MEASURES OF LOCATION AND SCALE FOR VELOCITIES IN CLUSTERS OF GALAXIES—A ROBUST APPROACH

TIMOTHY C. BEERS, KEVIN FLYNN, AND KARL GEBHARDT
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824
Received 14 September 1989; revised 9 April 1990

ABSTRACT

Recent observational evidence suggests that few clusters and groups of galaxies have achieved dynamical equilibrium, where a Gaussian distribution of radial velocities might be expected. The canonical estimation techniques, which either assume Gaussian parent populations or clip observed velocity distributions until the Gaussian assumption is satisfied, are not, in general, minimum variance estimators of the kinematic properties of such clusters. In addition, a detailed examination of the local kinematical properties of clusters requires the use of efficient statistical estimators which are insensitive to localized misbehavior in small datasets. For these reasons we suggest that the traditional methods of assigning cluster mean velocities, dispersions, and confidence intervals on these quantities are no longer adequate. In this paper we discuss alternative estimators of the kinematical properties of clusters of galaxies—estimators that are resistant in the presence of outliers, and robust for a broad range of non-Gaussian underlying populations. Because a number of different estimators may be used for any given quantity, we urge a change in the nomenclature to one that does not imply an underlying probabalistic model: we suggest C_u for the central location ("mean"), S_v for the scale ("dispersion"), and $IC_{u,v}$ and IS_v for the set of confidence intervals about C_u and S_v, respectively. The subscripts u and v indicate the methods used to obtain the sample estimate. Extensive simulations for a number of common situations realizable in small to large samples of cluster radial velocities allow us to identify minimum variance estimators. We also explore the estimation of confidence intervals using the jackknife and bootstrap resampling techniques, and compare these methods to simple formulas based on sample estimates of central location and scale. Our tests reveal that the family of location and scale estimators based on Tukey's biweight prove consistently superior for most applications. Confidence intervals on location based on the biweight also prove superior. Estimators of confidence intervals on scale require resampling—although bootstrapping is preferred, less computationally demanding estimators based on the jackknife of the biweight scale are shown to be adequate for most situations.

I. INTRODUCTION

Recent observational and theoretical evidence suggests that many clusters of galaxies have yet to achieve dynamical equilibrium, where an isotropic distribution of galaxy orbits is expected (see Fitchett 1988 and Sarazin 1987 for good reviews of this topic). Observations of subclusters within clusters vitiate models that assume a single gravitational potential well (Forman *et al.* 1981; Geller and Beers 1982). In the regions of massive D or cD galaxies a mix of galaxy orbits may exist: a bound, low-dispersion population soon to be consumed by the central galaxy as well as galaxies on predominantly radial orbits (Tonry 1985; Bothun and Schombert 1988; Bothun and Schombert 1989). "Central" substructure may exist in a number of clusters, clouding interpretation of the dynamical significance of the D/cD galaxy (Fitchett and Webster 1987; Fitchett and Merritt 1988; Oegerle, Fitchett, and Hoessel 1989). On larger scales, infalling spirals (or groups of spirals) may have yet to complete their first cluster crossing, and thus cannot be in equilibrium with the underlying mass distribution (Huchra 1985; Binggeli, Tammann, and Sandage 1987). Further complications are introduced by the likelihood that clusters may have formed at the interstitial locations of large-scale "bubbles" of luminous matter, with the result that radial-velocity surveys of clusters will include a correlated (rather than random) foreground/background contamination (de Lapparent *et al.* 1986). Consequently, the interpretation of redshift surveys in clusters of galaxies may be confounded by the presence of asymmetries, heavier-than-Gaussian tails, and outliers in velocity distributions.

Although advances in multifiber spectroscopy will soon make available large redshift samples for many clusters, it is clear that the behavior of statistical estimators in the small-n limit will continue to be an important issue. Critical evaluation of the statistical significance and dynamical importance of local properties in clusters will require examination of small subsamples of velocities (Dressler and Schectman 1988). Surveys of poor clusters and compact groups are constrained to quantify kinematic properties based on small numbers of velocities (Huchra and Geller 1982; Geller and Huchra 1983; Beers *et al.* 1984; Maia, DaCosta, and Latham 1989; Beers *et al.* 1990). High-redshift clusters will, for the foreseeable future, have a limited number of measured redshifts per cluster (Newberry, Kirshner, and Boroson 1988, and references therein).

Velocity dispersions in clusters commonly form the basis for dynamical estimates of cluster mass via application of the virial theorem. Because a virial mass estimate is proportional to the square of the velocity dispersion, small errors in the estimate of the dispersion result in correspondingly larger errors in the derived mass. Studies of the velocity-dispersion profiles in rich clusters require binning of velocities into small subsamples (Kent and Gunn 1982; Dressler and Schectman 1988; Ostriker *et al.* 1988), which may increase the effect of a few deviant velocities. Recent attempts to incorporate both velocity and spatial information to evaluate the existence of substructure (Dressler and Schectman 1988; West and Bothun 1990) require local estimates of the velocity dispersion. This procedure necessarily involves no more than a few galaxies (typically ten) with measured redshifts in a given subsample. The power of such tests could clearly

 0004-6256/90/010032-15$00.90

be improved with the adoption of efficient, as well as resistant, estimators.

The majority of the statistical tools conventionally used for the analysis of radial-velocity distributions in clusters of galaxies were first applied at a time when the apparently satisfactory agreement of number-density profiles with isothermal models and the paucity of kinematical information for most clusters led to the assumption that the observed radial velocities were drawn from a Gaussian parent population. The difficulties outlined above are sufficiently severe to prompt a new look at methods for quantifying cluster velocity distributions.

The mean and standard deviation are known to be optimally efficient (minimum variance) estimators of central location and scale when the underlying population is Gaussian. Unfortunately, both estimators offer *little or no resistance* in the presence of outliers because the tails of any sample are heavily weighted. Canonical estimates of confidence intervals about the central location and scale also suffer when the normality assumption is violated. Ideally, we would prefer estimators that are efficient for the Gaussian case, retain high efficiency in the presence of significant deviations from the normality assumption (such as heavy tails), and are not grossly distorted by the effects of outliers. Such estimators have the advantage that they could be applied to data suspected of containing outliers in an objective, rather than blatantly subjective, manner.

In Sec. II we review the procedures commonly used to quantify cluster velocity distributions. In Sec. III we discuss alternative estimators of location and scale. The ambitious task of assigning realistic confidence intervals on these estimators is discussed in Sec. IV. In Sec. V we compare these alternative estimators with the canonical methods by simulating a variety of commonly occurring situations. We present our recommendations in Sec. VI. Section VII presents a discussion and summary.

II. THE CANONICAL METHODOLOGY

Before we present suggested improvements, it is enlightening to review the procedures presently in common use.

Yahil and Vidal (1977) test the hypothesis that the observed radial-velocity distributions in clusters of galaxies are consistent with a Gaussian. Based on the limited velocity data available at that time, they find that the Gaussian is a good description except for contamination due to "field" galaxies. The tests that these authors apply include the *a* test, the *u* test, and the *W* test of non-normality, a powerful omnibus test due to Shapiro and Wilk (1965). Having identified a number of clusters that deviate from normality, Yahil and Vidal apply a "3-sigma" cleaning procedure to eliminate presumed field galaxy contamination. Chapman, Geller, and Huchra (1987), Chapman, Geller, and Huchra (1988), and Postman, Geller, and Huchra (1988) measure cluster means and dispersions after trimming from the tails of the velocity distribution until a χ^2 test is unable to reject the Gaussian hypothesis, a method recommended by Fernley and Bhavsar (1984) based on *N*-body simulations of clusters. The "adhoc" clipping procedure discussed by Zabludoff, Huchra, and Geller (1990) is a simplified form of the statistical procedure known as "gapping." Unfortunately, the method employed by these authors does not account for the fact that gaps in a given dataset, when used to assess the presence of outliers, should be properly weighted by their position within the data batch—a large gap in the central portion of a sample is clearly more significant than an equivalent gap in the sparsely populated tails (see Wainer and Thissen 1976 for a detailed discussion of this point).

As we show in Sec. VI, pruning procedures that compare to an assumed probablistic model (such as the Gaussian) are generally poorly behaved due to the nonresistance of the standard deviation and their insensitivity to symmetrically distributed outliers. In addition, there is often no clear indication that a given velocity sample differs from a Gaussian. The *a* test, *u* test, and *W* test, for example, are sensitive (with varying power) to different kinds of deviations, and can yield conflicting opinions as to whether clipping is indicated for a given cluster. Of course, because the underlying distribution may not be Gaussian at all, the application of these model-dependent techniques to quantify cluster velocity distributions is surely not an optimum procedure.

Quintana and Lawrie (1982) attempt to reduce the sample-to-sample variance of their estimates of velocity dispersions in clusters by *defining* the mean of a sample of cluster velocities to be coincident with the velocity of the D/cD galaxy (where present), then measuring the standard deviation about this value. While this method may appear reasonable for some clusters, it is ill-defined for clusters in which there is more than one D/cD galaxy (e.g., Coma, A754, A957), and will not apply at all when a cluster (or subcluster) does not possess a single dominant galaxy. Given the recent claims of displacements in velocity space of a number of cD galaxies (Bothun and Schombert 1988; Bower, Ellis, and Efstathiou 1988; Hill *et al.* 1988; Sharples, Ellis, and Gray 1988; Bothun and Schombert 1989), we are all the more hesitant to apply this technique.

In a widely referenced paper, Danese, De Zotti, and di Tullio (1980, hereafter referred to as DDd) take the Gaussian distribution for cluster velocities as given and discuss the application of classical statistical methods for calculating the expected errors in estimates of cluster means and dispersions. The procedures they employ, however, are strongly dependent on the normality of the underlying velocity distribution, and are inappropriate in the presence of even slight deviations from this assumption (Pearson 1931; Box 1953). In general, an incorrect normality assumption will produce *unnecessarily large* confidence intervals about the mean and *optimistically short* intervals about the dispersion.

III. ALTERNATIVE METHODS FOR ESTIMATION OF LOCATION AND SCALE

Recent developments in statistical research have provided a number of improved methods for quantifying the nature of datasets without requiring models for the underlying population. Many of these developments are summarized in the highly readable book, *Understanding Robust and Exploratory Data Analysis* (Hoaglin, Mosteller, and Tukey 1983, hereafter referred to as UREDA)—the astronomer employing robust analysis techniques can be pointed toward no better source. Rice (1988) presents a discussion of the pertinent statistical issues on a less technical level.

