

MEANS AS IMPROPER INTEGRALS

John E. Gray ¹ and Andrew Vogt ²
John E. Gray, Andrew Vogt

¹ Code B-31, Sensor Technology & Analysis Branch, Electromagnetic and Sensor Systems Department, Naval Surface Warfare Center Dahlgren, 18444 Frontage Road Suite 328, Dahlgren, VA 22448-5161 USA; john.e.gray@navy.mil

² Department of Mathematics and Statistics, Georgetown University, Washington, D. C. 20057-1233 USA; vogta@georgetown.edu

Correspondence: vogta@georgetown.edu; Tel.: 1-202-687-2529

Abstract

The aim of this work is to study generalizations of the notion of mean. Kolmogorov proposed a generalization based on an improper integral with a decay rate for the tail probabilities. This weak or Kolmogorov mean relates to the Weak Law of Large Numbers in the same way that the ordinary mean relates to the Strong Law. We propose a further generalization, also based on an improper integral, called the doubly weak mean, applicable to heavy-tailed distributions such as the Cauchy distribution and the other symmetric stable distributions. We also consider generalizations arising from Abel-Feynman type mollifiers that damp the behavior at infinity and alternative formulations of the mean in terms of the cumulative distribution and the characteristic function.

Law of Large Numbers; weak or Kolmogorov mean; Abel's Theorem; mollifiers; summation methods; stable distributions

60A05, 26E60, 60F05, 26A42, 34M30

1 Introduction

The mean is sometimes taken as the foundational principle for all of probability theory (see, for example, [19]). Nate Silver [16] in the President's Invited Address at JSM 2013 said that "the average is still the most useful statistical tool ever invented." Steven Levitt [10] in the Arthur M. Sackler Lecture at the National Academy of Sciences in March 2015 said "When I work with companies ..., I like to show a comparison of means and that's often more effective than very complicated things" Kosko [9] suggested, on the other hand, jettisoning the mean in favor of the median because of its shortcomings. The chief shortcoming is that the mean sometimes does not exist whereas the median always exists (even though it is multi-valued in some cases). Kosko cited the Cauchy distribution as a leading example where the mean fails to exist.

Here we study generalizations of the mean intended to extend its reach as much as possible. At the outset we note that these generalizations are only applicable to distributions with heavy tails at both ends:

$$\int_{-\infty}^0 xP(dx) = -\infty, \quad \int_0^{\infty} xP(dx) = \infty$$

must hold. If neither of these equations is true, then $E(X)$ exists. If only one of them holds, then $E(X)$ is $\pm\infty$. Thus our ideas will apply only to distributions heavy-tailed at both ends and indeed having "equally" heavy tails at both ends. This includes the symmetric stable distributions ([2],[3],[12]), a family arising in signal processing, among which are the normal and Cauchy distributions and which satisfy a Generalized Central Limit Theorem.

Three ways to extend the notion of mean come to mind.

One way, always available, is to transform the variable so that the mean exists. We apply a strictly increasing function f to the variable X and study $E_f(X) = f^{-1}(E(f(X)))$. The function f can be a power function, a log function, a logistic function, or even the function $\arctan x$. The function converts X into a bounded variable that necessarily has a finite mean. The difficulty with this approach is that the extensions do not recover the ordinary mean. If $E_f(X) = E(X)$ whenever $E(X)$ exists, then f is necessarily a linear transformation given by $f(x) = Cx + D, C \neq 0$, and $E_f(X)$ exists only when $E(X)$ exists.

A second way to extend the mean, an obvious and straightforward approach, is to replace the usual definition by an improper integral. Instead of integrating x over \mathcal{R} with respect to the probability distribution P of the random variable X we integrate over the

interval $[c - M, c + M]$ and determine what happens as M tends to ∞ . We follow this approach here.

We itemize the different cases that can result (Theorem 2.1) in this way, and give examples of each. We examine a weak mean proposed by Kolmogorov that corresponds to the Weak Law of Large Numbers, and we show that it is additive. We also consider two further generalizations - one with a weakening of the decay conditions that Kolmogorov imposed for his weak mean, which we call the *doubly weak* mean; and another, the *superweak* mean, when the improper integral exists for at least one choice of the center c . These topics occupy sections 2 and 3.

A third way to extend the notion of mean is to introduce what we call a “mollifier,” namely, a parameter-dependent multiplier $\phi_\lambda(x)$ such that $x \mapsto \phi_\lambda(x)x$ is integrable with respect to P and $\phi_\lambda(x)$ tends to 1 as λ tends to λ_0 . Then we consider what happens to $E(\phi_\lambda(X)X)$ as λ tends to λ_0 . This approach is examined in section 4. We note some dangers associated with mollifiers, consider some examples, and determine cases where this approach reduces to the previous approach. Richard Feynman sometimes used mollifier methods, and his approach motivated us. A well known theorem of Abel is also related to these methods.

In section 5 we compare and restate the results discussed earlier in terms of the cumulative distribution and the characteristic function of the variable X . In section 6 we offer a brief conclusion.

2 Improper Integrals extending the Mean

Consider a real-valued random variable X . We suppose that associated with X is a Borel probability measure P that takes each Borel subset A of the real numbers to ¹:

$$P(A) = \text{the probability that } X \text{ belongs to the set } A.$$

The *mean* of X , denoted by $E(X)$ or μ_X , is defined when x is integrable with respect to P to be:

$$E(X) = \int_{\mathcal{R}} x P(dx).$$

One direction of the remarkable Strong Law of Large Numbers (see Pollard [14, p. 78 and pp. 37-8]) states that if $\{X_n\}$ is a sequence of independent random variables with common distribution P and if there exists a constant m such that

¹We shall also use the notation $P(X \in A)$.

$$\frac{X_1 + \dots + X_n}{n} \text{ converges almost surely to } m$$

as n tends to ∞ , then each X_n has mean m . Here “almost surely” means outside a set of measure zero in the countably infinite product space induced by the measure P (see [14, pp. 99-102]). Simply put, if the sample mean of independent copies of X settles down to a specific number, then that number is $E(X)$. This can be regarded as the motivation for the transition from the sample mean to the mathematical mean $E(X)$. The general notion of mean is derived from the finitary notion of sample mean.

The other direction of the Strong Law of Large Numbers asserts that if $E(X)$ exists, then the sample mean of n identical independent copies of X converges almost surely to $E(X)$ as n tends to infinity. For a proof of both directions of the Strong Law, see [14, pp. 95-102, p. 105]. For an alternate proof due to N. Etemadi, see [14, pp. 106-7].

When x is not integrable with respect to P , the notion $E(X)$ above is inapplicable and we must rely on other notions of mean. The most obvious generalization is the following improper integral:

$$L(c) \stackrel{\text{def}}{=} \lim_{M \rightarrow \infty} \int_{[c-M, c+M]} x P(dx)$$

for a real number c .

By the Lebesgue Dominated Convergence Theorem [5, p. 172] this notion coincides with the ordinary mean when x is integrable with respect to P . In his great foundational work [6, p. 40], Kolmogorov noted this option in the case when $c = 0$ and observed that it does not require integrability of $|x|$. Indeed if X is a random variable obeying the Cauchy distribution $f(x) = 1/\pi(1 + x^2)$, then X satisfies $L(c) \equiv 0$ for any choice of c .

Two related notions of mean are:

$$\text{L-1) } \lim_{M \rightarrow \infty} \int_{[a-M, b+M]} x P(dx)$$

and

$$\text{L-2) } \lim_{\min\{M, K\} \rightarrow \infty} \int_{[a-M, b+K]} x P(dx),$$

where $a \leq b$.

It is easily seen that the expression in L-1 coincides with $L((a + b)/2)$ since

$$[a - M, b + M] = \left[\frac{a + b}{2} - \left(M + \frac{b - a}{2}\right), \frac{a + b}{2} + \left(M + \frac{b - a}{2}\right) \right].$$

As for L-2, we have the following result.

Proposition 2.1 *Let X be a random variable with probability measure P . Let a and b be real numbers with $a \leq b$. Then*

$$\lim_{\min\{K, M\} \rightarrow \infty} \int_{[a-M, b+K]} x P(dx)$$

exists if and only if x is integrable with respect to P . Furthermore, when this limit exists, it is equal to $E(X)$.

