Robust Quick Tests of Scale

by

Gary W. Oehlert*

School of Statistics
University of Minnesota
Technical Report No. 494
August 1987

*This research was supported in part by a University of Minnesota Faculty Summer Research Grant.

Robust Quick Tests of Scale

Gary W. Oehlert

Department of Applied Statistics

University of Minnesota

St. Paul, MN

ABSTRACT

Quasi-ranges and kth absolute deviations from the median are simple, robust, and resistant estimates of scale. Squares of these statistics have distributions that are approximately multiples of chi-squares, so that approximate tests and confidence intervals for scale can easily be formed. Two sample tests using quasi-ranges compare very favorably with several standard robust tests of scale when the underlying distribution is long tailed but are not competitive when the distribution is short tailed.

Keywords: robustness, scale, quasi-ranges, median absolute deviations

August 12, 1987

Robust Quick Tests of Scale

Gary W. Oehlert

Department of Applied Statistics

University of Minnesota

St. Paul, MN

1. Introduction

The classical F test for variances is known to be extremely sensitive to the assumption of nomality. Even modest deviations from normality can result in gross inaccuracies in the nominal significance level (Box 1953). As a result of this, several alternatives to the classical F test have been proposed. The better of these, for example, jackknifing log variance estimates, are fairly robust to the underlying distribution of the data, but are difficult to compute by hand. This paper studies two simple, robust families of procedures which are appropriate for one or two sample scale problems. Results may be expressed either as confidence intervals for a scale or the ratio of two scales, or as significance levels for tests.

Four desirable properties for any statistical procedure are robustness, resistance, efficiency, and simplicity. Robustness means that the validity of a procedure, for example the coverage rate of a confidence interval, is constant across a broad range of parent distributions for the data. Robustness is a property of a procedure and the generating distributions, not an individual data set. Resistance means that the inference drawn from a particular set of data does not change much if some of the data are changed arbitrarily. Resistance is generally measured by the breakdown bound, the greatest fraction of the data that can be changed arbitrarily while leaving the result essentially unchanged. Efficiency is standard. We always want shorter confidence intervals or more powerful tests. Until fairly recently, efficiency was the ultimate criterion, and robustness and resistance played only a minor role. Simplicity means that we prefer procedures that can be performed with a minimum of effort, preferably by hand.

There is no procedure for scales that has all four properties. The classical F test is neither robust nor resistant, though it is efficient at the model of normality. The Box-Andersen (Box and Andersen 1955) and Lemmer (1978) procedures are robust, simple, and fairly efficient, but not resistant. (The Lemmer procedure is resistant if location is known.) Miller (1968) has shown that jackknifed logged sample variances can form the basis of one or two sample procedures for scale that are robust and as efficient at the Box-Andersen procedure, but they are difficult to compute and not resistant. A-estimates (Lax 1985) are excellent robust, resistant estimates of scale, and Tukey (1980) has suggested jackknifing logged A-estimates (specifically the bisquare). These should be robust, resistant, and efficient, but are even more difficult to compute. Shoemaker and Hettmansperger (1982) present an asymptotic approximation to the distribution of A-estimates (which they call mid-variances) that is fairly accurate even at n=10.

Conover, *et al.* (1981) have made a comprehensive comparison of tests for homogeneity of variances in the k-sample problem. They recommend modifications of the Levene (1960) and Flinger and Killeen (1976) tests. The Levene test is an analysis of variance on the absolute deviations from the group medians and is not resistant. The Flinger and Killeen test is a linear rank statistic and thus resistant, but the rank scores are percentage points of the standard normal and may be tedious to compute.

This article studies one and two sample scale procedures of two types. The first type is based on quasi-ranges (David 1981). Let $x_1, x_2, \cdots x_n$ be an independent identically distributed sample from a distribution F, and let $x_{(1)}, x_{(2)}, \cdots x_{(n)}$ be the associated order statistics. The distribution F is assumed to be continuous and symmetric; the center of symmetry may be taken to be zero without loss of generality. For a given r, a quasi-range is a difference of the form $W_r = x_{(n+1-r)} - x_{(r)}$. Taking r=1 yields the usual range; taking r=$\lfloor n/4 \rfloor$ yields one form of the interquartile range, where $\lfloor x \rfloor$ is the greatest integer less than or equal to x. Mosteller (1946) gives many results on the use of quasi-ranges for estimating the scale of a Gaussian distribution.

The second type of procedure is based on the median absolute deviation. Let m = $\lfloor n/2 \rfloor$; $x_{(m)}$ is called the low median of the sample. Let $y_i = |x_{(i)} - x_{(m)}|$ and let $y_{(i)}$, i = 1, ... , n-1 be the ordered y_i values excluding y_m, which is always zero. Then $D_k = y_{(k)}$ is a measure of scale that we will call the kth absolute deviation. Taking k near n/2 gives the usual median absolute deviation. (Note that the low median is used here for simplicity. For even n, it has been traditional to define the median to be $(x_{(m)} + x_{(m+1)})/2$, the average of the low and high medians.)

There is some ambiguity in how scale is defined across several distributional shapes. This is because if σ is a scale parameter, then $k\sigma$ is also a scale parameter for any positive k. Thus W_r and W_r/π both estimate valid scale parameters for any distribution. For a normal distribution, σ is usually taken to be the standard deviation. It is traditional to use estimates of standard deviation that are square roots of unbiased estimates of variance, and this tradition will be continued here. Let $\mathbf{W}_r = \rho_{r,n}\, W_r$ be an estimate of scale, with $\rho_{r,n}$ a constant chosen so that $\mathbf{W}_r^2$ is an unbiased estimate of variance when sampling from a standard normal. Thus, $1/\rho_{r,n}^2 = E\ W_r^2$ when the data are sampled from a standard normal. For distributions other than normal, take as the natural scale $(\sigma = 1)$ that scale in which $\mathbf{W}_r^2$ is unbiased for σ^2. To illustrate this, consider a large n with r near n/4. Then $\rho_{r,n}$ is approximately equal to 1/1.35, and the natural scale for the uniform will be uniform (0,2*1.35). Similarly define $\mathbf{D}_k$ as a rescaled version of D_k.

