Scientific
Research
Publishing

A Note on "Limit Distributions of Self-Normalized Sums" Using Cauchy-Generated Samples

Jan Vrbik

Department of Mathematics and Statistics, Brock University, St. Catharines, Canada
Email: jvrbik@brocku.ca

Abstract

In this case study, we would like to illustrate the utility of characteristic functions, using an example of a sample statistic defined for samples from Cauchy distribution. The derivation of the corresponding asymptotic probability density function is based on [1], elaborating and expanding the individual steps of their presentation, and including a small extension; our reason for such a plagiarism is to make the technique, its mathematical tools and ingenious arguments available to the widest possible audience.

Keywords

Self-Normalized Sum, Cauchy Distribution, Characteristic Functions, Fourier Transform, Padé Approximation

1. Introduction

The problem of finding the distribution of self-normalized sum of a Cauchy-generated sample has been considered since at least 1969 (see, for example, [2]), but solved (by proposing a concrete numerical algorithm) only in 1973 by [1], our key reference. After that, the research has been focused mainly on proving general statements concerning self-normalized sums, without dealing with specific distributions such as Cauchy—see, for example, [3].

In this article, we demonstrate how *characteristic functions* (CHF) are used in Statistics to find a distribution of a specific *sample statistic* (a function of individual observations). We do this by using a single comprehensive example, namely finding the $n \to \infty$ limit of the probability density function (PDF) of

$$\frac{\sum_{i=1}^{n} X_i}{\sqrt{\sum_{i=1}^{n} X_i^2}} \tag{1}$$

where $X_1, X_2, \cdots, X_n$ is a random independent sample (RIS) of size n from a Cauchy distribution with zero median. This goal has already been achieved (in a more general setting) by [1]; our main (rather pedagogical) reason for extending their presentation is to make it accessible to graduate and advance undergraduate students.

First we recall that the CHF of a random variable X is defined by

$$\varphi_x(s) = \mathbb{E}\left(e^{isX}\right) = \int_{-\infty}^{\infty} \cos(sx) \cdot f(x) \, \mathrm{d}x + i \int_{-\infty}^{\infty} \sin(sx) \cdot f(x) \, \mathrm{d}x \tag{2}$$

where $f(x)$ stands for the X's PDF, $\mathbb{E}$ for the corresponding expected value, and i is the purely imaginary unit; note that the real part of $\varphi_x(s)$ is always an *even* (an alternate name is *symmetric*) function of s, while the purely imaginary part is *odd*. Also note that when $f(x)$ is even, the corresponding CHF must be *real* (its purely imaginary part becomes zero because the integrand is odd).

Similarly

$$\varphi_{xy}(s,t) = \mathbb{E}\left(e^{isX+itY}\right) = \iint_{\substack{x\text{-}y \text{ plane}}} e^{isx+ity} \cdot f(x,y) \, \mathrm{d}x\mathrm{d}y \tag{3}$$

(where $f(x,y)$ is the *joint* PDF of X and Y) defines the *joint* CHF of two (not necessarily independent) random variables X and Y; when they *are* independent, we have

$$\varphi_{xy}(s,t) = \varphi_x(s) \cdot \varphi_y(t) \tag{4}$$

(product of the individual CHFs). This implies that (when independent), the CHF of $X + Y$ is given by $\varphi_x(s) \cdot \varphi_y(s)$, further implying that the CHF of a *sample mean* of n independent X_i is given by

$$\varphi_x\left(\frac{s}{n}\right)^n \tag{5}$$

When two random variables are combined into one (e.g. defining $U = \dfrac{X}{Y}$), the CHF of U is given by

$$\varphi_U(s) = \mathbb{E}\left(e^{is \cdot \frac{X}{Y}}\right) = \iint_{\substack{x\text{-}y \text{ plane}}} e^{is \cdot \frac{x}{y}} \cdot f(x,y) \, \mathrm{d}x\mathrm{d}y \tag{6}$$

Under very general conditions, knowing a generating function enables us to reverse the process and find the corresponding PDF thus:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} \varphi_x(s) \, \mathrm{d}s \tag{7}$$

To learn more about the mathematical details of transforming PDF into CHF and back the reader may like to consult [4].