Proof: If x is integrable on \mathcal{R} , the limit exists and equals the mean of x by Lebesgue's Dominated Convergence Theorem. Conversely, if the limit exists, then

$$0 \leq \int_{(b+K, b+K']} x P(dx) < \epsilon$$

for $K < K'$, both sufficiently large, and any given ϵ . Likewise

$$-\epsilon < \int_{[a-M', a-M)} x P(dx) \leq 0$$

for $M < M'$, both sufficiently large. Fatou's Lemma or Levi's Theorem [5, p. 172] thus implies both that

$$0 \leq \int_{(b+K, \infty)} x P(dx) \leq \epsilon \quad \text{and} \quad -\epsilon \leq \int_{(-\infty, a-M)} x P(dx) \leq 0.$$

Thus x is integrable on $[0, \infty)$ as well as $(-\infty, 0]$, and so is integrable on $\mathcal{R} = (-\infty, \infty)$. ■

Returning now to $L(c)$, we stipulate that $-\infty \leq L(c) \leq \infty$. This gives us flexibility in characterizing what can happen. A necessary but not sufficient condition for $L(c)$ to be finite, we note in passing, is that for each positive real number p

$$\lim_{M \rightarrow \infty} M |P(M \leq X - c \leq M + p) - P(-M - p \leq X - c \leq -M)| = 0.$$

Lemma 2.1 *Let X be a random variable with probability measure P , and let c_1 and c_2 be real numbers with $c_1 < c_2$. Then there are three possibilities:*

i) *If $L(c_1)$ exists in $[-\infty, \infty]$, then*

$$L(c_1) \leq \liminf_{M \rightarrow \infty} \int_{[c_2-M, c_2+M]} x P(dx);$$

ii) *If $L(c_2)$ exists in $[-\infty, \infty]$, then*

$$L(c_2) \geq \limsup_{M \rightarrow \infty} \int_{[c_1-M, c_1+M]} x P(dx);$$

iii) *If $L(c_1)$ and $L(c_2)$ both exist in $[-\infty, \infty]$, then*

$$L(c_1) = L(c_2).$$

Proof: Suppose $c_1 < c_2$. Then

$$\begin{aligned} \int_{[c_2-M, c_2+M]} x P(dx) &= \int_{[c_1-M, c_1+M]} x P(dx) \\ &+ \int_{(c_1+M, c_2+M]} x P(dx) + \int_{[c_1-M, c_2-M)} (-x) P(dx). \end{aligned} \quad (1)$$

The second and third terms on the right side of equation (1) are both non-negative for M sufficiently large, and i) and ii) of Lemma 2.1 follow at once.

In the case of iii), note that i) and ii) imply that if both $L(c_1)$ and $L(c_2)$ exist, then $L(c_1) \leq L(c_2)$. Suppose then that $L(c_1)$ and $L(c_2)$ both exist with $c_1 < c_2$ and $-\infty \leq L(c_1) < L(c_2) \leq \infty$. Then there is a positive number K such that $K < L(c_2) - L(c_1)$, and for M sufficiently large:

$$\begin{aligned} K &< \int_{(c_1+M, c_2+M]} x P(dx) + \int_{[c_1-M, c_2-M)} (-x) P(dx) \\ &\leq (c_2 + M)P((c_1 + M, c_2 + M]) + (M - c_1)P([c_1 - M, c_2 - M)) \\ &\leq (M + d)(P((c_1 + M, c_2 + M] \cup [c_1 - M, c_2 - M))) \end{aligned}$$

where $d = \max\{|c_2|, |c_1|\}$. Thus

$$\frac{K}{M + d} < P((c_1 + M, c_2 + M] \cup [c_1 - M, c_2 - M)).$$

Now replace M by $M_j = M + j(c_2 - c_1)$ for each non-negative integer j to get:

$$\frac{K}{M_j + d} < P((c_1 + M_j, c_2 + M_j] \cup [c_1 - M_j, c_2 - M_j)).$$

Summing over these inequalities and noting that $c_2 + M_j = c_1 + M_{j+1}$ and $c_1 - M_j = c_2 - M_{j+1}$, we obtain:

$$\infty = \sum_{j=0}^{\infty} \frac{K}{M + d + j(c_2 - c_1)} \leq P((c_1 + M, \infty) \cup (-\infty, c_2 - M)) \leq 1,$$

which is a contradiction. Thus, the possibility $L(c_1) < L(c_2)$ is eliminated, and $L(c_1) = L(c_2)$. ■

Theorem 2.1 *Let X be a random variable with probability measure P . Then exactly one of these possibilities holds:*

- i) $L(c)$ does not exist in $[-\infty, \infty]$ for any real number c ;*
- ii) $L(c)$ exists in $(-\infty, \infty)$ for exactly one real number c ;*
- iii) $L(c)$ exists in $[-\infty, \infty]$ for all real numbers c and is independent of c ;*
- iv) there is a number c_0 such that $L(c) = \infty$ for $c > c_0$, $L(c)$ does not exist for $c < c_0$, and $L(c_0)$ equals ∞ or does not exist;*
- v) there is a number c_0 such that $L(c) = -\infty$ for $c < c_0$, $L(c)$ does not exist for $c > c_0$, and $L(c_0)$ equals $-\infty$ or does not exist.*

Proof: Suppose $L(c) = \infty$ for some number c . Let $c_0 = \inf \{c : L(c) = \infty\}$. By Lemma 2.1 part i) $L(c) = \infty$ for all $c > c_0$. If $c_0 = -\infty$, then the situation is described by Theorem 2.1 part iii). If $c_0 > -\infty$, part iii) of Lemma 2.1 implies that the situation is given by Theorem 2.1 part iv).

Likewise, suppose $L(c) = -\infty$ for some number c . Let $c_1 = \sup \{c : L(c) = -\infty\}$. If $c_1 = \infty$, then Theorem 2.1 part iii) applies. If $c_1 < \infty$, part iii) of Lemma 2.1 implies that Theorem 2.1 part v) holds.

Now suppose $L(c) \neq \pm\infty$ for all real numbers c . Then either $L(c)$ does not exist for any c and Theorem 2.1 part i) holds, or $L(c)$ exists for exactly one c and Theorem 2.1 part ii) holds, or $L(c)$ exists at two or more points and is a finite number.

In the latter case we can find c_1 and c_2 with $c_1 < c_2$ so that by Lemma 2.1 part iii) $L(c_1) = L(c_2) \in \mathcal{R}$. In this case the last two terms in equation (1) each tend to 0 as M tends to infinity. By changes of variable we conclude that:

$$\lim_{M \rightarrow \infty} \int_{(M, p+M]} x P(dx) = 0. \quad (2)$$

and

$$\lim_{M \rightarrow \infty} \int_{[-M-p, -M]} x P(dx) = 0. \quad (3)$$

hold where $p = c_2 - c_1$. It follows immediately that (2) and (3) also hold when $0 < p < c_2 - c_1$. Likewise, if (2) and (3) hold for p , they also hold for $2p, 3p, \dots$, and indeed for np where n is any fixed positive integer. Accordingly, (2) and (3) hold for all positive real numbers p .

Applying this result to the two terms on the far right side of equation (1), we find that

$$L(c) \equiv L(c_1) = L(c_2)$$

for all real numbers c , and thus Theorem (2.1) part iii) holds. ■

We give some examples to illustrate that each of the possibilities enumerated in Theorem 2.1 can occur. For simplicity we use discrete random variables in most of our examples.

Example 2.1. Consider a random variable X whose probability measure is of the form

$$P = \sum_{n=1}^{\infty} \left(\frac{1}{2^{2n}} \delta_{2^{2n}} + \frac{1}{2^{2n-1}} \delta_{(-2^{2n-1})} \right)$$

where δ_z is the (Dirac) probability measure whose value is 1 on any Borel subset of \mathcal{R} that contains the real number z and whose value is zero on the remaining Borel subsets of \mathcal{R} . The sum of the nonzero values is 1, so this obviously defines a probability measure. However the integral of x over the interval $[c - M, c + M]$ is the difference between the size of the first set and the size of the second set below:

$$\text{The size of the set } \{n : 1 \leq n, 2^{2n} \leq (c + M)\} = \lfloor \frac{\log(c + M)}{2 \log 2} \rfloor$$

$$\begin{aligned} \text{The size of the set } \{n : 1 \leq n, 2^{2n-1} \leq (M - c)\} = \\ \lfloor \frac{\log(M - c)}{2 \log 2} + \frac{1}{2} \rfloor \end{aligned}$$

where $\lfloor \cdot \rfloor$ is the floor function. For fixed c and sufficiently large M the difference of the above quantities can assume the values 0 and -1 and the integral does not settle down to either one. This is an instance of Theorem 2.1, part i).

Example 2.2. A random variable X can also be defined with probability measure of the form

$$P = \sum_{n=1}^{\infty} \frac{1}{2^{n+1}} (\delta_{2^n} + \delta_{(-2^n)}).$$

So P is concentrated at the points $\pm 2^n$ and assigns probability $1/(2^{n+1})$ to these points. For this measure, $L(0)$ equals 0 by symmetry. However, $L(c)$ does not exist for other choices of c . If c is positive, the integral of x over the closed interval $[c - M, c + M]$ reduces by cancellation to its integral over the half-open interval $(M - c, M + c]$. This integral oscillates between 0 and $\frac{1}{2}$ for large M depending on whether 2^n is in the interval $(M - c, M + c]$ or not. Similar behavior occurs when $c < 0$. This example is an instance of Theorem 2.1, part ii).

