PERGAMON

Mathematical and Computer Modelling 34 (2001) 921–936

MATHEMATICAL AND COMPUTER MODELLING

www.elsevier.nl/locate/mcm

The Mean and Median Absolute Deviations

T. PHAM-GIA
Département de Mathématiques et de Statistique
Université de Moncton, Moncton, New Brunswick, Canada E1A 3E9

T. L. HUNG
Hue University, Viet Nam

(Received and accepted May 2000)

Abstract—In this article, we present a survey of important results related to the mean and median absolute deviations of a distribution, both denoted by MAD in the statistical modelling literature and hence creating some confusion. Some up-to-date published results, and some original ones of our own, are also included, along with discussions on several controversial issues.

Keywords—Mean absolute deviation, Median absolute deviation, Standard deviation, Sampling distributions, Contamination, Robustness, Estimation, Skewness, Gini index, Peters formula, Asymptotics.

1. INTRODUCTION

The L^2-norm is very well present in the probabilistic and statistical literatures, most evidently in the basic measure of dispersion (variance) in statistics: the standard deviation of a real random variable X with finite mean

$$\mu = \int_{-\infty}^{\infty} x\,dF(x), \quad \sigma(X) = \|X-\mu\|_2 = \left[E(X-\mu)^2\right]^{1/2} = \left(\int_{-\infty}^{\infty} (x-\mu)^2\,dF(x)\right)^{1/2}.$$

Also, the validity of the orthogonal relation $\|X+Y\|_2^2 = \|X\|_2^2 + \|Y\|_2^2$, for X and Y independent centered variables, is the prime reason for the widespread use of this norm in sampling theory and in an important topic of statistics: the analysis of variance. Due to its convenience in differentiation, and hence in optimization, the L^2-norm is also preponderant in estimation, for example in the method of least squares, and for the same reason the square-error loss function is widely used in statistical decision theory. This seemingly very useful character of the L^2-norm has several drawbacks however, for example, when the design ceases to be orthogonal in the analysis of variance.

On the other hand, although playing a dominant role in mathematical functional analysis, the L^1-norm has seen relatively few applications in statistics and statistical modelling. Here, the

Research partially supported by NSERC Grant 41-505 (Canada). The authors wish to warmly thank several colleagues that have helped to improve the content and the presentation of this paper. Thanks are also due to Professor N. Turkkan for setting up several complex computer programs.

0895-7177/01/$ - see front matter Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX
PII: S0895-7177(01)00109-1

presence of the absolute value complicate computations and difficulties encountered in deriving sampling distributions with this norm have further reduced its role.

Recent research efforts on robust statistical modelling and inference have given the L^1-norm, and the median of a distribution, some renewed popularity (see, e.g., articles in [1]). First, in curve fitting and regression, the L^1-norm, when applied to a discrete set of points, leads to the minimization of a sum of absolute distances, in an approach known under a variety of names, the most frequently used ones being the least absolute deviation (LAD) or the least absolute errors (LAE). A ground breaking result by Basset and Koenker [2] states that, for linear models, the LAE estimator has strictly smaller asymptotic confidence ellipsoids than the least squares estimator, for any distribution for which the sample median is a more efficient location than the sample mean. Reference [3] gives a comprehensive and fairly up-to-date coverage of topics related to this domain, and Portnoy and Koenker [4] give a comparative overview of both L^1 and L^2 approaches. On the other hand, Bai *et al.* [5] obtained some encouraging results on analysis of variance based on LAD.

Second, for a continuous or discrete random variable, the counterpart of the standard deviation in the L^1-norm is the mean absolute deviation (MAD), denoted by $\delta_1(X) = \|X-\mu\|_1 = \int_{-\infty}^{\infty} |x-\mu|\, dF(x) = E(|X-\mu|)$. The same concept can be used when Md, the median of the distribution, is considered, instead of μ, and we have the mean absolute deviation about the population median, or $\delta_2(X) = E(|X - Md|)$, which has several applications in statistics and other related domains.

Third, while the above measures are still based on expectations (or "averages"), another concept based completely on the median (or "middle value") gives rise to the median absolute deviation, also denoted by MAD. In this article, we set $\lambda(X) = \text{Median}\,(|X - Md(X)|)$, and λ has applications in a growing number of domains, mostly in issues related to robustness. There is considerable confusion in the literature on the abbreviation MAD, with several distinct statistical measures all called MAD. In this article, we will focus on the above three dispersion measures, and their related sampling statistics, and will use a distinct notation and a distinct name for each of them for the study of their behaviours and applications.

In Section 2, we review the main properties of the mean absolute deviation, and in Section 3 basic results on the sampling distribution of d_n, the sample mean absolute deviation about the sample mean, are presented, together with a comparative study on the advantages in using S_n and d_n. In Section 4, we look into the estimation of the population mean absolute deviations, from both point and interval viewpoints. Some uses of the mean absolute deviations, in statistics, in experimental physics, and in other domains are presented in Section 5. In Section 6, we look into the asymptotic expressions of some statistics associated with different mean absolute deviations. The dispersion function, as a generalization of the mean absolute deviation, is presented in Section 7, and finally, Section 8 deals with symmetrization of a random variable and the median absolute deviation.

2. THE MEAN ABSOLUTE DEVIATIONS

2.1. Inequalities

A basic inequality in probability theory states that if X is a random variable and g is a nonnegative Borel function on R, which is even and nondecreasing on $(0,\infty)$, then for $a > 0$, we have [6]

$$\frac{Eg(X) - g(a)}{\text{a.s.}\sup g(X)} \leq P(|X| \geq a) \leq E\left(\frac{g(X)}{g(a)}\right)$$

(a.s. means almost surely). Considering $X - c$, where c is a constant, instead of X, and taking $g(x) = x^r$, where r is a positive integer, we have the classical Markov inequality $P(|X-c| \geq \epsilon) \leq E(|X-c|^r)/\epsilon^r$. For $c = \mu$, $r = 2$, and $\epsilon = kV_r^{1/r}$, where V_r is the r^{th} absolute moment of X, i.e.,

$V_r = E(|X - \mu|^r)$, we obtain the Pearson inequality [7, p. 110]

$$P\left(|X - \mu| \geq kV_r^{1/r}\right) \leq \frac{1}{k^r}.$$

For $r = 2$, we then have the well-known Chebyshev inequality

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}, \qquad \text{or} \qquad P(|X - \mu| \geq k\|X - \mu\|_2) \leq \frac{1}{k^2}. \tag{1}$$

For $c = \mu$, $r = 1$, and $\epsilon = k\delta_1(X)$, we have the lesser-known case concerning the L^1-norm

$$P(|X - \mu| \geq k\delta_1(X)) < \frac{1}{k}, \qquad \text{where } \delta_1 = \delta_1(X) = E(|X - \mu|) = \|X - \mu\|_1$$

is called the mean absolute deviation (MAD) of X (about its mean).

Similarly, if Md is the median of the distribution of X, we have $P(|X - Md| \geq k\delta_2(X)) < 1/k$, where $\delta_2 = \delta_2(X) = \|X - Md\|_1$ is the mean absolute deviation of X about its median. Since the median minimizes the average absolute distance [8, p. 232], we have $\delta_2 \leq \delta_1$. By Lyapunov's inequality [9, p. 103], we also have $\delta_1 < \sigma$ and hence, we always have $\delta_2 \leq \delta_1 \leq \sigma$ for any distribution.

2.2. The Basic Meaning of the Mean Absolute Deviations

We can see that δ_1 offers a direct measure of the dispersion of X about its mean, of the first degree in $|X|$, unlike the standard deviation which, according to [7, p. 54], is obtained by using "*the device of squaring and then taking the square root of the resultant sum, which may appear a little artificial*".