Below we begin with a number of definitions required for the present discussion. We have drawn freely from many sections of UREDA in preparing the following comments.

a) Resistance, Robustness, and Efficiency

The notions of resistance, robustness, and efficiency are central to an appreciation of why the canonical methodology

should be viewed as suspect, and thus bear some discussion.

Resistance implies an insensitivity to localized misbehavior in data. A resistant estimator, such as the sample median, changes little when a small part of the data is replaced by new (possibly very different) numbers. In contrast, the sample mean is clearly nonresistant. Resistant methods depend on the bulk of the data and pay little attention to outliers.

Robustness implies insensitivity to the assumed nature of the population from which data are drawn. A robust estimator will work well (that is, exhibit a small variance) without relying on a detailed match of the data to the underlying probabalistic model (if indeed one even has such a model). Viewed from a robust perspective, models that initially compare to a Gaussian model prior to the pruning of data are clearly undesirable. The kinematic information contained in a velocity distribution should best be summarized by statistics that do not depend explicitly on the analyst's opinion of whether or not the Gaussian assumption is satisfied.

Efficiency refers to the quality of information one can derive from data employing a given method. In data-limited applications efficiency is often of primary importance. An efficient method will extract the required information from the data in hand, whereas an inefficient method may require a doubling or even tripling of the sample before an equivalent result can be obtained. We could improve efficiency, for example, by decreasing the variance of an estimator, or by producing a shorter confidence interval. Techniques for measuring the efficiency of estimators of central location, scale, and associated confidence intervals are discussed in Sec. V*e* below.

b) Order Statistics, Ranks, and Depths

In the estimators that follow, we make use of the so-called "order statistics" of a sample, defined from the data sorted in ascending order, i.e.,

$$x_i, x_{i+1}, \ldots, x_n, \tag{1}$$

where x_i is the ith smallest observation, and is referred to as the ith order statistic of the sample. The rank of a given observation is defined in one of two ways—counting up from the smallest value or counting down from the largest value. In the first case we are measuring the observation's upward rank; in the latter, its downward rank. Note that according to this definition x_i has an upward rank i and downward rank $n + 1 - i$. The depth of a given observation is then defined as the smaller of its upward and downward rank.

Using the notion of depth one can extract a number of summary statistics for a given sample. The most familiar such statistic is the sample median (M). For n odd, the median's depth is defined as $(n + 1)/2$. For n even, the median is taken as halfway between x_k and x_{k+1}:

$$M = x_{k+1}, \quad n = 2k + 1,$$
$$M = \tfrac{1}{2}(x_k + x_{k+1}), \quad n = 2k. \tag{2}$$

Another useful, though not as familiar, set of summary statistics are the fourths of a data sample. The fourths are defined such that

$$\text{depth of fourth} = \frac{[\text{depth of median}] + 1}{2}, \tag{3}$$

where the brackets $[x]$ stand for the largest integer not greater than x. The difference in the fourths of a dataset (as counted from each end of the sample) is often referred to as the Inter-Quartile Range (IQR).

c) Resistant and Robust Estimators of Location

1) The median

Perhaps the classic example of a resistant location estimator is the sample median M, defined above. The median has the (sometimes desirable) characteristic of putting zero weight on all but one (or two) central order statistics. The median thus achieves a high breakdown bound, defined as the largest possible fraction of observations for which there exists a bound on the change of the estimate when that fraction is altered without restriction (Hampel 1968). For the median, the maximum breakdown bound is 1/2 (for the sample mean the breakdown bound is identically zero). The price one pays for such a high breakdown bound is sensitivity to rounding and grouping of the data. Nevertheless, for tiny samples ($n \leqslant 5$) there is often no better choice. In the discussion below, the notation C_M will indicate the estimate of central location provided by the median.

2) The broadened median, trimean, and trimmed means

We would prefer estimators of central location that incorporate the inherent resistance of the sample median while improving somewhat on its reliance on one or two data points. One useful measure which accomplishes this goal is the so-called broadened median, or BM (Andrews *et al.* 1972). The BM estimator averages the sample median and either one or two order statistics on either side of the median, depending on sample size. Using the conventions of Rosenberger and Gasko (1983):

> For n odd, BM is the average of the three central order statistics when n is in the range $5 \leqslant n \leqslant 12$; the five central order statistics for $n \geqslant 13$. For n even, BM is a weighted average of the central four order statistics for $5 \leqslant n \leqslant 12$ with weights (1/6, 1/3, 1/3, and 1/6); for $n \geqslant 13$ it is the weighted average of the central six order statistics with weights 1/5 to the central four and weights 1/10 to the end ones of the six.

We use the notation $C_{\rm BM}$ to indicate the central location estimator obtained from the broadened median.

The trimean (TRI) provides an alternative, and computationally simpler, measure of central location to BM. The TRI estimator is based on an average of the sample median and the lower and upper fourths:

$$\text{TRI} = \tfrac{1}{4}(F_l + 2M + F_u), \tag{4}$$

where F_l and F_u represent the lower and upper fourths, respectively. $C_{\rm TRI}$ is the trimean estimator of central location.

For many groups and clusters of galaxies, we are concerned about the influence of a few extreme velocities on our estimate of central location. A useful technique in such cases is to simply calculate the mean of the dataset after a certain percentage (α) of the data have been trimmed from the tails. For example, a 20% trimmed mean, T_{20}, of a sample of size 10 is the straight average of the six remaining observations after setting aside the two largest and two smallest data points. We would like to know which fraction α is optimal for a variety of Gaussian and non-Gaussian situations. Simulations of samples with $n = 10$ and $n = 20$ for contaminated normal distributions (Rosenberger and Gasko 1983) indicate that for $n = 10$ a trim between 10% and 20% produces a minimum variance in the estimator. For $n = 20$, between 5% and 10% behaves well. For long-tailed distributions such as the logistic or Cauchy, somewhat heavier trimming

is indicated, on order 25%. The 25% trimmed mean, T_{25}, is nothing more than the average of the data within the central quartiles, and is often referred to as the midmean or MID. Below we indicate the central location estimated with the trimmed mean as C_α. The midmean location estimate is $C_{\rm MID}$.

3) The biweight

The biweight location estimator was suggested by Tukey after the Princeton Robustness Study (Andrews *et al.* 1972) as a possible improvement for non-Gaussian or contaminated normal distributions, and comes from the broad family of estimators known as M estimators of location. The defining characteristic of M estimators is the minimization of a function of the deviations of each observation from the estimate of location. In particular, the biweight uses a so-called redescending Ψ function, which reaches a maximum at some data-defined distance from the center of a distribution, and falls continuously to zero a farther distance out. The biweight location estimator requires an auxiliary estimate of scale. The auxiliary scale estimator in general use is the MAD (median absolute deviation from the sample median), defined in Sec. III*d1* below. The definition of the biweight location estimator is

$$C_{\rm BI} = M + \frac{\Sigma_{|u_i|<1}(x_i - M)(1-u_i^2)^2}{\Sigma_{|u_i|<1}(1-u_i^2)^2}, \quad (5)$$

where M is the sample median and u_i are given by

$$u_i = \frac{(x_i - M)}{c\,\mathrm{MAD}}. \quad (6)$$

The constant c is known as the "tuning constant" and is chosen to give $C_{\rm BI}$ high efficiency for a broad range of distributions. The best balance of efficiency for location estimation is found for $c = 6.0$, which includes data up to four standard deviations from the central location (Mosteller and Tukey 1977).

What we define above is the so-called *one-step* biweight estimator. One might seek to improve the estimate of central location by iteration—taking M as a first guess, calculating the u_i from the data, obtaining $C_{\rm BI}$, substituting the $C_{\rm BI}$ estimate in the place of M in the above formulas, and continuing the process to convergence. Convergence is generally rapid, and usually requires no more than a few steps. For most samples the one-step estimator is sufficient.

As shown by the simulations of Goodall (1983), $C_{\rm BI}$ retains high efficiency ($>75\%$) for Gaussian as well as non-Gaussian distributions, even for n as small as 5. For $n \geqslant 10$, $C_{\rm BI}$ is better than 80% efficient. By way of comparison, the sample mean is 100% efficient for Gaussian distributions but rapidly drops to zero efficiency in the presence of even modest deviations from normality.

Any number of location estimators can be defined along the lines of the above examples, and one of those suggested above may not be the optimal estimator in any particular situation. However, all are superior to the sample mean, C_μ, both in terms of resistance and robustness. Given that a pure Gaussian population is rarely expected in observed cluster velocity distributions, we are encouraged to employ one or more of the above estimators.

d) Resistant and Robust Estimators of Scale

The standard deviation (S_σ) is far and away the most commonly used estimator for the scale of a dataset. Unfortunately, the standard deviation lacks both resistance and robustness. The careful worker usually will report several estimates of S_σ based on the inclusion or exclusion of a small number of potential outliers. Estimates of physical quantities such as a virial mass, or correlations of cluster properties such as x-ray luminosity with velocity dispersion (Quintana and Melnick 1982), will regrettably include additional scatter or bias due to an individual observer's choice of a "preferred" dispersion.

Fortunately, there are a number of simple alternative measures of scale available. In all cases, the estimators defined below are either asymptotically equivalent to S_σ or can be normalized to an identical metric.

1) The median absolute deviation

The median absolute deviation from the sample median, or MAD, is defined as

$$\mathrm{MAD} = \mathrm{median}(|x_i - M|). \quad (7)$$

MAD retains resistance because of the use of the median rather than an average of the second moments as with the standard deviation. Andrews *et al.* (1972) demonstrate the superiority of MAD over a number of simple alternative scale estimators. In particular, MAD possesses a breakdown bound of 1/2, and thus is tolerant to some asymmetry in the sample. For a standard normal population, $N(0,1)$, MAD is equal to 0.6745. We normalize to the same metric as S_σ by defining $S_{\rm MAD} = \mathrm{MAD}/0.6745$.

Simulations for samples of size $n = 20$ (Iglewicz 1983) indicate that $S_{\rm MAD}$ is particularly suited for estimates from heavy-tailed distributions, where its efficiency exceeds 90%. Unfortunately, for samples that are Gaussian, or nearly so, the efficiency of $S_{\rm MAD}$ drops precipitously, to approximately 40%.