Example 2.3. Now consider a random variable X having a probability density (with respect to Lebesgue measure on the real line) of the form

$$f(x) = \begin{cases} \frac{A}{1+Cx^a} & \text{if } x \geq 0 \\ \frac{B}{1+D|x|^b} & \text{if } x < 0 \end{cases}$$

where a and b are numbers in $(1, 2)$ and $A, B, C,$ and D are suitable positive constants that guarantee that the density integrates to 1. This random variable satisfies part iii) of Theorem 2.1. It is easy to see that $L(c) \equiv \infty$ for all c or $-\infty$ for all c according as $b > a$ or $a > b$. When $a = b$, then $L(c) \equiv +\infty$ if $AD - BC > 0$ and $L(c) \equiv -\infty$ if $AD - BC < 0$. When $a = b$ and $AD = BC$, then $L(c)$ is a real number independent of c as in part iii) of Theorem 2.1.

If we let $a = b = 2$ and $(A, C) = (B, D) = (\frac{1}{\pi}, 1)$, we obtain the Cauchy distribution, which also illustrates part iii) of Theorem 2. with $L(c) \equiv 0$.

Example 2.4. The probability measure

$$P = \sum_{n=1}^{\infty} \left(\frac{2^n}{3^{n+1}} \delta_{3^n} + \frac{2^{n-1}}{3^{n+1}} \delta_{(-3^n)} \right)$$

illustrates part iv) of Theorem 2.1. If $c \geq 0$, the integral of x over $[c - M, c + M]$ is given by:

$$\sum_{\{n : 1 \leq n, 3^n \leq c+M\}} \frac{2^n}{3} - \sum_{\{n : 1 \leq n, 3^n \leq M-c\}} \frac{2^{n-1}}{3},$$

and this expression has the value $(2^{n_0} - 1)/3$ or $(2^{n_0} + 2^{n_0-1} - 1)/3$ where $n_0 = \lfloor \log(c + M)/\log 3 \rfloor$ for large M . Since M and n_0 tend to infinity together, it follows that $L(c) \equiv \infty$ for $c \geq 0$.

On the other hand, if $c = -d$ where $d > 0$, the integral of x over $[c - M, c + M] = [-M - d, M - d]$ is given by:

$$\sum_{\{n : 1 \leq n, 3^n \leq M-d\}} \frac{2^n}{3} - \sum_{\{n : 1 \leq n, 3^n \leq M+d\}} \frac{2^{n-1}}{3},$$

and this reduces to $(2^{n_0} - 1)/3$ or to $(-1)/3$ for large M where $n_0 = \lfloor \log(M + d)/\log 3 \rfloor$ depending on whether a positive integer lies in the interval $(\log(M - d)/\log 3, \log(M + d)/\log 3]$ or not. Thus $L(c)$ does not exist for $c < 0$.

Example 2.5. Another example for part iv) of Theorem 2.1 is the case where the probability measure is given by:

$$P = K \sum_{n=1}^{\infty} \left(\frac{2^n}{3^n + (1/n)} \delta_{3^n + (1/n)} + \frac{2^{n-1}}{3^n} \delta_{(-3^n)} \right).$$

Here K is a suitably chosen positive normalizer, which is easily seen to be in the interval $(1/3, 1)$. For $c > 0$, the integral of x over $[c - M, c + M]$ is

$$K \left(\sum_{\{n : 1 \leq n, (3^n + 1/n) \leq M+c\}} 2^n - \sum_{\{n : 1 \leq n, 3^n \leq M-c\}} 2^{n-1} \right),$$

and this is greater than or equal to $K(2^{n_0} - 1)$ for M sufficiently large where $n_0 = \lfloor \log(M - c)/\log 3 \rfloor$. Since n_0 and M tend to infinity together, $L(c) \equiv \infty$ for all $c > 0$. When $c = 0$, the integral of x over $[-M, M]$ reduces to $K(2^{n_0} - 1)$ or to $-K$ where n_0 is the largest integer such that $3^{n_0} + (1/n_0) \leq M$ and the first or second reduction occurs according as $M < 3^{n_0+1}$ or not. Thus $L(0)$ does not exist. By Theorem 2.1 $L(c)$ does not exist for $c < 0$.

Other cases arising in Theorem 2.1, such as part v), are obtained by modifying the examples above, e.g., replacing X by $-X$ or by $X + a$.

3 Weak Means

An implication of Theorem 2.1 is that if $L(c)$ exists for more than one choice of c and is finite for one c , then $L(c)$ exists for all c , is finite, and is independent of c . The case of the Cauchy distribution shows that this can happen without the ordinary mean existing. Accordingly for a random variable X , we define the *doubly weak mean* of X , denoted by $E_{ww}(X)$, to be the common value of $L(c)$ for all c when this common value exists and is in $(-\infty, \infty)$. In Theorem 3.3 below we will consider alternative characterizations of the doubly weak mean.

An intermediate notion exists due to Kolmogorov [6, pp. 64-66] between the ordinary mean and the doubly weak that motivates our terminology. The *weak mean* of X , denoted by $E_w(X)$, is defined as follows: $E_w(X)$ is the quantity $L(0)$ provided the latter exists in $(-\infty, \infty)$ and

$$\lim_{n \rightarrow \infty} nP(|X| > n) = 0. \quad (4)$$

The following theorem is due to Kolmogorov. It indicates that existence of the weak mean coincides precisely with existence of a number for which the Weak Law of Large Numbers holds.

Theorem 3.1 (Kolmogorov, 1928) *Let X be a random variable. Suppose that X_1, \dots, X_n, \dots are independent identically distributed copies of X with P_n the n -fold product distribution. Then there is a real number m such that for each $\epsilon > 0$*

$$\lim_{n \rightarrow \infty} P_n\left(\left|\frac{X_1 + \dots + X_n}{n} - m\right| > \epsilon\right) = 0$$

if and only if X has weak mean $E_w(X) = m$.

Proof: See [6, p. 65], [7], and [8, Theorems XII and XIII]. ■

Proposition 3.1 *Let X be a random variable.*

- i) If X has a mean, then X has a weak mean and $E(X) = E_w(X)$;*
- and*
- ii) if X has a weak mean, then X has a doubly weak mean and $E_w(X) = E_{ww}(X)$.*

Proof: In the case when X has a mean, then the identity function $x \mapsto x$ is integrable with respect to the probability measure P on the real line. In particular the tail integrals

$$\int_{[n, \infty)} x P(dx) \quad \text{and} \quad \int_{(-\infty, -n]} x P(dx)$$

tend to zero as n tends to infinity. Since the absolute values of these integrals are larger respectively than $nP(X \geq n)$ and $nP(X \leq -n)$, it follows that $\lim_{n \rightarrow \infty} nP(|X| > n) = 0$. Likewise by Lebesgue's Dominated Convergence Theorem, $L(0) = E(X)$. Now suppose X has a weak mean. If $c_1 < c_2$ and $\epsilon > 0$ are given, then for a sufficiently large M ,

$$\begin{aligned}
 0 \leq \int_{(c_1+M, c_2+M]} x P(dx) &\leq (c_2 + M)P((c_1 + M, c_2 + M]) \\
 &\leq (c_2 + M)P(|X| > \lfloor c_1 + M \rfloor) \\
 &\leq \frac{c_2 + M}{\lfloor c_1 + M \rfloor} \epsilon.
 \end{aligned}$$

For sufficiently large M , the right side is as close to ϵ as we like. Thus

$$\lim_{M \rightarrow \infty} \int_{(c_1+M, c_2+M]} x P(dx) = 0.$$

Similarly,

$$\lim_{M \rightarrow \infty} \int_{[c_1-M, c_2-M)} x P(dx) = 0.$$

Accordingly from equation (1) in the proof of Lemma 2.1, it follows that whenever one of $L(c_2)$ or $L(c_1)$ exists and is finite, the other exists and is equal to it. Since $L(0) = m$, then $L(c)$ exists for all c , $L(c) \equiv m$, and m is the doubly weak mean of X . ■

Kolmogorov in [6, p. 66] gives an example where the Weak Law holds but the Strong Law does not. Kolmogorov’s example is a random variable X whose probability distribution P is given by:

$$P(X \in A) = \int_A \frac{C}{(|x| + 2)^2(\ln(|x| + 2))} dx$$

where C is a suitable normalizing constant and A is any Borel set in the reals. Cauchy random variables have $L(c)$ existing for all c , independent of c , but violate the Weak Law by not decaying rapidly enough at infinity. These examples demonstrate that the notions of mean, weak mean, and doubly weak mean are strictly distinct.