We have $\delta_1(X + a) = \delta_1(X)$ and $\delta_1(aX) = |a|\delta_1(X)$ for any real a. Interpretationwise, δ_1 can be easily seen to be "the average of the variation of X about its mean, irrespective of the sign", while the practical meaning of the standard deviation remains unclear, as the quadratic mean of the variation of X about μ. For X uniform on $[0, 1]$, δ_1 has the very natural value of $1/4$ whereas $\sigma = 1/(2\sqrt{3})$, and does not lend itself to any interpretation. Furthermore, we then have $P(|X - \mu| < \delta_1) = 1/2$, a very logical equality, while $P(|X - \mu| < \sigma) = 1/\sqrt{3}$ is, again, difficult to interpret.

The first equality can be intuitively understood as follows. For the uniform distribution, there is 50% probability that X is within its mean of the "average distance" about that same mean.

For most common distributions, δ_1 can be obtained in closed form [10, Table 1]. Indeed, for the binomial distribution Diaconis and Zabell [11] reported four equivalent forms for δ_1. For the Pearson family, defined by $f'(x)/f(x) = (x + a)/(b_0 + b_1x + b_2x^2)$, with mean μ and variance σ^2, we have

$$\delta_1 = 2C\sigma^2 f(\mu), \qquad \text{where } C = \frac{4\beta_2 - 3\beta_1}{6(\beta_2 - \beta_1 - 1)},$$

with β_1 and β_2 being the usual coefficients of skewness and kurtosis [12]. For the Pearson Type III, or gamma distributions, we have $C = 1$ and δ_1 is "twice the variance times the value of the density at its mean". This property also holds for the normal and for several discrete distributions, like the Poisson, the binomial, and the negative binomial where the mean is replaced by its integral part. A similar relation holds for the beta distribution [13], and some other closed form summation for common distributions, of which δ_1 is a special case, are found to have interesting relations with classical orthogonal polynomials [11]. For the normal $N(\mu, \sigma^2)$ we have $\delta_1 = \delta_2 = \sigma\sqrt{2/\pi} \simeq 0.8\sigma$.

For the Student distribution t_n, with density

$$f(t; n) = \left[\sqrt{n}\beta\left(\frac{1}{2}, \frac{n}{2}\right)\right]^{-1}\left(1 + \frac{t^2}{n}\right)^{n+1/2}, \qquad -\infty < t < \infty, \quad n \geq 2,$$

we have $\delta_1(T_n) = (n/\pi)^{1/2}\,\Gamma((n-1)/2/\Gamma(n/2))$ which converges toward $\to \sqrt{2/\pi}$, as expected. For several common distributions, although δ_1 does not come in closed form, δ_1/σ does and we have to limit ourselves to the study of this ratio.

The population median Md is the location parameter such that $P(X \ge Md) \ge 1/2$ and $P(X \le Md) \ge 1/2$. For simplicity, we will suppose Md unique here. Then, we have $Md = F^{-1}(1/2)$, where F^{-1} is the quantile function defined by $F^{-1}(t) = \inf\{x : F(x) \ge t\}$, $t \in [0,1]$. For symmetrical distributions we have $\delta_1 = \delta_2$. Since for asymmetrical distributions the median is more representative of the "center" than the mean, δ_2 readily provides a meaningful dispersion measure related to that center, while in the L^2 norm very few applications have been found for its counterpart $[E(X - Md)^2]^{1/2}$, whose meaning is also unclear.

Since the median, usually, is not available in closed form, no analytic expression of δ_2 can be given for skewed distributions, except for the chi-square, where Causey [14] reported that for $X \sim \chi^2_n$, where n is the number of degrees of freedom, we have $\delta_2 = 2na_k + 1$ if $n = 2k+1$ and $\delta_2 = 2nb_k + 1$ if $n = 2k+2$, where $a_k = \sqrt{2/\pi}(Md)^{k-1/2}\exp(-Md/2)/\prod_{j=1}^{k}(2j-1)$ and $b_k = (Md/2)^k \exp(-Md/2)/k!$. Using the quantile function F^{-1}, δ_2 can also be represented as $\int_0^{1/2}[F^{-1}(1-\alpha) - F^{-1}(\alpha)]\,d\alpha$, and most of the discussion on δ_1 above can be carried over to δ_2 which has applications in issues related to skewed distributions. For example, a measure of skewness proposed by Groeneveld and Meeden [15] is $\beta = (\mu - Md)/\delta_2$.

3. THE SAMPLE MEAN ABSOLUTE DEVIATIONS AND THEIR DISTRIBUTIONS

Because of the widespread use of the standard deviation, it is informative to have a comparative study of the respective distributions of the sample standard deviation and of the sample mean absolute deviations about the mean and about the median.

3.1. Distribution of the Sample Standard Deviation

For a sample of size n taken from a population with mean μ and variance σ^2, let $S_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2/n$ be the sample variance, where $\bar{X}_n = \sum_{i=1}^n X_i/n$ is the sample mean. We know that $E(S_n^2) = \sigma^2(n-1)/n$ and if X is normal, we have $nS_n^2/\sigma^2 \sim \chi^2_{(n-1)}$ and hence, $\mathrm{Var}(S_n^2) = 2\sigma^4(n-1)/n^2$, with the density of S_n^2 given by a gamma distribution, $f(s^2) = K(n)(s^2)^{(n-3)/2}e^{-ns^2/2\sigma^2}$, $s^2 > 0$, where $K(n)$ is the normalizing constant depending on n, $K(n) = [(2\sigma^2/n)^{(n-1)/2}\Gamma((n-1)/2)]^{-1}$.

Hence, for the sample standard deviation S_n, we have $f(s) = 2\lambda(n)(s)^{n-2}e^{-ns^2/2\sigma^2}$, $s > 0$, since $\sqrt{n}S_n/\sigma$ has a chi-distribution. Then, S_n is a biased estimator of σ, with $E(S_n) = A(n)\sigma$ and $\mathrm{Var}(S_n) = B(n)\sigma^2$, where

$$A(n) = \left(\frac{2}{n}\right)^{1/2} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}, \qquad n \ge 2, \qquad \text{and} \tag{2}$$

$$B(n) = 1 - \frac{1}{n} - [A(n)]^2. \tag{3}$$

Asymptotically, we have the following convenient expression:

$$B(n) = \frac{1}{2n} - \frac{1}{8n^2} - \frac{3}{16n^3} - \frac{157}{1024n^4} - \cdots.$$

Using the relations

$$\frac{\Gamma(p)}{\Gamma(p+1/2)} = \frac{\Gamma(2p)\Gamma(1/2)}{(2^{2p-1}[\Gamma(p+1/2)]^2)}, \tag{4}$$

and $\Gamma(1/2) = \sqrt{\pi}$, we have obtained the interesting expression of $A(n)$, $n \ge 2$, in terms of n, which does not seem to be available anywhere in the literature. We have

$$A(2) = \frac{1}{\sqrt{\pi}}, \qquad A(3) = \sqrt{\frac{\pi}{6}},$$

for n even, $n \geq 2$,

$$A(n) = \frac{(n-2)\cdots 4.2}{(n-3)\cdots 3.1}\sqrt{\frac{2}{n\pi}},$$

and for n odd, $n \geq 3$,

$$A(n) = \frac{(n-2)\cdots 3.1}{(n-3)\cdots 4.2}\sqrt{\frac{\pi}{2n}}.$$

3.2. Distribution of the Sample Mean Absolute Deviations About the Mean

Let us consider first the case when μ is known, and let us define $d_n^* = \sum_{i=1}^{n} |X_i - \mu|/n$. We then have $E(d_n^*) = \delta_1$ and $\mathrm{Var}(d_n^*) = (\sigma^2 - \delta_1^2)/n$.

