SOME DISTRIBUTIONS OF SAMPLE MEANS

George W. Brown and John W. Tukey

RCA Laboratories and *Princeton University*

1. **Summary.** It is shown that certain monomials in normally distributed quantities have stable distributions with index 2^{-k}. This provides, for $k > 1$, simple examples where the mean of a sample has a distribution equivalent to that of a fixed, arbitrarily large multiple of a single observation. These examples include distributions symmetrical about zero, and positive distributions.

Using these examples, it is shown that any distribution with a very long tail (of average order $\geq x^{-3/2}$) has the distributions of its sample means grow flatter and flatter as the sample size increases. Thus the sample mean provides *less* information than a single value. Stronger results are proved for still longer tails.

2. **Introduction.** This paper derives and exploits certain elementary expressions for stable distributions. The practicing statistician may be interested in the general discussion of results, going as far as Section 5. The reader interested in probability theory may be interested in

(i) the simple monomials in normally distributed quantities which are shown to be stable (Section 7)

(ii) the resulting bounds on the densities of these stable distributions (Section 8)

(iii) Theorem A, which forms a partial converse to the Central Limit Theorem.

It should be pointed out that examples of stable chance quantities arising from infinite series (Khintchine 1937, [2], [3]) and integrals (Levy 1935, [4]) are already known. These results form a natural part of broader investigations into

(i) the relative value of the mean, the median, and their competitors

(ii) the properties and distributions of simple functions of normally distributed quantities.

3. **Stable distributions.** One of the typical properties of the normal distribution with zero mean is that the distribution of the mean of a sample of n has the same shape but is compressed by the factor $\sqrt{n}$. The Cauchy distribution is well-known for the property that the mean of a sample of n has the same distribution as a single observation.

Statisticians have not widely appreciated the fact that there are symmetric, smooth distributions for every positive $\lambda \leq 2$, with the property that the distribution of the mean of a sample of n has the same shape as the original distribution but is spread out in the ratio $n^{(1-\lambda)/\lambda}$. These are the symmetric stable distributions of index λ.

It is interesting to note that if $\lambda = .001$, then the mean of a sample of two is 2^{999} times as variable as the mean of a sample of one. For small λ the means become unduly variable with a rapidity which is difficult to comprehend.

4. **Outline of results.** Section 7 is devoted to the proof that certain monomials in normal variables are stable of index 2^{-k} for integral k. Both symmetrical and positive cases are shown to exist. For $k = 0$, the symmetrical case is the familiar Cauchy distribution, which is the distribution of Student's "t" on one degree of freedom, while the positive case for $k = 1$ is the distribution of Snedecor's "F" on ∞ and 1 degrees of freedom.

In Section 8 it is shown that the symmetrical stable distribution of index λ has a density which is

(i) bounded by a constant

(ii) bounded by a constant times $|x|^{-1-\lambda}$, for the values $\lambda = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \cdots$, for which elementary examples are available. It is conjectured that this is true for all $\lambda \leq 2$.

In section 9 it is shown that, if a distribution has one long tail in the sense that

$$\varliminf_{x \to \infty} |x|^{1+\lambda} P\{x < X \leq x + h\} > 0, \tag{1.1}$$

for some h and one of the above values of λ (the $\varliminf$ may be taken either as $x \to +\infty$ or as $x \to -\infty$), then the distribution of the sum of a sample of n spreads out as fast as for a stable distribution with the same value of λ. This may be restated for the mean as follows:

(i) *A distribution has a long tail of order* $|x|^{-(1+\lambda)}$ *if* (1.1) *holds for some* $h > 0$ *and choice of sign for* x.

(ii) *If the distribution has a density* $f(x)$, *then* (1.1) *is a consequence of*

$$f(x) \geq \frac{A}{1 + |x|^{1+\lambda}}, \qquad A > 0. \tag{1.2}$$

(iii) *The distribution of the mean of a sample of* n *will be said to spread out as fast as* n^k, *if the distance between any two percentage points for the mean of a sample of* n *is ultimately larger than a fixed multiple of* n^k.