Since the Strong Law and the Weak Law correspond precisely to the mean and the weak mean, it is natural to wonder if another such law corresponds to the doubly weak mean.

Theorem 3.2 (Doubly Weak Law of Large Numbers) *Let X be a random variable. For c in \mathcal{R} and $M > 0$, let ${}^cX^M$ denote the random variable defined by:*

$${}^cX^M = \begin{cases} X & \text{if } |X - c| \leq M \\ 0 & \text{else.} \end{cases}$$

Suppose that ${}^cX_1^M, \dots, {}^cX_n^M, \dots$ are independent identically distributed copies of ${}^cX^M$ with P_n their n -fold product distribution. If

there is a real number m such that for each $\epsilon > 0$ and for $c = c_1$ and $c = c_2$ for distinct real numbers c_1 and c_2 , and for M sufficiently large

$$\lim_{n \rightarrow \infty} \hat{P}_n(|\frac{{}^c X_1^M + \dots + {}^c X_n^M}{n} - m| > \epsilon) = 0, \quad (5)$$

then X has doubly weak mean $E_{ww}(X) = m$.

Conversely, if X has doubly weak mean $E_{ww}(X) = m$, then equation (5) holds for each $\epsilon > 0$ for each $c \in \mathcal{R}$ and for M sufficiently large.

Proof: Suppose that ${}^c X^M$ is as above and (5) holds. Note that the probability distribution \hat{P} of ${}^c X^M$ is given by $\hat{P}(A) = P(A \cap [c - M, c + M]) + \delta_0(A)(1 - P([c - M, c + M]))$ for each Borel set A . The variable ${}^c X^M$ is a bounded variable and hence has a mean given by:

$$E({}^c X^M) = \int_{[c-M, c+M]} x \hat{P}(dx).$$

Furthermore this variable obeys the Weak Law of Large Numbers, i. e.,

$$\lim_{n \rightarrow \infty} \hat{P}_n(|\frac{{}^c X_1^M + \dots + {}^c X_n^M}{n} - E({}^c X^M)| > \epsilon) = 0 \quad (6)$$

for every $\epsilon > 0$, every $c \in \mathcal{R}$, and every $M > 0$. Combining (5) and (6), we conclude that

$$\lim_{n \rightarrow \infty} \hat{P}_n(|E({}^c X^M) - m| \leq 2\epsilon) = 1$$

for $c \in \{c_1, c_2\}$ and M sufficiently large. Since the inequality $|E({}^c X^M) - m| \leq 2\epsilon$ does not depend on n , we conclude that it must be true for M sufficiently large. But ϵ is an arbitrary positive number. So

$$\begin{aligned} L(c) &= \lim_{M \rightarrow \infty} \int_{[c-M, c+M]} x P(dx) \\ &= \lim_{M \rightarrow \infty} \int_{[c-M, c+M]} x \hat{P}(dx) \\ &= \lim_{M \rightarrow \infty} E({}^c X^M) = m \end{aligned} \quad (7)$$

for $c \in \{c_1, c_2\}$, that is to say, $L(c_1)$ and $L(c_2)$ exist and are equal to the real number m . By Theorem 2.1 $L(c) \equiv m$ for every real numbers c , and m is the doubly weak mean of X .

To prove the converse, note that since $L(c) \equiv E_{ww}(X)$ for all c then for each $c \in \mathcal{R}$, (7) holds. Thus for any $\epsilon > 0$, any c , and for any M sufficiently large (dependent on both ϵ and c)

$$|E({}^c X^M) - m| \leq \frac{\epsilon}{2},$$

$$\hat{P}_n(|E({}^c X^M) - m| \leq \frac{\epsilon}{2}) = 1 \text{ for all } n \geq 1.$$

However, combining the above result with (6), with ϵ , we obtain (5) with ϵ replaced by $\frac{\epsilon}{2}$. Since ϵ is an arbitrary positive number, the factor of 2 is irrelevant. Thus (5) holds for all ϵ , all c , and M sufficiently large. ■

The doubly weak mean can be characterized in a different manner as follows, based on the argument in Theorem 2.1.

Theorem 3.3 *The random variable X has a doubly weak mean if and only if any of the following equivalent conditions holds:*

- i) $L(c)$ exists in $(-\infty, \infty)$ for all real numbers c and is independent of c ; or*
- ii) $L(c)$ exists in $(-\infty, \infty)$ for two distinct real numbers c ; or*
- iii) $L(c)$ exists in $(-\infty, \infty)$ for some real number c and for every positive real number p $\lim_{M \rightarrow \infty} MP(M < |X| \leq M + p) = 0$; or*
- iv) $L(c)$ exists in $(-\infty, \infty)$ for some real number c and there exists a real number $p \geq 1$ such that $\lim_{n \rightarrow \infty} nP(n < |X| \leq n + p) = 0$.*

Of course n above denotes an integer variable, while M is a real variable. Any one of these conditions can be taken as defining when a doubly weak mean exists, in which case the doubly weak mean is the (common) value of $L(c)$. The closest in spirit to Kolmogorov's definition for the weak mean (cf. equation (4)) is condition iv).

Proof: That i) and ii) are equivalent is a consequence of Theorem 2.1. If ii) holds, it also follows from Theorem 2.1 that $-\infty < L(c_1) = L(c_2) < \infty$ for real numbers c_1 and c_2 with $c_1 < c_2$. The proof of Theorem 2.1 then shows that equations (2) and (3) hold for any positive number p . Since

$$0 \leq MP_X(M < X \leq M + p) \leq \int_{(M, p+M]} xP(dx)$$

$$0 \geq -MP(-M > X \geq -M - p) \geq \int_{[-M-p, -M)} xP(dx),$$

it follows that ii) implies iii). Condition iv) is a special case of iii).

We must still show that iv) implies i) or ii). The condition that

$$\lim_{n \rightarrow \infty} nP(n < |X| \leq n + p) = 0$$

for some $p \geq 1$ implies the same for any smaller $p > 0$. For a number k larger than p in the interval $(jp, (j+1)p]$ with j a positive integer, choose n_0 so large that for $n \geq n_0$

$$0 \leq nP(n < |X| \leq n + p) < \frac{\epsilon}{j+1}.$$

Then

$$\begin{aligned} nP(n < |X| \leq n + k) &\leq nP(n < |X| \leq n + (j+1)p) \\ &\leq (j+1) \frac{\epsilon}{j+1} = \epsilon \end{aligned}$$

for $n \geq n_0$. Thus iv) holds for any $p > 0$.

If $L(c)$ exists for some number c and d is another number, larger than c without loss of generality, then in imitation of equation (1)

$$\begin{aligned} \int_{d-M}^{d+M} xP(dx) &= \int_{c-M}^{c+M} xP(dx) + \int_{c+M}^{d+M} xP(dx) \\ &\quad - \int_{c-M}^{d-M} xP(dx). \end{aligned} \quad (8)$$

However, for a positive number $M > \max(-c, d)$

$$\begin{aligned} 0 &\leq \int_{c+M}^{d+M} xP(dx) \leq (d+M)P(c+M \leq X \leq d+M) \\ &\leq n_1 P(n_1 < |X| \leq n_1 + p) + pP([n_1 < X]) \end{aligned}$$

and

$$\begin{aligned} 0 &\leq \left| \int_{c-M}^{d-M} xP(dx) \right| \leq (M-d)P(c-M \leq X \leq d-M) \\ &\leq n_2 P(n_2 < |X| \leq n_2 + p) + pP(X < -n_2), \end{aligned}$$

where n_1 and n_2 are integers such that $n_1 < c + M \leq n_1 + 1$ and $n_2 < M - d \leq n_2 + 1$, and $p = d - c + 1$.

For M sufficiently large, n_1 and n_2 are as large as we like, but with p fixed, the right sides of these two inequalities tend to 0. According

in the limit as M tends to ∞ in equation (8), we obtain $L(d) = L(c)$. This establishes ii) and hence i). ■

The doubly weak mean may appear to be the last possibility for generalizing the mean. But Theorem 2.1 suggests yet another generalization. We say that X has a *superweak mean* if $L(c)$ exists in $[-\infty, \infty]$ for some real number c . Example 2.2 is a case where a finite superweak mean exists but not a doubly weak or weak mean. We have thus covered every case in Theorem 2.1 since Example 2.1 shows that there are case where even the superweak mean does not exist.

We are now in a position to treat symmetric stable distributions.