This definition of scale has the disadvantage that "natural" scale depends on sample size for nonGaussian distributions. (However, this dependence of the estimand on sample size is not uncommon for robust estimates of scale, for example, trimmed standard deviations and biweight A-estimates.) The magnitude of the change in "natural" scale will be discussed in Section 2.

One sample confidence intervals for scale based on these statistics will be formed analogously with traditional chisquare intervals. The scale statistic $\mathbf{W}_r$ (or $\mathbf{D}_k$) is distributed as σ times a random variable with c.d.f. G. The distribution G depends on the parent distribution of the data, n and r (or k). (A robust procedure has weak dependence on the parent distribution.) Let η_l and η_u be the lower and upper $\alpha/2$ percent points of G. Then a $(1-\alpha)$ percent confidence interval for σ is

$$\frac{\mathbf{W}_r}{\eta_u} \le \sigma \le \frac{\mathbf{W}_r}{\eta_l} \tag{1.1}$$

For the two sample problem, suppose the sample sizes are n and n', and call the scale estimates from the two independent samples $\mathbf{W}_r$ and $\mathbf{W'}_{r'}$. The ratio $\mathbf{W}_r/\mathbf{W'}_{r'}$ has a distribution H (which depends on n, n', r, and r') and lower and upper $\alpha/2$ percent points ζ_l and ζ_u. A $(1-\alpha)$ percent confidence interval for σ/σ' is

$$\frac{\mathbf{W}_r}{\mathbf{W'}_{r'}\zeta_u} \le \frac{\sigma}{\sigma'} \le \frac{\mathbf{W}_r}{\mathbf{W'}_{r'}\zeta_l} \ . \tag{1.2}$$

Confidence intervals using $\mathbf{D}_k$ are formed analogously.

Confidence intervals and tests based on **W** and **D** are simple and resistant with breakdown bounds equal to the breakdown bounds of the scale estimates involved. For quasi-ranges, the breakdown bound is (r-1)/n; for kth absolute deviations, the breakdown bound is $min(m, (n-1-k))/n$. Thus for the interquartile range and median absolute deviation, the breakdown bounds are approximately 0.25 and 0.5 respectively. Robustness and efficiency of these procedures are controlled by r and k. Large values of r (small values of k) make these procedures more robust, while small values of r (large values of k) make these procedures more efficient. Hence, compromise values of r and k are needed.

The performance of one and two sample scale procedures will be evaluated by examining the distributions of the scale estimates and ratios of scale estimates for various values of n, r, and k, and for four parent distributions spanning the range from short tails to very long tails: uniform, normal, slacu, and slash. The slash and slacu are models for long tailed data (see, for example, Rosenberger and Gasko 1983) defined by a normal divided by an independent uniform (0,1) and cube root of a uniform respectively. The slash and slacu have tails like t-distributions with 1 and 3 degrees of freedom respectively, but their central portions are more nearly Gaussian than t-distributions.

2. Quasi-ranges

A quasi-range is a symmetric difference of order statistics

$$W_r = x_{(n+1-r)} - x_{(r)}$$

with distribution given in David (1981):

$$f_W(w) = C \int_{-\infty}^{\infty} F_X^{r-1}(x) f_X(x) \, [\, F_X(x+w) - F_X(x)]^{n-2r} f_X(x+w) \, [\, 1 - F_X(x+w)\,]^{r-1} \, dx \ , \qquad (2.1)$$

where C is a proportionality constant. Suppose that r/n approaches p as n increases ($0 < p < 1$), and let ξ_p be such that $F_X(\xi_p) = p$. Then $\sqrt{n}(W_r - 2\xi_p)$ converges in distribution to a normal with mean 0 and variance $2p(1-2p)/f_X(\xi_p)^2$. (See David (1981) or Mosteller (1946).) It is convenient to approximate this normal distribution by a multiple of a chisquare distribution with the same mean and variance. The degrees of freedom for the chisquare (called the equivalent degrees of freedom) summarize the tightness of the distribution. The larger the equivalent degrees of freedom, the shorter one sample confidence intervals for scale will be.

Figure 1 shows the asymptotic equivalent degrees of freedom divided by n for quasi-ranges as a function of p and the parent distribution. Large values of equivalent degrees of freedom correspond to high efficiency, and similar values across distributions correspond to robustness. It is obvious from Figure 1 that fairly large values of p, say about 0.35, are needed to get acceptable robustness asymptotically, while efficiency is always maximized for p less than 0.25. Hence, asymptotically one must choose between robustness and efficiency.

Figure 1 about here.

For small sample sizes, the exact distribution can be computed for any parent distribution F via numerical integration. (The uniform may be integrated in closed form, of course.) One may then compute the upper and lower percentage points of the distribution and the confidence interval associated with a given F, n, and r. Table 1 gives the lower and upper multiples of W_r used in constructing central 95% confidence intervals for σ for several values of F, n, and r. The values of r are taken to be $\lceil 0.25n \rceil$ and $\lceil 0.35n \rceil$ (where $\lceil x \rceil$ is the least integer greater than or equal to x). These r values will illustrate the difference between the usual interquartile range (p=0.25) and the more robust p=0.35.

Table 1 about here

Table 1 illustrates the tradeoffs between robustness and efficiency. Confidence intervals are considerably shorter when using small r's than when using large r's, so efficiency requires a smaller value of r. In fact, 99% intervals for small r's are often shorter than 95% intervals for larger r's. Table 1 also gives the true coverage rate for nonGaussian distributions when the Gaussian lower and upper cutoffs are used. There is much less variation in this coverage rate when a larger r is used, so that use of Gaussian cutoffs will lead to larger errors in the coverage rate if the interquartile range is used. However, the confidence intervals for larger r's are so much longer than those for smaller r's that most statisticians would prefer the shorter intervals and accept some error in the stated coverage rate.