If μ is unknown, let us define the sample mean absolute deviation (about the sample mean) by $d_n = (\sum_{i=1}^{n} |X_i - \bar{X}_n|)/n$. Only for the normal case can the distribution of d_n be computed precisely, but by a recurrence relation. As established by Godwin [16,17] (see also the Appendix), this relation is quite complex. However, we have, for the mean and variance of d_n

$$E(d_n) = \delta_1\sqrt{1 - \frac{1}{n}}, \tag{5}$$

and

$$\mathrm{Var}(d_n) = \frac{2\sigma^2}{\pi}\,\frac{n-1}{n^2}J(n), \qquad \text{where } J(n) = \left[\frac{\pi}{2} + \sqrt{n(n-2)} - n + \arcsin\left(\frac{1}{n-1}\right)\right]. \tag{6}$$

We know that $J(n)$ converges toward $(\pi - 2)/2$ as $n \to \infty$. Hence, $E(d_n/C(n)) = \sigma$, where

$$C(n) = \sqrt{\left(1 - \frac{1}{n}\right)\frac{2}{\pi}}.$$

Also,

$$\mathrm{Var}\,\frac{d_n}{C(n)} = \frac{\sigma^2}{n}J(n) \tag{7}$$

converges toward $(\sigma^2/n)[(\pi - 2)/2]$.

If the population median, $Md = Md(X)$, is known, let $d_n^{*'} = \sum_{i=1}^{n} |X_i - Md|/n$. Then, we have

$$E\left(d_n^{*'}\right) = \delta_2, \qquad \text{and} \qquad \mathrm{Var}\left(d_n^{*'}\right) = \frac{\sigma^2 + (Md - \mu)^2 - \delta_2^2}{n}.$$

However, if Md is unknown, let md_n be the sample median (a precise definition of md_n is given in Section 8) and let $d_n' = \sum_{i=1}^{n} |X_i - md_n|/n$. The distribution of d_n', however, is untractable, even for the normal case, and we have to make asymptotic considerations to derive its approximate distribution when n is large.

A priori, we can consider the following ratios relating the above statistics:

$$W_n = \frac{d_n}{s_n}, \qquad W_n' = \frac{d_n'}{s_n},$$

$$h_n = \sqrt{n}\frac{\bar{X}_n - \mu}{d_n}, \quad h_n' = \sqrt{n}\frac{\bar{X}_n - \mu}{d_n'}, \quad Y_n = \sqrt{n}\frac{md_n - \mu}{d_n}, \quad \text{and} \quad Y_n' = \sqrt{n}\frac{md_n - \mu}{d_n'}.$$

Only two of the above statistics have been studied in detail, W_n by Geary [18] and h_n by Herrey [19].

3.3. Using μ, δ_1 as Location-Scale Parameters for the Normal

In the L^1-norm approach, δ_1 should be used instead of the standard deviation σ. Using μ, δ_1 as location-scale parameters for the normal distributions, $X \sim N(\mu, \delta_1)$ has as density

$$f(x) = \frac{1}{\pi\delta_1} \exp\left\{-\frac{1}{\pi}\left(\frac{x-\mu}{\delta_1}\right)^2\right\}, \qquad -\infty < x < \infty, \tag{8}$$

a much more symmetrical and simpler form than the one with σ^2. Here, $\delta_1 = \sqrt{2/\pi}$ corresponds to the common standard case $\sigma = 1$. The standard (in the L^1-norm) normal variate V, with $E(V) = 0$ and $\delta_1(V) = 1$, has as density $f(v) = e^{-(v/\sqrt{\pi})^2}/\pi$. Hence, $V = (X-\mu)/\delta_1$. Percentiles of V are then those of the normal with $\mu = 0$ and $\mathrm{Var}(X) = \pi/2$ and can be obtained from common normal tables via a linear transformation on the variable.

However, although δ_1 has its origin back to de Moivre [11] its applications in statistical inference are quite limited due to mathematical difficulties encountered in exact sampling theory. Classical inference is anchored around the variance, which enjoys the property $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$, when X and Y are independent, a property certainly not shared by δ_1 and hence, unlike the standard deviation, we do not have $\delta_1(\bar{X}_n) = \delta_1/\sqrt{n}$ concerning the mean of a sample of size n, except for the normal case. The exact expression of $\delta_1(\bar{X}_n)$ is, in general, highly complex except in the case where the distribution is regenerative. This is the case of the gamma (α, β) distribution, for example, with density: $f(x;\alpha,\beta) = x^{\alpha-1}\exp(-x/\beta)/(\beta^\alpha\Gamma(\alpha))$, $x > 0$. We then have $\bar{X}_n \sim \Gamma(n\alpha, \beta/n)$ and since $\delta_1(X) = 2\alpha^\alpha\beta/[e^\alpha\Gamma(\alpha)]$, we have

$$\delta_1\left(\bar{X}_n\right) = \frac{2\beta}{n}\frac{(n\alpha)^{n\alpha}}{\exp(n\alpha)\Gamma(n\alpha)} \neq \frac{\delta_1(X)}{\sqrt{n}}.$$

3.4. Independence of $\bar{X}_n$ and d_n

The independence of $\bar{X}_n$ and S_n^2 for a normal distribution is a basic result in statistics. In fact, it can be proven that $\bar{X}_n$ and $g(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ are independent for any function g. The proof can be obtained directly, or can be a consequence of Basu's theorem, since $\bar{X}_n$ is completely sufficient and $(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ is ancillary, as pointed out by Boos and Hughes-Oliver [20]. Hence, $\bar{X}_n$ and d_n are independent, and whereas the ratio $(\bar{X}_n - \mu)/(S_n/\sqrt{n-1})$ can be shown to have a t_{n-1} distribution independent of S_n, the ratio $H_n = (\bar{X}_n - \mu)/(d_n/\sqrt{n})$ can also be shown to have a distribution independent of d_n [19].

However, the precise density of H_n can only be obtained by a complex recurrence integral relation, which uses the distribution of d_n obtained by Godwin [16]; see the Appendix. This density is symmetrical with respect to the vertical axis, and converges toward the standard form of the density, $f(v)$, considered above. Distributions of H_n and their critical values at different confidence levels can be obtained by using a computer program available from the authors, or by consulting tabulated values given by Herrey [21].

To our knowledge, the precise sampling distributions of W_n', h_n', Y_n, and Y_n', for n finite, have not been determined, due to the difficulty in establishing the value of the sample median, but asymptotic expansions for most of them are given by Babu and Rao [22].

3.5. Other Results Concerning d_n

To compare the relative magnitudes of S_n and d_n for the normal distribution, Geary [18] studied the distribution of the Studentized d_n, or Geary's ratio $W_n = d_n/S_n$, which also has a highly complex expression. Basically, the density of W_n is centered at 0.789 and gets tighter about that value as n increases. Geary gave approximate percentiles of the W_n, according to various values of n. Due to the complexity of the exact distribution of d_n in the normal case, some authors have given approximations to this distribution. Cadwell [23], for example, gave an approximate density to d_n/σ, using $(\chi_n^2/C)^\alpha$, where n, C, and α are determined by matching the first three moments. He also used a chi-square distribution to give an approximate distribution to the average of k independent sample mean deviations, each of size n from the same normal population, denoted $\bar{m}(k,n)$. He found that $c\{\bar{m}(k,n)/\sigma\}^{1.8} \sim \chi_v^2$, where v and c are given by two formulae. As consequences, we have, approximately, $\sum_{i=1}^k c_i(m_i/\sigma)^{1.8} \sim \chi_\mu^2$, with $\mu = \sum_{i=1}^k v_i$ and $[\sum c_i m_i^{1.8}/\sum c_i'(m_i')^{1.8}][\sum v_{i'}/\sum v_i] \sim F_{\mu,u'}$ for two independent sets of k and k' independent sample mean deviations.