Proposition 3.2 *Let X be a symmetric stable distribution with location parameter a . Then X has a doubly weak mean $E_{ww}(X) = a$.*

Proof: In the notation of [12, p. 14] when the characteristic exponent α is in $(1, 2]$ the ordinary mean $E(X) = a$ exists. See [12, p. 22]. We assume the symmetry parameter $\beta = 0$. When $\alpha = 1$, we are dealing with the Cauchy distribution, which obviously satisfies iii) and iv) of Theorem 3.3. So it suffices to consider cases when $0 < \alpha < 1$. Furthermore, the variable X can be taken to have location parameter $a = 0$ and scale parameter $\gamma = 1$ since we can replace X by $(X - a)/\gamma^{1/\alpha}$ without loss of generality. Then note that the density function for such an X is symmetric about 0 and hence $L(0) = 0$.

It thus suffices to show that $MP(M < X \leq M + p)$ tends to 0 as M tends to ∞ for a fixed $p > 0$. Using equation 2.9 of [12, p. 16], we obtain:

$$\begin{aligned} MP(M \leq X \leq M + p) &= M \int_M^{M+p} \left[\frac{1}{\pi x} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \Gamma(\alpha k + 1) x^{-\alpha k} \sin \frac{k\alpha\pi}{2} \right] dx \\ &= Mp \left[\frac{1}{\pi M'} \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!} \Gamma(\alpha k + 1) (M')^{-\alpha k} \sin \frac{k\alpha\pi}{2} \right] \\ &\leq \frac{Mp}{\pi M'} \sum_{k=1}^{\infty} (M')^{-\alpha k} = \frac{Mp}{\pi M'} \frac{(M')^{-\alpha}}{1 - (M')^{-\alpha}} \\ &= \frac{Mp}{M'((M')^{\alpha} - 1)}, \end{aligned}$$

which tends to 0 as M tends to ∞ . Here M' is a number in $(M, M + p)$ that depends on M and p and is guaranteed to exist by the integral form of the Mean Value Theorem. ■

Another proposition that can be established is the following.

Proposition 3.3 *Let X and Y be random variables both of which have a weak mean. Then $X + Y$ has a weak mean and $E_w(X + Y) = E_w(X) + E_w(Y)$.*

Proof: This can be shown directly from the definition of the weak mean and equation (4) at the beginning of section 3. However, here we will use the Weak Law. Let $(X + Y)_1, \dots, (X + Y)_n, \dots$ be a sequence of independent copies of $X + Y$ with associated product measure P_n for the first n of these variables. Let X_1, \dots, X_n, \dots be a sequence of independent copies of X , Y_1, \dots, Y_n, \dots a sequence of independent copies of Y , and without loss of generality suppose $X_1 + Y_1, \dots, X_n + Y_n, \dots$ are independent. Let Q_n, Q'_n , and Q_{2n} be the product measure spaces associated respectively with the first n terms of these sequences. Then

$$\begin{aligned}
 & P_n\left(\left|\frac{(X + Y)_1 + \dots + (X + Y)_n}{n} - E_w(X) - E_w(Y)\right| < \epsilon\right) = \\
 & \quad Q_{2n}\left(\left|\frac{X_1 + Y_1 + \dots + X_n + Y_n}{n} - E_w(X) - E_w(Y)\right| < \epsilon\right) \geq \\
 & \quad Q_{2n}\left(\left|\frac{X_1 + \dots + X_n}{n} - E_w(X)\right| < \frac{\epsilon}{2}, \left|\frac{Y_1 + \dots + Y_n}{n} - E_w(Y)\right| < \frac{\epsilon}{2}\right) \\
 & = Q_{2n}\left(\left|\frac{X_1 + \dots + X_n}{n} - E_w(X)\right| < \frac{\epsilon}{2}\right) + \\
 & \quad Q_{2n}\left(\left|\frac{Y_1 + \dots + Y_n}{n} - E_w(X)\right| < \frac{\epsilon}{2}\right) - \\
 & \quad Q_{2n}\left(\left|\frac{X_1 + \dots + X_n}{n} - E_w(X)\right| < \frac{\epsilon}{2} \text{ or } \left|\frac{Y_1 + \dots + Y_n}{n} - E_w(X)\right| < \frac{\epsilon}{2}\right) \\
 & \geq Q_{2n}\left(\left|\frac{X_1 + \dots + X_n}{n} - E_w(X)\right| < \frac{\epsilon}{2}\right) \\
 & + Q_{2n}\left(\left|\frac{Y_1 + \dots + Y_n}{n} - E_w(X)\right| < \frac{\epsilon}{2}\right) - 1 \\
 & = Q_n\left(\left|\frac{X_1 + \dots + X_n}{n} - E_w(X)\right| < \frac{\epsilon}{2}\right) \\
 & + Q'_n\left(\left|\frac{Y_1 + \dots + Y_n}{n} - E_w(X)\right| < \frac{\epsilon}{2}\right) - 1.
 \end{aligned}$$

Since X and Y have weak means the last two lines converges to $1 + 1 - 1 = 1$ as n tends to ∞ . Accordingly the first expression does also, and our result follows by Theorem 3.1. \blacksquare

A similar result is not available for doubly weak means except for special cases (e.g., linear combinations of independent copies of a symmetric stable distributions).

4 Mollifiers

Richard Feynman was famous for his integration techniques, some of which are recorded in the book of Mathews and Walker [11], based on lectures Feynman gave at Cornell. Feynman's ideas, as noted earlier, partly motivated our investigation.

Mollifiers are used to aid approximation of the delta function and to smooth functions, but another use is to regularize behavior at $\pm\infty$. Mollifiers fall under the heading of summation methods [18]. The mollifier can then be used to reinterpret integrals, or renormalize them, in a manner that makes them finite. This method is used to "evaluate" the integrals of $\sin bx$ and $\sin x/x$ on $[0, \infty)$ in [11, p. 60 and p. 91].)

This idea can be used to generalize the notion of mean. We introduce a function $\phi_\lambda(x)$ that depends on a parameter λ so that $x \mapsto \phi_\lambda(x)x$ is integrable with respect to P for $\lambda \neq \lambda_0$ and $\phi_\lambda(x) \rightarrow 1$ for a. e. x as $\lambda \rightarrow \lambda_0$. Then we define:

$$L(\phi) \stackrel{def}{=} \lim_{\lambda \rightarrow \lambda_0} E(\phi_\lambda(X)X)$$

In the case of the means $L(c)$ discussed in earlier sections, the multiplier can be taken to be

$$\phi_\lambda(x) = \chi_{[c-1/|\lambda|, c+1/|\lambda|]}(x) \quad (9)$$

where χ_A is the characteristic function of the set A , $|\lambda| = 1/M$, and $\lambda_0 = 0$. Then $L(\phi)$ is the same as $L(c)$.

However, there are dangers associated with mollifiers that the following example illustrates.

Example 4.1 Define a function $\phi_{\lambda,D}$ for λ and D in \mathcal{R} by:

$$\phi_{\lambda,D}(x) = \begin{cases} e^{-|\lambda|x} & \text{if } x \geq 0 \\ e^{(|\lambda|x)(1 + \pi D|\lambda|x)} & \text{if } x < 0, \end{cases}$$

Evidently $\phi_{\lambda,D}$ is a well-behaved function, integrable and dying off at $\pm\infty$. Also $\{\phi_{\lambda,D}\}$ converges pointwise to the constant function identically equal to 1 as λ tend to $\lambda_0 = 0$ with D fixed.

Suppose we use this family of functions as a mollifier to determine a mean for a variable obeying the Cauchy distribution. Let $m(\lambda, D)$ be defined by:

$$m(\lambda, D) = \int_{-\infty}^{\infty} \phi_{\lambda,D}(x) \frac{x}{\pi(1+x^2)} dx = \int_{-\infty}^0 \frac{D\lambda e^{|\lambda|x} x^2}{1+x^2} dx = Dm(\lambda, 1).$$

Now

$$\begin{aligned}
1 &= e^{|\lambda|x}|_{-\infty}^0 = \int_{-\infty}^0 |\lambda|e^{|\lambda|x} dx \\
&\geq \int_{-\infty}^0 \frac{|\lambda|e^{|\lambda|x}x^2}{1+x^2} dx = m(\lambda, 1) = \int_0^{\infty} \frac{|\lambda|e^{-|\lambda|x}x^2}{1+x^2} dx \\
&\geq \int_K^{\infty} \frac{|\lambda|e^{-|\lambda|x}x^2}{1+x^2} dx \geq \frac{K^2e^{-|\lambda|K}}{1+K^2}
\end{aligned}$$

for any positive real number K . Thus

$$1 \geq \limsup_{\lambda \rightarrow 0} m(\lambda, 1) \geq \liminf_{\lambda \rightarrow 0} m(\lambda, 1) \geq \frac{K^2}{1+K^2}.$$

Letting K tend to infinity, we find that $\lim_{\lambda \rightarrow 0} m(\lambda, 1) = 1$.