The situation for two sample problems is similar. The distribution of the ratio of two independent W_r variables is easily calculated via numerical integration. This distribution then forms the basis for tests or confidence intervals on the ratio of two scales. Table 2 gives the critical values for .05 and .01 one sided tests of $\sigma_1/\sigma_2 = 1$ versus $\sigma_1/\sigma_2 > 1$ for each of the sampling distributions along with the sizes obtained when using the Gaussian critical values. Analogous with the one sample case, the critical values for large r's are considerably greater than for small r's, with 0.01 critical values for small r's being only slightly larger than 0.05 critical values for large r's.

Table 2 about here

Another interesting feature of Table 2 is that the deviation of the true sizes from nominal is consistently smaller for the two sample problem than for the one sample problem. This compensation appears to be related to that which occurs in t-tests. When sampling from distributions with tails longer than Gaussian tails, the distributions of the numerator and denominator of the t-test both deviate more from their Gaussian-parent sampling distributions than does their ratio.

As was mentioned in Section 1, the natural scale depends on the sample size for nonGaussian distributions. Figure 2 shows the natural scale as a function of sample size for n=10(1)40, $r=\lceil n/4 \rceil$, and the four example distributions. The saw-tooth nature of the plots is the result of r incrementing once for every four increments in n. For moderate tailed distributions (Gaussian and slacu), the sample size dependence is minimal and can generally be ignored. For sample sizes greater than about 25, the uniform and slash distributions also have scales nearly equal to 1. However, for sample sizes less than about 25, the uniform and slash distributions have scales which differ markedly from 1, the uniform scales being less than one and the slash scales being greater. Thus, when comparing scales for small sample sizes and extreme (long or short) tails, some sample size adjustment may be needed unless the sample sizes are exactly equal.

Figure 2 about here

The results given so far are based on the exact sampling distributions of W_r, but approximations to these exact distributions and approximations to $\rho_{r,n}$ are needed to make routine use of W_r feasible. Fortunately, a simple approximation based on the chisquare distribution is adequate. We approximate W_r^2 as σ^2 times a chisquare divided by its degrees of freedom, in exact analogy with the sample variance. The first two moments of W_r^2 are computed via numerical integration; $\rho_{r,n}^{-2}$ is the expectation of W_r^2 and the degrees of freedom are chosen so that W_r and the approximating multiple of a chisquare have the same variance. For n=10(1)40 and $r=\lceil n/4 \rceil$, the degrees of freedom and ρ are well approximated by the formulae:

$$df = 0.290 + 0.971(n+1-2r) - 0.119n \quad , \tag{2.2}$$

and

$$\rho_{r,n}^{-2} = -2.66 + 8.96(n+1-2r)/n + 1.51/n \quad . \tag{2.3}$$

Note that for the interquartile range, r is approximately n/4 resulting in approximately 0.4n degrees of freedom for the distribution of a squared scale. This should be contrasted with n-1 degrees of freedom for s^2 when sampling from a Gaussian distribution. Figure 3 shows the percent error in the approximate chisquare quantiles of W_r as a function of the cumulative approximate probability for n=10(10)40. Except for the extreme tails and n=10, the relative error is less than 1%.

Figure 3 about here

The approximation of W_r^2 by a chisquare suggests the approximation of the ratio of two W's by the square root of an F-distribution, with degrees of freedom computed separately for numerator and denominator. Figure 4 shows the percent error in the approximate F quantiles of the ratio of two W's as a function of the cumulative approximate probability for n=10(10)40. Again, except for the extreme tails and n=10, the relative error is less than 1%.

Figure 4 about here

3. Kth Absolute Deviations

Let $m = \lfloor n/2 \rfloor$; then $x_{(m)}$ is called the low median of the sample. Denote the deviations from the low median by $y_i = |x_{(i)} - x_{(m)}|$ and let $y_{(i)}$, i = 1, ... , n-1 be the ordered deviations omitting y_m, which is always zero. $D_k = y_{(k)}$ is the kth absolute deviation. If $k/n \rightarrow p$, then the results of Sen (1968) may be used to show that $\sqrt{n}(D_k - \xi_{(1+p)/2})$ converges in distribution to a normal with mean zero and variance $p(1-p)/2f_X(\xi_{(1+p)/2})^2$. Taking the case where (n-k)/2r converges to one, we see that kth deviations and quasi-ranges are asymptotically equivalent. Thus, the results of Figure 1 hold true for kth deviations when (k/n ->) p for the kth deviation is translated to (1-p)/2, the corresponding value for r/n in a quasi-range. In particular, we must compromise between robustness and efficiency in kth deviations, with small values of k corresponding to robustness and large values of k corresponding to efficiency.

The small sample distribution of a kth deviation (H(s)) is more complicated than that of a quasi-range, but it may still be written in closed form using the expression given in David (1981) for nonidentically distributed order statistics. Conditional on $x_{(m)}$, the deviations above and below $x_{(m)}$ are independent, with those below having a distribution $G_1(x) = (F_X(x_{(m)}) - F_X(x_{(m)} - x))/F_X(x_{(m)})$, and those above having a distribution $G_2(x) = (F_X(x_{(m)} + x) - F_X(x_{(m)}))/(1 - F_X(x_{(m)}))$. There are n1 = m-1 values from G_1 and n2 = n-m values from G_2. We may express $h(s|x_{(m)})$ as

$$h(s \mid x_{(m)}) = \sum_{i=max(0,k-1-n2)}^{min(k-1,n1-1)} \frac{n1!}{i!(n1-i-1)!} \frac{n2!}{(k-1-i)!(n2-k+1+i)!} \times \tag{3.1}$$

$$G_1(s)^i G_2(s)^{k-1-i} g_1(s)(1-G_1(s))^{n1-1-i}(1-G_2(s))^{n2-k+1+i}$$

$$+ \sum_{i=max(0,k-1-n1)}^{min(k-1,n2-1)} \frac{n1!}{(k-1-i)!(n1-k+1+i)!} \frac{n2!}{i!(n2-i-1)!} \times$$

$$G_1(s)^{k-1-i} G_2(s)^i g_2(s)(1-G_1(s))^{n1-k+1+i}(1-G_2(s))^{n2-i-1}$$

Of course, $x_{(m)}$ is implicit in the definitions of G_1 and G_2. The density of the kth deviation unconditional on the low median is found by integrating the conditional density against the density of the low median. This integration may be performed numerically.