2) The f spread and f pseudosigma

The f spread is defined as the difference between the upper and lower fourths of a data set:

$$f = F_u - F_l, \quad (8)$$

and is highly recommended for quick hand calculation. For large n the f spread of a standard normal distribution is 1.349, thus we define an f "pseudosigma" as $S_f = f/1.349$. (See Beers and Gebhardt 1990 for a discussion of additional pseudosigmas and correction factors for the small-n case). The S_f estimator has moderate efficiency ($>50\%$) for non-Gaussian situations but is still only 40% efficient in the case of the Gaussian distribution. One appropriate use of S_f is to alert the user to strongly non-Gaussian behavior (such as one or several outliers) which might distort the canonical standard deviation.

3) The biweight

The simple scale estimators discussed above are not grossly inefficient, but it is desirable to do better. Although computationally more complex, the biweight estimate of scale has proven to be superior in many respects. The biweight scale estimator is defined as

$$S_{\rm BI} = n^{1/2} \frac{\left[\Sigma_{|u_i|<1}(x_i - M)^2(1-u_i^2)^4\right]^{1/2}}{\left|\Sigma_{|u_i|<1}(1-u_i^2)(1-5u_i^2)\right|}, \quad (9)$$

where u_i are defined as above, and the tuning constant c is set equal to 9.0. Like the biweight location estimator, $S_{\rm BI}$ may be calculated with an iterative procedure or stopped after one step. The results are similar in either case. The biweight scale asymptotically approaches S_σ when the sample is taken from a Gaussian population. Lax (1985) shows that in the Gaussian situation ($n = 20$) the biweight scale estimator obtains 87% efficiency (more than twice the efficiency of $S_{\rm MAD}$). For the deviations from Gaussian distribution that he simulated, the biweight never dropped below 86% efficiency. Goodall (1983) suggests that the biweight also performs well for samples as small as $n = 10$.

4) The gapper—a scale estimator based on gaps

Wainer and Thissen (1976) present an interesting alternative scale estimator based on the gaps between order statistics. For the order statistics $x_i, x_{i+1}, \ldots, x_n$, with gaps defined by

$$g_i = x_{i+1} - x_i, \quad i = 1,\ldots,n-1 \quad (10)$$

and a set of approximately Gaussian weights:

$$w_i = i(n-i), \quad (11)$$

they obtain a robust estimator of scale:

$$S_G = \frac{\sqrt{\pi}}{n(n-1)} \sum_{i=1}^{n-1} w_i g_i. \quad (12)$$

The preliminary work of Wainer and Thissen suggests that $S_{\rm G}$ possesses efficiency exceeding 90% for Gaussian samples as small as $n = 10$, and moderately high efficiency for a variety of contaminated Gaussian samples.

IV. ALTERNATIVE METHODS FOR CONFIDENCE INTERVAL ESTIMATION

As stressed by Efron and Tibshirani (1986), the estimation of a confidence interval for a given statistic is a fundamentally more ambitious undertaking than the specification of that statistic's expected value. Indeed, the development of satisfactory methodology for confidence interval estimation is still an effort very much in progress at the present time. The particular method one chooses to employ will depend, to a great extent, on the effort one is willing to expend and on the demands of the application at hand. The tradeoff is generally an increase in computation in exchange for a decrease in assumptions concerning the nature of the underlying population from which one is sampling. It is important to recognize the degree to which a given confidence interval depends on these assumptions.

For choosing optimum sample estimates of location and scale, one is primarily concerned with the issue of variance. For confidence interval estimation, one must also consider the degree to which a given method provides the quoted coverage. For example, to what extent does a quoted 1σ interval cover the true 68% range of variability in the estimated statistic? The answer to this question cannot, in general, be obtained analytically, but rather must be obtained by conducting large and time-consuming Monte Carlo experiments. We have chosen not to fully explore the coverage issue in the present paper, but rather to concentrate on the variance of simple confidence interval estimators. To some extent, the various bootstrap methods allow one to assume that the quoted coverage is indeed being attained (Efron 1987), but even this is still a matter of debate among statisticians, particularly for small to moderate sample sizes (Schenker 1985).

We proceed in our discussion below in order of increasing computational effort, beginning with interval estimates based on sample estimates of location and scale described above, and concluding with a discussion of two popular resampling techniques, the jackknife and the bootstrap.

a) Simple formulas for Location Confidence Intervals

The classical method for estimating population $100(1-\alpha)\%$ confidence intervals on the central location of a symmetric, Gaussian-distributed set of data is to compute

$$IC_{u,v} = C_u \pm \frac{t_{n-1} S_v}{\sqrt{n}}, \quad (13)$$

where t is the $100(1-\alpha/2)$ percentage point of the t distribution for $n-1$ degrees of freedom, and C_u and S_v are estimators of location and scale. In their study, DDd apply this method to obtain confidence intervals on C_μ for clusters of galaxies, under the assumption of a Gaussian distribution of velocities. When the data come from contaminated or heavier-tailed distributions than Gaussian, however, C_μ and S_σ are no longer efficient estimators of location and scale; the resulting confidence interval will include considerable scatter from sample to sample. In addition, the nonresistance of S_σ for estimation of scale may arbitrarily inflate the estimate of confidence intervals about the central location.

Iglewicz (1983) discusses a number of alternative methods for assigning robust confidence intervals of the form $C \pm t^* S/\sqrt{n}$, calculated using the improved estimators of location and scale discussed above. This simple form is useful because the sampling distribution of central location estimators is close to Gaussian, even when the parent population itself is quite far from normality (a consequence of the Central Limit Theorem). In order to assure that the quoted coverage is attained, the t^* are chosen to give intervals of *at least* $100(1-\alpha)\%$ for a variety of Gaussian and non-Gaussian distributions. Simulations for sample sizes $n = 10$ through $n = 100$ indicate that the values t^* are well approximated in the following manner. For the pair of estimators C_M and S_f, take $t^* = t_{n-1}/1.075$. For the biweight estimators, $C_{\rm BI}$ and $S_{\rm BI}$, Mosteller and Tukey (1977) recommend $t^* = t_{0.7(n-1)}$, which indicates the t value with $[0.7(n-1)]$ degrees of freedom. Thus the robust confidence intervals given by

$$IC_{M,f} = M \pm \frac{t_{n-1} S_f}{1.075\sqrt{n}} \quad (14)$$

and

$$IC_{\rm BI,BI} = C_{\rm BI} \pm \frac{t_{0.7(n-1)} S_{\rm BI}}{\sqrt{n}} \quad (15)$$

are recommended. For convenient hand calculation, the former is superior. For more efficient but computation-intensive implementation, the latter is preferred. When sampling from a Gaussian distribution, the robust confidence intervals represented by Eqs. (14) and (15) *tend to be longer* than those obtained by the standard methods. For non-Gaussian data, however, the robust intervals can be *substantially shorter* than the standard results.

One may also choose to estimate confidence intervals on the location directly from the sample by employing resampling techniques such as the jackknife or bootstrap, described in detail below.

b) Simple Formulas for Scale Confidence Intervals

In order to assess the confidence intervals on an estimate of scale, one might ask what the expected variance of a given scale estimator is when the data come from a population with a known analytic form. In the case of a Gaussian population, the variance of the standard deviation S_σ closely follows a χ^2 distribution (Hoel 1971). By assuming that velocities in clusters of galaxies are distributed normally, DDd apply the classical formula for obtaining confidence intervals about S_σ:

$$IS_{\sigma\pm} = [(\nu/\chi^2_{\mp})^{1/2} - 1]S_\sigma, \tag{16}$$

where ν is the number of degrees of freedom $(n-1)$.

Unlike sampling distributions for estimates of central location, the sampling distribution of a scale estimator is quite sensitive to the nature of the parent population. For example, as discussed by Miller (1968), the interval given by Eq. (16) relies upon a theoretical variability which is *only* valid in the Gaussian situation. Because few cluster velocity samples will be drawn from Gaussian parent populations, we cannot rely on the classical theory to give reasonable confidence limits. Also note that the classical procedure expressed in Eq. (16) assumes that we have made a reasonable estimate of the scale S_σ, which of course is fraught with nonresitance of its own.

Because of the sensitivity of scale intervals to the form of the parent population from which the data are drawn, simple formulas such as Eq. (16) should not be expected to perform well for the variety of populations expected for distributions of cluster velocities. Direct assessment via resampling is indicated.

c) Resampling Intervals for Location and Scale

1) The statistical jackknife and variations

The statistical jackknife was introduced as a rough and ready tool for assessing the stability of sample statistics for which no asymptotic theory exists, or in cases where the data analyst has little or no idea what the underlying parent population resembles. The jackknife has been employed by a number of workers to estimate errors in astrophysical quantities, for example, viral masses and mass-to-light ratios (Bothun *et al.* 1983; Beers *et al.* 1984; Ostriker *et al.* 1988).

The jackknife produces an estimator based on n "pseudovalues." Each pseudovalue of a statistic is calculated after the removal of one datum from the sample. That is, we obtain the set of pseudovalues y_{*j} such that

$$y_{*j} = n\, y_{\rm all} - (n-1)y_j, \quad j = 1,...,n, \tag{17}$$

where $y_{\rm all}$ is that value of the statistic calculated using all the data, and y_j is the statistic calculated after excision of the jth point. The average of y_{*j} is taken as the jackknife estimate of the statistic:

$$y_* = \frac{\Sigma y_{*j}}{n}. \tag{18}$$

The variance of y_* is given by

$$s^2_* = \frac{\Sigma y^2_{*j} - (1/n)(\Sigma y_{*j})^2}{n(n-1)}. \tag{19}$$

Tukey (1958) proposed the central assumptions of the jackknife technique, that the pseudovalues of Eq. (17) could be treated as approximately independent and identically distributed random variables for a variety of situations. If this assumption holds true, then the statistic s_* should have an approximate t distribution with $n-1$ degrees of freedom. A jackknife estimate of location interval thus takes the form

$$IC_{J_u} = C_u \pm t_{n-1}s_*. \tag{20}$$

When the sample mean is jackknifed, simple algebra reveals that s^2_*, the jackknife estimate of its variance, is identically equal to the estimate obtained from the sample standard deviation S^2_σ/n. This property suggests that jackknife intervals for other robust location estimators, such as $C_{\rm BI}$, might be profitably employed.