3.6. Advantages of Using d_n over S_n

Although the use of S_n seems to have overwhelming advantages, and S_n appears to be the sample scale statistic mostly used in the literature (among others such as the range, the interquartile range, etc.), there are still several merits for d_n.

(a) A simpler and more robust sample measure of dispersion

Containing only first degree terms d_n gives less importance to extreme values, and is hence more robust and is also simpler to compute than S_n. Population-wise, it is often suggested that a multiple of δ_1 can provide a more robust measure of spread than the standard deviation [7, p. 59].

(b) A better estimator of scale

Comparing (2) and (3) with (5) and (6), respectively, we can see that, although both d_n and S_n are asymptotically unbiased estimators of δ_1 and of σ, respectively, $E(d_n)$ converges toward δ_1 faster than $E(S_n)$ does toward σ. Similarly, $\mathrm{Var}(d_n)$ goes to zero faster than $\mathrm{Var}(S_n)$ does. Hence, d_n performs better as a point estimator for δ_1 than S_n does for σ. This point seemed to be missed by several authors who used d_n to estimate σ instead of δ_1. For example, Davies and Pearson [24] provided correction coefficients for d_n for the estimation of σ. On the other hand, Fisher (see discussion in [19]) used (3) and (7) to conclude that the asymptotic relative efficiency (ARE) of $d_n/C(n)$ compared to S_n, in estimating σ, is only $1/(\pi-2)$ (i.e., $\mathrm{ARE}(d_n/C(n)|S_n|) = \lim_{n\to\infty}\mathrm{Var}(S_n)/\mathrm{Var}(d_n/C(n)) \to 1/(\pi-2)$). Naturally, in the estimation of σ, S_n is the square root of a second degree expression and should perform better than d_n, which contains only first degree terms. However, as respective unbiased estimators of δ_1 and of σ separately, the statistics $d_n/(\sqrt{1-1/n})$ and $S_n/A(n)$ give $\mathrm{ARE}(d_n/\sqrt{1-1/n}\,|(S_n/A(n))|) = (B(n)/2[A(n)]^2)\cdot n\pi/J(n) > 1$.

Therefore, it is important to notice here that in order to benefit from the fact that d_n is a better estimator of scale, one has to use δ_1, instead of σ, in all related analysis (see Section 5).

(c) A better estimator in the event of contamination

In general, there is contamination when the observations, instead of coming from a single distribution, come from the mixture of two or several distributions having the same mean, but with different variances. For mathematical tractability, we usually consider normal distributions only. For example, Teichroew [25] considered the case when these variances are gamma distributed, derived the marginal distribution of X, and showed that the variance of the latter increases as the shape parameter of the gamma distribution decreases. In the presence of contamination, the estimation of the population scale parameters changes drastically.

Let us consider the case where the population distribution is $\alpha N(\mu,\sigma^2) + (1-\alpha)N(\mu,9\sigma^2)$, with $0 \le \alpha \le 1$.

In an exhaustive investigation of the above distribution, Tukey [26] found that although for $\alpha = 1$ (no contamination) the ARE of the mean absolute deviation to the standard deviation, $\mathrm{ARE}(d_n \mid S_n) = \lim_{n\to\infty}\mathrm{Var}(S_n)/\mathrm{Var}(d_n)$, is 0.88, this measure quickly rises to 1 for $\alpha = 0.992$

and becomes much larger as α decreases. Hence, for any significantly contaminated population $(1-\alpha \geq 0.008)$, d_n can be a much better estimator of scale than S_n.

Since a similar conclusion can be reached when using $\bar{X}_n$ to estimate μ, Tukey concluded that "nearly imperceptible nonnormalities may make conventional relative efficiencies of scale and location entirely useless" and that "for some contaminated populations, d_n will be a better estimate of scale in large samples than S_n". Concerning Fisher's rejection of d_n and Eddington's remark that "for the errors commonly occurring in practice, the mean absolute error is a safer criterion of accuracy than the mean square error, especially if doubtful observations have been rejected", Tukey supported Eddington's conclusion rather than Fisher's because of two reasons:

(1) Fisher's analysis did not seem to be carried out far enough.
(2) His assumption was that the sole purpose of scaling is to estimate σ.

Huber [27] in his preliminary discussion about robustness also cited the above example and provided numerical values for ARE, according to different values of α.

More generally, by considering the contaminated distribution $\alpha N(\mu,\sigma)^2 + (1-\alpha)N(\mu,k\sigma^2)$, we can show [9, p. 584] that, if $k(1-\alpha) \approx 0$ (very small contamination), $E[(S'_n - \sigma)^2] \approx (\sigma^2/2n)\cdot A(k,\alpha)$, where $S' = S(n/(n-1))^{-1/2}$ and $A(k,\alpha) = (1+1.5(1-\alpha)k^2)$ while $\sqrt{(\pi/2)}d_n$ has approximately $\alpha(\sigma^2/2n)(\pi-2)$ as variance, when n is sufficiently large. Hence, the second unbiased estimator of σ is indeed the better one in this case.

Concerning d'_n, at the present time no distribution of this statistic has been derived, and consequently, no result on its behavior can be presented.

4. ESTIMATION OF THE MEAN ABSOLUTE DEVIATIONS

When δ_1 is a scale parameter of a distribution, on the same basis as σ, the question related to its direct estimation from data is of interest. However, when δ_1 is a function of the parameters of the distribution, its estimation can be obtained through the estimation of these parameters.

(a) Point estimation of δ_1

For the Pearson system, Suzuki [28] has developed a general formula to estimate δ_1, based on the four first sample central moments m_i, $i=1,\dots,4$. The function $H(m) = H(m_1,m_2,m_3,m_4)$ gives efficient estimators of δ_1, with $\sqrt{n}(H(m) - H(\mu))$ asymptotically normal, where $\mu = (\mu_1,\mu_2,\mu_3,\mu_4)$. More specifically,

$$m_2\sqrt{\frac{2}{\pi}}, \qquad \frac{m_3}{m_2}\frac{\left(4m_2^2/em_3^2\right)^{4m_2^3/m_3^2}}{\Gamma\left(4m_2^3/m_3^2\right)},$$

and

$$\frac{2\left[2m_2m_4\left(m_4-3m_2^2\right)\right]^{1/2}}{\left[3\left(m_4-m_2^2\right)\beta\left(1/2,\left(2m_4-3m_2^2\right)/\left(m_4-3m_2^2\right)\right)\right]},$$

are efficient estimators of δ_1 for the three distributions $N(\mu,\sigma^2)$, χ^2_ℓ, and t_ℓ, respectively. Suzuki [28] also gave the variances of these estimators, but their expressions are very complex.

(b) Interval estimation

In classical interval estimation of a distribution parameter, the $(1-\alpha)$ 100% confidence interval is often obtained via the distribution of a "pivotal quantity", whose percentage points are used in the computation of the confidence interval end points. For the normal case, we have the following result on δ_1, which is similar to the one stating that $nS_n^2/\sigma^2 \sim \chi^2_{(n-1)}$.