Analogous with Table 1, Table 3 gives the lower and upper multiples of $\mathbf{D}_k$ used in constructing central 95% and 99% confidence intervals for σ for the values of F and n used in Table 1. The values of k were chosen to yield kth deviations with distributions most similar to the quasi-ranges used in Table 1. The k giving the most similar distribution is k = n+1-2r; that is, if a quasi-range spans j gaps, then the most similar kth deviation has k equal to j.

Table 3 about here

Matched kth deviations and quasi-ranges produce similar confidence intervals, though the kth deviation intervals tend to be slightly longer and slightly more robust than corresponding quasi-range intervals. Differences between the two procedures are most pronounced at the 99% rate and decrease as sample size increases. Quasi-ranges and k-th deviations are essentially equivalent at n=40.

4. Power and alternative tests

Since the quasi-range and kth deviations procedures produce very similar results, only the quasi-range will be used when examining power and making comparisons with other tests.

Table 4 gives the powers for one-sided two sample tests when Gaussian critical values are used and σ_1/σ_2 takes the values 1.0, 1.5, 2.0, and 2.5. An interesting feature of Table 4 is that the power of the quasi-range test is very stable across the different distributions. Thus the quasi-range procedures are approximately distribution free in terms of both size and power.

Table 4 about here

Table 5 gives the estimated powers for the Box-Andersen, jackknifed log sample variance (Miller), jackknifed log bisquare A-estimate (Tukey), Levene, Flinger-Killeen, and quasi-range tests in the one-sided two sample problem for equal sample sizes when σ_1/σ_2 takes the values 1.0 and 2.0. These powers were estimated by Monte Carlo using 5000 simulations for each sample size and parent distribution. The Levene and Flinger-Killeen tests maintain a stable size, though they tend to be conservative. The Box-Andersen and jackknifed tests are liberal with observed sizes ranging from about 5% for the uniform to about 13% for the slash. The quasi-range sizes vary between 3% for the uniform and 10% for the slash.

Table 5 about here

As noted before, the power of the quasi-range test is approximately constant across the four distributions. This is not true for the other tests which have much more power for the short tailed distributions than the long tailed distributions. The power of the quasi-range test is clearly inferior for the uniform and Gaussian distributions, competitive but on the low end for the slacu, and clearly superior for the slash. Taking a smaller value of r for the quasi-range procedure would make it less inferior for the shorter tailed distributions and less superior for the slash, but would also decrease the robustness of the p-values across distributions.

5. Conclusions

One and two sample test procedures based on quasi-ranges and kth deviations are simple, robust, and resistant. They are more efficient than many recommended scale procedures for long tailed distributions, competitive for slightly long tailed distributions, and poor for distributions with tails as short or shorter than the Gaussian. As many real data sets have tail lengths approximately like the slacu or even longer (Huber 1981, page 3), quasi-range and kth deviations procedures appear very attractive.

There are few differences in performance between quasi-ranges and kth deviations. Quasi-ranges are slightly more efficient, but kth deviations are slightly more robust. Kth deviations are also more resistant than quasi-range.

The distributions of quasi-ranges near the interquartile range may be approximated as square roots of multiples of chisquare distributions. The quality of this approximation improves rapidly with sample size, and the parameters of the approximating density are easily estimated using equations (2.2) and (2.3).

Acknowledgements. This research was supported in part by a University of Minnesota Summer Research Fellowship. The numerical integration for the computation of quasi-range densities was also performed in Margaret Bellamy's 1983 Princeton University Senior Thesis under the author's direction.

REFERENCES

Box, G. E. P. (1953), "Nonnormality and tests on variances," *Biometrika,* 40, 318-335.

Box, G. E. P. and Andersen, S. L. (1955), "Permutation theory in the derivation of robust criteria and the study of departures from assumption," *Journal of the Royal Statistical Society, Series B,* 17, 1-26.

Conover, W. J., Johnson, M. E., and Johnson, M. M. (1981), "A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data," *Technometrics,* 23, 351-361.

David, H. A. (1981), *Order Statistics, 2nd Edition,* New York: Wiley.

Flinger, M. A. and Killeen, T. J. (1976), "Distribution-Free Two-Sample Tests for Scale," *Journal of the American Statistical Association,* 71,210-213.

Huber, P. J. (1981), *Robust Statistics,* New York: Wiley.

Lax, D. A. (1985), "Robust Estimators of Scale: Finite-Sample Performance in Long-Tailed Symmetric Distributions," *Journal of the American Statistical Association,* 80, 736-741.

Lemmer, H. H. (1978), "A Robust Test for Dispersion," *Journal of the American Statistical Association,* 73, 419-421.

Levene, H. (1960), "Robust Tests for Equality of Variances," in *Contributions to Probability and Statistics,* ed. I. Olkin, Palo Alto, Calif: Stanford University Press.

Miller, R. G. (1968), "Jackknifing Variances," *Annals of Mathematical Statistics,* 39, 567-582.

Mosteller, F. (1946), "On some useful 'inefficient' statistics," *Annals of Mathematical Statistics,* 17, 377-408.