The jackknife technique is also useful for constructing approximate confidence intervals of scale estimators:

$$IS_{J_v} = S_v \pm t_{n-1}s_*. \tag{21}$$

One concern when obtaining confidence intervals for scale estimators from Eq. (21), is that the pseudovalues calculated by jackknifing a given scale estimator can become negative, which is unphysical. We can avoid this annoyance by jackknifing the logarithm of the scale estimator rather than the estimator itself. The resulting intervals will be asymmetric, consistent with the fact that scale estimators are unbounded in the positive sense but remain bounded by zero at the low end. The procedure is similar to that defined above, except one finishes by taking the antilog. The confidence interval is then

$$IS_{JL_v} = \text{antilog}(\log S_v \pm t_{n-1}s_{*L}), \tag{22}$$

where s_{*L} is obtained by jackknifing $\log S_v$. For a worked example of these calculations, see Mosteller and Tukey (1977). As noted by Mosteller and Tukey, the use of the logarithm may introduce a bias in the estimate, but one that is at least partially compensated for by the bias-reduction character of the jackknife (Miller 1974).

One technical point bears mentioning. The jackknife confidence intervals are actually the intervals about the jackknife estimate y_* [Eq. (18)], rather than about the estimator one is studying. In practice, the difference between the jackknife estimators and the estimators C_u or S_v is small; this slight inconsistency should not discourage the use of the jackknife in the specification of approximate confidence intervals (see Yang and Robinson 1986).

2) Bootstrap intervals

Rather than making assumptions concerning the form of the sampling distribution for a given statistic, as does the jackknife, the bootstrap technique relies on the generation of large numbers of *nonindependent* samples (based on random draws with replacement) from the original dataset. An empirical sampling distribution is obtained by calculating the statistic of interest for each of the "bootstrapped" datasets. To the extent that the bootstrap version of a statistic's sampling distribution matches the real (asymptotic) sampling distribution, one can use the bootstrap to construct confidence intervals on the estimated value of this statistic (see Efron and Tibshirani 1986 for a review of the theory and development of this technique). Below we discuss four boot-

strap confidence interval techniques and comment on their application for velocity data in clusters and groups of galaxies.

i) Standard intervals. If we are willing to assume that the bootstrap sampling distribution is unbiased and normal in shape, confidence intervals can be obtained directly from an estimate of the standard deviation of this distribution. This assumption is essentially the same as made by the jackknife, except that the jackknife makes an explicit attempt to correct for the effects of bias. The standard bootstrap intervals for location and scale are given by

$$IC_{B_u} = C_u \pm zs_b \tag{23}$$

and

$$IS_{B_v} = S_v \pm zs_b, \tag{24}$$

respectively, where z is the $100(1 - \alpha/2)$ percentage point of a standard normal variate. We use the normal deviate rather than a deviate from the t distribution as before because t asymptotically approaches z in the limit of large numbers of bootstrap samples. Standard intervals are useful when one does want to carry out a large amount of bootstrap resampling; after all, the number of bootstraps merely has to be great enough to beat down the sampling noise in s_b. Efron (1987) suggests that as few as 25 bootstraps can give reasonable results; 100 replications are recommended. An astrophysical example of the Standard Interval calculation can be found in Barrow, Bhavsar, and Sonoda (1984), and Ling, Frenk, and Barrow (1986). These authors obtain Standard Intervals to assess errors in the determination of correlation coefficients and correlation lengths from catalogs of galaxy positions.

ii) Percentile intervals. One may generalize the above result by abandoning the assumption that the sampling distribution is normal, and obtaining confidence intervals directly from the empirical sampling distribution itself. That is, one takes the Percentile Intervals IC_{P_u} and IS_{P_v} to be defined by the bootstrap estimate $\alpha/2\%$ in from each endpoint of the sampling distribution. Clearly, stable estimates of percentile intervals will require considerably more bootstrap sampling than the standard intervals. Efron (1987) suggests that 500–1000 bootstrap samples are required to obtain accurate percentile intervals.

iii) Bias-corrected percentile intervals. The Bias-Corrected Percentile Intervals, IC_{BC_u} and IS_{BC_v} attempt to correct for the problem of bias in the bootstrap sampling distribution. Simply put, if the median of the bootstrap sampling distribution is not equal to the maximum-likelihood estimator of the statistic of interest, the percentile method will not yield correct coverage. The bias-corrected method shifts the regular percentile intervals by an amount proportional to the difference between these two quantities. Efron (1987) introduces a refinement to this procedure which increases its realm of application, particularly for the case of small to moderate n and for samples drawn from asymmetric parent populations. The refined intervals $IC_{BC_u^*}$ and $IS_{BC_v^*}$ are second-order asymptotically correct for a small price in the form of increased computation.

V. A COMPARISON OF ESTIMATORS

In the sections above, we have defined a number of alternative estimators for quantifying the nature of velocity distributions for clusters and groups of galaxies. We now wish to compare the relative efficiencies of these estimators by examining their behavior when applied to simulated datasets that cover the range of deviations from normality we might expect in velocity distributions for typical clusters and groups of galaxies. It is important at this point to emphasize that we are not attempting to recover the input parameters of the simulations. Rather, we are simply evaluating the statistical behavior of each estimator. The "true" parameters that correspond to a given situation will not, in general, be revealed by the use of particular estimators of location and scale—a maximum-likelihood analysis of a completely specified physical model would be required.

a) Case A—Superposed Gaussians of Identical Location

One situation we would like to simulate is the case where a simple Gaussian distribution $G(\mu,\sigma)$ is observed in the presence of a broader Gaussian of identical location but differing scale $G(\mu,k\sigma)$. That is

$$G'_A(\mu,\sigma,k,\alpha) = \alpha G(\mu,\sigma) + (1 - \alpha)G(\mu,k\sigma), \tag{25}$$

where α and $(1 - \alpha)$ represent the relative fraction of values taken from the two distributions ($10\% \leqslant \alpha \leqslant 50\%$), and k is a number between 2 and 5. Such a mixture might be realized in several astrophysical situations—(a) observations of a subcluster within a cluster (or cluster within a supercluster), (b) observations of cluster galaxies in the presence of correlated foreground/background contamination, or (c) observations of galaxies in the same cluster but with differing orbital distributions (e.g., ellipticals versus spirals in the Virgo Cluster). For a given cluster, the presence of a long-tailed symmetric component may not be readily recognizable, but at the very least we should employ estimators whose efficiencies are not unduly influenced by its existence.

b) Case B—q wild

A second instance that occurs in real clusters is the q-wild situation. An otherwise uncontaminated Gaussian distribution $G(\mu,\sigma)$ is observed in the presence of a small number (q) of observations from a much broader, identically located Gaussian population $G(u,k\sigma,q)$. Here q is taken to be between 1 and 3, and k is chosen between 5 and 10. The q-wild situation will usually occur in small datasets, such as groups or subclusters, where the observer is typically forced to resort to a subjective censoring of suspected outliers in order to report "reasonable" estimates of location and scale.

$$G'_B(\mu,\sigma,k,q) = G(\mu,\sigma) + G(\mu,k\sigma,q). \tag{26}$$

c) Case C—The General Case

There are many scenarios that we could posit to corrupt a simple Gaussian velocity distribution of cluster velocities, such as a mixture of galaxies with kinematically different central locations, or the superposition of a cluster distribution on the general cosmological background. Of course, even without these complications, we have argued that there is no good reason to expect that a simple Gaussian is representative of the underlying velocity distribution in most clusters. Long asymmetric tails, for example, present a particular challenge to most estimators. An optimal resistant and robust estimator should perform well even when a substantial fraction of the data is displaced to one side or the other from the bulk of the batch. Thus we would like to include in our simulations a formulation for a family of deviations from

normality that accounts for the problems of increasing asymmetry and tail weight.

Hoaglin and Peters (1979) discuss one such function. If Z represents a random variable from a standard normal distribution, then

$$Y = g^{-1}(e^{gZ} - 1)e^{\frac{1}{2}hZ^2} \tag{27}$$

is said to follow the Tukey $T(h,g)$ distribution. The parameters (h,g) determine the heaviness and symmetry of the tails, respectively. $T(0,0)$ is the standard normal. $T(h,0)$ is symmetric and increases in elongation with increasing h. For example, $T(0.97,0)$ is approximately Cauchy. $T(0,g)$ is log-normal and is increasingly skewed as g increases.

We can allow for an arbitrary choice of location and scale here and in the distributions above by forming

$$X = C + SY, \tag{28}$$

where C and S are the desired location and scale, respectively. For the simulations discussed below, we arbitrarily set $C = 5000$ and $S = 100$.

d) The Simulations

We examine 72 different situations, exploring the parameter space of the three cases outlined above. For each situation, we draw one thousand random samples of size 5, 10, 20, 50, 100, and 200, representative of the typical sizes of velocity samples in groups or clusters.

For each realization, 12 estimators of location C_u are obtained: C_μ, $C_{3\sigma}$, C_M, C_5, C_{10}, C_{20}, C_{MID}, C_{30}, C_{40}, C_{BM}, C_{TRI}, and C_{BI}. Six estimators of scale S_v are obtained: S_σ, $S_{3\sigma}$, S_f, S_{MAD}, S_G, and S_{BI}.

The formulas of Sec. IV*a* are used to construct 68% (1σ) confidence intervals on C_μ, C_M, and C_{BI} based on sample estimates of location and scale. In addition, we obtain jackknifed and refined bias-corrected bootstrap estimates of confidence intervals about the biweight location estimator C_{BI}. For each situation, one thousand bootstrap replications are used to define the empirical sampling distribution.

We obtain 68% confidence intervals about the scale S_σ from the classical formula based on the sample standard deviation IS_σ [Eq. (16)]. We calculate jackknife intervals about S_σ, S_{BI}, and S_G. Intervals are also obtained by jackknifing $\log S\sigma$, $\log S_{BI}$, and $\log S_G$. Refined bias-corrected bootstrap intervals are calculated for the biweight estimator of scale S_{BI}.

e) Relative Efficiencies of the Estimators

We calculate the variance of each statistic from the one thousand realizations simulated for each situation. In order to compare the relative efficiency of the estimators of central location C_u, we obtain the optrim-efficiency:

$$\epsilon_C = \frac{\text{var } C_{u_0}}{\text{var } C_u} \times 100\%, \tag{29}$$

where var C_{u_0} represents the minimum variance estimator tested.