THEOREM 1. *Let* $X \sim N(\mu, \delta_1)$ *and* $d_n = \sum_{i=1}^{n} |X_i - \bar{X}_n|/n$. *Then,* $nd_n/\delta_1 \sim g_n$, *with* g_n *obtained from Godwin's density.*

PROOF. See the Appendix.

Hence, for an observed value of d_n, the $100(1-\alpha)\%$ confidence interval for δ_1 is obtained from the double inequality $g_{n,\alpha/2} \leq nd_n/\delta_1 \leq g_{n,1-\alpha/2}$, with the values of the percentage points of g_n to be computed using the expression of g_n given in the Appendix (a computer program is available from the authors). For example, if six observations are taken from a normal population with MAD δ_1, with an observed value of d_6, we have the 90% confidence interval for δ_1, $6d_6/8.705 \leq \delta_1 \leq 6d_6/2.708$, where 2.708 and 8.705 are, respectively, the 5^{th} and 95^{th} percentiles of g_6, obtained from our computer program.

For nonnormal distributions, we can apply the above result to obtain an approximate confidence interval for δ_1 when n is large.

5. SOME USES OF THE MEAN ABSOLUTE DEVIATIONS

In this section, we will present some uses of δ_1 in current statistical issues.

5.1. Sample Size Determination

Since there is a simple relation between δ_1 and σ for the normal case, the sample size problem when using δ_1, regarding mean estimation and hypothesis testing with control over the two types of error, can be solved in the usual way and some advantages can even be gained when using δ_1. Let us consider the case where we have to compute the sample size required to satisfy a bound condition on the length of a confidence interval of the mean, in a normal distribution.

(a) *Case when* δ_1 *is known.* In order for the $(1-\alpha)$ 100% confidence interval for μ to be shorter than L, it suffices to take $n > (2V_{\alpha/2}\delta_1/L)^2 + 1$, where $V_{\alpha/2}$ is the $\alpha/2^{\text{th}}$ percentile of the standard normal variate V considered above.

(b) *Case when* δ_1 *is unknown.* If the ratio $\delta_1/(L/2)$ can be given, the above formula can still be applied. This is the case where the meaning of δ_1 can be effectively used and presents a clear advantage over $\sigma/(L/2)$ since $\delta_1/(L/2)$ clearly means the "*ratio of the average variation over half of the desired variation*", which is easier to interpret and evaluate. Working with σ, the ratio $\sigma/(L/2)$, often used in textbooks, does not have that clear meaning, and, in fact, any meaning at all in the case of a skewed population distribution, even in the large sample case.

5.2. Peters Formula in Applied Physics

In experimental physics, when the number of observations is small and extreme observations may be willingly deleted, it is justified to use d_n. First, as pointed out by Brown [29, p. 178] the aim of most physical experiments is to obtain a numerical value for some quantity, under the form $A \pm b$, with b called either precision, or error, although the last term can lead to confusion.

There are several types of errors that can contribute to the value of b. They are mistakes, gross errors, constant errors, systematic errors, and random errors. When only errors of random nature are considered, they are supposed to have a normal distribution. Four types of precision, which are related to various confidence levels or to various multiples of the standard deviation, are frequently considered:

(1) The precision index h (or $[\sigma\sqrt{2}]^{-1}$).
(2) The standard deviation (or root mean square error) σ.
(3) The average error δ_1, or "true average error", which is zero if account is taken of signs.
(4) The probable error r, or quartile, which is defined as being such that the probability of making an error in the range $-r$ to $+r$ is the same as of making any numerically larger error.

We then have, taking σ and the normal distribution as reference,

$$\delta = 0.798\sigma, \qquad \frac{1}{h} = \sigma\sqrt{2}, \qquad \text{and} \qquad r = 0.674\sigma.$$

The precision frequently used in experimental physics is the probable error r. With a 50% confidence level, we have the classical Peters formula, $r = 0.845347 d_n/\sqrt{n-1}$, which was obtained via a very simple argument by Peters (see [19]), and used mostly for small values of n. However, this formula is inaccurate if the coefficient 0.84537 does not change with the sample size since it came from the characteristic value of the standard normal distribution, $z_{0.75} = 0.67449$ (case $n = \infty$) multiplied by $\sqrt{2/\pi}$. The precise formula is $\bar{X}_n \pm h_{n,\alpha/2}(d_n/\sqrt{n})$, where $h_{n,\alpha/2}$ can be obtained from the tabulated values given by Herrey [21] (or the computer program obtained from the authors). For example, for $n = 8$, the 50% Peters formula is $\bar{X}_8 \pm 0.921(d_8/\sqrt{8})$, whereas the 95% formula is $\bar{X}_8 \pm 3.933(d_8/\sqrt{8})$ and gives an interval about 4.27 times larger.

Herrey [21] has found that the two confidence intervals, based, respectively, on d_n and on S_n (we then have $\bar{X}_n \pm t_{n-1,\alpha/2}(S_n/\sqrt{n-1})$), give practically the same answers.

5.3. Other Applications in Fields Associated with Statistics

There are several applications of δ_1 and δ_2 in fields closely associated with statistics. For example, in econometrics, δ_1 is closely associated with the Lorenz curve, where $\delta_1/2\mu$ is the maximum vertical distance between that curve and the first diagonal, occurring at the abscissa $F(\mu)$. It is closely related to other measures like the Gini index and the Pietra ratio used in measuring income distributions [30]. A similar conclusion applies for δ_2 on the Lorenz curve, where $\delta_2/2\mu$ is the vertical distance between the diagonal and the curve, occurring at the abscissa $F(Md) = 1/2$. Since the Lorenz curve has close relations with the total time on test transform curve in reliability theory [10], corresponding results for this curve in relation with δ_1 can also be established. In engineering statistics, the main use of δ_1 is to characterize extreme value distributions [31] and in Bayesian statistics, Pham-Gia *et al.* [13] have used it to elicit the beta prior distribution.

Concerning δ_2, in descriptive statistics it has been used effectively to measure skewness and kurtosis and to partially order distributions [15]. For example, the Bowley coefficient of skewness of a distribution is $b_1 = (Q_3 + Q_1 - 2Q_2)/(Q_3 - Q_1)$, where Q_i, $1 \le i \le 3$, are the quartiles of the distribution. It can be extended to $b_2 = \{F^{-1}(1-\alpha) + F^{-1}(\alpha) - 2Md\}/[F^{-1}(1-\alpha) - F^{-1}(\alpha)]$, where F^{-1} is the quantile function. Noticing that $\delta_2 = \int_0^{1/2}[F^{-1}(1-\alpha) - F^{-1}(\alpha)]\,d\alpha$, another measure of skewness can be defined very simply by $b_3 = (\mu - Md)/\delta_2$ and obtained by integrating the numerator and denominator of b_2 from 0 to 1/2. This measure, b_3, maintains the $<_c$ ordering, i.e., if $F_X <_c F_Y$ then $b_3(X) < b_3(Y)$, where, for two distributions F and G, we write $F <_c G$ (i.e., Fc-precedes G) if $G^{-1}(F(x))$ is convex, i.e., $G(x)$ is at least as skewed to the right as $F(x)$. On the other hand, Bickel and Lehman [32] found that δ_2 outperforms the trimmed deviations for thin-tailed distributions when the standard asymptotic variance is taken as comparison criterion. In decision theory, δ_2 is closely associated with the absolute error loss function [33], in either a two-person or a multiperson game, and in economics (income distribution and Lorenz curve) results similar to those related to δ_1 are valid for δ_2, when the median, instead of the mean, is considered in heavily skewed distributions.