2. Joint CHF and Its Limit

Assume a RIS of size n from a Cauchy distribution with the following PDF

$$g(x) = \frac{1}{\pi(1+x^2)} \tag{8}$$

The joint CHF of X and X^2 (seen as *distinct* random variables) is then given by

$$\int_{-\infty}^{\infty} \exp\left(ixs + ix^2 t\right) g(x)\,\mathrm{d}x \tag{9}$$

This implies that the joint CHF of

$$U_n \overset{\mathrm{def}}{=} \frac{\sum_{i=1}^{n} X_i}{n} \tag{10}$$

and of

$$V_n^2 \overset{\mathrm{def}}{=} \frac{\sum_{i=1}^{n} X_i^2}{n^2} \tag{11}$$

is (replace s by $\dfrac{s}{n}$ and t by $\dfrac{t}{n^2}$ in (9), and raise the result to the power of n)

$$\varphi_{U_n, V_n^2}(s,t) = \left\{1 + \int_{-\infty}^{\infty}\left[\exp\left(\frac{ixs}{n} + \frac{ix^2 t}{n^2}\right) - 1 - \frac{ixs}{n}\right] g(x)\,\mathrm{d}x\right\}^n \tag{12}$$

where adding and subtracting 1 facilitates taking the $n \to \infty$ limit; similarly, subtracting $\dfrac{ixs}{n}$ (which does not change the value of the integral since it integrates, in the principal-value sense, to 0) helps to make subsequent integrals converge. Note that principal value of an integral implies replacing $\int_{-\infty}^{\infty} \cdots \mathrm{d}x$ by $\lim_{R\to\infty} \int_{-R}^{R} \cdots \mathrm{d}x$; this is tacitly assumed from now on.

Replacing x by ny makes the previous expression into

$$\left\{1 + \int_{-\infty}^{\infty}\left[\exp\left(iys + iy^2 t\right) - 1 - iys\right] g(ny)\,n\,\mathrm{d}y\right\}^n \tag{13}$$

whose $n \to \infty$ limit is

$$\exp\left\{\int_{-\infty}^{\infty}\left[\exp\left(iys + iy^2 t\right) - 1 - iys\right] \frac{\mathrm{d}y}{\pi y^2}\right\} \tag{14}$$

since $g(ny) \to \dfrac{1}{n^2 u^2}$ and $\left(1 + \dfrac{A}{n}\right)^n \to \exp(A)$. The last displayed expression is thus the characteristic function of U and V^2 (our notation for the corresponding *limits* of U_n and V_n^2).

Note that only the tail behaviour of (8) was relevant in the end.

3. Finding CHF of (1)

Replacing t by $is^2 w$ (with w *positive*) results in

$$\varphi_{U,V^2}\left(s, is^2 w\right)$$
$$= \exp\left\{\int_{-\infty}^{\infty}\left[\exp\left(iys - s^2 w y^2\right) - 1 - iys\right] \frac{\mathrm{d}y}{\pi y^2}\right\} = \exp\left[s\psi(w)\right] \tag{15}$$

where (after the $ys = z$ substitution)

$$\psi(w) \overset{\text{def}}{=} \int_{-\infty}^{\infty} \left[\exp(iz - wz^2) - 1 - iz \right] \frac{dz}{\pi z^2}$$
$$= -\sqrt{\frac{2}{\pi}} Q\left(\frac{1}{\sqrt{2w}}\right) - 2\sqrt{\frac{w}{\pi}} \exp\left(-\frac{1}{4w}\right) \qquad (16)$$

with

$$Q(\tau) \overset{\text{def}}{=} \int_0^\tau \exp\left(-\frac{z^2}{2}\right) dz \qquad (17)$$

Note that the value of $\psi(w)$ is thus always *negative* (this remains true for the real part of $\psi(w)$ after we make w *complex*; this becomes consequential later on).

Proof.

$$\psi'(w) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \exp(iz - wz^2) dz$$

$$= -\frac{1}{\pi} \exp\left(-\frac{1}{4w}\right) \int_{-\infty}^{\infty} \exp\left(-w\left(z - \frac{i}{2w}\right)^2\right) dz$$

$$= -\frac{1}{\pi} \exp\left(-\frac{1}{4w}\right) \int_{-\infty - i/2w}^{\infty - i/2w} \exp(-wt^2) dt$$

$$= -\frac{1}{\pi} \exp\left(-\frac{1}{4w}\right) \int_{-\infty}^{\infty} \exp(-wt^2) dt$$

$$= -\frac{\exp\left(-\frac{1}{4w}\right)}{\sqrt{\pi w}}$$

since moving the path of the t integration does not change the integral's value (there are no singularities between the two paths) and the integrand tends to 0 sufficiently fast within the same strip when w tends to plus or minus infinity.

