{\Gamma(\frac{1}{\alpha})\sin(\frac{\pi}{\alpha})} \tag{34}$$

This ratio is approximately 1.055 for $\alpha=1.9$ and increases towards 1.5 as α approaches 1.5. Overestimation should not be surprising as we will try to explain 'fat tails' by higher standard deviation rather than by leptokurtic shape of the underlying distribution.

7 From (12), using $\Gamma(1\text{-}x)\,\Gamma(x)=\pi/\sin(\pi x)$

Appendix

This appendix lists the main properties of the Hilbert transform. Properties 1-5 are standard and are presented without proof (see [4], for example). Property 6 is a generalization of the Hilbert transform of a function multiplied by its argument and is presented with proof.

1. Rearranging the integral in definition 1, gives us alternative formulas for the Hilbert transform,

$$Hf = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{f(t-x)}{x} dx = \frac{1}{\pi} \lim_{\delta \to 0+} \int_{\delta}^{+\infty} \frac{f(t-x) - f(t+x)}{x} dx \qquad \text{(A1)}$$

where the latter is often useful for numerical evaluations.

2. The Hilbert transform is linear, i.e. for constants a_1 and a_2

$$H(a_1 f_1 + a_2 f_2) = a_1 Hf_1 + a_2 Hf_2 \qquad \text{(A2)}$$

3. For the Fourier transform of f

$$Ff \equiv \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(x) e^{-i\omega x} dx$$

H and F are related as,

$$F(H^n f) = (-i \operatorname{sgn}(\omega))^n Ff \qquad \text{(A3)}$$

Note that for n = 2 and from linearity of F, we have $H^2 = -I$, or $H^{-1} = -H$.

4. If f is n-times differentiable and $Hd^n f/dt^n$ exists, the Hilbert transform and differentiation are mutually exchangeable, i.e.

$$\frac{\partial^n}{\partial t^n} Hf = H \frac{\partial^n}{\partial x^n} f(x) \qquad \text{(A4)}$$

5. The Hilbert transform can be understood as a convolution $f*(1/\pi t)$ allowing us to extend the definition to distributions (generalized functions). H applied to delta function is

$$H\delta = \frac{1}{\pi t} \text{ and } H \frac{\partial^n}{\partial x^n} \delta = \frac{\partial^n}{\partial t^n} \frac{1}{\pi t} \qquad \text{(A5)}$$

6. Let us define

$$g(x) \equiv \sum_{k=-\infty}^{+\infty} a_k x^k$$

where we assume that the sum converges for any x (or the sum contains only a finite number of non-zero elements), then

$$H(gf) = gHf + \sum_{k=-1}^{-\infty} t^k \sum_{j=k}^{-\infty} a_j \zeta_{kj} - \sum_{k=0}^{+\infty} t^k \sum_{j=k+1}^{+\infty} a_j \zeta_{kj} \text{ for constants } \zeta_{kj} = \frac{1}{\pi} P \int_{-\infty}^{+\infty} x^{j-k-1} f(x) dx \qquad \text{(A6)}$$

Proof:

Following the definition, we have

$$H(gf) = \frac{1}{\pi} P \int_{-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} a_k \frac{x^k f(x)}{t-x} dx$$

and if we change the order of summation and integration (both are limits in fact), we obtain

$$H(gf) = \sum_{k=-1}^{-\infty} a_k \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{x^k f(x)}{t-x} dx + a_0 Hf + \sum_{k=1}^{+\infty} a_k \left(\frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{(x^k - t^k) f(x)}{t-x} dx + t^k Hf \right)$$

Using identities

$$\frac{t^k - x^k}{t-x} = \sum_{j=0}^{k-1} t^{k-1-j} x^j \quad \text{and} \quad \frac{1}{x^k (t-x)} = \frac{1}{t^k (t-x)} + \sum_{j=1}^{k} \frac{1}{t^{k-j+1} x^j}$$

the last formula transforms into

$$H(gf) = \sum_{k=-\infty}^{+\infty} a_k t^k Hf + \sum_{k=-1}^{-\infty} a_k \sum_{j=-1}^{k} t^{k-j-1} \frac{1}{\pi} P \int_{-\infty}^{+\infty} x^j f(x) dx - \sum_{k=1}^{+\infty} a_k \sum_{j=0}^{k-1} t^{k-j-1} \frac{1}{\pi} P \int_{-\infty}^{+\infty} x^j f(x) dx$$

and rearranging such that we put together all terms corresponding to the same power of t, we obtain

$$H(gf) = gHf + \sum_{k=-1}^{-\infty} t^k \sum_{j=k}^{-\infty} a_j \frac{1}{\pi} P \int_{-\infty}^{+\infty} x^{j-k-1} f(x) dx - \sum_{k=0}^{+\infty} t^k \sum_{j=k+1}^{+\infty} a_j \frac{1}{\pi} P \int_{-\infty}^{+\infty} x^{j-k-1} f(x) dx$$

which completes the proof.

Two interesting identities follow from A(6) for g(x)=x and g(x)=x^{-1},

$$H(tf) = tHf - \frac{1}{\pi} P \int_{-\infty}^{+\infty} f(x) dx \tag{A7}$$

$$H(t^{-1} f) = t^{-1} Hf + t^{-1} \frac{1}{\pi} P \int_{-\infty}^{+\infty} \frac{f(x)}{x} dx \tag{A8}$$

which reduces for odd function f to H(tf)=tHf and even function f to H(t^{-1}f)=t^{-1}Hf, respectively.

References

1. Rao, R.C., Linear Statistical Inference and Its Applications, John Wiley & Sons, 1973
2. Zolotarev V.M., One-Dimensional Stable Distributions, Amer Mathematical Society, 1986
3. Mandelbrot, B.B., Fractals and Scaling in Finance, Springer-Verlag 1997
4. Johansson M., The Hilbert Transform, Master Thesis, Växjö University, available at http://w3.msi.vxu.se/exarb/mj_ex.pdf