6. ASYMPTOTIC CONSIDERATIONS

Although several sampling distributions considered here cannot be given in a closed form for n finite, where n is the sample size, we have the following results for $n \to \infty$.

THEOREM 2 [22]. *If F is differentiable in a neighborhood of μ with derivative $f(\mu) > 0$ and variance σ^2, then:*

(1) *If $F(\mu) \neq 1/2$, $\sqrt{n}(d_n - d_n^*)/[2F(\mu) - 1] \to N(0,\sigma^2)$ in distribution.*
(2) *If $F(\mu) = 1/2$, $\sqrt{n}(d_n - d_n^*) \to f(\mu)U^2 - UV$ in distribution,*

where (U,V) is bivariate normal with mean zero and covariance matrix $\sum = \begin{pmatrix} \sigma^2 & \delta_1 \\ \delta_1 & 1 \end{pmatrix}$.

Concerning the sample median md_n and d'_n, we have another result, subject, however, to a Lipschitz condition on F, about the median Md.

THEOREM 3. *Let $|F(x) - F(Md)| \leq c|x - Md|^\beta$ for some $c > 0$ and $0 < \beta \leq 1$ and let F have a unique median Md. If $\sqrt{n}(md_n - Md)$ is bounded in probability, then $\sqrt{n}(d'_n - \delta_2) \to N(0,\xi^2)$ in distribution, where $\xi^2 = E(X - Md)^2 - \delta_2^2$.*

Concerning d'_n and $d_n^{*'}$, we have the following theorem.

THEOREM 3′. *If F has a continuous derivative f in a neighborhood of Md and $f(Md) > 0$, then $n(d_n^{*'} - d'_n) = n(md_n - Md)^2(f(Md) + o(1)) + O(n^{-1/4}(\log n)^2)$. As a consequence, $(n/4)f(Md)(d_n^{*'} - d'_n) \to \chi_1^2$, in distribution.*

Proofs of the above theorems can be found in [22], where Edgeworth expansions for the distributions of W_n, W'_n, h_n, and h'_n are also given.

7. THE DISPERSION FUNCTION: A GENERALIZATION OF THE MEAN ABSOLUTE DEVIATION

More generally, the mean absolute deviation of a random variable about an arbitrary point a, denoted $\delta_a(X)$, has been studied in [13], where it was related to the first moment distribution of F, denoted Φ, and used for the elicitation of the prior distribution, in a Bayesian context. More generally, Munoz-Perez and Sanchez-Gomez [34] considered the following case.

For $X \in L^1$, let $D_X(u) = E(|X - u|)$ be a function of $u \in R$, called dispersion function of X. We have a very simple relation between $F(x)$ and D_X at points of continuity of the latter $F(u) = (1 + D'_X(u))/2$.

THEOREM 4. *D_X has the following properties:*

(1) *D_X is differentiable and convex;*
(2) *$\lim_{u\to\infty} D_X(u) = 1$ and $\lim_{u\to-\infty} D_X(u) = -1$;*
(3) *$\lim_{u\to\infty}(D_X(u) - u) = -E(X)$ and $\lim_{u\to\infty}(D_X(u) + u) = E(X)$.*

Conversely, a function D_X having the above properties is the dispersion function of a unique distribution. For a proof, see [34]. Considering the degenerate distribution G_u at a point u, i.e., $G_u(x) = I_{[u,\infty)}(x)$, it can be established that $D_X(u) = \int_{-\infty}^{\infty} |F_X(x) - G_u(x)|\, dx = \int_0^1 |F_X^{-1}(t) - u|\, dt$, where $F_X^{-1}(t)$ is the quantile function, as defined previously.

Hence, D_X is the L^1 distance between F_X and G_u, or equivalently between F_X^{-1} and $G_u^{-1} = u$. Then, the variance σ^2 is the L^1-distance between $D_X(u)$ and $G_\mu(u)$, i.e., we have $\int_{-\infty}^{\infty} | D_X(u) - G_\mu(u)|\, du = \sigma^2$, where $G_\mu(u) = |\mu - u|$.

8. THE MEDIAN ABSOLUTE DEVIATION

For the median absolute deviation (about the median), results presented in the literature are mostly asymptotic in nature. Here, we only discuss the most important ones.

Let Md be the median of the distribution F. For simplicity, let us consider the case F is strictly increasing and there is a unique value for its median. For symmetric distributions we naturally have $Md = \mu$. For finite populations, or a sample, we define $md_n = \text{median}(X_1, \ldots, X_n) = X_{k+1:n}$ if $n = 2k + 1$, and

$$md_n = aX_{k:n} + (1 - a)X_{k+1:n}, \qquad 0 \leq a \leq 1, \quad \text{if } n = 2k$$

(in general, we take $a = 1/2$), where $X_{i:n}$, $i = 1, \ldots, n$, is the ordered sample.

8.1. The Population Median (Md) and the Population Median Absolute Deviation (or MAD) Denoted by λ

Since the median cannot usually be defined in closed form, there are relatively few analytical results concerning the median. For positively skewed distributions we have $mo \leq Md \leq \mu$, where mo is the mode. The reverse order applies if the distribution is negatively skewed. In general, we can say that for an arbitrary population, the median is a highly robust measure of location.

An important application of the median is in centering and symmetrization of random variables. In the statistical literature, a centered variable Y is usually centered at the population mean, i.e., $Y = X - \mu_X$, and hence $E(Y) = 0$. Using the same concept with the median Md, we have median-centered random variables, an idea that, according to [6, p. 255], "*not only completes the first one, but also tends to replace it altogether*".

A random variable W is said to be symmetric if $\forall w$, $P(W < w) = P(W > -w)$, or equivalently, $F(-w+0) = 1 - F(w)$. Then, $Md(W) = 0$ and if the density $f(w)$ exists, it is symmetrical w.r.t. the vertical axis. In symmetrization, we assign to any r.v. X, with median $Md(X)$, the r.v. $X^s = X - Y$, where Y is independent of X but has the same distribution. We then have X^s symmetric, with $E(X^s) = Md(X^s) = 0$, $E(|X^s|) = \delta_1(X^s) = \delta_2(X^s) = \Delta$, where Δ is the classical mean-difference of X, which is "the average of the differences between all possible pairs of values taken by X, regardless of sign". Moreover, ch.f. (X^s) is $|\Psi|^2$, where Ψ is the characteristic function of X.

Without knowing the value of $Md(X)$, probabilities of quantities relating X to $Md(X)$ can still be related to those of X^s itself, and to those of $X - a$, $\forall a \in R$. For example, we have $P(|X - Md(X)| \geq \epsilon) \leq 2P(|X^s| \geq \epsilon) \leq 4P(|X - a| \geq \epsilon/2)$ and $E(|X - Md(X)|^r) \leq 2E(|X^s|^r) \leq 4C_r(E|X-a|^r)$, where $C_1 = 1$ and $C_r = 2^{r-1}$ for $r > 1$, which gives, for $r = 1$,

$$\delta_2(X) \leq 2\Delta \leq 4\delta_a(X), \qquad \forall a \in R$$

(see [6, p. 257]).

In the theory of Lorenz curves, we known that $\Delta = 4\mu_X A$, where A is the area between the Lorenz curve of X and the first diagonal.

We define the population median absolute deviation (MAD) by $\lambda = \text{Median}(|X - Md|)$, which reduces to $\sigma\Phi^{-1}(3/4)$ for the normal case, where Φ is the cumulative standard normal distribution.

8.2. The Sample Median md_n and Sample Median Absolute Deviation (or mad_n) Denoted by ℓ_n

Let $X_1, \ldots, X_n$ be a random sample from a distribution F and let us define md_n as in the case of a finite population.