Therefore, $\psi(w)$ is given by

$$-\int \frac{\exp\left(-\frac{1}{4w}\right)}{\sqrt{\pi w}} dw + C = \sqrt{\frac{2}{\pi}} \int \frac{\exp\left(-\frac{z^2}{z}\right)}{z^2} dz + C \qquad (18)$$

$$= -\sqrt{\frac{2}{\pi}} \frac{\exp\left(-\frac{z^2}{z}\right)}{z} - \sqrt{\frac{2}{\pi}} \int \exp\left(-\frac{z^2}{z}\right) dz \qquad (19)$$

$$= -2\sqrt{\frac{w}{\pi}} \exp\left(-\frac{1}{4w}\right) - \sqrt{\frac{2}{\pi}} Q\left(\frac{1}{\sqrt{2w}}\right) \qquad (20)$$

using the $w = \frac{1}{2z^2}$ substitution. Note that $C = 0$, since $Q(\infty) = \sqrt{\frac{\pi}{2}}$ and

$$\psi(0) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\exp(iz) - 1 - iz}{z^2} dz = -1 \qquad (21)$$

This is proved by replacing the path of integration (the real axis) by a clockwise half circle of radius R, centered on the origin and denoted $\mathcal{R}$ (legitimate,

since the integrand has no singularities between the two paths), and then evaluating

$$\lim_{R \to \infty} \int_{\mathcal{R}} \frac{\exp(iz)}{z^2} dz - \lim_{R \to \infty} \int_{\mathcal{R}} \frac{1+iz}{z^2} dz = 0 + \lim_{R \to \infty} \int_0^\pi \frac{i - Re^{it}}{Re^{2it}} \cdot e^{it} dt = -\pi \qquad (22)$$

The first integral is equal to 0 due to Jordan's lemma. In the second integral, we have traded the clockwise half circle for counter-clockwise (by changing the integrand's sign).

Evaluating (15) at two distinct values of w (let us denote them w and w_0), dividing the difference by s, and integrating over s from 0 to infinity yields

$$\int_0^\infty \int_0^\infty \int_{-\infty}^\infty \exp(isu) \frac{\exp(-s^2 v^2 w) - \exp(-s^2 v^2 w_0)}{s} \cdot f(u,v) dudvds$$
$$= \int_0^\infty \frac{\exp[s\psi(w)] - \exp[s\psi(w_0)]}{s} ds \qquad (23)$$

where $f(u,v)$ is the joint PDF of U and V; note that its support is the whole upper half of the u-v plane.

Replacing s by $\dfrac{t}{v}$ (note that v cancels out from $\dfrac{ds}{s} = \dfrac{dt}{t}$) changes the left hand side of (23) to

$$\int_0^\infty \int_0^\infty \int_{-\infty}^\infty \exp\left(it\frac{u}{v}\right) \frac{\exp(-t^2 w) - \exp(-t^2 w_0)}{t} \cdot f(u,v) dudvdt \qquad (24)$$

$$= \int_0^\infty \varphi(t) \cdot \frac{\exp(-t^2 w) - \exp(-t^2 w_0)}{t} dt \qquad (25)$$

where $\varphi(t)$ is the characteristic function of $\dfrac{U}{V}$, the variable whose distribution we seek; the last expression is then equal to the right-hand side of (23).

4. Converting to PDF

Differentiating the resulting equation, *i.e.* (25) = (23), with respect to w then yields (after a sign reversal)

$$\int_0^\infty t\varphi(t) \cdot \exp(-t^2 w) dt = -\psi'(w) \int_0^\infty \exp[s\psi(w)] ds = \frac{\psi'(w)}{\psi(w)} \qquad (26)$$

(to the last integrand, $\psi(w)$ is just a negative constant).

Finally, multiplying both sides of the previous equation by

$$\frac{\exp\left(\dfrac{y^2}{4w}\right)}{\pi\sqrt{\pi w}} \qquad (27)$$

and integrating over w from 0 to $i \cdot \infty$ (a notation which implies following the positive *imaginary* axis) results in

$$\frac{1}{\pi} \int_0^\infty \varphi(t) \cdot \exp(-ity) dt = \frac{1}{\pi^{3/2}} \int_0^{i\cdot\infty} \frac{\exp\left(\dfrac{y^2}{4w}\right)}{\sqrt{w}} \cdot \frac{\psi'(w)}{\psi(w)} dw \qquad (28)$$

where $\sqrt{w}$ denotes the corresponding *principal* value, and (16) is now extended to complex arguments.