(iv) THEOREM A. *If the distribution of* X *has at least one long tail of order* $|x|^{-(1+\lambda)}$, *where* $\lambda = 1, \frac{1}{2}, \frac{1}{4}, \cdots$, *then the distribution of the mean of a sample of* n *values of* X *spreads out as fast as* $n^{(1-\lambda)/\lambda}$.

Section 10 presents a simple example of a distribution symmetric about zero with such long tails that

(i) the distribution of the sample mean spreads out faster than any power of n,

(ii) the median of a sample of any size fails to have finite moments of positive order, integral or fractional.

5. **Consequences for applied statistics.** The basic consequences of these results for applied statistics can be summarized in the following statements.

(a) The positions that the Cauchy distribution is an isolated case, or else an extreme example of pathology, are now untenable.

(b) The use of the mean of a sample as a measure of location (or, when dealing with positive distributions fixed at zero, as a measure of scale) implies a belief that the tails of the underlying distribution are not too long.

(c) It is probable that the relative efficiencies of mean and median are greatly affected by the length of the tail.

The importance of this last statement lies in the fact that direct empirical evidence about tail length is very hard to obtain. The mean is well known to be more efficient when the underlying distribution is normal. Normality of the tails of practical distributions is rarely based on firm empirical evidence. In these practical cases, greater efficiency of the mean should often not be assumed without empirical confirmation.

It may be argued that the results of this paper apply to the limit as $n \to \infty$ and to the behavior of the distribution near infinity, while the practical problems involve moderate values of n and the behavior of the distribution near its 5%, 1%, 0.1%, 95%, 99%, and 99.9% points. This is undoubtedly true, but the authors believe, and have some evidence to confirm, the following correspondence principle:

If certain mathematical tails imply certain asymptotic behavior, then similar practical tails imply similar behavior in moderate samples.

Here "mathematical tails" refers to behavior at infinity while practical tails run from the 5% to the 0.1% point and from the 95% to the 99.9% point.

It is of some interest to point out that Snedecor's "F" provides applications of Theorem A. If N values of F are averaged, where each was obtained on n_1 and n_2 degrees of freedom, then as N increases

(i) if $n_2 > 2$, the average converges to 1 (i.e. all percent points converge to 1), by the Central Limit Theorem

(ii) if $n_2 = 2$, the percent points of the average stay a finite distance away from each other, by Theorem A

(iii) if $n_2 = 1$, the percent points of the average separate from each other at least as fast as a constant times $\sqrt{N}$, by Theorem A.

The consequences of Theorem A follow from the asymptotic density of F, which is a constant times $F^{-1-\frac{1}{2}n_2}$.

6. **Notation and terminology.** Chance quantities (random variables) will be denoted by capitals and their values by lower case letters. The same letter will generally be used, so that x will frequently be a value of X.

The letter S, with or without indices, represents a standard deviate (normally distributed quantity with zero mean and unit variance). Unless other wise specified all sets of chance quantities will be assumed to be independent.

Cumulative distribution functions will be referred to simply as "cumulatives" and will be denoted by capitals. Probability density functions will be referred to as "densities" and will be denoted by the corresponding lower case letters.

The convolution of two cumulatives F and G will be denoted by $F*G$. It is the cumulative of sums of two independent values, one from each distribution.

7. Special stable distributions. Cauchy (1853, [1]) recognized that distributions with characteristic functions of the form

$$e^{-\alpha|u|^\lambda}$$

were stable. A distribution is stable if whenever k and l are positive and A and B are independent chance quantities distributed according to the same law, then $kA + lB$ is distributed like a fixed multiple of A. It is known (Lévy 1937, [5], pp. 94 ff.) that any stable distribution has a characteristic function of the form

$$e^{-(\alpha+i\beta \operatorname{sgn} u)|u|^\lambda},$$

where $0 < \lambda \leq 2$, $\alpha > 0$, and $|\beta| \leq |\alpha \tan \frac{1}{2}\pi\lambda|$. Each stable distribution thus has an index λ such that $kA + lB$ and $(k^\lambda + l^\lambda)^{1/\lambda}A$ have the same distribution when A and B are a sample of two from the given distribution.