Hence the mollifier-induced mean of the standard Cauchy distribution is:

$$\begin{aligned}
L(\phi) &= \lim_{\lambda \rightarrow 0} \int_{-\infty}^{\infty} \phi_{\lambda, D}(x) \frac{x}{\pi(1+x^2)} dx = \lim_{\lambda \rightarrow 0} m(\lambda, D) \\
&= \lim_{\lambda \rightarrow 0} Dm(\lambda, 1) = D \lim_{\lambda \rightarrow 0} m(\lambda, 1) = D.
\end{aligned}$$

But D is arbitrary and depends on the choice of the mollifier!

The underlying problem is that the choice of mollifier is asymmetric. There are two issues: symmetry and a center of symmetry. The doubly weak mean discussed in the previous section addresses these two issues by making use of the family of improper integrals (9) and requiring that $E_{ww}(X) = L(c)$ be the same regardless of the center c .

This motivates the following definition. We call $\{\phi_\lambda\}$ a *mollifier for X with center c* if and only if for each real number λ in some neighborhood of 0 we are given a function $x \mapsto \phi_\lambda(x)$ taking \mathcal{R} into \mathcal{R} with the following properties:

1. $x \mapsto \phi_\lambda(x)x$ is integrable with respect to P for $\lambda \neq 0$,
2. $\lim_{\lambda \rightarrow 0} \phi_\lambda(x) = 1$ for each x in \mathcal{R} ,
3. $\phi_\lambda(c-x) = \phi_\lambda(c+x)$ for all $x \geq 0$, and
4. $\phi_\lambda(c) = 1$ for all λ .

Without loss of generality we have taken the limiting value of λ to be 0 in this definition. Furthermore, since $\phi_\lambda(c)$ tends to 1 as λ tends to 0, if $\{\phi_\lambda\}$ does not satisfy 4 we can replace this family by $\{\frac{\phi_\lambda}{\phi_\lambda(c)}\}$.

Examples of mollifiers, in addition to (9), include:

$$e^{-|\lambda x|}, e^{-|\lambda|x^2}, \frac{\sin \lambda x}{\lambda x}, \text{ and } \cos \lambda x.$$

The examples in (9) have center c , while those just mentioned have center 0 for suitable X 's. In general a mollifier with center c can be created by taking one with center 0, call it $\phi_\lambda(x)$, and replacing it by $\phi_\lambda(x - c)$. However, we cannot be sure that the new mollifier will satisfy the property that $x \mapsto x\phi_\lambda(x - c)$ remains integrable with respect to P (although it will be with respect to P_Y where $Y = X - c$).

Let $\{\phi_\lambda\}$ be a mollifier for the random variable X , and as before let

$$L(\phi) = \lim_{\lambda \rightarrow 0} \int_{-\infty}^{\infty} \phi_\lambda(x) x P(dx).$$

It is natural to consider how $L(\phi)$ behaves in relation to the family $\{L(c) : c \in \mathcal{R}\}$ described in Theorem (2.1). We content ourselves with the following theorem and a few examples.

Theorem 4.1 *Let $\{\phi_\lambda\}$ be a mollifier for the random variable X with center c . Suppose that for each $\lambda \neq 0$ $\phi_\lambda(x)$ is an absolutely continuous function of x on each finite subinterval of \mathcal{R} . Suppose also that there is a positive constant K such that for each $\lambda \neq 0$ $x \mapsto \phi_\lambda(x)$ has variation bounded by K on \mathcal{R} . If $L(c)$ exists for X and some number c , then $L(\phi)$ exists and is equal to $L(c)$.*

Proof: Without loss of generality we assume that $c = 0$. Consider the following integral identities where $0 < M < M'$.

$$\begin{aligned} & \int_M^{M'} x(1 - \phi_\lambda(x)) P(dx) + \int_{-M'}^{-M} x(1 - \phi_\lambda(x)) P(dx) = \\ & \int_M^{M'} x \left(\int_0^x -\phi'_\lambda(t) dt \right) P(dx) + \int_{-M'}^{-M} x \left(\int_x^0 \phi'_\lambda(t) dt \right) P(dx) \\ & = \int_0^{M'} -\phi'_\lambda(t) \left(\int_{\max(t, M)}^{M'} x P(dx) \right) dt \\ & + \int_{-M'}^0 \phi'_\lambda(t) \left(\int_{-M'}^{\min(-M, t)} x P(dx) \right) dt = \\ & \int_0^{M'} -\phi'_\lambda(t) \left(\int_{\max(t, M)}^{M'} x P(dx) + \int_{-M'}^{-\max(M, t)} x P(dx) \right) dt \end{aligned} \tag{10}$$

Here we have used differentiability of ϕ_λ , integrability of ϕ'_λ , and Fubini's Theorem. In the last line of (10) we have made a change of variables from t to $-t$ and used the identity $\phi'_\lambda(-t) = -\phi'_\lambda(t)$.

The last line of (10) has absolute value less than or equal to:

$$\int_0^{M'} |\phi'_\lambda(t)| \left| \int_{\max(t,M)}^{M'} x P(dx) + \int_{-M'}^{-\max(M,t)} x P(dx) \right| dt. \quad (11)$$

Since $L(0)$ is finite the second expression in absolute values in this integral can be made smaller than $\epsilon/2K$ where ϵ is a given positive number ϵ for M sufficiently large with $M \leq t \leq M'$. But ϕ_λ has variation bounded by K . So the entire integral in (11) is smaller than $\epsilon/2$.

Let us also require that M be so large that

$$|L(0) - \int_{-M}^M x P(dx)| < \epsilon/6. \quad (12)$$

Now pick a positive number δ so that $|\lambda| < \delta$ implies that

$$\left| \int_{-M}^M x(1 - \phi_\lambda(x)) P(dx) \right| < \epsilon/6 \quad (13)$$

This can be done since for fixed M the integrand in (13) is bounded due to the bounded variation of ϕ_λ , and hence the Lebesgue Dominated Convergence Theorem implies that the integral inside the absolute value tends to 0 as λ tends to 0.

Finally given a nonzero λ in $(-\delta, \delta)$ choose M' larger than M so that

$$\left| \int_{-\infty}^{\infty} x \phi_\lambda(x) P(dx) - \int_{-M'}^{M'} x \phi_\lambda(x) P(dx) \right| < \epsilon/6. \quad (14)$$

Combining the estimate for the integrals in (10) resulting from (11) with (12), (13), and (14), we obtain that for $0 < |\lambda| < \delta$

$$|L(0) - \int_{-\infty}^{\infty} x \phi_\lambda(x) P(dx)| < \epsilon/2 + 3\epsilon/6 = \epsilon,$$

which establishes the theorem. ■

Theorem 4.1 indicates that some mollifiers are guaranteed to yield the same answer for the mean as our earlier techniques. One question, though, is whether they will sometimes give an answer when the previous methods do not.

A theorem of Abel asserts that a power series in z that converges at a point z_0 on the unit circle is absolutely convergent at each point in the open unit disk. Moreover, the analytic function $f(z)$ that it

converges to has the property that $\lim_{z \rightarrow z_0} f(z) = f(z_0)$ provided z tends to z_0 non-tangentially from the interior of the unit disk. In particular, $\lim_{r \rightarrow 1^-} \sum a_n r^n = \sum a_n$ provided the latter converges.

A well-known counterexample to the converse of Abel's Theorem is the series $\sum_{n \geq 0} (-1)^n$. This series diverges since its partial sums oscillate between 1 and -1 . However, $\sum_{n \geq 0} (-1)^n r^n$ converges to $1/(1+r)$ for $|r| < 1$, and, as r tends to 1^- , tends to $1/2$. The quantity r^n serves as a kind of mollifier for the original series and enables a kind of convergence.

Example 4.2 We can use the above to find a counterexample to the converse of Theorem 4.1, i. e., a case where $L(\phi)$ exists even though $L(c)$ fails to exist for all c . Let X be a random variable taking the values

$$x_n = n^3 + (-1)^{|n|} n^2$$

for each non-zero integer n with $P(X = x_n) = \frac{K}{n^2}$ where $\frac{1}{K} = 2(\sum_{\{n : n \geq 1\}} \frac{1}{n^2}) = \frac{\pi^2}{3}$.

Let $\theta : [0, \infty) \rightarrow [0, \infty)$ be defined in the following way: for each positive integer n , set $\theta(x) = n$ if $n^3 - n^2 \leq x \leq n^3 + n^2$, and let θ take the interval $[n^3 + n^2, (n+1)^3 - (n+1)^2]$ onto the interval $[n, n+1]$ in an increasing (and smooth) fashion. A consequence of this definition is that $\theta(|x_n|) = |n|$ for non-zero integers n .