Rosenberger, J. L. and Gasko, M. (1982), "Comparing Location Estimators: Trimmed Means, Medians, and Trimean," in *Understanding Robust and Exploratory Data Analysis,* eds. D. C. Hoaglin, F. Mosteller, and J. W. Tudey, New York: Wiley.

Sen, P. K. (1968), "Asymptotic Normality of Sample Quantiles for m-Dependent Processes," *Annals of Mathematical Statistics,* 39, 1724-1730.

Shoemaker, L. H. and Hettmansperger, T. P. (1982), "Robust estimates and tests for the one- and two-sample scale models," *Biometrika,* 69, 47-53.

Tukey, J. W. (1980), unpublished lecture notes, Department of Statistics, Princeton University.

Table 1. Confidence intervals and coverages rates based on W_r

				95%			99%	
n	r	distribution	lower	upper	Gcover	lower	upper	Gcover
10	3	uniform	0.65	2.55	975	0.59	3.73	996
		Gaussian	0.60	2.74	950	0.51	3.99	990
		slacu	0.58	2.80	942	0.48	4.09	986
		slash	0.48	3.51	879	0.32	5.15	952
10	4	uniform	0.54	4.55	958	0.47	8.37	993
		Gaussian	0.52	4.64	950	0.43	8.37	990
		slacu	0.52	4.71	948	0.42	8.56	989
		slash	0.48	5.23	929	0.34	9.49	977
20	5	uniform	0.73	1.70	981	0.68	2.08	997
		Gaussian	0.68	1.82	950	0.61	2.25	990
		slacu	0.67	1.85	941	0.59	2.28	986
		slash	0.57	2.13	868	0.44	2.65	948
20	7	uniform	0.64	2.27	963	0.57	3.08	994
		Gaussian	0.62	2.34	950	0.54	3.17	990
		slacu	0.61	2.35	948	0.53	3.19	989
		slash	0.58	2.48	928	0.47	3.38	979
30	8	uniform	0.75	1.57	976	0.70	1.86	997
		Gaussian	0.71	1.66	950	0.64	1.97	990
		slacu	0.70	1.67	945	0.63	1.99	988
		slash	0.64	1.81	895	0.53	2.16	963
30	11	uniform	0.66	2.05	959	0.58	2.66	993
		Gaussian	0.64	2.09	950	0.56	2.71	990
		slacu	0.64	2.09	950	0.56	2.72	990
		slash	0.62	2.15	937	0.53	2.80	984
40	10	uniform	0.78	1.44	979	0.73	1.64	997
		Gaussian	0.74	1.51	950	0.68	1.74	990
		slacu	0.73	1.52	943	0.67	1.75	987
		slash	0.68	1.63	888	0.58	1.89	961
40	14	uniform	0.70	1.76	961	0.63	2.14	994
		Gaussian	0.68	1.79	950	0.61	2.19	990
		slacu	0.68	1.79	950	0.61	2.19	990
		slash	0.66	1.84	936	0.57	2.25	984

Gcover is 1000 times the coverage rate obtained using the Gaussian endpoints for each n,r combination.

Table 2. Two sample critical values and sizes based on W_r

n	r	distribution	95% cutoff	95% Gsize	99% cutoff	99% Gsize
10	3	uniform	2.29	38	3.45	7
		Gaussian	2.47	50	3.78	10
		slacu	2.54	55	3.92	12
		slash	3.18	97	5.42	31
10	4	uniform	3.59	48	6.80	9
		Gaussian	3.66	50	6.94	10
		slacu	3.72	52	7.08	11
		slash	4.07	63	7.98	14
20	5	uniform	1.66	30	2.09	5
		Gaussian	1.79	50	2.32	10
		slacu	1.83	57	2.39	12
		slash	2.16	105	3.04	37
20	7	uniform	2.14	44	3.02	8
		Gaussian	2.21	50	3.15	10
		slacu	2.23	52	3.18	10
		slash	2.37	65	3.46	15
30	8	uniform	1.56	33	1.91	5
		Gaussian	1.65	50	2.06	10
		slacu	1.68	55	2.10	12
		slash	1.85	88	2.40	27
30	11	uniform	1.98	45	2.68	8
		Gaussian	2.02	50	2.76	10
		slacu	2.03	51	2.76	10
		slash	2.10	59	2.89	13
40	10	uniform	1.44	30	1.69	4
		Gaussian	1.52	50	1.82	10
		slacu	1.55	56	1.86	12
		slash	1.69	93	2.11	30
40	14	uniform	1.74	44	2.20	8
		Gaussian	1.78	50	2.28	10
		slacu	1.78	51	2.28	10
		slash	1.84	60	2.38	13

Gsize is 1000 times the size of the test using the Gaussian critical value.

Table 3. Confidence intervals and coverages rates based on D_k

n	k	distribution	95% lower	95% upper	Gcover	99% lower	99% upper	Gcover
10	5	uniform	0.61	2.65	962	0.51	3.91	992
		Gaussian	0.59	2.80	950	0.50	4.09	990
		slacu	0.57	2.87	943	0.48	4.23	987
		slash	0.49	3.55	882	0.36	5.29	958
20	11	uniform	0.73	1.70	981	0.68	2.09	997
		Gaussian	0.68	1.83	950	0.60	2.26	990
		slacu	0.66	1.87	940	0.58	2.31	986
		slash	0.60	2.06	894	0.48	2.57	962
30	15	uniform	0.75	1.58	976	0.69	1.87	997
		Gaussian	0.71	1.66	950	0.64	1.98	990
		slacu	0.70	1.68	943	0.63	2.00	987
		slash	0.65	1.80	901	0.56	2.16	968
40	21	uniform	0.78	1.44	979	0.73	1.65	997
		Gaussian	0.74	1.51	950	0.68	1.74	990
		slacu	0.73	1.53	942	0.67	1.77	987
		slash	0.68	1.64	892	0.58	1.90	962

Gcover is 1000 times the coverage rate obtained using the Gaussian endpoints for each n,k combination.