Proof. Following [5], we write

$$\int_0^{i \cdot \infty} \frac{\exp\left(\dfrac{y^2}{4w} - t^2 w\right)}{\sqrt{w}} \, dw = \sqrt{\frac{2yi}{t}} \int_0^{\infty} \exp\left(-\frac{ity}{2}\left(z^2 + \frac{1}{z^2}\right)\right) dz$$

$$= e^{-ity} \sqrt{\frac{2yi}{t}} \int_0^{\infty} \exp\left(-\frac{ity}{2}\left(z - \frac{1}{z}\right)^2\right) dz \qquad (29)$$

after introducing

$$w = \frac{iz^2 y}{2t} \qquad (30)$$

A further $z \mapsto \dfrac{1}{z}$ substitution makes it into

$$e^{-ity} \sqrt{\frac{2yi}{t}} \int_0^{\infty} \exp\left(-\frac{ity}{2}\left(z - \frac{1}{z}\right)^2\right) \frac{dz}{z^2} \qquad (31)$$

Adding (29) and (31), which are identical in terms of their value, and dividing by 2 yields

$$e^{-ity} \sqrt{\frac{yi}{2t}} \int_0^{\infty} \exp\left(-\frac{ity}{2}\left(z - \frac{1}{z}\right)^2\right)\left(1 + \frac{1}{z^2}\right) dz \qquad (32)$$

Finally, introducing $q = \left(z - \dfrac{1}{z}\right)$ results in

$$e^{-ity} \sqrt{\frac{yi}{2t}} \int_{-\infty}^{\infty} \exp\left(-\frac{ity}{2} q^2\right) dq = e^{-ity} \cdot \frac{\sqrt{\pi}}{t} \qquad (33)$$

assuming that, to evaluate the last integral, we first replace t by $t - i\varepsilon$ (to make it converge), and then take the $\varepsilon \to 0$ limit of the answer (Cauchy-type integration).

It is well known that the *real part* of the left hand side (and, therefore, of the right-hand side) of (28) yields the desired (clearly *symmetric*) PDF of $\dfrac{U}{V}$, say $f(y)$, due to $\varphi(t)$ being real (and thus automatically symmetric as well). This then yields

$$f(y) = \frac{1}{\pi^{3/2}} \operatorname{Re} \int_0^{i \cdot \infty} \frac{\exp\left(\dfrac{y^2}{4w}\right)}{\sqrt{w}} \cdot \frac{\psi'(w)}{\psi(w)} \, dw$$

$$= \frac{\sqrt{2}}{\pi^{3/2}} \operatorname{Re} \int_0^{i \cdot \infty} \frac{\exp\left(\dfrac{y^2}{4w}\right)}{(2w)^{3/2}} \cdot \frac{dw}{1 + \dfrac{1}{\sqrt{2w}} \exp\left(\dfrac{1}{4w}\right) Q\left(\dfrac{1}{\sqrt{2w}}\right)} \qquad (34)$$

based on (16) and (18). The problem is that there is no simple analytic answer to the last integral; worse yet, its integrand is highly oscillatory, preventing us from direct numerical integration.

In an attempt to break this impasse, we introduce the following substitution

$$w = \frac{1}{2\tau^2} \tag{35}$$

getting

$$f(y) = \frac{\sqrt{2}}{\pi^{3/2}} \mathrm{Re} \int_0^{\exp(-i\pi/4)\cdot\infty} \frac{\exp\left(\dfrac{y^2\tau^2}{2}\right) \mathrm{d}\tau}{1 + \tau \exp\left(\dfrac{\tau^2}{2}\right) Q(\tau)} \tag{36}$$

where τ now follows the complex ray at $-45°$.

Unfortunately, even after this substitution, special measures are still needed to facilitate accurate numerical integration of the integral. For one, it becomes necessary to separately deal with the $y^2 < 1$ and $y^2 > 1$ regions.

5. Case of $y^2 < 1$

When $y^2 < 1$, it is legitimate to rotate the ray of the last integration to the positive *real* axis, since there are no singularities of the integrand between the old and the new path, and the function decreases sufficiently fast towards infinity in that segment. We then get

$$f(y) = \int_0^{\infty} \frac{\sqrt{\dfrac{2}{\pi^3}} \exp\left(\dfrac{y^2\tau^2}{2}\right) \mathrm{d}\tau}{1 + \tau \exp\left(\dfrac{\tau^2}{2}\right) Q(\tau)} \tag{37}$$

where τ is *real*, and the integrand is "well behaved" (no oscillations). Results of numerical evaluation are displayed in **Figure 1** (we are showing only the $y > 0$ portion of the graph; visualize a mirror image of this curve when $-1 < y < 0$).