This section exhibits, for every integral k, simple monomials of standard deviates which have stable distributions of index 2^{-k}.

(7.1) THEOREM: *Let S, S_0, S_1, S_2, $\cdots$ be a sequence of independent standard deviates. Then*

(i) $C_0 = S/S_0$ *and* $P_0 = 1$
are stable of index $1 = 2^{-0}$.

(ii) $C_1 = S/S_0S_1^2 = C_0/S_2^2$ *and* $P_1 = 1/S_1^2 = P_0/S_1^2$
are stable of index $\frac{1}{2} = 2^{-1}$.

(iii) $C_2 = S/S_0S_1^2S_2^{2^2} = C_1/S_2^{2^2}$
and $P_2 = 1/S_1^2S_2^{2^2} = P_1/S_2^{2^2}$
are stable of index $\frac{1}{4} = 2^{-2}$.

(iv) *in general,* $C_k = C_{k-1}/S_k^{2k}$ *and* $P_k = P_{k-1}/S_k^{2k}$
are stable of index 2^{-k}.

The C_k are a sequence of symmetrically distributed chance quantities which are here presented as monomials in normally distributed chance quantities and whose stability properties imply for $k \geq 1$ that the distributions of means of samples spread out as the sample size increases. The P_k are a similar sequence, all of whose values are positive.

The stability properties of the C_k follow, directly, by means of elementary composition properties of characteritic functions, from

(7.2) LEMMA: *The characteristic function of C_k is*

$$E(e^{itC_k}) = \exp(-2\,|\tfrac{1}{2}t|^{2^{-k}}).$$

PROOF: The case $k = 0$ is the familiar Cauchy distribution. Denoting the normal cumulative by $N(s)$, it is seen that

$$\begin{aligned} E(e^{itC_0}) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{its/s_0}\, dN(s)\, dN(s_0) \\ &= \int_{-\infty}^{\infty} \exp\,(-\tfrac{1}{2}t^2/s_0^2)\, dN(s_0) \\ &= e^{-|t|}. \end{aligned}$$

The second definite integral is well known (e.g. Formula 495 in B. O. Pierce's table). Assuming the result for $k-1$, write

$$E(e^{itC_k}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\,(itC_{k-1}/s_k^{2^k})\, dF_{k-1}(C_{k-1})\, dN(s_k)$$
$$= \int_{-\infty}^{\infty} \exp\,(-2\,|\,\tfrac{1}{2}t\,|^{2^{-k+1}}/s_k^2)\, dN(s_k)$$
$$= \exp\,(-2\,|\,\tfrac{1}{2}t\,|^{2^{-k}}),$$

precisely as in the derivation for $k = 0$.

The stability properties of the P_k follow, by completely analogous use of the moment generating function, from

(7.3) LEMMA: *The moment generating function of P_k is*

$$E(e^{-tP_k}) = \exp(-2(\tfrac{1}{2}t)^{2^{-k}}),\ t \geq 0.$$

PROOF: The trivial case $k = 0$ is verified directly, since $P_0 \equiv 1$. The induction from $k-1$ to k is identical with the derivation of (7.2), as is seen by writing

$$E(e^{-tP_k}) = \int_{-\infty}^{\infty}\int_{0}^{\infty} \exp\,(-tP_{k-1}/s_k^{2^k})\, dG_{k-1}(P_k)\, dN(s_k)$$
$$= \int_{-\infty}^{\infty} \exp\,(-2(\tfrac{1}{2}t)^{2^{-k+1}}/s_k^2)\, dN(s_k)$$
$$= \exp\,(-2(\tfrac{1}{2}t)^{2^{-k}})\,.$$

In order to verify the stability properties, consider distributions with characteristic functions of the form $exp(-d\,|\,t\,|^{\lambda})$. If A and B are independently distributed according to this distribution, then

$$E(e^{it(lA+mB)}) \quad = \quad E(e^{itlA})E(e^{itmB}) \quad = \quad e^{-l(l^{\lambda}+m^{\lambda})|t|^{\lambda}}$$

for $l,\ m \geq 0$. Parallel application of the moment generating function yields precisely analogous results.