Table 4. 1000 x power of two sample W_r test using Gaussian cutoffs

n	r	distribution	σ_1/σ_2 1.0	1.5	2.0	2.5
10	3	uniform	38	152	330	515
		Gaussian	50	175	344	508
		slacu	55	185	353	512
		slash	97	234	380	510
20	5	uniform	30	277	652	871
		Gaussian	50	304	621	827
		slacu	57	315	623	823
		slash	105	351	597	768
30	8	uniform	33	361	770	941
		Gaussian	50	372	732	911
		slacu	55	381	734	910
		slash	88	398	701	871
40	10	uniform	30	480	899	988
		Gaussian	50	474	856	972
		slacu	56	483	855	971
		slash	93	485	810	943

Table 5. 1000 times sizes and powers of nominal 5% tests

n	distribution	BA	J log v	J log bwt	Levene	FK	IQR
			Sizes				
10	uniform	62	33	80	35	39	38
	Gaussian	64	48	75	36	29	50
	slacu	57	62	81	33	27	55
	slash	133	142	109	18	37	97
20	uniform	56	35	71	37	41	30
	Gaussian	61	53	72	46	40	50
	slacu	51	70	74	36	36	57
	slash	125	160	93	18	39	105
30	uniform	53	40	84	35	43	33
	Gaussian	65	59	70	47	49	50
	slacu	49	69	67	39	38	55
	slash	122	166	114	15	39	88
			Power for $\sigma_1/\sigma_2 = 2$				
n	distribution	BA	J log v	J log bwt	Levene	FK	IQR
10	uniform	772	772	595	403	219	330
	Gaussian	590	545	467	307	158	344
	slacu	485	471	417	237	146	353
	slash	349	329	312	62	124	380
20	uniform	987	991	913	851	709	652
	Gaussian	877	867	792	691	517	621
	slacu	674	666	670	539	454	623
	slash	321	332	444	96	287	597
30	uniform	1000	1000	973	975	935	770
	Gaussian	973	972	928	894	793	732
	slacu	770	755	818	743	701	734
	slash	301	323	544	115	432	701

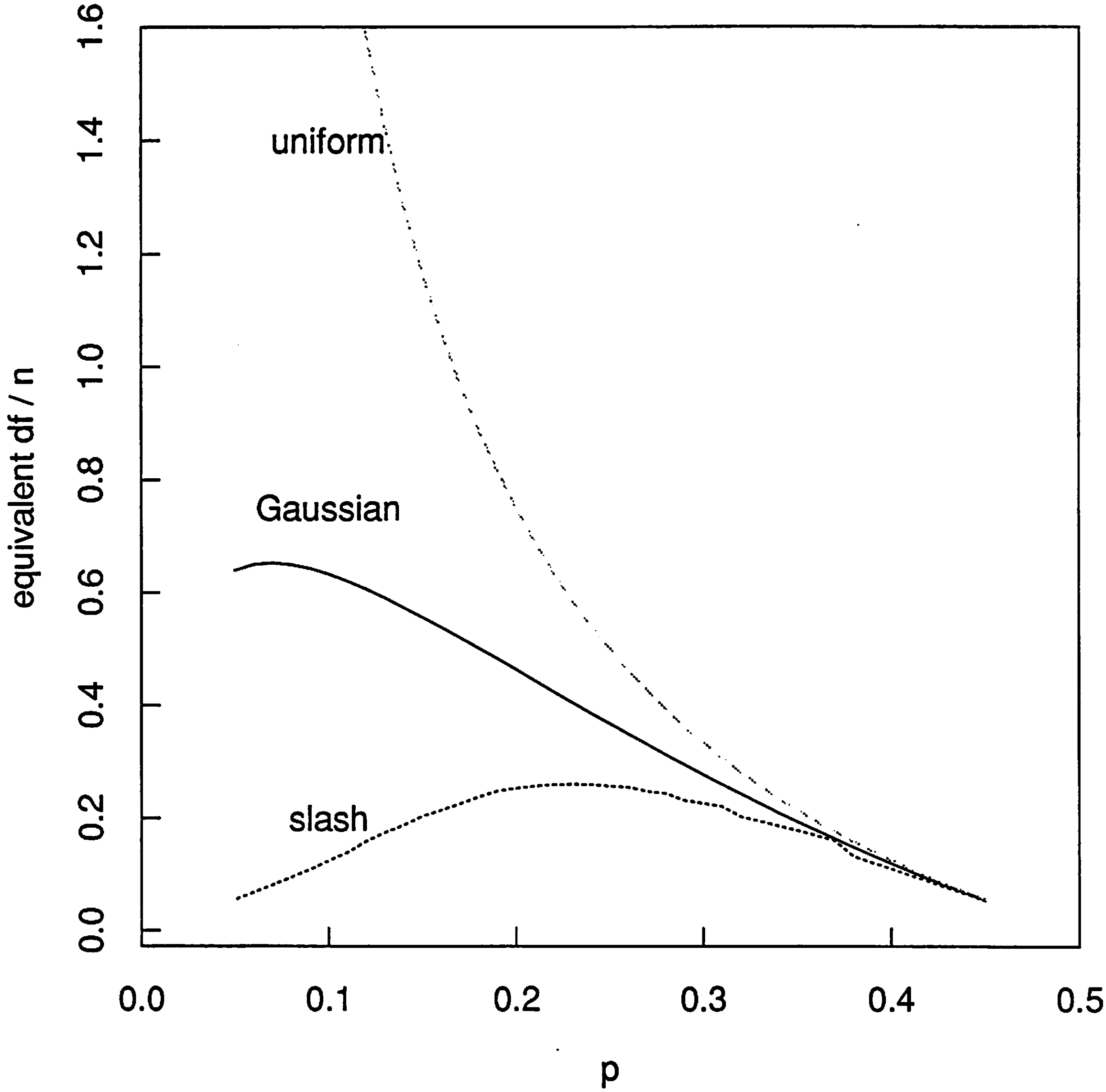
BA	Box Andersen procedure
J log v	Jackknifed log variance (Miller)
J log b	Jackknifed log biweight A-estimate
Levene	Levene procedure using absolute deviations from medians
FK	Flinger-Killeen linear rank test
IQR	Quasi-range test with $r = \lceil n/4 \rceil$

Figure 1. Quasi-range asymptotic equivalent degrees of freedom divided by sample size for uniform, Gaussian, and slash parent distributions and $r/n=p$.