To investigate the nature of this singularity we notice that, for large τ, the integrand of (37) tends to
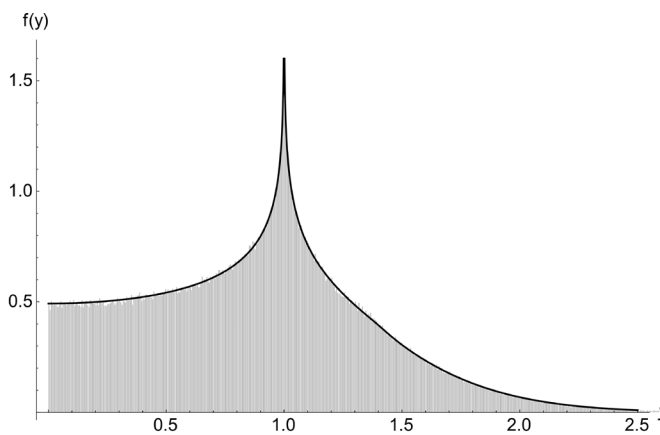
Figure 1. Resulting PDF for $y^2 < 1$. Note the logarithmic-type singularity at $y = \pm 1$.

$$\frac{2}{\pi^2 \tau} \cdot \exp\left(-\frac{\tau^2\left(1-y^2\right)}{2}\right) \tag{38}$$

since $Q(\tau) \xrightarrow[\tau \to \infty]{} \sqrt{\pi/2}$. Integrating the last function over τ from 0 to infinity yields

$$\frac{1}{\pi^2} \cdot \Gamma\left(0, \frac{1-y^2}{2}\right) \tag{39}$$

where the second factor is a special case of the incomplete gamma function defined by

$$\Gamma(0, z) \stackrel{\text{def}}{=} -\gamma - \ln z - \sum_{k=1}^{\infty} \frac{(-z)^k}{k \cdot k!} \tag{40}$$

and γ is Euler's gamma. This indicates that, by adding $\ln\left(1-y^2\right)\pi^{-2}$ to $f(y)$ makes the resulting function finite (and thus amendable to a simple, Padé-type approximation, constructed later on).

6. Case of $y^2 > 1$

The situation becomes somehow more difficult when $y^2 > 1$. Now, we can rotate the $-45°$ ray of the (36) integration to $-90°$ (the negative imaginary axis), since the integrand decreases sufficiently fast towards infinity in this range of directions. The new path contributes 0 to the real part of (36), since the integrand (along the $-90°$ ray) is real and $d\tau$ is purely imaginary. But moving from $-45°$ to $-90°$ we have crossed infinitely many *singularities* located between the two rays; these can be found *numerically* (a rather tedious procedure) by solving

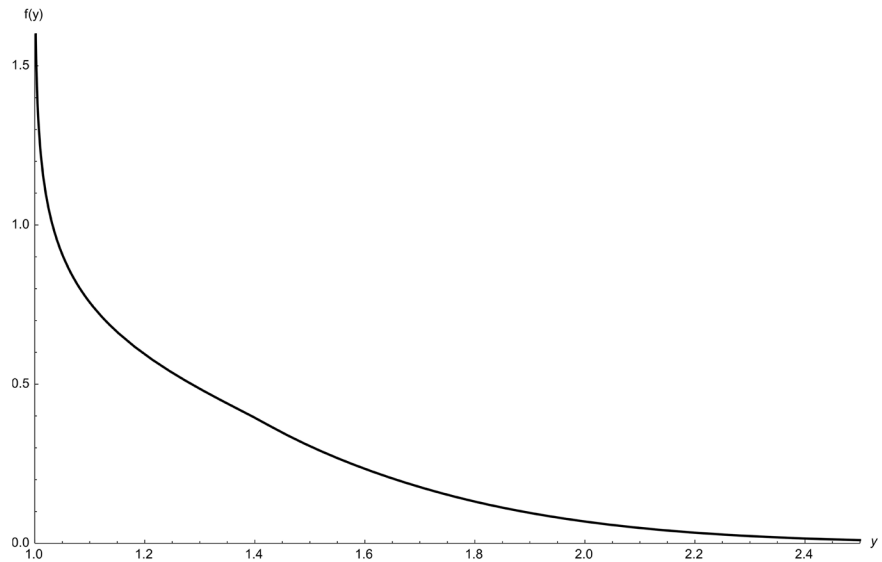$$1 + \tau \exp\left(\frac{\tau^2}{2}\right) Q(\tau) = 0 \tag{41}$$

The location of the first 50 of these (they are all *simple* poles, slowly converging to the $\operatorname{Im}\tau = -\operatorname{Re}\tau$ line) are shown in **Figure 2**.