Next we define $\phi_\lambda : \mathcal{R} \rightarrow [0, 1]$ by $\phi_\lambda(x) = e^{-|\lambda|(\theta(|x|)-1)}$ for x in \mathcal{R} . Evidently $\{\phi_\lambda\}$ satisfies conditions 2, 3, and 4 of the definition of a mollifier for X with center 0. Furthermore ϕ_λ can even be taken to be C^∞ (but not analytic!). To confirm that $x \mapsto x\phi_\lambda(x)$ is integrable with respect to the probability measure P associated with X , note that this measure is given by:

$$P = \sum_{n=1}^{\infty} \frac{K}{n^2} (\delta_{x_n} + \delta_{(x_{-n})}).$$

Then

$$\begin{aligned} \int_{\mathcal{R}} |x\phi_\lambda(x)| P(dx) &= K \sum_{n=1}^{\infty} \frac{e^{-|\lambda|(\theta(|x_n|)-1)} (|x_n| + |x_{-n}|)}{n^2} \leq \\ K \sum_{n=1}^{\infty} 2 \frac{n^3 + n^2}{n^2} r^{n-1} &= \frac{2K(2-r)}{(1-r)^2} < \infty \end{aligned}$$

where $r = e^{-|\lambda|}$ with $\lambda \neq 0$.

If we now try to calculate the improper integral $L(c)$ for X for some number c , we obtain:

$$\begin{aligned} \int_{c-M}^{c+M} x P(dx) &= \sum_{\{n : c-M \leq x_n \leq c+M, n \neq 0\}} \frac{Kx_n}{n^2} \\ &= K \left(\sum_{\{n : c-M \leq x_n \leq c+M, n \neq 0\}} n \right) \\ &+ K \left(\sum_{\{n : c-M \leq x_n \leq c+M, n \neq 0\}} (-1)^{|n|} \right). \end{aligned}$$

To see that $L(c)$ does not exist for any c , consider the following. The first sum in the last expression, for M sufficiently large, will be the sum of consecutive nonzero integers from n_0 to n_1 with $n_0 < 0 < n_1$ and the second sum will be $((-1)^{n_0} + (-1)^{n_0+1} + \dots + (-1)^{n_1} - (-1)^0)$ where the power $(-1)^0$ is added and then subtracted. The second sum will thus be either 0, -1 , or -2 , depending on the parity of n_0 and n_1 . The first sum will oscillate among 0, an increasingly large positive number, and an increasingly large negative number. Unless the first sum is eventually equal to 0 for all sufficiently large M , there is no way the combined sum can have a finite limit. But if the first sum is identically zero for large M , then $n_0 = -n_1$ and the second sum equals 0 or -2 depending on the parity of n_1 . If M yields one parity for n_1 a slight increase in M will reverse the parity. So the sum will oscillate between 0 and -2 .

Now consider the existence of $L(\phi)$ where we use the mollifier ϕ_λ just defined.

$$\begin{aligned} \int_{-\infty}^{\infty} x \phi_\lambda(x) P(dx) &= \sum_{\{n : n \neq 0\}} \frac{Kx_n r^{|n|-1}}{n^2} \\ &= K \left(\sum_{\{n : n \neq 0\}} nr^{|n|-1} + \sum_{\{n : n \neq 0\}} (-1)^{|n|} r^{|n|-1} \right) = \frac{-2K}{1+r}. \end{aligned}$$

As λ tends to 0, r tends to 1 and the above tends to $L(\phi) = -K$. Hence $L(\phi)$ exists even though $L(c)$ exists for no c .

5 Cumulative Distributions and Characteristic Functions

Two familiar alternative formulas for the ordinary mean are:

$$\int_{-\infty}^{\infty} (1 - F(t-) - F(-t)) dt \quad (15)$$

and

$$-i \frac{d}{dt} E(e^{itX}) \Big|_{t=0}. \quad (16)$$

Here $F(t) \stackrel{\text{def}}{=} P(X \leq t)$ for t in \mathcal{R} and $F(t-) = P(X < t) = \lim_{x \rightarrow t-} F(x)$. We review these formulas below and relate them to our notions of mean.

In the case of the cumulative distribution we have the following theorem.

Theorem 5.1 *Let X be a random variable with probability distribution P and cumulative distribution F . Then*

i) $E(X)$ exists if and only if $1-F(t-)$ and $F(-t)$ are in $L^1([0, \infty))$ with respect to Lebesgue measure. In this case

$$E(X) = \int_0^\infty (1 - F(t-) - F(-t)) dt.$$

ii) If $\lim_{M \rightarrow \infty} M(P(c+M < X) - P(c-M > X)) = 0$ for some real number c , then $L(c)$ exists if and only if

$\lim_{M \rightarrow \infty} \int_0^M (1 - F(t-) - F(-t)) dt$ exists. In this case $L(c)$ is given by the limit of the integral.

iii) If $L(c)$ and $\lim_{M \rightarrow \infty} \int_0^M (1 - F(t-) - F(-t)) dt$ both exist for some real number c and are equal, then

$$\lim_{M \rightarrow \infty} M(P(c+M < X) - P(c-M > X)) = 0.$$

Proof: In order for $E(X)$ to exist, it is necessary and sufficient that x be integrable with respect to P on $(-\infty, \infty)$. In particular this is equivalent to x being integrable with respect to P on both $[0, \infty)$ and $(-\infty, 0]$, with

$$\begin{aligned} \int_0^\infty x P(dx) &= \int_0^\infty \int_0^x 1 dt P(dx) = \\ \int_0^\infty P(X \geq t) dt &= \int_0^\infty (1 - F(t-)) dt \end{aligned}$$

and

$$\begin{aligned} \int_{-\infty}^0 x P(dx) &= - \int_{-\infty}^0 \int_x^0 1 dt P(dx) = \\ - \int_{-\infty}^0 P(t \geq X) dt &= - \int_0^\infty F(-t) dt \end{aligned}$$

where Fubini's Theorem [5, p. 386] has been applied. The statement of Fubini's Theorem implies that integrability of x occurs if and only if $1 - F(t-)$ and $F(-t)$ are both integrable.

Turning to ii) and iii) of Theorem 5.1, for large M consider

$$\begin{aligned} \int_{c-M}^{c+M} x P(dx) &= \int_0^{c+M} x P(dx) + \int_{c-M}^0 x P(dx) \\ &= \int_0^{c+M} \int_0^x 1 dt P(dx) - \int_{c-M}^0 \int_x^0 1 dt P(dx) \\ &= \int_0^{c+M} P(c+M \geq X \geq t) dt - \int_{c-M}^0 P(t \geq X \geq c-M) dt \quad (17) \end{aligned}$$

where again Fubini's Theorem has been used. This entire expression differs from

$$\int_0^{M_1} (1 - F(t-) - F(-t)) dt = \int_0^{M_1} (P(X \geq t) - P(-t \geq X)) dt,$$

where we have introduced a new variable M_1 , by the quantity

$$\begin{aligned} & - \int_0^{c+M} P(X > c+M) dt + \int_{c-M}^0 P(c-M > X) dt \\ & + \int_{M_1}^{c+M} P(X \geq t) dt - \int_{M_1}^{M-c} P(-t > X) dt = \\ & -MP(X > c+M) + MP(c-M > X) \\ & -cP(X > c+M) - cP(c-M > X) \\ & + \int_{M_1}^{c+M} P(X \geq t) dt - \int_{M_1}^{M-c} P(-t > X) dt. \end{aligned}$$

If we suppose that $\lim_{K \rightarrow \infty} \int_0^K (1 - F(t-) - F(-t)) dt$, then the combination of the last two terms tends to 0 as M and M_1 tend to ∞ since $\int_{M-c}^{M+c} P(t > X) dt$ tends to 0 as M gets large. Likewise the previous two terms with multiplier c tend to 0 as M gets large. So $L(c)$ exists and equals $\lim_{M_1 \rightarrow \infty} \int_0^{M_1} (1 - F(t-) - F(-t)) dt$ if and only if $\lim_{M \rightarrow \infty} M(P(X > c+M) - P(c-M > X)) = 0$.