8. **Some auxiliary results.** It is the purpose of this section to establish some results concerning stable distributions. It will be convenient to state and prove some of these lemmas in general form.

(8.1) LEMMA: *If X has a density $f(x)$ satisfying*

$$f(x) \leq A\,|\,x\,|^{-\alpha},$$

then X has finite negative moments of orders down to $-(1-\alpha)$.

PROOF: If $-(1-\alpha) < \beta < 0$, then

$$|\,x\,|^{\beta}f(x) \leq A\,|\,x\,|^{-\alpha+\beta},$$

with $-\alpha+\beta > -1$. Now

$$\int_{-\infty}^{\infty} |\,x\,|^{\beta}f(x)\,dx \leq \int_{-\infty}^{-1} f(x)\,dx + \int_{-1}^{1} |\,x\,|^{\beta} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx$$
$$\leq \int_{-\infty}^{\infty} f(x)\,dx + \int_{-1}^{1} A\,|\,x\,|^{-\alpha+\beta}\,dx < \infty,$$

which proves the lemma.

(8.2) LEMMA: *If X has a density $f(x)$ satisfying*

$$f(x) \leq A \mid x \mid^{-\alpha}$$

and if Y has a density $g(y)$ and a finite negative moment of order $-(1-\alpha)$, then the density $h(x)$ of XY satisfies

$$h(x) \leq A_1 \mid x \mid^{-\alpha}.$$

PROOF: The density $h(x)$ satisfies

$$\begin{aligned} h(x) &= \int_{-\infty}^{\infty} \{f(x/t)g(t)/\mid t \mid\} \, dt \\ &\leq \int_{-\infty}^{\infty} A \mid t \mid^{\alpha} \mid x \mid^{-\alpha} g(t) \mid t \mid^{-1} dt \\ &= \left\{ \int_{-\infty}^{\infty} A \mid t \mid^{-(1-\alpha)} g(t) \, dt \right\} \mid x \mid^{-\alpha} = A_1 \mid x \mid^{-\alpha}. \end{aligned}$$

(8.3) LEMMA: *The density $h_k(y)$ of*

$$Y_k = S(S_1)^2(S_2)^{2^2} \cdots (S_k)^{2^k},$$

where, $S, S_1, S_2, \cdots S_k$ are independent standard deviates, satisfies

$$h_k(y) \leq A \mid y \mid^{-1+2^{-k}},$$

and hence Y_k has finite negative moments of all orders down to -2^{-k}.
PROOF: Let $g_k(x)$ be the density of

$$X_k = (S_k)^{2^k},$$

then

$$g_k(x) = (2\pi)^{-\frac{1}{2}} 2^{-k} \exp(-\tfrac{1}{2} x^{2^{1-k}}) x^{-1+2^{-k}},$$

whence

$$g_k(x) \leq A_1 \mid x \mid^{-1+2^{-k}}.$$

For $k = 0$ this is the desired result; the other cases follow by induction, using $Y_k = X_k Y_{k-1}$ and lemma (8.2). The final statement of the lemma then follows from lemma (8.1).
(8.4) THEOREM: *For $\lambda = 2^{-k}$, the density $m_\lambda(x)$ of C_k satisfies*

$$(*) \quad m_\lambda(x) \leq A \mid x \mid^{-(1+2^{-k})} = A \mid x \mid^{-(1+\lambda)},$$

and also

$$(**) \quad m_\lambda(x) \leq A_2 .$$

PROOF: By definition, $C_k = S/Y_k$. By lemma (8.3) the density of Y_k satisfies

$$h_k(y) \leq A_1 \mid y \mid^{-1+2^{-k}}.$$

The density of $1/Y_k$ satisfies

$$\begin{aligned} l_k(z) &= \frac{1}{z^2} h_k(1/z) \\ &\leq |z|^{-2} A_1 |z|^{1-2^{-k}} = A_1 |z|^{-(1+2^{-k})}. \end{aligned}$$

Since S has a finite moment of order 2^{-k}, it follows from lemma (8.2) that the density of S/Y_k satisfies the desired relation (*). Since S has finite moments of all positive orders, so does S^{2^k} and therefore Y_k. Thus $1/Y_k$ has moments of all negative orders, including -1. Since the density of S is bounded, lemma (8.2) implies the same for S/Y_k and hence for C_k. This completes the proof of the theorem.