In general, the variance of a scale estimator for a data batch will depend on the central location. To avoid this dependence in comparisons of the variances of S_v we employ (following Iglewicz 1983):

$$\epsilon_s = \frac{\text{var } (\ln S_{v_0})}{\text{var } (\ln S_v)} \times 100\%, \tag{30}$$

where var$(\ln S_{v_0})$ represents the minimum variance estimator tested.

An optimum confidence interval for a given estimator is the shortest interval still providing the quoted coverage. It is outside the scope of the present investigation to establish that a given interval achieves coverage. Rather, we *assume* that the refined bias-corrected bootstrap intervals on location and scale ($IC_{BC^*_u}$ and $IS_{BC^*_v}$) provide accurate measures of the true confidence interval (as suggested by Efron 1987). As discussed below, the biweight estimators of location and scale are superior for a wide variety of situations, so we choose to compare our derived confidence intervals to the bootstrapped interval lengths of these estimators. Again following Iglewicz(1983), we obtain relative efficiencies for a set of interval estimators using:

$$\epsilon_{CI} = \left(\frac{\text{average BC}^*_{BI} \text{ interval length}}{\text{average interval length}}\right)^2 \times 100\%. \tag{31}$$

A similar calculation is used to obtain ϵ_{SI}.

Note that we have defined relative efficiency for confidence intervals such that intervals that provide, on the average, less coverage than the refined bias-corrected bootstrap intervals will have an apparent efficiency *greater* than 100%. This should not be taken as an indication that they are better than our standard of comparison, but rather that they should be considered suspect due to their failure to provide coverage.

f) Results of the Simulations

In Table I we summarize the input parameters, or range of parameters, used for each situation.

To choose the best estimator, one would not expect to refer to a table of efficiencies for every different situation. Rather, one would like to judge which estimators are best suited for a variety of situations. Small differences in relative efficiency ($\leqslant 10\%$) will not be noticeable in the application of these estimators, thus we have "graded" each estimator on a familiar four-point scale. Estimators of central location and scale that achieve relative efficiency $90\% \leqslant \epsilon \leqslant 100\%$ for a given situation are given a score of 4.0, those with $80\% \leqslant \epsilon < 90\%$ are assigned a grade of 3.0, those with $70\% \leqslant \epsilon < 80\%$ are given a 2.0, those with $60\% \leqslant \epsilon < 70\%$ are assigned a 1.0, and those estimators with efficiencies below 60% are given a score of 0. The overall score for a given estimator in each of cases A, B, and C above is obtained from a straight average of its score in each situation. The results are presented in Table II.

Relative efficiencies of confidence intervals need to be considered somewhat differently from the relative efficiencies of sample estimates of location and scale. Indeed, the desire to obtain an acceptable shorthand method of deriving satisfactory intervals encourages us to be somewhat more lenient in applying our decision criteria. For this reason, and because of the limited resolution provided on confidence intervals from only one thousand realizations, we choose to grade the interval efficiencies in the following manner. Intervals with efficiencies in the range $81\% \leqslant \epsilon < 100\%$ (corresponding to an actual ratio of interval length within 10% of the bootstrap biweight interval) are given a grade of 4.0, those with $64\% \leqslant \epsilon < 81\%$ (interval ratio between 10% and

TABLE I. Simulations of cluster radial-velocity distributions.

Situation	Parameters
A_1	$(k = 2, \alpha = 0.10)$
A_2	$(k = 2, \alpha = 0.20)$
A_3	$(k = 2, \alpha = 0.50)$
A_4	$(k = 3, \alpha = 0.10)$
A_5	$(k = 3, \alpha = 0.20)$
A_6	$(k = 3, \alpha = 0.50)$
A_7	$(k = 5, \alpha = 0.10)$
A_8	$(k = 5, \alpha = 0.20)$
A_9	$(k = 5, \alpha = 0.50)$
B_1	$(k = 5, q = 1)$
B_2	$(k = 5, q = 2)$
B_3	$(k = 5, q = 3)$
B_4	$(k = 7.5, q = 1)$
B_5	$(k = 7.5, q = 2)$
B_6	$(k = 7.5, q = 3)$
B_7	$(k = 10, q = 1)$
B_8	$(k = 10, q = 2)$
B_9	$(k = 10, q = 3)$
C_1	$(h = 0, g = 0)$ (Gaussian)
C_2	$(h = 0, 0.10 \leq g \leq 0.25)$
C_3	$(h = 0, 0.50 \leq g \leq 0.95)$
C_4	$(0.10 \leq h \leq 0.25, g = 0)$
C_5	$(0.10 \leq h \leq 0.25, 0.10 \leq g \leq 0.25)$
C_6	$(0.10 \leq h \leq 0.25, 0.50 \leq g \leq 0.95)$
C_7	$(0.50 \leq h \leq 0.75, g = 0)$
C_8	$(0.50 \leq h \leq 0.75, 0.10 \leq g \leq 0.25)$
C_9	$(0.50 \leq h \leq 0.75, 0.50 \leq g \leq 0.95)$

20%) are assigned a grade of 3.0, those with $49\% \leqslant \epsilon < 64\%$ (interval ratio between 20% and 30%) are assigned a 2.0, those with $36\% \leqslant \epsilon < 49\%$ (interval ratio between 30% and 40%) are assigned a 1.0, and those estimators with efficiencies below 36% are given a score of 0.

As noted above, the efficiencies of confidence intervals may exceed 100%, indicating that they are shorter than our standard of comparison. We penalize these estimators for their lack of coverage by assigning scores as follows. Estimators with efficiencies in the range $100\% < \epsilon \leqslant 121\%$ are given a 3.5, those with $121\% < \epsilon \leqslant 144\%$ are assigned a grade of 2.5, those with $144\% < \epsilon \leqslant 169\%$ are given a 1.5, those with $169\% < \epsilon \leqslant 256\%$ are assigned a 0.5, and those estimators with efficiencies exceeding 256% are given a score of 0. Scores for estimators of confidence intervals are presented in Table III.

The overall grades assigned an estimator should provide a good indication of its utility in application to real data. Grades above 3.0 should be considered superior. Grades above 2.0 are adequate, while grades in the range 1.5–2.0 should be considered marginal. Estimators with grades below 1.5 indicate that for much of the parameter space explored in each case, the estimator has consistently high variance compared to its competition, and should be generally avoided.

In choosing among estimators, one should also consider the stability of efficiency across the range of parameters for each case. This information is encoded in the form of profile plots in Figs. 1 and 2. For each estimator, the profile plot codes the relative efficiency in the form of height above the horizontal axis. For each case, the situations are considered in the order given in Table I, which corresponds roughly to increasing asymmetry and tail weight. The shading of the profile plots represents the overall score discussed above. A superior score (3.0 or above) is given the darkest shading. Adequate scores (2.0–3.0) are given a light shading. Marginal scores (1.5–2.0), as well as unacceptable ($\leqslant 1.5$) scores, are left unshaded. Large variations of efficiency from situation to situation are clearly reflected in Figs. 1 and 2 as jagged profile plots. Estimators of roughly equivalent efficiency (indicated by a similar shading) that have flat profiles are preferred over those with jagged behavior.

As noted above, the relative efficiencies of confidence intervals sometimes exceed 100%. For these situations, the relative heights of the profiles in Fig. 2 encode the absolute difference between the measured efficiency and 100%.

VI. RECOMMENDED ESTIMATORS

The choice of an estimator should be guided, foremost, by the sample size n. For a given sample size the robustness indicated by the grades in Tables II and III should be considered, as well as the stability of efficiency indicated by the profile plots shown in Figs. 1 and 2. Among estimators of roughly equivalent robustness and stability of relative efficiency, ease of computation might be considered important for some applications. Below we discuss our results for the tiny-, small-, intermediate-, and large-n cases.

a) Tiny n (n=5)

For sample sizes as small as $n = 5$, one is not blessed with a wide variety of choices. Simple robust estimators of location, such as $C_{\rm M}$, $C_{\rm MID}$, and $C_{\rm TRI}$ appear quite satisfactory. $C_{\rm M}$ is particularly useful in maintaining low variance for case B contamination, and is a very good performer over the broad range of case C tails and asymmetry. It is instructive to observe the rapid degradation of the mean C_μ in the face of case A contamination (Fig. 1), as well as for case B (where the mean never achieves greater than 54% relative efficiency). For case C, C_μ is a superior estimator only for the Gaussian situation, C_1, and the mild asymmetry situation, C_2. As tail weight or asymmetry are increased, C_μ fails relative to its competitors $C_{\rm M}$, $C_{\rm MID}$, $C_{\rm BM}$, $C_{\rm TRI}$, and $C_{\rm BI}$. We are encouraged to see $C_{\rm BI}$ perform so well in the tiny-n case.

We now consider the estimators of confidence intervals on location. The canonical estimator $IC_{\mu,\sigma}$ is superior for case A contamination, and adequate for case C, although it breaks down rather rapidly for the extreme-asymmetry and heavy-tail situations. The estimator $IC_{{\rm M},J}$ consistently underestimates the lengths of the interval relative to that obtained by the bootstrap of the biweight location, and should probably be avoided for the tiny-n case. It is of some concern that $IC_{\sigma,\mu}$ underestimates the interval length in the Gaussian case, C_1. The biweight interval $IC_{\rm BI,BI}$ is superior for all three cases. The jackknifed biweight interval $IC_{J_{\rm BI}}$ is adequate or superior, but suffers somewhat from excessive variation in efficiency across situations.

Only two estimators of scale can be recommended in the tiny-n case, the canonical standard deviation S_σ and the gapper $S_{\rm G}$.

The confidence intervals on scale obtained by jackknifing the biweight appear adequate for all three cases, although the variation of efficiency for case B contamination is rather severe. The canonical interval IS_σ is even a little better be-

haved. It is worth noting that, for much of the parameter space we explore, the interval estimates on scale all have a somewhat larger variance than the bootstrap intervals we compare them to.

b) Small n (n = 10)

The median is an adequate or superior estimator of central location for the small-n situation, but the biweight is even better. Among the simple estimators, C_{MID} and C_{TRI} are consistent performers, as are the moderately to heavily trimmed estimators C_{20} and C_{30}. Not surprisingly, the mean is a poor estimator for distributions that differ from Gaussian. The clipped estimator $C_{3\sigma}$ has lower variance than the mean, but possesses considerably higher variance than its competitors.