For normal F, the efficiency is $\text{Var}(\bar{X}_n)/\text{Var}(md_n) = 2/\pi = (\delta_1/\sigma)^2 = 0.637$. However, it is found that this efficiency is higher in sampling from other distributions, reaching 2 for the Laplace and ∞ for the Cauchy. Under imperceptible contamination raised by Tukey [26], this efficiency is practically one and it is higher for scale-contaminated normal distributions and other heavy tailed distributions.

For a general distribution F with density f continuous at Md, md_n is asymptotically normal, with mean Md and variance $[4nf^2(Md)]^{-1}$, which reduces to $((\sigma^2/n) \cdot \pi/2)^{-1}$ for the normal case. ARE$(md_n \mid \bar{X}_n)$ is hence $4\sigma^2 f^2(Md)$.

The estimation of the population median from an ordered sample of observations is an important topic of statistics, about which several results have been reported in [35]. In general, the sample median is an estimator of location which is less sensitive to heavy tails than $\bar{X}_n$. It also plays a central role in the distributions of order statistics and in nonparametric statistical methods.

If X is normal, then $\bar{X}_n$ and $md_n - \bar{X}_n = \text{median}(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$ are independent, with $\text{Var}(md_n - \bar{X}_n) \leq \text{Var}(md_n)$. This result is the counterpart of the one already mentioned, on the independence between $\bar{X}_n$ and $g(X_1 - \bar{X}_n, \dots, X_n - \bar{X}_n)$.

The sample median absolute deviation (from the sample median) is mad_n, denoted by $\ell_n = \text{median}\,(|X_i - md_n|,\ i = 1, \dots, n)$.

According to Hampel [36] ℓ_n is the natural nonparametric estimator of the "probable error" r of a single observation already encountered in the section on experimental physics. It is also the scale equivalent of the median and is the most robust estimate of scale. He also proved that ℓ_n is superior to the sample interquartile range and that the two are equivalent when the population distribution is symmetric. Possible uses of ℓ_n include a rough but fast scale estimate in cases where no higher accuracy is required (see [36]).

Let us consider the case where there is a unique median Md for the distribution F, and let us consider the r.v. $W = |X - Md|$ and suppose that W too has a unique median, denoted as λ previously (we have $G(\lambda) = 1/2$, where $G(x) = P(W \leq x)$). Hall and Welsh [37] then proved that ℓ_n converges almost surely toward λ if F is continuous at $Md \pm \lambda$. Furthermore, a central limit-type result is valid for ℓ_n.

THEOREM 5 (Hall and Welsh). *Suppose that $F'(Md)$ exists and is positive and that $F'(Md \pm \lambda) + x$ exists for x in a neighborhood of the origin and is continuous at $x = 0$. If $g(x, Md) = F'(Md + x) + F'(Md - x)$ is positive on a neighbourhood of λ, then $\sqrt{n}(\ell_n - \lambda) \to N(0, \sigma^2)$ in distribution, where $\sigma^2 = S + V$, with S and V being functions of F, F', λ, and Md.*

As a consequence, considering the population median Md, if $g(\lambda, Md)$ is positive, then $\sqrt{n}(\ell_n^* - \lambda)$ converges in distribution toward $N(0, [4g^2(\lambda, Md)]^{-1})$, where $\ell_n^* = \text{median}\,(|X_i - Md|,\ i = 1, \dots, n)$. Furthermore, if $F(Md + \lambda) = 1 - F(Md - \lambda)$ and $F'(Md - \lambda)$ then $\sqrt{n}(\ell_n - \ell_n^*) \to 0$ in probability and $\sqrt{n}(\ell_n - Q_n) \to 0$ in probability, where $Q_n = (\hat{\xi}_{3/4} - \hat{\xi}_{1/4})/2$ is the semi-interquartile range, with $\hat{\xi}_p$ being the sample p^{th} quantile, $0 < p < 1$. The asymptotic equivalence of ℓ_n and the sample semi-interquartile range for $n \to \infty$ is hence established.

For a normal population we also have these results:

(1) $\bar{X}_n - md_n \to 0$, a.s.;

(2) $\ell_n - S_n\phi^{-1}(3/4) \to 0$, a.s., for $n \to \infty$, where Φ is the standard normal cdf.

Furthermore, there are interesting asymptotic properties of the distribution of the couple (md_n, ℓ_n), which is a robust counterpart of $(\bar{X}_n, S_n^*)$, where $S_n^* = \sqrt{(n/(n-1))}S_n$. Some of these results parallel those of $(\bar{X}_n, S_n^*)$. For example, let us consider the following known property of the latter couple.

THEOREM 6. *Let X be such that $E(X^4)$ is finite. Then, $(\bar{X}_n, S_n^*)$ is asymptotically binormal, i.e., $\sqrt{n}(\bar{X}_n - \mu, S_n - \sigma) \to N(0, \sigma_1^2, \sigma_2^2, \sigma_{12})$ in distribution, where the covariance matrix is $\sigma_1^2 = \sigma^2$, $\sigma_2^2 = [E(x-\mu)^4 - \sigma^4]/4\sigma^2$, and $\sigma_{12} = E[(x-\mu)^3]/2\sigma$. Hence, $\bar{X}_n$ and S_n are asymptotically independent iff $E(X-\mu)^3 = 0$, i.e., for symmetrical distributions.*

Correspondingly, for (md_n, ℓ_n), we have the following theorem.

THEOREM 7 (Falk). *Suppose F is continuous, differentiable at $F^{-1}(1/2)$, and at $F^{-1}(1/2) \pm \lambda$, and that $f(F^{-1}(1/2)) > 0$. Then, $\sqrt{n}(md_n - F^{-1}(1/2), \ell_n - \lambda) \to N(0, p_1, p_2, p_{12})$ in distribution, where p_1, p_2, and p_{12} are functions of f, F, F^{-1}, and λ.*

PROOF. See [38] where complex expressions of p_1, p_2, and p_{12} are given.

From the above theorem, we have the asymptotic independence of md_n and ℓ_n if a certain condition is verified, a condition which is satisfied when X is symmetric w.r.t. its unique median $F^{-1}(1/2)$.

A similar result can be proved for the sample interquartile range $X_{[3n/4]:n} - X_{[n/4]:n}$, where $[x] = \inf\{k \in N,\ k \geq x\}$ is the right integral neighbour of x. The above theorem can also be seen as an extension of the Theorem 5 by Hall and Welsh [37].

9. CONCLUSION

The purpose of this article is to give an up-to-date and unified presentation of results related to several dispersion measures and sampling statistics, all denoted by MAD in the literature. Difficulties in establishing analytic expressions for the former and exact sampling distributions for the latter have prevented the full development of these topics for the past several decades. Hopefully, with the renewed interest in robustness and in the L^1-approach to statistics, new progress will be made.

APPENDIX

Godwin [16] gave the following expression for the density of the sample mean absolute deviation d_n, denoted m here, for a standard normal $N(0,1)$ population (mean zero and variance unity). Let $G_0(x) = 1$, and

$$G_1(x) = \int_0^x \exp\left[-\frac{t^2}{2r(r+1)}\right] G_{r-1}(t)\,dt, \qquad r = 1, 2, \dots.$$

Setting

$$g_r(x) = G_r(x) \exp\left[-\frac{x^2}{2(r+1)}\right],$$

we have

(a)

$$f_n^*(m) = \frac{n^{3/2}}{2^{(n+1)/2}(\pi)^{(n-1)/2}} \sum_{k=1}^{n-1} \binom{n}{k} g_{n-k-1}\left(\frac{nm}{2}\right), \tag{A1}$$

where n is the sample size.