We can then evaluate (36) using Cauchy integral theorem; note that, at each of these poles (say at τ_p), the τ derivative of $1 + \tau \exp\left(\frac{\tau^2}{2}\right)Q(\tau)$ equals to $-\frac{1}{\tau_p}$, since

$$\left.\frac{d\left(1 + \tau \exp\left(\frac{\tau^2}{2}\right)Q(\tau)\right)}{d\tau}\right|_{\tau=\tau_p} \tag{42}$$

$$= \exp\left(\frac{\tau_p^2}{2}\right)Q(\tau_p) + \tau_p^2 \exp\left(\frac{\tau_p^2}{2}\right)Q(\tau_p) + \tau_p \exp\left(\frac{\tau_p^2}{2}\right)\exp\left(-\frac{\tau_p^2}{2}\right)$$

and

Figure 2. First 50 roots of (41).

$$1 + \tau_p \exp\left(\frac{\tau_p^2}{2}\right) Q(\tau_p) = 0 \qquad (43)$$

implying

$$\exp\left(\frac{\tau_p^2}{2}\right) Q(\tau_p) = -\frac{1}{\tau_p} \qquad (44)$$

This means that each pole (with the exception of the first one—see the next paragraph), contributes

$$\sqrt{\frac{2}{\pi^3}} \operatorname{Re}\left[2\pi i \cdot \tau_p \exp\left(\frac{\tau_p^2 y^2}{2}\right)\right] \qquad (45)$$

to (36).

The pole on the imaginary axis needs to be avoided by an infinitesimal *half* circle, which means that it contributes only one *half* of the above amount, namely

$$\sqrt{\frac{2}{\pi}} \cdot 1.30693 \exp\left(-\frac{1.30693^2 y^2}{2}\right) \qquad (46)$$

The last function yields the asymptotic behaviour of $f(y)$ as $y^2 \to \infty$, and provides an excellent approximation (its maximum absolute error is about 2×10^{-6}) to this PDF when $|y| > 1.8$. Unfortunately, to build a numerical solution for the full range of $y^2 > 1$ values by adding contributions of sufficiently many of these poles (as done by [1]) is rather tedious, as finding hundreds of these poles (needed to reach a good accuracy, especially when y^2 approaches 1) is a non-trivial task, and the resulting convergence is quite slow.

Nevertheless we would like to mention that, by adding the contribution of the second pole, namely

$$-2\sqrt{\frac{2}{\pi}}\operatorname{Im}\left(\tau_2\exp\left(\frac{\tau_2^2 y^2}{2}\right)\right) \tag{47}$$

where $\tau_2 = -1.84906 - 3.45208\mathrm{i}$, to (46) results in an equally accurate approximation (its error is less than 2×10^{-6}) in the $|y|>1.7$ tails of $f(y)$.

7. Alternate Solution

We are now going to backtrack and deal directly with (36). Even though its integrand is still highly oscillatory (with a frequency which increases as $\tau \to \infty \cdot \mathrm{e}^{-\mathrm{i}\cdot\pi/4}$), we can mitigate the problem by dividing the range of integration into two segments: first we integrate from 0 to $1-\mathrm{i}$ (which does not pose any numerical difficulty), thus getting what we will call the *first* component of (36), then from $1-\mathrm{i}$ to infinity (along the $-45°$ ray), getting the *second* component.

To carry out the latter integration, we first note that, for large τ, $Q(\tau)$ can be expanded in the following manner

$$Q(\tau)\simeq\sqrt{\frac{\pi}{2}}-\exp\left(-\frac{\tau^2}{2}\right)\cdot\left(\frac{1}{\tau}-\frac{1}{\tau^3}+\frac{3}{\tau^5}-\frac{5\cdot3}{\tau^7}+\cdots\right) \tag{48}$$