9. **Distributions with a long tail.** The purpose of this section is to prove

(9.1) THEOREM: *If D has a cumulative $F(x)$ such that for some $h > 0$, either*

$$\varliminf_{x \to +\infty} \frac{F(x+h) - F(x)}{|x|^{-(1+\lambda)}} > 0, \quad or \quad \varliminf_{x \to -\infty} \frac{F(x+h) - F(x)}{|x|^{-(1+\lambda)}} > 0,$$

where $\lambda = 2^{-k}$ for $k = 0, 1, 2, \cdots$, and if $k_n(\alpha)$ is the α-point (100α percent point) of the distribution of sums of n independent values of D, then

$$\varliminf_{n} \frac{K_n(\alpha_1) - K_n(\alpha_2)}{n^{1/\lambda}} > 0,$$

whenever $\alpha_1 > \alpha_2$.

We begin with some lemmas.

(9.2) LEMMA: *If*

$$\left.\begin{aligned} F(x) &= \beta F'(x) + (1-\beta)F''(x), \\ G(x) &= \beta F'(x) + (1-\beta)\mathbf{1}(x), \end{aligned}\right\} \qquad 0 \leq \beta \leq 1$$

where $F'(x)$ is a cumulative symmetric about zero and unimodal, $F''(x)$ is a cumulative symmetric about zero, and $\mathbf{1}(x)$ is the cumulative concentrated at zero (whence $F(x)$ and $G(x)$ are cumulatives), and if $F_n(x)$ and $G_n(x)$ are the cumulatives of sums of samples of n from $F(x)$ and $G(x)$ respectively, then

$$\begin{aligned} F_n(x) &\leq G_n(x), \quad x > 0, \\ F_n(x) &\geq G_n(x), \quad x < 0. \end{aligned}$$

PROOF: We begin with the case $n = 2$, where

$$F_2 = \beta^2 F'*F' + 2\beta(1-\beta)F'*F'' + (1-\beta)^2 F''*F''$$

and

$$G_2 = \beta^2 F'*F' + 2\beta(1-\beta)F' + (1-\beta)^2\mathbf{1}.$$

The lemma will have been proved for $n = 2$ if we can show that

$$\begin{aligned} F'*F''(x) &\leq F'(x), x > 0, \\ F'*F''(x) &\geq F'(x), x < 0. \end{aligned}$$

Now, if $x > 0$,

$$\begin{aligned} F'*F''(x) &= \int_{-\infty}^{\infty} F'(x - s)\, dF''(s) \\ &= \int_0^{\infty} \{F'(x - s) + F'(x + s)\}\, dF''(s) \\ &\leq 2 \int_0^{\infty} F'(x)\, dF''(s) = F'(x), \end{aligned}$$

where the first equality follows from the symmetry of F', the inequality follows from the unimodality of F', and the last equality follows from the symmetry of F''. The inequality is reversed if $x < 0$.

For general n,

$$F_n = \sum_k \binom{n}{k} \beta^k (1 - \beta)^{n-k} F'_k * F''_{n-k},$$

$$G_n = \sum_k \binom{n}{k} \beta^k (1 - \beta)^{n-k} F'_k,$$

where F'_k (the convolution of k copies of F') is the cumulative for sums of k independent values from F', and F''_k is similarly related to F''. Since F'_k is unimodal and symmetric and since F''_{n-k} is symmetric, the same argument can be applied term by term to complete the proof of the lemma. The requirement that F'' be symmetric could be replaced by the formally weaker condition that $F''_k(0) = \frac{1}{2}$ for all k.