Percentage points computed by Hartley [39] for $1 \le r \le 10$ and Godwin [17] provided another proof of the above expression, using geometric considerations. He also provided expressions for the moments of m, based on a recurrence relation, and proposed some approximations for (A1). For large n, the distribution of m approaches

$$N\left(\sqrt{\frac{2}{\pi}}, \frac{n-2}{\pi n}\right).$$

(b) The density of $H_n = (\bar{X}_n - \mu)/m$, as given by Herrey [19], is

$$f_n(h) = \int_0^\infty \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{h^2 v^2}{2}\right] f^*(v) v\,dv, \tag{A2}$$

where $f_n^*(v)$ is given by (A1). Critical values for $f_n(h)$ are given by Herrey [21].

(c) Let $X \sim N(0, \delta_1)$ and let $W = X/(\delta_1\sqrt{\pi/2})$. We then have

$$d_n(X) = \sum_{i=1}^n \frac{|X_i - \bar{X}_n|}{n} = \delta_1 \sqrt{\frac{\pi}{2}}\, d_n(W).$$

Let $Y = n d_n(X)/\delta_1$. Then, $Y = n\sqrt{\pi/2}\, d_n(W)$. Hence, by (A1), we have the density of Y as

$$g_n(y) = f_n^*\left(\frac{y}{\left(n\sqrt{\pi/2}\right)}\right) \frac{\sqrt{2}}{(n\sqrt{\pi})}, \qquad y > 0. \tag{A3}$$

REFERENCES

1. Y. Dodge, Editor, *Statistical Data Analysis Based on the L^1-Norm and Related Methods*, North-Holland, Amsterdam, (1987).
2. J.G. Bassett and R. Koenker, Asymptotic theory of least absolute error regression, *Journ. Amer. Stat. Assoc.* **73**, 618–622, (1978).
3. P. Bloomfield and W.S. Steiger, *Least Absolute Deviations: Theory, Applications and Algorithms*, Birkhäuser, Boston, (1983).
4. S. Portnoy and R. Koenker, The Gaussian hare and the Laplacian tortoise: Computability of square-error versus absolute-error estimators, *Statistical Science* **2** (4), 279–300, (1997).
5. Z.D. Bai, C.R. Rao and Y.Q. Yin, Least absolute deviations analysis of variance, *Sankhya, Series A* **52**, 166–177, (1990).
6. M. Loève, *Probability Theory I*, 4^{th} edition, Springer-Verlag, New York, (1977).
7. A. Stuart and J.K. Ord, *Kendall's Advanced Theory of Statistics*, Vol. I, Oxford University Press, New York, (1987).
8. M. De Groot, *Optimal Statistical Decisions*, McGraw-Hill, New York, (1970).
9. V.K. Rothagi, *An Introduction to Probability Theory and Mathematical Statistics*, John Wiley and Sons, New York, (1976).
10. T. Pham-Gia and N. Turkkan, The Lorenz and T_4-curves: A unified approach, *IEEE Trans. on Reliability*, 76–84, (1994).
11. P. Diaconis and S. Zabell, Closed form summation for classical distributions: Variations on theme of de Moivre, *Statistical Science* **6** (3), 284–302, (1991).
12. A.R. Kamat, A property of the mean deviation of the Pearson-type distributions, *Biometrika* **53**, 287–289, (1966).
13. T. Pham-Gia, N. Turkkan and Q.P. Duong, Using the mean deviation to determine the prior distribution, *Stat. and Prob. Letters* **13**, 373–381, (1992).
14. B.D. Causey, Expected absolute departure of chi-square from its median, *Commun. Statist.-Simula.* **15** (1), 181–183, (1986).
15. R.A. Groeneveld and G. Meeden, Measuring skewness and kurtosis, *The Statistician* **33**, 391–399, (1984).
16. H.J. Godwin, On the distribution of the estimate of mean deviation obtained from samples from a normal population, *Biometrika* **33**, 254–257, (1943–1946).
17. H.J. Godwin, A further note on the mean deviation, *Biometrika* **34**, 304–309, (1947).
18. R.C. Geary, Moments of the ratio of the mean deviation to the standard deviation for normal samples, *Biometrika* **28**, 295–305, (1936).
19. E.M. Herrey, Confidence intervals based on the mean deviation of a normal sample, *Journ. Amer. Stat. Assoc.*, 257–269, (1965).
20. D.D. Boos and Hughes-Oliver, Applications of Basu's theorem, *The American Statistician* **52** (3), 218–221, (1998).
21. E.M. Herrey, Percentage points of H-distributions for computing confidence limits or performing t-tests by way of the mean absolute deviation, *Journ. Amer. Stat. Assoc.* **66**, 187–188, (1971).
22. G.J. Babu and C.R. Rao, Expansions for statistics involving the mean absolute deviation, *Ann. Inst. Statist. Math.* **44**, 387–403, (1992).
23. J.H. Cadwell, The statistical treatment of mean deviation, *Biometrika* **41**, 12–18, (1954).
24. O.L. Davies and E.S. Pearson, Methods of estimating from samples the population standard deviation, *Journ. Royal Stat. Soc., Suppl. 1*, 76–93, (1934).
25. D. Teichroew, The mixture of normal distributions with different variances, *Ann. of Math. Stat.* **28**, 510–513, (1957).
26. J.W. Tukey, A survey of sampling from contaminated distributions, In *Contributions to Probability and Statistics*, (Edited by I. Olkin), Paper 39, pp. 448–485, The Stanford Press, Stanford, CA, (1960).
27. P.J. Huber, *Robust Statistical Procedures*, Soc. Ind. & Appl. Math., (1979).
28. G. Suzuki, A consistent estimator for the mean deviation of the Pearson-type distribution, *Annals Inst. Stat. Math.* **17**, 271–285, (1966).
29. A.F. Brown, *Statistical Physics*, American Elsevier, New York, (1970).
30. T. Pham-Gia and N. Turkkan, Determination of the beta distribution from its Lorenz curve, *Mathl. Comput. Modelling* **16** (2), 73–84, (1992).
31. E.J. Gumbel, *Statistics of Extremes*, Columbia University Press, (1958).
32. P.J. Bickel and E.L. Lehman, Descriptive statistics for nonparametric models III. Dispersion, *Ann. Statist.* **6**, 1139–1158, (1976).
33. T. Pham-Gia, Some applications of the Lorenz curve in decision analysis, *Amer. Journ. of Math. and Manag. Sciences* **15**, 1–34, (1995).
34. J. Munoz-Perez and Sanchez-Gomez, A characterization of the distribution function: The dispersion function, *Stat. and Prob. Letters* **10**, 235–239, (1990).
35. D.F. Andrews, P.J. Bickel, F.R. Hampel, P.J. Huber, W.H. Rogers and J.W. Tukey, *Robust Estimates of Location: Survey and Advances*, Princeton University Press, Princeton, NJ, (1972).
36. F.R. Hampel, The influence curve and its role in robust estimation, *Journ. Amer. Stat. Assoc.* **69**, 383–393, (1974).

37. P. Hall and A.H. Welsh, Limit theorems for the median deviation, *Ann. Inst. Statist. Math., Part A* **37**, 27–36, (1985).
38. M. Falk, Asymptotic independence of median and MAD, *Stat. and Prob. Letters* **34**, 341–345, (1997).
39. H.O. Hartley, Note on the calculation of the distribution of the estimate of mean deviation in normal samples, *Biometrika* **33**, 257–265, (1943–1946).