Substituting this into the integrand of (36) and further expanding in powers of $\frac{1}{\tau}$ (up to and including the 6th power) while keeping the two exponential terms fixed (*i.e. not* expanding in τ) yields

$$\left(\left(\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\tau}-\frac{2\exp\left(-\frac{\tau^2}{2}\right)}{\tau^4\pi}+\frac{6\exp\left(-\frac{\tau^2}{2}\right)}{\tau^6\pi}\right)\cdot\exp\left(-\frac{\tau^2\left(y^2-1\right)}{2}\right)\right) \tag{49}$$

which can be τ-integrated, from $1-\mathrm{i}$ to $\infty(1-\mathrm{i})$, *analytically*; then we *numerically* integrate (over the same line segment) the difference between the integrands of (36) and (49)—since the resulting integrand now approaches zero (as τ increases) very quickly, the integration range becomes effectively finite ($1-\mathrm{i}$ to $6-6\mathrm{i}$ already achieves very high accuracy), thus eliminating any troublesome oscillation. Adding the real part of the last two answers (and multiplying by $\sqrt{2\pi^{-3}}$) then provides the second component of $f(y)$.

This technique yields the graph of **Figure 3** (drawn for $y>1$; visualize a mirror image of this curve when $y<-1$):

8. Monte-Carlo Verification

One can always get a good idea about a distribution of *any* sample statistic (however complicated and inaccessible to analytic treatment) by actually generating a random sample with the required properties by a computer, and using it to compute the desired statistic's value; this is then repeated as many times as possible, displaying the results in a histogram.

We have done this with our (1), using $n=100$ and generating one million of

its random values. Plotting the resulting histogram, together with the theoretical asymptotic PDF of the last three sections, yields **Figure 4**; visual comparison clearly indicates a good agreement between the two answers. The tiny discrepancy still discernible (mainly in the $0 < y < 1$ range) is due to the fact that $n = 100$ is not large enough to have reached the $n \to \infty$ limit yet (trying to make the sample size substantially higher would run up against our computer's capacity).

Note that, to be more economical, we have folded the negative and positive parts of the distribution into a single graph, effectively plotting the PDF of the *absolute value* of (1).

9. Accurate Approximation

First we have to remember the we already have an excellent approximation to $f(y)$ in the $|y| > 1.7$ region, given by the sum of (46) and (47); now we have to find a similarly accurate way of dealing with $|y| \leq 1.7$.
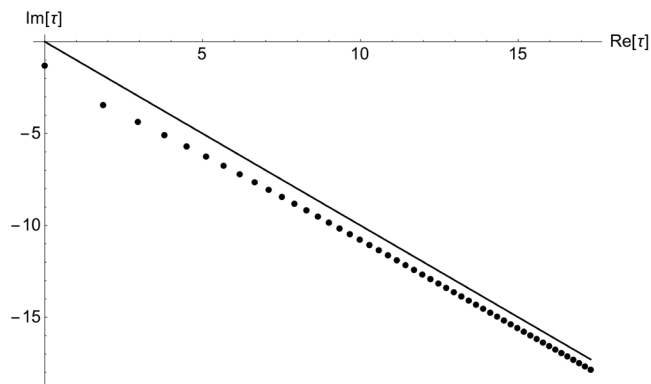
Figure 3. Resulting PDF when $y^2 > 1$.
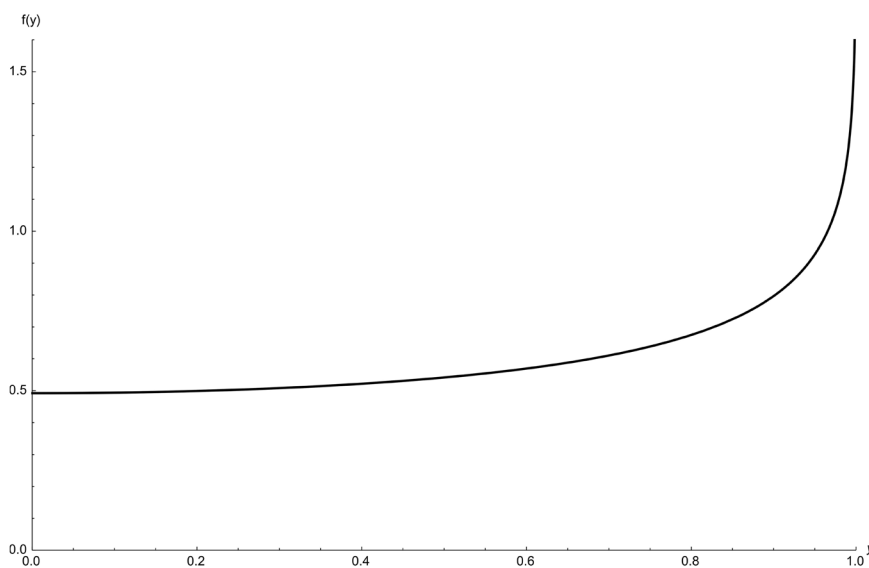
Figure 4. Comparing theoretical and empirical PDF's.