(9.3) Lemma: *If*

$$F(x) = \beta F_{(\lambda)}(x) + (1 - \beta)\mathbf{1}(x), \qquad 0 \leq \beta \leq 1,$$

where $F_{(\lambda)}(x)$ *is the cumulative of* C_k, *with* $\lambda = 2^{-k}$, *and if* $K_n(\alpha)$ *is as defined in* (9.1), *then*

$$\lim n^{-1/\lambda} K_n(\alpha) = \beta^{1/\lambda} K_{(\lambda)}(\alpha),$$

where $K_{(\lambda)}(\alpha)$ *is the* α*-point for* $F_{(\lambda)}(x)$.

Proof: Let F_n and $F_{(\lambda)n}$ be the cumulatives of sums of n from F and $F_{(\lambda)}$ respectively, whence

$$F_{(\lambda)n}(x) = F_{(\lambda)}(n^{1/\lambda} x).$$

Then

$$\begin{aligned} F_n(x) &= \sum_k \binom{n}{k} \beta^k (1 - \beta)^{n-k} F_{(\lambda)k}(x) \\ &= \sum_k \binom{n}{k} \beta^k (1 - \beta)^{n-k} F_{(\lambda)}(k^{1/\lambda} x). \end{aligned}$$

The characteristic function of $(n\beta)^{1/\lambda}x$ is

$$E(e^{it(n\beta)^{1/\lambda}x}) = \sum_k \binom{n}{k} \beta^k (1 - \beta)^{n-k} \exp(- d \mid (n\beta)^{-1/\lambda} k^{1/\lambda} t \mid^\lambda,$$

$$= \sum_k \binom{n}{k} \beta^k (1 - \beta)^{n-k} \exp\left(- \frac{dk}{n\beta} \mid t \mid^\lambda \right),$$

where the characteristic function associated with $F_{(\lambda)}(x)$ is $exp(-d \mid t \mid^\lambda)$. Thus we have to deal with

$$\exp\left(- \frac{dk}{n\beta} \mid t \mid^\lambda \right)$$

where k has a binomial distribution with mean $n\beta$ and variance $n\beta(1 - \beta)$, so that $k/n\beta$ converges stochastically to unity. This implies that

$$\lim E(e^{it(n\beta)^{1/\lambda}x}) = e^{-d|t|^\lambda}$$

uniformly in every finite interval, whence $(n\beta)^{1/\lambda}X$ converges stochastically to C_k, which completes the proof of the lemma.

(9.4) LEMMA: *If the symmetric cumulative $F(x)$ has a density $f(x)$, and if constants c_1 and c_2 exist such that*

$$f(x) \geq \min (c_1, c_2 \mid x \mid^{-(1+\lambda)}),$$

where $\lambda = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \cdots$, *then, if* $\alpha \neq \frac{1}{2}$,

$$\underline{\lim} \mid n^{-1/\lambda} K_n(\alpha) \mid > 0.$$

PROOF: According to theorem (8.4) there are constants d_1 and d_2 such that the density of C_k is bounded by $\min (d_1, d_2 \mid x \mid^{-(1+\lambda)})$. Hence

$$\frac{F(x) - \beta F_{(\lambda)}(x)}{1 - \beta}$$

is monotone when $\beta = \min (c_1/d_1, c_2/d_2)$, and hence is a distribution function. By lemma (9.2) the α-points of F lie outside those of $\beta F_{(\lambda)}(x) + (1 - \beta)\mathbf{1}(x)$, and these, by lemma (9.3), increase at least as fast as $An^{-1/\lambda}$.