The confidence intervals provided by $IC_{BI,BI}$ are superior for all three cases. The estimators $IC_{M,f}$ and $IC_{J_{BI}}$ are marginal or superior alternatives.

Among estimators of scale for small n, the canonical standard deviation is superior for case A contamination, fails miserably for case B, and is only adequate for case C. The gapper or biweight estimators of scale are the preferred alternatives. The 3σ clipper cannot be recommended.

The jackknife of the gapper is a superior estimate of confidence intervals on the scale for case A, and adequate for cases B and C, whereas the canonical estimator IS_{σ} is only adequate or marginal. The jackknife of the standard deviation $IS_{J_{\sigma}}$ is superior for case A, and adequate for case C, but only marginal for case B. The jackknife of the biweight is adequate or marginal for cases A and C, but is markedly poor in case B, failing to provide the necessary coverage to handle the q-wild case.

c) Intermediate n ($20 \leqslant n \leqslant 50$)

The intermediate-n case will apply for most velocity samples of clusters now and in the near future. The clipped estimator $C_{3\sigma}$ is superior for this range of sample sizes only for case B contamination, is adequate for case A contamination, and cannot be recommended for the samples with the asymmetric or heavy tails of case C. Lightly trimmed estimators of location C_{10} or C_{20} are better behaved when one considers all three cases of contamination. The resistance of C_{MID} is also encouraging. The median, C_M, appears particularly good for cases with large asymmetry and tail weight, as do the heavily trimmed estimators C_{30} and C_{40}. The best estimator over the range $20 \leqslant n \leqslant 50$ is the biweight location C_{BI}.

The confidence intervals about the central location are clearly dominated by the jackknife of the biweight $IC_{J_{BI}}$, or the computationally simpler biweight estimate $IC_{BI,BI}$. The estimator $IC_{M,f}$ provides a superior or adequate alternative in all cases, and might be preferred for some applications because it always provides coverage.

TABLE II. Efficiency scores for estimators of central location and scale.

N	Case	C_{μ}	$C_{3\sigma}$	C_M	C_5	C_{10}	C_{20}	C_{MID}	C_{30}	C_{40}	C_{BM}	C_{TRI}	C_{BI}	S_{σ}	$S_{3\sigma}$	S_f	S_{MAD}	S_G	S_{BI}
5	A	2.3	...	3.2	...	...	3.9	4.0	4.0	...	3.9	4.0	3.8	4.0	...	0.2	0.3	4.0	0.2
	B	0.0	...	3.7	...	...	1.8	2.2	2.3	...	1.8	2.3	2.7	3.6	...	0.9	1.2	4.0	0.4
	C	1.6	...	2.8	...	...	3.2	3.2	3.3	...	3.2	3.3	3.6	3.4	...	1.4	1.4	3.9	1.2
10	A	1.9	2.4	3.0	...	3.1	3.7	3.7	3.4	3.0	3.6	3.6	3.4	3.2	0.3	0.8	0.9	3.9	1.8
	B	0.0	1.9	2.9	...	1.3	3.0	3.3	3.4	2.9	3.4	2.7	4.0	1.0	0.6	2.3	3.3	2.4	3.2
	C	1.2	2.2	3.6	...	2.3	3.2	3.4	3.8	3.6	3.7	3.0	3.6	2.0	0.1	1.8	2.2	2.8	2.9
20	A	1.6	3.1	1.3	2.2	3.0	3.0	2.8	2.6	2.0	2.0	2.8	3.6	2.9	1.0	0.6	0.3	3.4	2.3
	B	0.0	3.9	1.3	1.9	3.6	3.9	3.3	3.0	2.2	2.3	3.2	4.0	0.0	1.3	1.2	0.9	0.7	4.0
	C	1.6	1.8	2.7	2.3	2.3	2.4	2.7	3.1	3.1	3.2	2.1	2.6	1.3	0.6	1.4	2.1	2.2	3.6
50	A	1.4	2.3	1.4	2.4	2.8	2.9	3.1	2.9	2.1	1.9	2.4	3.4	2.7	1.1	0.4	1.0	3.6	3.6
	B	0.1	4.0	1.2	2.8	2.9	2.2	2.2	2.3	2.1	2.0	2.1	3.4	0.0	3.3	0.0	0.8	0.4	4.0
	C	1.3	1.8	2.1	2.4	2.6	3.3	3.8	3.6	3.0	2.2	2.3	3.4	1.0	1.4	0.6	1.0	1.7	3.7
100	A	1.0	2.2	1.0	1.8	2.0	2.3	2.0	1.9	1.6	1.1	2.6	3.2	2.4	0.4	0.7	0.6	3.7	3.1
	B	0.2	4.0	0.4	3.0	2.0	1.3	1.4	1.9	2.0	0.9	0.9	2.2	0.0	3.0	0.0	1.1	0.9	4.0
	C	0.9	2.0	1.0	2.0	2.3	3.2	3.4	3.6	2.4	1.0	1.8	3.7	0.9	2.0	0.9	1.3	1.3	3.9
200	A	0.6	2.2	1.3	1.2	1.7	1.2	1.1	0.9	0.4	1.2	2.3	3.8	1.0	0.8	0.8	0.6	2.8	3.6
	B	1.1	4.0	1.0	3.0	1.9	1.0	1.0	1.0	2.0	0.9	1.0	2.0	0.0	2.0	1.9	2.0	1.8	3.0
	C	0.7	1.7	0.2	1.6	2.0	3.2	3.6	3.9	2.4	0.3	0.7	3.4	0.1	1.4	0.2	0.7	1.1	4.0

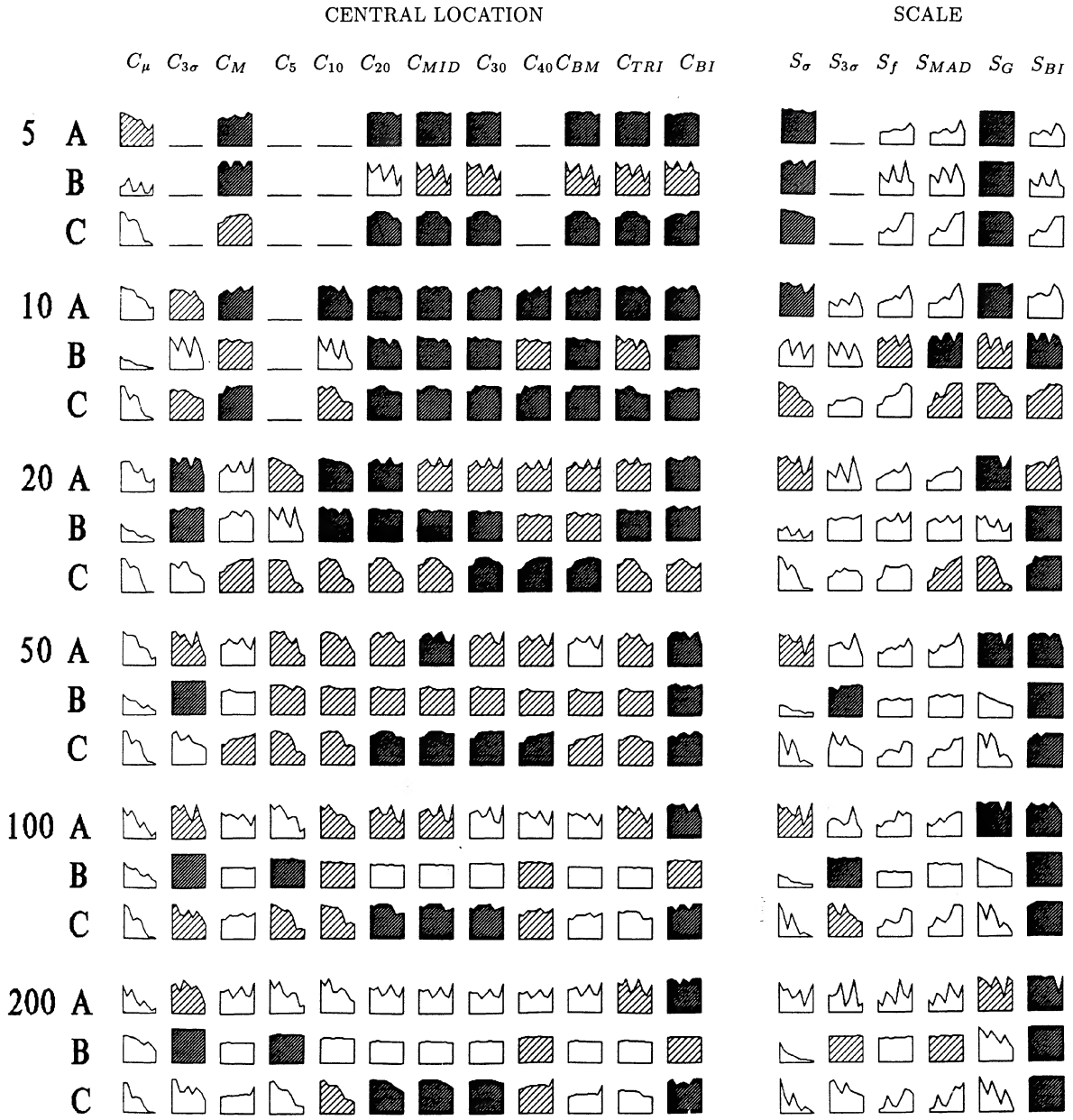

FIG. 1. Profile plots of relative efficiency for estimators of central location and scale. The plots are grouped by sample size n. The shading of each profile corresponds to the overall efficiency score listed in Table II. The highest-efficiency estimators are assigned the darkest shading.