We already know that, by adding

$$\frac{\ln\left|1-y^2\right|}{\pi^2} \tag{50}$$

to $f(y)$ removes the corresponding singularity when $y^2 \leq 1$; we can easily verify that the same is true when $y^2 > 1$, since one of the terms in the analytic result of the (49) integration (once we take its real part and multiply by $\sqrt{2\pi^{-3}}$) equals to

$$\frac{1}{\pi^2}\cdot\operatorname{Re}\Gamma\left(0,\mathrm{i}\cdot\frac{y^2-1}{2}\right) \tag{51}$$

where

$$\operatorname{Re}\Gamma\left(0,\mathrm{i}\cdot z\right)=-\gamma-\ln z-\sum_{k=1}^{\infty}\frac{\left(-1\right)^k z^{2k}}{2k\cdot\left(2k\right)!} \tag{52}$$

whose singular part is given by (50). This proves that

$$f(y)+\frac{\ln\left|1-y^2\right|}{\pi^2} \tag{53}$$

is singularity-free for *all* values of *y*; unfortunately, that still does not make it sufficiently "smooth", as the next paragraph indicates.

There is yet another subtle issue causing great difficulty when trying to build an approximate formula for (53): the function has a small (hardly noticeable) kink (a discontinuous second derivative) at $y^2 = 2$; this is not an artifact of the new technique—the same (rather surprising) phenomenon can be confirmed by the old technique of adding residues. Luckily, we can identify yet another term of the (49) contribution responsible for this discontinuity, and remove it by further subtracting the offending term from (53), thus getting

$$f(y)+\frac{\ln\left|1-y^2\right|}{\pi^2}-\frac{2\operatorname{Re}\left(2-y^2\right)^{3/2}\left(3y^2-11\right)}{15\pi^2}. \tag{54}$$

This function is (finally!) sufficiently (even though not perfectly—tiny issues remain with higher derivatives at $y^2 = 2, y^2 = 3$ etc.) smooth to fit it (in the $|y| < 1.7$ range) by a ratio of two polynomials (by minimizing the total error squared). This results in the following approximation

$$\frac{0.666457-0.82876y^2+0.401154y^4-0.0800733y^6+0.00203206y^{10}}{1-0.572265y^2+0.0327094y^4+0.0207377y^6+0.00042241y^{10}+0.000261666y^{14}}. \tag{55}$$

Its absolute error never exceeds 6×10^{-6}, which is more than sufficient for any practical application. The form of this expression and the individual powers of *y* have been chosen somehow arbitrarily; we do not claim that our choices are optimal.

Do not forget that (55) is approximating (54); a corresponding adjustment has to be made to convert it into an approximation for $f(y)$.

10. Conclusion

The aim of this article was to demonstrate that finding the distribution of a relatively simple sample statistic requires a skillful use of characteristic functions and a whole gamut of sophisticated mathematical techniques, including real and complex analysis, Fourier transform, and curve fitting. We hope that students of Statistics can benefit from the ingenuity of the authors of the original derivation of this distribution as presented in [1], and from some extra details included in this article; the latter includes a different numerical approach to building the resulting PDF, expressing it in the form of an accurate Padė-type approximation (discovering an interesting discontinuity in the process), and verifying the answer by Monte Carlo simulation.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Logan, B.F., Mallows, C.L., Rice, S.O. and Shepp, A.L. (1973) Limit Distributions of Self-Normalized Sums. *The Annals of Probability*, **1**, 788-809.
https://doi.org/10.1214/aop/1176996846

[2] Efron, B. (1969) Student's t-Test under Symmetry Conditions. *Journal of the Americal Statistical Association*, **64**, 1278-1302.
https://doi.org/10.1080/01621459.1969.10501056

[3] Spătaru, A. (2014) Convergence and Precise Asymptotics for Series Involving Self-Normalized Sums. *Journal of Theoretical Probability*, **29**, 267-276.
https://doi.org/10.1007/s10959-014-0560-1

[4] Sneddon, I.N. (2010) Fourier Transforms. Dover Publications.

[5] Viola, M. How Does One Derive the Following Formula of Integration?
https://math.stackexchange.com/q/3323687