(9.5) LEMMA: *If the density of D exists and equals $f(x)$, and if either*

$$\underline{\lim}_{x \to +\infty} \mid x \mid^{1+\lambda} f(x) > 0,$$

or

$$\underline{\lim}_{x \to -\infty} \mid x \mid^{1+\lambda} f(x) > 0,$$

where $\lambda = 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \cdots$, *then, for* $\alpha_1 > \alpha_2$,

$$\underline{\lim}_{n \to \infty} n^{-1/\lambda} \{K_n(\alpha_1) - K_n(\alpha_2)\} > 0.$$

PROOF: Let D_1 and D_2 be independent with the distribution of D. Then $D_1 - D_2$ has a symmetric density given by

$$g(x) = \int_{-\infty}^{\infty} f(x + s)f(s)ds.$$

If

$$\underline{\lim}_{x \to +\infty} |x|^{1+\lambda} f(x) > 0,$$

then for suitable h and $\epsilon > 0$,

$$f(x) \geq \epsilon |x|^{-(1+\lambda)}, \text{ for all } x \geq h.$$

Therefore, for $x \geq 0$, writing $\gamma = -(1 + \lambda)$,

$$g(x) \geq \int_h^{h+1} f(x + s)f(s)ds \geq \epsilon^2 |h + 1 + x|^{\gamma} |h + 1|^{\gamma} = b_1 |b_2 + x|^{\gamma}.$$

Now

$$b_1 |b_2 + x|^{\gamma} \geq \min \{b_1 2^{\gamma} b_2^{\gamma}, b_1 2^{\gamma} |x|^{\gamma}\}$$

and hence, for $x \geq 0$ and suitable $c_1 > 0$, $c_2 > 0$,

$$g(x) \geq \min \{c_1, c_2 |x|^{\gamma}\}.$$

Since $g(x)$ is symmetric, this is also true for $x < 0$. If

$$\underline{\lim}_{x \to -\infty} |x|^{1+\lambda} f(x) > 0,$$

then a similar argument proves the same result.

Let $K_{\delta_n}(\alpha)$ be the α-point for the sum of n values of $D_1 - D_2$ and $K_n(\alpha)$ be the α-point for the sum of n values of D. The most elementary relation between these functions is

$$| K_{\delta_n}(\tfrac{1}{2} \pm \tfrac{1}{2}(\alpha_1 - A_2)^2) | \leq | K_n(\alpha_1) - K_n(\alpha_2) |.$$

To see this, observe that the sum of a sample of n values of $D_1 - D_2$ is the difference of the sums of two independent samples of n values of D, and that there is a probability of $(\alpha_1 - \alpha_2)^2$ that both of these sums will fall between $K_n(\alpha_1)$ and $K_n(\alpha_2)$. Thus the intervals $(-| K_n(\alpha_1) - K_n(\alpha_2) |, 0)$ and $(0, | K_n(\alpha_1) - K_n(\alpha_2) |)$ are each occupied by the difference with probability $\geq \frac{1}{2}(\alpha_1) - \alpha_2)^2$. Since $K_{\delta_n}(\frac{1}{2}) = 0$, the relation follows. Hence, if $\alpha_1 > \alpha_2$,

$$\underline{\lim}\, n^{-1/\lambda}\{K_n(\alpha_1) - K_n(\alpha_2)\} \geq \underline{\lim}\, n^{-1/\lambda} K_{\delta_n}(\tfrac{1}{2} \pm \tfrac{1}{2}(\alpha_1 - \alpha_2)^2)$$

and by lemma (9.4) applied to the distribution of $D_1 - D_2$ this latter $\underline{\lim}$ is positive, which completes the proof of the lemma.

With the ground prepared, it is now possible to complete the

PROOF OF THE THEOREM: Let h be chosen so that

$$\underline{\lim}_{x \to +\infty} |x|^{1+\lambda} (F(x + h) - F(x)) > 0.$$

This can always be done, if X is replaced by $-X$ when necessary. Let U have the uniform distribution on the interval (0, 1) and consider the variable $D + hU$. This variable has a density given by

$$g(x) = \frac{F(x + h) - F(x)}{h},$$

and, therefore,

$$\underline{\lim}_{x \to +\infty} |x|^{1+\lambda} g(x) > 0.$$

Let $K_n(\alpha)$ be the α-point for the sum of a sample of n values of D, and let $K_n^*(\alpha)$ be the α-point for the sum of a sample of n values of $D + hU$. Since $|hU| \leq h$, it follows that

$$|K_n(\alpha) - K_n^*(\alpha)| \leq nh.$$

Therefore, if $1/\lambda > 1$ and $\alpha_1 > \alpha_2$,

$$\underline{\lim}_{x \to \infty} n^{-1/\lambda} \{K_n(\alpha_1) - K_n(\alpha_2)\} = \underline{\lim}_{x \to \infty} n^{-1/\lambda} \{K_n^*(\alpha_1) - K_n^*(\alpha_2)\},$$

and by lemma (9.5) the latter $\underline{\lim}$ is positive.