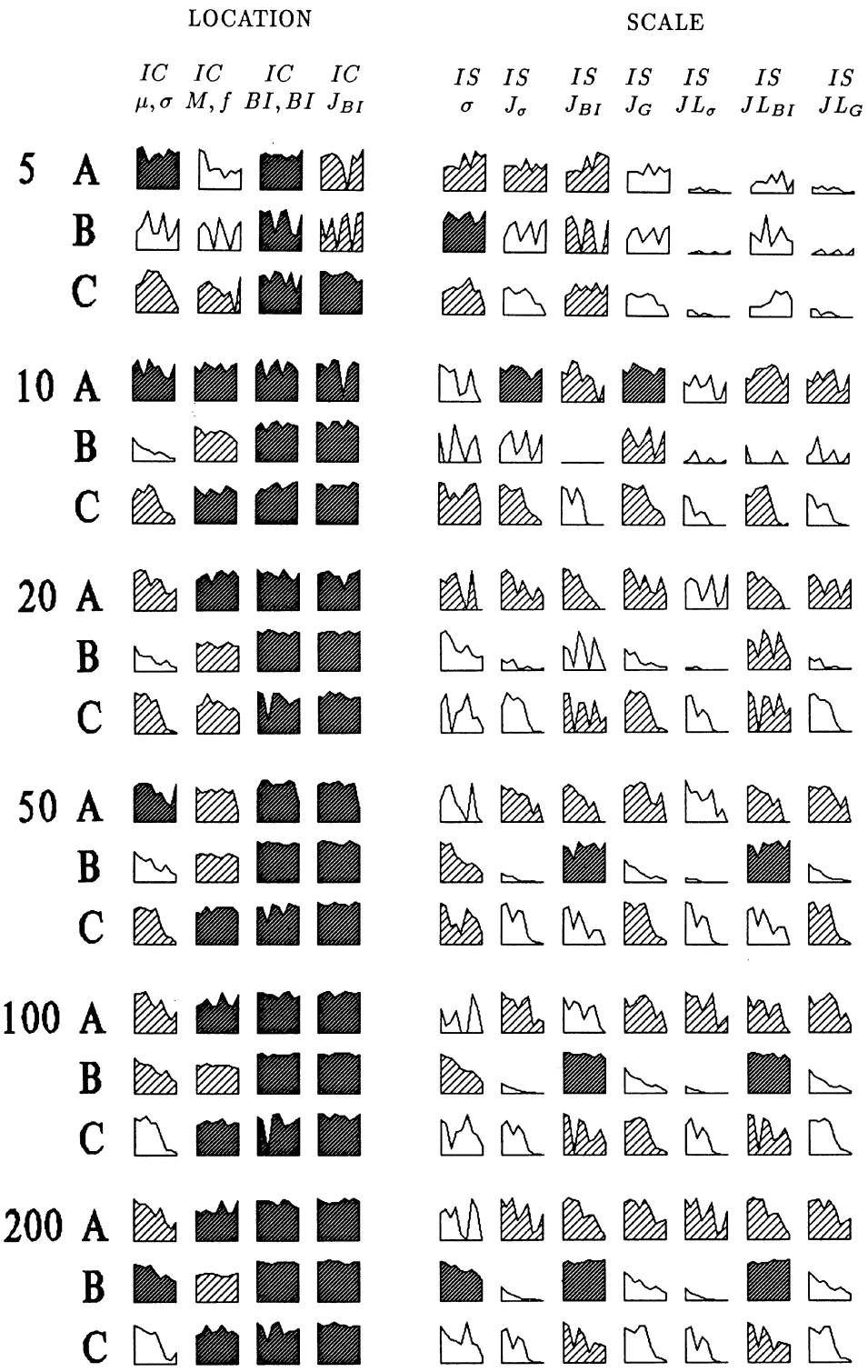

FIG. 2. Profile plots of relative efficiency for estimators of confidence intervals on central location and scale. The plots are grouped by sample size n. The shading of each profile corresponds to the overall efficiency score listed in Table III. The highest-efficiency estimators are assigned the darkest shading.

TABLE III. Efficiency scores for estimators of confidence interval length.

N	Case	$IC_{\mu,\sigma}$	$IC_{M,f}$	$IC_{BI,BI}$	$IC_{J_{BI}}$	IS_σ	IS_{J_σ}	$IS_{J_{BI}}$	IS_{J_G}	IS_{JL_σ}	$IS_{JL_{BI}}$	IS_{JL_G}
5	A	3.4	1.8	3.3	2.9	2.9	2.2	2.6	1.9	0.0	0.3	0.0
	B	1.9	1.6	3.1	2.3	3.4	1.9	2.2	1.3	0.0	1.2	0.0
	C	2.2	2.1	3.1	3.4	2.2	1.4	2.2	0.9	0.0	0.8	0.0
10	A	3.2	3.5	3.2	3.2	1.9	3.2	2.3	3.3	1.3	2.8	2.0
	B	0.3	2.7	3.6	3.7	1.4	1.4	0.1	2.2	0.0	0.3	0.3
	C	2.0	3.4	3.3	3.8	2.9	2.1	1.8	2.6	0.7	2.2	1.2
20	A	2.6	3.7	3.4	3.5	2.0	2.2	2.1	2.7	1.9	2.2	2.2
	B	0.6	2.4	3.8	3.9	1.9	0.0	1.7	0.4	0.0	2.1	0.0
	C	2.0	2.8	3.2	3.9	1.6	1.8	2.0	2.1	1.2	2.6	1.8
50	A	3.2	2.9	3.4	3.5	1.7	2.2	2.2	2.7	1.8	2.2	2.6
	B	1.2	2.7	3.7	4.0	2.2	0.0	3.2	0.4	0.0	3.5	0.1
	C	2.1	3.6	3.4	3.8	2.2	1.7	1.9	2.2	1.5	1.9	2.1
100	A	2.7	3.3	3.8	3.8	1.4	2.2	1.9	2.3	2.3	2.1	2.2
	B	2.1	2.9	3.6	3.9	2.4	0.0	3.8	0.4	0.0	3.9	0.3
	C	1.9	3.2	3.1	3.9	1.9	1.3	2.1	2.0	1.2	2.2	1.9
200	A	2.4	3.2	3.8	3.9	1.8	2.2	2.4	2.7	2.2	2.4	2.6
	B	3.1	2.8	3.6	3.9	3.3	0.0	3.8	1.0	0.0	3.9	0.8
	C	1.9	3.2	3.2	4.0	1.8	1.1	2.2	1.9	1.0	2.2	1.9

The biweight estimator of scale S_{BI} is clearly the estimator of choice over the sample range $20 \leqslant n \leqslant 50$. The canonical standard deviation S_σ is adequate only for case A contamination, and cannot be recommended for cases B and C. The clipped scale estimate $S_{3\sigma}$ cannot be recommended for case A or case C, and only climbs to superior in case B as n approaches 50. Although the gapped estimate S_G is superior for case A contamination, its poor performance for case B and marginal to adequate performance for case C discourage its general use.

The jackknife confidence interval on the biweight, $IS_{J_{BI}}$, is adequate or marginal in all cases; the jackknife log of the biweight scale, $IS_{JL_{BI}}$, is even somewhat better behaved. The estimator IS_{J_σ} is only adequate for case A contamination, and is unacceptable or marginal for cases B and C. The canonical interval IS_σ is barely adequate or marginal for most cases.

d) Large n (100⩽n⩽200)

The large-n situation will only be realized in practice for quantifying the global properties for the few clusters that have complete redshift results for their bright members and appear free of substructure which would point toward analyses of smaller subsamples of velocities. To our knowledge, the only published sample of velocities that satisfies both of these criteria is that for A2670 (Sharples, Ellis, and Gray 1988).

The clipped estimator $C_{3\sigma}$ is superior for this range of sample sizes only for case B contamination, and marginal or adequate for case A and case C. The heavily trimmed estimators of location C_{20}, C_{MID}, or C_{30} are superior estimators for case C, but do not provide adequate resistance for case B contamination. Once again, the biweight location estimator C_{BI} can be recommended as providing the best combination of resistance and efficiency across all three cases of contamination considered in this study.

The biweight confidence interval $IC_{BI,BI}$ once again is superior for all cases; the jackknife of the biweight $IC_{J_{BI}}$ is even better, as it exhibits less variance of efficiency for case C. The estimator $IC_{M,f}$ provides an adequate or superior alternative to the canonical estimator $IC_{\mu,\sigma}$, which, although adequate or marginal in most cases, degrades rapidly in the face of increasing asymmetry or tail weight.

The biweight estimator of scale S_{BI} is the estimator of choice for $n \geqslant 100$. The canonical standard deviation S_σ cannot be recommended. The clipped scale estimate $S_{3\sigma}$ cannot be recommended for case A, is superior or adequate for case B, and adequate or marginal for case C. Although the

TABLE IV. Recommended estimators for location, scale, and confidence intervals.

Sample size n	C_u	$IC_{u,v}$	S_v	IS_v
5	C_M, C_{MID}, C_{TRI}	$IC_{\mu,\sigma}, IC_{BI,BI}$	S_σ, S_G	$IS_\sigma, IS_{J_{BI}}$
	C_{BI}	$IC_{J_{BI}}$		
10	C_M, C_{20}, C_{MID}	$IC_{M,f}, IC_{BI,BI}$	S_G, S_{BI}	IS_{J_G}
	C_{30}, C_{TRI}, C_{BI}	$IC_{J_{BI}}$		
20–50	$C_{10}, C_{20}, C_{MID}, C_{30}$	$IC_{M,f}, IC_{BI,BI}$	S_{BI}	$IS_{J_{BI}}, IS_{JL_{BI}}$
	C_{40}, C_{TRI}, C_{BI}	$IC_{J_{BI}}$		
100–200	C_{BI}	$IC_{M,f}, IC_{BI,BI}$	S_{BI}	$IS_{J_{BI}}, IS_{JL_{BI}}$
		$IC_{J_{BI}}$		

gapped estimator S_G is superior, or nearly so, for case A contamination, its poor performance for cases B and C discourage its general use for large n.

The jackknifed biweight interval, $IS_{J_{BI}}$, or the jackknife log of the biweight, $IS_{JL_{BI}}$, provide adequate or superior estimates of the scale confidence interval for all cases. The estimator IS_{J_σ} is only adequate for case A contamination, and unacceptable for cases B and C. The canonical interval IS_σ is marginal or unacceptable for all but case B contamination, where it achieves an adequate or superior score.

VII. SUMMARY AND DISCUSSION

Table IV summarizes the results of our study. For situations we expect to confront in the analysis of radial-velocity data in clusters of galaxies, the estimators of location C_u, scale S_v, and their associated confidence levels $IC_{u,v}$ and IS_v listed in Table IV all proved to be adequate or superior (in the language of our rudimentary grading system). Although there might be some disagreement over particular cases, we believe our grading scheme gives a fair picture of the efficiency and robustness of a given statistic. Obviously, if one has *a priori* knowledge of the parent population (e.g., when sampling from an input dynamical model), a more refined choice of estimators might be made. To paraphrase Tukey, it is not so important *which* robust estimator is used, but that one *is* used. Especially in a field where data are difficult and expensive to obtain, the use of efficient estimators should be highly encouraged.