Writing $M_1 = c + M_2$, we can express the difference described above as:

$$\begin{aligned} & \int_0^{c+M} P(c+M \geq X \geq t) dt - \int_{c-M}^0 P(t \geq X \geq c-M) dt \\ & - \left(\int_0^{c+M_2} P(X \geq t) dt - \int_{-c-M_2}^0 P(t \geq X) dt \right) = \\ & -M_2P(X > c+M_2) + M_2P(c-M_2 > X) \\ & -cP(X > c+M_2) + cP(c-M_2 > X) \\ & + \int_{-c-M_2}^{c-M_2} P(t \geq X \geq c-M_2) dt \end{aligned}$$

$$\begin{aligned}
 & + \int_0^{c+M} P(c+M \geq X \geq t) dt - \int_{c-M}^0 P(t \geq X \geq c-M) dt \\
 & - \left(\int_0^{c+M_2} P(c+M_2 \geq X \geq t) dt - \int_{c-M_2}^0 P(t \geq X \geq c-M_2) dt \right).
 \end{aligned}$$

As M and M_1 , and hence M_2 , tend to infinity, in case $L(c)$ exists, the last two lines tend to 0 (compare with (17)). The two lines above these also tend to 0. Accordingly $\int_0^{M_1} (1 - F(t-) - F(-t)) dt$ tends to $L(c)$ as M_1 tends to infinity if and only if $\lim_{M_2 \rightarrow \infty} M_2 P(X > c + M_2) - M_2 P(c - M_2 > X) = 0$. ■

Theorem 5.1 suggests the possibility that the difference between $1 - F(t-)$ and $F(-t)$ might be in $L^1([0, \infty))$ even though neither function individually is, or even more generally that the difference may not be integrable but still might have a finite improper integral on $[0, \infty)$. In these cases, provided

$$\lim_{M \rightarrow \infty} MP(c + M < X) - MP(c - M > X) = 0, \tag{18}$$

for some c , then X has a finite superweak mean given by (15). If, in addition to a finite improper integral, (18) holds for all c , then X has a doubly weak mean given by Theorem 5.1 part ii. If X has a weak mean, then the Kolmogorov condition $\lim_{n \rightarrow \infty} nP_X(|X| > n) = 0$ ensures that (18) holds for all c , and accordingly the weak mean is automatically given by (15).

Example 5.1 Example 2.2 can be reused again. Since the distribution in that example is symmetric about 0, $L(0)$ exists and equals 0, and $M(P_X(X > M) - P_X(-M > X)) \equiv 0$ for all $M \geq 0$. Hence

$$\lim_{M \rightarrow \infty} \int_0^M (1 - F(t-) - F(-t)) dt = 0.$$

Since $L(c)$ does not exist for any non-zero c , by Theorem 5.1 equation (18) cannot be true for any non-zero c .

Now let us consider the expression (16) based on the derivative of the characteristic function. A natural question is to ask when this derivative exists, i.e., when the limit as t tends to 0 exists for the expression

$$\int_{-\infty}^{\infty} \frac{e^{itx} - 1}{it} P_X(dx).$$

Note that $\frac{e^{itx} - 1}{it}$ is an integrable function of x with respect to P for t non-zero and real and that $|\frac{e^{itx} - 1}{it}| = |\int_0^x e^{itu} du| \leq |x|$ for

such t . Furthermore we can decompose the integral into a real and imaginary part arriving at:

$$\int_{-\infty}^{\infty} \frac{e^{itx} - 1}{it} P(dx) = \int_{-\infty}^{\infty} \frac{\sin tx}{t} P(dx) + i \int_{-\infty}^{\infty} \frac{1 - \cos tx}{t} P(dx).$$

The real part of this is precisely the expression one would obtain with the mollifier $\frac{\sin tx}{tx}$ applied to x and the imaginary part presumably under some circumstances will tend to 0 as t tends to 0. A more detailed look is permitted by the following expansion where we consider the integral from $-M$ to M instead of from $-\infty$ to ∞ and apply Fubini to get:

$$\begin{aligned} & \int_{-M}^M \frac{\sin tx}{t} P(dx) + i \int_{-M}^M \frac{1 - \cos tx}{t} P(dx) \\ &= \int_{-M}^M \int_0^x \cos tu \, du \, P(dx) + i \int_{-M}^M \int_0^x \sin tu \, du \, P(dx) \\ &= \int_0^M (\cos tu) (P(M \geq X \geq u) - P_X(-u \geq X \geq -M)) \, du \\ &+ i \int_0^M (\sin tu) (P(M \geq X \geq u) + P(-u \geq X \geq -M)) \, du. \quad (19) \end{aligned}$$

This discussion is summarized in the following theorem due to Pitman [13].

Theorem 5.2 *Let X be a random variable with probability distribution P and cumulative distribution F . Then X has a weak mean if and only if the derivative of the characteristic function of X exists at $t = 0$, in which case $E_w(X)$ is given by that derivative, i.e.,*

$$\begin{aligned} E_w(X) &= \lim_{t \rightarrow 0} \int_{-\infty}^{\infty} \frac{e^{itx} - 1}{it} P(dx) \\ &= \lim_{t \rightarrow 0} \int_{-\infty}^{\infty} \frac{\sin tx}{tx} x P(dx) \\ &= \lim_{t \rightarrow 0} \lim_{M \rightarrow \infty} \int_0^M (\cos tu) (P(M \geq X \geq u) - P(-u \geq X \geq -M)) \, du. \end{aligned}$$

Proof: See [13]. ■

An implication of Theorem 5.2 is that the imaginary parts of the expressions represented in (19) all go to 0 in the appropriate

limits. Note that the second expression for $E_w(X)$ in Theorem 5.2 is obtained by a mollifier.

6 Conclusion

The legitimacy and importance of the Cauchy distribution stem in part from the fact that it is the quotient of two independent standard normal random variables. It has application in physics under the name of the Lorentz distribution. Indeed long-tailed and counter-intuitive distributions are increasingly important in recent times in financial mathematics, the study of natural and man-made disasters, and computer network analysis (see Gumble [4], Resnick[15], or Taleb [17]).

The stable distributions, as discussed in [3] and [2], are the limits of sums of independent random variables and necessarily arise in limit processes of renewal theory and random walks. They also arise in signal processing [12].

If there is only one tail then the mean is often infinity. However, the considerations here may still apply to the surrogate variable $\log X$ where X has values between 0 and ∞ . The tails that result are often asymmetric, but it sometimes happen that $\log X$ is heavy-tailed at both ends. For this to happen it suffices if $P(X \leq t) = C_1/(\log(1/t))^\alpha$ with $0 < \alpha \leq 1$ for $0 < t \leq t_1 < 1$ and $P(X \geq t) = C_2/(\log t)^\beta$ with $1 \geq \beta > 0$ for $t \geq t_2 > 1$ with suitable positive constants C_1 and C_2 .

Extending the notion of mean to the Cauchy distribution, other stable distributions, surrogate variables and other heavy-tailed cases, and investigating these extensions, are ways of moving us beyond the standard framework into realms where engineers and scientists may benefit from mathematical insight.

References

- [1] WHITTLE, P., *Probability via Expectation*, Springer-Verlag, New York (1992).
- [2] SILVER, N., *President's Invited Address, JSM 2013, Montreal, Canada* (2013).
- [3] LEVITT, S., Thinking Differently About Big Data *Arthur M. Sackler Lecture, National Academy of Sciences, March 26, 2015* (2015).
- [4] KOSKO, B., The Sample Mean *The Edge: What have you changed your mind about? Why?*
URL: <http://www.edge.org/responses/what-have-you-changed-your-mind-about-why>, (2008).

- [5] FELLER, W.E., *An Introduction to Probability Theory and Its Applications*, Vol. II, John Wiley & Sons, Inc., New York (1966).
- [6] GNEDENKO, B.V. and KOLMOGOROV, A.N., *Limit Distributions for Sums of Independent Random Variables*, translated and annotated by K.L. Chung, Addison-Wesley, Cambridge MA (1954).
- [7] NIKIAS, C. L. and SHAO, M., *Signal Processing with Alpha-Stable Distributions and Applications*, John Wiley & Sons, New York (1995).
- [8] POLLARD, D., *A User's Guide to Measure Theoretic Probability*, Cambridge University Press, Cambridge, U. K. (2002).
- [9] HEWITT, E. and STROMBERG, K., *Real and Abstract Analysis*, Springer-Verlag, New York (1965).
- [10] KOLMOGOROV, A. N., Über die Summen durch den Zufall bestimmter unabhängiger Grössen. *Math. Ann.* **99** 309–319. (1928).
- [11] KOLMOGOROV, A. N., *Foundations of the Theory of Probability*, Chelsea Publishing Company, New York (1950).
- [12] KOLMOGOROV, A. N., Bemerkung zu meiner Arbeit “Über die Summen zufälliger Grössen.” *Math. Ann.* **102** 484–488 (1929).
- [13] MATHEWS, J. and WALKER, R. L., *Mathematical Methods of Physics*, Benjamin, New York (1970).
- [14] VOLKOV, I. I. *Summation methods*, Encyclopedia of Mathematics.
URL: http://www.encyclopediaofmath.org/index.php?title=Summation_methods&oldid=11209.
- [15] PITMAN, E. J. G., On the derivation of a characteristic function at the origin *Ann. Math. Stat.* **27** 1156–1160.(1956)
- [16] GUMBEL, E. J., *Statistics of Extremes*, Dover, Mineola, New York (2004).
- [17] RESNICK, S. I., *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Soringer Science Business Media LLC, New York (2007).
- [18] TALEB, N. N., *The Black Swan: The Impact of the Highly Improbable*, Random House, New York (2010).
- [19] DE FINETTI, B., *Probability, Induction, and Statistics*, Wiley, New York (1972).