The case of $\lambda = 1$ requires a slightly more delicate argument. The sum of a sample of n values of hU is asymptotically normally distributed, and hence it is less than $A_\beta n^{\frac{1}{2}}$, for a suitable A_β, with probability β. Therefore

$$K_n(\alpha\beta) \leq K_n^*(\alpha\beta) \leq K_n(\alpha) + A_\alpha n^{\frac{1}{2}}$$

and the same process yields the desired conclusion.

10. **A distribution with very long tails.** A somewhat pathological example is provided by the symmetric cumulative

$$F(x) = \frac{1}{ln(e^2 + |x|)}, \qquad x \leq 0,$$

$$F(x) = 1 - \frac{1}{ln(e^2 + |x|)}, \ x \geq 0,$$

which has the density

$$f(x) = \frac{1}{(e^2 + |x|)\{ln(e^2 + |x|)\}^2}.$$

Since

$$\lim_{x \to \infty} |x|^{1+\lambda} f(x) = \infty \quad \text{for all} \quad \lambda > 0,$$

it follows from theorem (9.1) that the distribution of the sum of a sample of n values of X spreads out faster than any power of n. The same must therefore be true of the mean of a sample of n. There is clearly no use in taking any kind of mean of such a sample.

There will, of course, be something to gain by taking the median of a sample of $2n + 1$, since the distribution of the median always shrinks together as $n \to \infty$, and whenever, as is true here, the density is finite and continuous at the population median, the distributions of the sample medians shrink toward the population median.

This does not prevent some pathology, however, since the cumulative for the median of $2n + 1$ takes the form

$$\frac{(2n+1)!}{(n!)^2(n+1)} \{F(x)\}^{n+1}\{1 + P(F(x))\},$$

where $P(t)$ is a polynomial of degree n with no constant term. Thus, for large negative values of x, the cumulative for the median is asymptotically

$$\frac{(2n+1)!}{(n!)^2(n+1)} \cdot \frac{1}{\{ln(e^2 + |x|)\}^n}$$

and the corresponding density is asymptotically

$$\frac{(2n+1)!\,n}{(n!)^2(n+1)\{ln(e^2 + |x|)\}^{n+1}(e^2 + |x|)}$$

and it follows that the median has no moments of any positive order, integral or fractional. This is true no matter how large the sample used!

REFERENCES

[1] A. L. CAUCHY, "Sur les résultats moyens d'observations de même nature, et sur les résultats les plus probables," *Comptes Rendus*, Vol. 37 (1853), pp. 197 ff.

[2] A. KHINTCHINE, "Exemples des variables aléatoires obéissant à des lois stables," *Bull. Univ. Etat a Moscou*, Sect. A Mathematique et Mecanique, Vol. 1 (1937), Fasc. 5, pp. 6–8, (*Math. Zbl.*, Vol. 18 (1938), pp. 317–318.).

[3] A. KHINTCHINE, "Zur Theorie der unbeschränkt teilbaren Verteilungsgesetze," *Recueil Mathematique* (*Matematiceski Sbornik*), N.S. Vol. 2 (1937), pp. 79–119 (in particular chapter 3).

[4] P. LÉVY, "Propriétés asymptotiques des sommes de variables aléatoires indépendantes ou enchainées." *Jour. Math. Pures et Appl.*, Vol. 100 (Vol. 14 of 9th series) (1935), pp. 347–402.

[5] PAUL LÉVY, *Théorie de l'Addition des Variables Aléatoires*, Gauthier Villars, Paris, 1937.