It is noteworthy that the clipped estimators $C_{3\sigma}$ and $S_{3\sigma}$ do not, in general, provide low enough variance to be useful for cluster velocity analysis. The 3σ-clipped estimators fail because the initial estimate of σ is dominated by the presence (or absence) of outliers, leading to a substantial sample-to-sample variance. Apparently, if one is going to clip, the sample-independent clippers such as C_{10}, C_{20}, C_{MID},... should be employed. Estimators that dilute the effects of outliers in a continuous fashion, as in the biweight, have been shown to be more generally useful.

Every investigation of cluster velocity distributions is faced with the vexing problem of how to assign realistic confidence intervals on estimates of location and scale. There seems little question that the refined bias-corrected bootstrap intervals $IC_{BC^*_u}$ and $IC_{BC^*_v}$ provide the best available empirical estimates. They are, however, computationally expensive, and require a minimum of 500–1000 bootstrap replications to yield reliable results. Our tests indicate that a number of less computation-intensive methods are reasonable alternatives.

Confidence intervals about the location can be obtained directly from sample estimates of location and scale using the simple formulas of Sec. IV*a*, based on the C_M and S_f or C_{BI} and S_{BI}. A jackknife estimator such as $IC_{J_{BI}}$ provides useful intervals but requires no more than n sampling steps. An asymptotically equivalent procedure to the jackknife, the bootstrap Standard Interval, requires at most one hundred bootstrap replications, and may prove useful for some applications. It is worth emphasizing that, with the exception of tiny-n samples, the canonical procedure for forming a location confidence interval based on C_μ and S_σ is not a low-variance estimator of interval length.

Because sampling distributions of scale estimators are generally long tailed and asymmetric, confidence intervals on the scale of a data batch should ideally be bootstrapped. One can obtain approximate intervals, however, with less computational effort. For $n \leqslant 10$ our tests indicate that the classical procedure is adequate, but that several jackknife estimators are better. For larger samples the jackknifed biweight and log biweight estimators are best.

As larger and more detailed samples of velocities in clusters become available, we should look to increasingly sophisticated statistical analysis as a means of extracting the maximum information possible. The present paper represents only a first step in this direction.

Thanks to several colleagues, in particular M. Fitchett and M. West, for their patience in awaiting the results of this work, and for encouraging us to bring it to fruition on a finite timescale. We also acknowledge the Department of Physics and Astronomy at Michigan State for the generous allotment of computer resources which made this work possible.

Note added in proof: The authors will make available, upon request, standard FORTRAN source code for calculation of the estimators and confidence intervals discussed in this paper. Please send a $5\frac{1}{4}''$ or $3\frac{1}{2}''$ floppy disk along with any request.

REFERENCES

Andrews, D. F., Bickel, P. J., Hampel, F. R., Rogers, W. H., and Tukey, J. W. (1972). *Robust Estimates of Location: Survey and Advances* (Princeton University, Princeton).

Barrow, J. D., Bhavsar, S. P., and Sonoda, D. H. (1984). Mon. Not. R. Astron. Soc. **210**, 19P.

Beers, T. C., and Gebhardt, K. (1990). In preparation.

Beers, T. C., Gebhardt, K., Geller, M. J., and Huchra, J. P. (1990). In preparation.

Beers, T. C., Geller, M. J., Huchra, J. P., Latham, D. W., and Davis, R. J. (1984). Astrophys. J. **283**, 33.

Binggeli, B., Tammann, G. A., and Sandage, A. (1987). Astron. J. **94**, 251.

Bothun, G. D., Geller, M. J., Beers, T. C., and Huchra, J. P.(1983). Astrophys. J. **268**, 47.

Bothun, G. D., and Schombert, J. M. (1988). Astrophys. J. **335**, 617.

Bothun, G. D., and Schombert, J. M. (1990). Preprint.

Bower, R. G., Ellis, R. S., and Efstathiou, G. F. (1988). Mon. Not. R. Astron. Soc. **234**, 725.

Box, G. E. P. (1953). Biometrika **40**, 318.

Chapman, G. N. F., Geller, M. J., and Huchra, J. P. (1987). Astron. J. **94**, 571.

Chapman, G. N. F., Gellere, M. J., and Huchra, J. P. (1988). Astron. J. **95**, 999.

Danese, L., De Zotti, G., and di Tullio, G. (1980). Astron. Astrophys. **82**, 322 (DDd).

Dressler, A., and Schectman, S. (1988). Astron. J. **95**, 985.

Efron, B.(1987). J. Am. Stat. Assoc. **82**, 171.

Efron, B., and Tibshirani, R. (1986). Stat. Sci. **1**, 54.

Fernley, J. A., and Bhavsar, S. P. (1984). Mon. Not. R. Astron. Soc. **210**, 883.

Fitchett, M. J. (1988). In *The Minnesota Lectures on Clusters of Galaxies and Large-Scale Structure*, edited by J. M. Dickey (Brigham Young University Press, Provo), p. 143.

Fitchett, M. J., and Merritt, D. (1988). Astrophys. J. **335**, 18.

Fitchett, M. J., and Webster, R. L. (1987). Astrophys. J. **317**, 653.

Forman, W., Bechtold, J., Blair, W., Giacconi, R., Van Speybroeck, L., and Jones, C. (1981). Astrophys J. Lett. **243**, L133.

Geller, M. J., and Beers, T. C. (1982). Publ. Astron. Soc. Pac. **94**, 421.

Geller, M. J., and Huchra, J. P. (1983). Astrophys. J. Suppl. **52**, 61.

Goodall, C. (1983). In *Understanding Robust and Exploratory Data Analysis*, edited by D. C. Hoaglin, F. Mosteller, and J. W. Tukey (Wiley, New York), p. 339.

Hampel, F. R. (1968). Ph.D. thesis, University of California, Berkeley.

Hill, J. M., Hintzen, P., Oegerle, W. R., Romanishin, W., Lesser, M. P., Eisenhammer, J. D., and Batuski, D. J. (1988). Astrophys. J. Lett. **322**, L23.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (1983). *Understanding Robust and Exploratory Data Analysis* (Wiley, New York) (UREDA).

Hoaglin, D. C., and Peters, S. C. (1979). In *Proceedings of Computer Science Statistics: Twelfth Annual Symposium on the Interface*, edited by J. F. Gentleman (Springer, New York), p. 418.

Hoel, P. G. (1971). *Introduction to Mathematical Statistics* (Wiley, New York).

Huchra, J. P. (1984). In *Proceedings of the ESO Workshop on the Virgo Cluster*, edited by B. Binggeli and O. Richter (ESO, Garching), p. 181.

Huchra, J. P., and Geller, M. J. (1982). Astrophys. J. **257**, 423.

Iglewicz, B. (1983). In *Understanding Robust and Exploratory Data Analysis*, edited by D. C. Hoaglin, F. Mosteller, and J. W. Tukey (Wiley, New York), p. 404.

de Lapparent, V., Geller, M. J., and Huchra, J. P. (1986). Astrophys. J. Lett. **302**, L1.

Kent, S. M., and Gunn, J. E. (1982). Astron. J. **87**, 945.

Lax, D. (1985). J. Am. Stat. Assoc. **80**, 736.

Ling, E. N., Frenk, C. S., and Barrow, J. D. (1986). Mon. Not. R. Astron. Soc. **223**, 21P.

Maia, M. A. G., DaCosta, L. N., and Latham, D. W. (1989). Astrophys. J. Suppl. **69**, 809.

Miller, R. G. (1986). Ann. Math. Stat. **39**, 567.

Miller, R. G. (1974). Biometrika **61**, 1.

Mosteller, F., and Tukey, J. W. (1977). In *Data Analysis and Regression* (Addison Wesley, Reading, MA), p. 133.

Newberry, M. V., Kirschner, R. P., and Boroson, T. A. (1988). Astrophys. J. **335**, 629.

Oegerle, W. R., Fitchett, M. J., and Hoessel, J. G. (1989). Astron. J. **97**, 627.

Ostriker, E. C., Huchra, J. P., Geller, M. J., and Kurtz, M. J. (1988). Astron. J. **96**, 1775.

Pearson, E. S. (1931). Biometrika **21**, 114.

Postman, M., Geller, M. J., and Huchra, J. P. (1988). Astron. J. **95**, 267.

Quintana, H., and Lawrie, D. G. (1982). Astron. J. **87**, 1.

Quintana, H., and Melnick, J. (1982). Astron. J. **87**, 972.

Rice, J. A. (1988). *Mathematical Statistics and Data Analysis* (Wadsworth, Belmont, CA).

Rosenberger, J. L., and Gasko, M. (183). In *Understanding Robust and Exploratory Data Analysis*, edited by D. C. Hoaglin, F. Mosteller, and J. W. Tukey (Wiley, New York), p. 297.

Sarazin, C. L. (1987). In *Proceedings of the NATO Advanced Research Workshop: Cooling Flows in Clusters of Galaxies*, edited by A. Fabian (Reidel, Dordrecht), p. 1.

Schenker, N. (1985). J. Am. Stat. Assoc. **80**, 360.

Shapiro, S. S., and Wilk, M. B. (1965). Biometrika **51**, 484.

Sharples, R. M., Ellis, R. S., and Gray, P. M. (1988). Mon. Not. R. Astron. Soc. **231**, 479.

Tonry, J. L. (1985). Astrophys. J. **291**, 45.

Tukey, J. W. (1958). Ann. Math. Stat. **29**, 614.

Wainer, H., and Thissen, D. (1976). Psychometrika **41**, 9.

West, M. J., and Bothun, G. D. (1990). Astrophys. J. **350**, 36.

Yahil, A., and Vidal, N. V. (1977). Astrophys. J. **214**, 347.

Yang, M. C. K., and Robinson, D. H. (1986). *Understanding and Learning Statistics by Computer* (World Scientific, Singapore), p. 154.

Zabludoff, A. I., Huchra, J. P., and Geller, M. J. (1990). Astron. J. (in press).