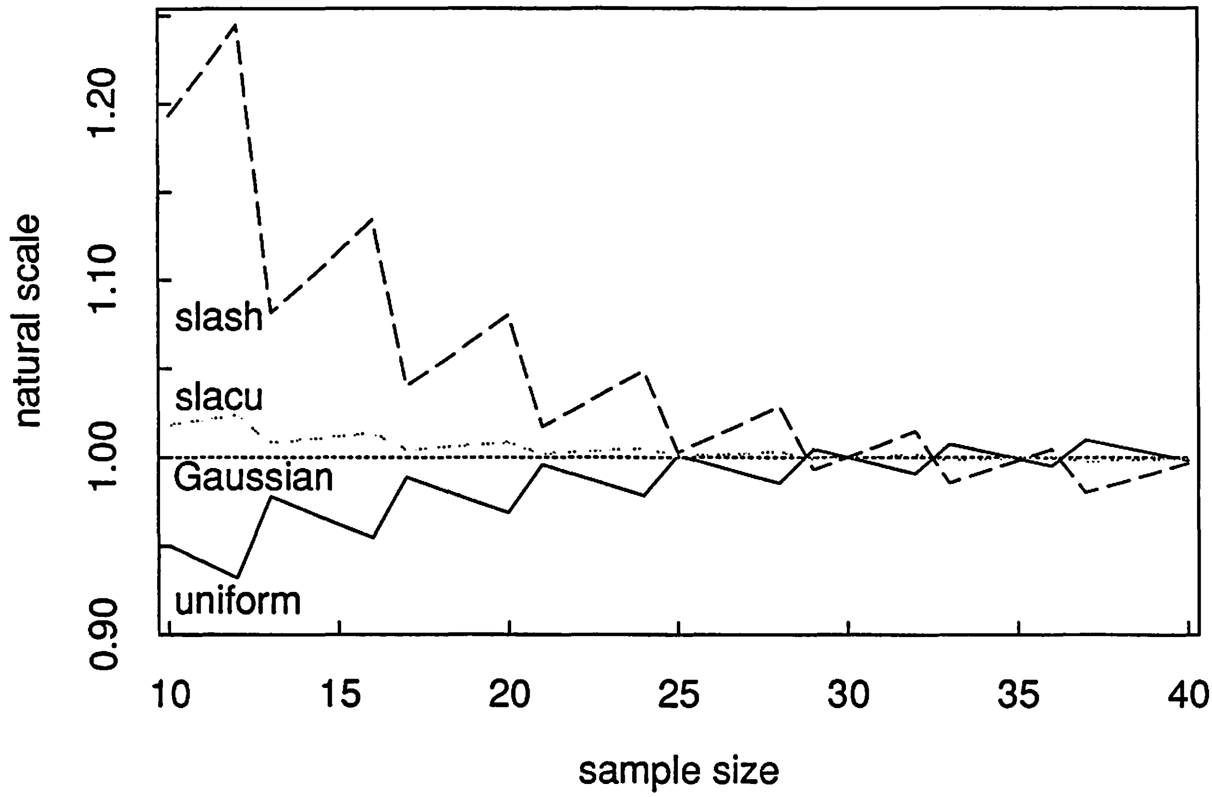
Figure 2. Quasi-range based natural scale for uniform, Gaussian, slacu, and slash distributions as a function of sample size for $r=\lceil n/4 \rceil$. Scales have been standardized to match at n=30.

Figure 3. Percent error in chisquare based approximate quantiles for quasi-ranges with $r=\lceil n/4 \rceil$ and Gaussian parent distribution as a function of the cumulative approximating probability and sample size. Solid - n=10, short dashes - n=20, dotted - n=30, long dashes - n=40.

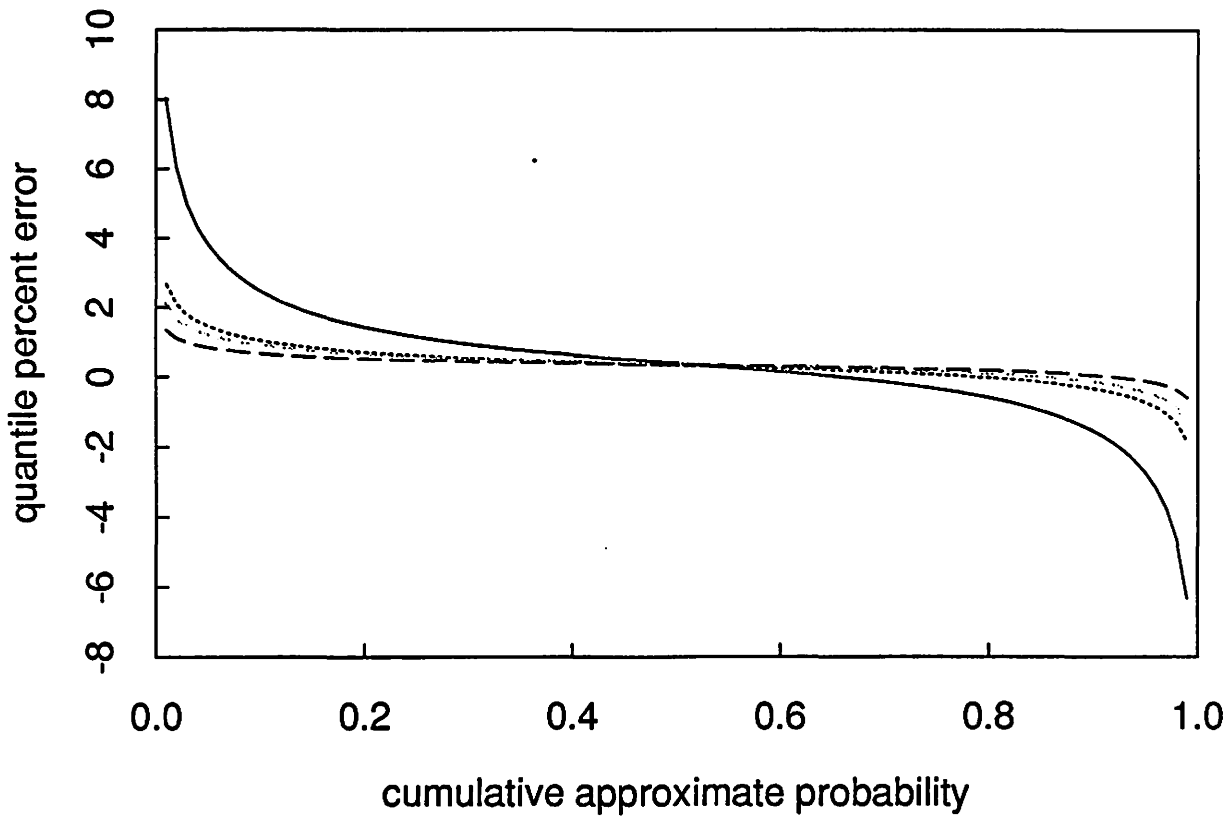
Figure 4. Percent error in F based approximate quantiles for ratios of equal sample size quasi-ranges with $r=\lceil n/4 \rceil$ and Gaussian parent distribution as a function of the cumulative approximating probability and sample size. Solid - n=10, short dashes - n=20, dotted - n=30, long dashes - n=40.



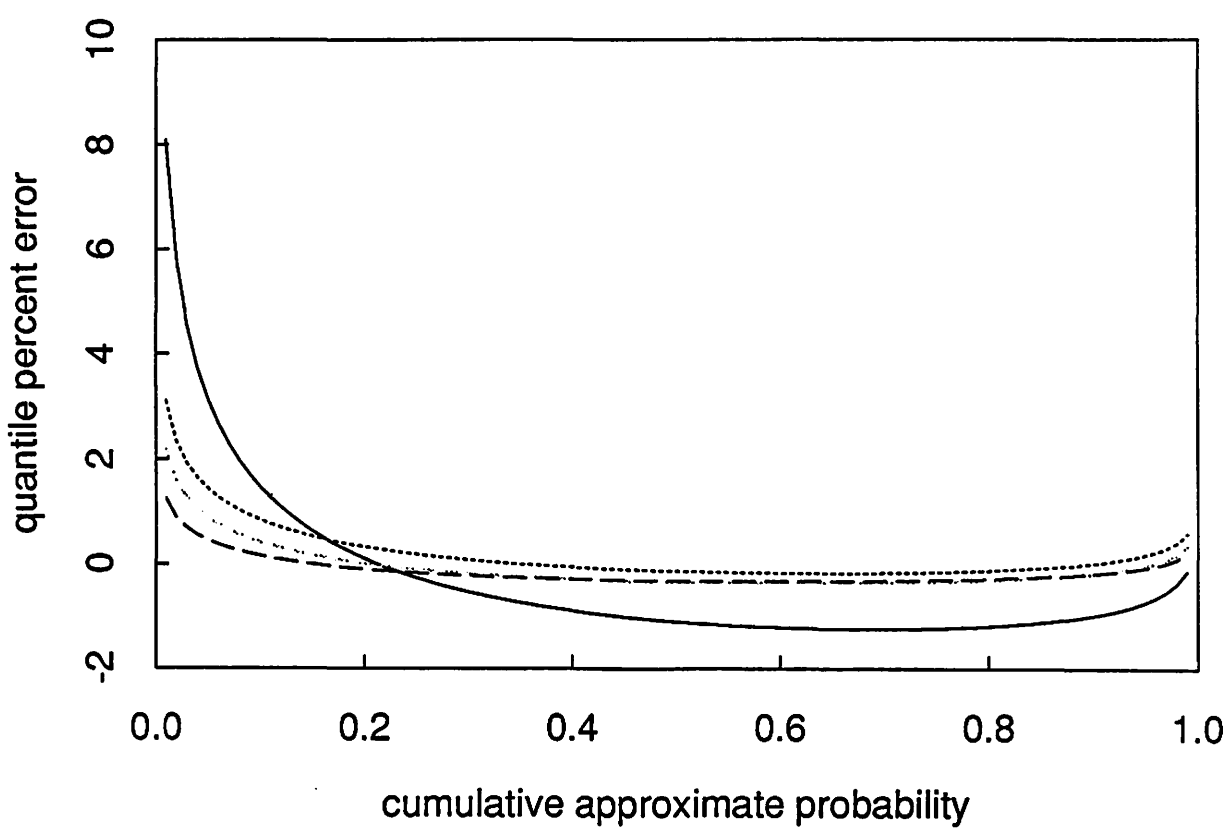
uniform
Gaussian
slash
equivalent df / n
p
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
0.0
0.1
0.2
0.3
0.4
0.5



natural scale
1.20
1.10
1.00
0.90
slash
slacu
Gaussian
uniform
10
15
20
25
30
35
40
sample size



10
8
6
4
2
0
-2
-4
-6
-8
quantile percent error
0.0
0.2
0.4
0.6
0.8
1.0
cumulative approximate probability



10
8
6
4
2
0
-2
quantile percent error
0.0
0.2
0.4
0.6
0.8
1.0
cumulative approximate probability