REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S): ARO 23360.12 MA
6a. NAME OF PERFORMING ORGANIZATION: Princeton University / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION: U. S. Army Research Office
6c. ADDRESS (City, State, and ZIP Code): Fine Hall, Washington Road, Princeton, NJ 08544-1000	7b. ADDRESS (City, State, and ZIP Code): P. O. Box 12211, Research Triangle Park, NC 27709-2211
8a. NAME OF FUNDING/SPONSORING ORGANIZATION: U. S. Army Research Office / 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: DAAL03-86-K-0073
8c. ADDRESS (City, State, and ZIP Code): P. O. Box 12211, Research Triangle Park, NC 27709-2211	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Procedures for separations within batches of values,
I. The orderly tool kit and some heuristics

12. PERSONAL AUTHOR(S) Thu Hoang and John W. Tukey

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Technical	FROM _____ TO _____	March 1989	105

16. SUPPLEMENTARY NOTATION
The view, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

First of several reports studying techniques to answer: "Is it desirable to treat this batch of values together, or to separate them into at least two parts?" (called below "the canyon problem").

This report focuses on tools to help answer the question, primarily on "the orderly tool kit" in which the simple general properties of order statistics from any reasonable distribution are identified and arranged for use.

Once we order the values of a set of observations, our attention needs to shift from values to gaps - - differences of adjacent ordered values. Rescaling gaps toward a common scale allows easy combination into measures of spread. Like other measures of spread, these measures of spread are well described in terms of "degrees of freedom", which can be defined for both linear and quadratic expressions. (Linear d.f. will be slightly less than 4 times quadratic d.f.) It can be natural and useful to order the rescaled gaps, when we now deal with gaps of rescaled gaps. (Various quantitative matters are also addressed.)

Application of the orderly tool kit to the canyon problem leads to test statistics for whose distribution a well-chosen F-distribution is a natural approximation. While the orderly tool kit provides a rough idea of the power of the procedure, simulation will be required to anchor and modify the approximation.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE (Include Area Code) / 22c. OFFICE SYMBOL

19. Abstract (cont'd)

More careful consideration of what we are trying to do in the canyon problem suggests strongly that other test statistics, for whose distribution we have no natural approximation, are almost certain to be more powerful.

KEYWORDS: canyon problem, correction for granularity, degrees of freedom (as measures of scale), e-leaps, focused knowledge, gaps (among order statistics), leapalls, leaps (= rescaled gaps), multiple comparisons (as analogy), orderly tool kit, rescaled gaps, stretches, sum leaps, tightness of spread estimates

ARO 23360.12-MA

Procedures for separations within batches of values, I. The orderly tool kit and some heuristics

by

Thu Hoang* and John W. Tukey**

*Universite Rene Descartes
Laboratoire de Statistique Medicale
45, Rue des Saints-Peres
75270 Paris Cedex 06
Paris, France

and

**Princeton University
Fine Hall
Washington Road
Princeton, NJ 08544-1000

Technical Report No. 293
Department of Statistics
Princeton University 08544-1000

March 1989

Prepared in connection with research at Princeton University sponsored by the Army Research Office (Durham), DAAL03-86-K-0073.

TABLE OF CONTENTS

Procedures for separations with batches of values, I. The orderly tool kit and heuristics

Thu Hoang and John W. Tukey***

*Universite Rene Descartes
Laboratoire de Statistique Medicale
45, Rue des Saints-Peres
75270 Paris Cedex 06
Paris, France
and
**Princeton University
Fine Hall
Washington Road
Princeton, NJ 08544

A. Overview.

This is the first of a group of technical reports, directed toward the question "Is it desirable to treat this batch of values together, or to separate them into at least two parts?". This question arose most prominently for us in connection with the analysis of multicenter clinical trials, where we anticipate that centers are truly different, at least to a degree, but need not - - in fact should not - - be analyzed separately, one by one. It may be, however, that at least one analysis should treat them in two or three groups!

Most of the present report is concerned with the tools that seem appropriate for attacking this question. Since these tools are based upon order statistics, it is natural to refer to them as "the orderly tool kit". The simple aspects of order statistics involve, first, their typical values and then, more importantly, the distributions of the corresponding gaps, multiples of gaps, and other functions of gaps. Typical values are well approximated by

Prepared in connection with research at Princeton University sponsored by the Army Research Office (Durham), DAA03-86-K-0073.

$F^{-1}((3i-1)/(3n+1))$, whose values we call *working values*. Gaps and multiples of gaps, including those we call leaps (chosen to have similar average values), are distributed roughly exponentially and roughly independently.

Whenever our concern is even partially with spread, once we sort a set of values, our focus of concern almost automatically changes *from values to leaps*. In the present problem, among others, it is natural to look at the shorter of the observed leaps as a basis for assessing undisturbed scale. This means sorting the leaps themselves - - and, hence, focusing on the leaps of the leaps, which we will call e-leaps.

An important tool in managing indicators of spread/scale is the notion of "d.f." - - classically thought of as "degrees of freedom" but in the present context better thought of as "degrees of firmness" - - defined by some sort of matching of distribution of the quantity of concern with the distribution of some multiple of χ^2 "on the appropriate df". (Notice that "zero" is to be preserved.) We use d.f. to describe the distribution of any always positive quantity, in particular, *both* quantities *quadratic* in the observations *and* quantities *linear* in the ordered observations. (The d.f. for the square-root of some quantity will be roughly 4 times the d.f. for the quantity itself.) One reason why d.f. are convenient is that exponential distributions have d.f. = 2, so that individual leaps will have (linear) d.f. close to 2 and sums of (sufficiently nearly) independent leaps will have (linear) d.f. close to 2 times the number of summands.

If the leaps were exactly i.i.d., the gaps (from which the leaps came) would

be independent and exponential with differing scales. Thus the sum of the gaps - - the range - - would have fewer d.f. than the leapall - - the sum of all leaps. Thus the leapall should often be a helpful replacement for the range. Indeed, for average gaps spaced like a Gaussian, dropping something like 7% of gaps recovers about half the difference between df for range and df for leapall. Merely rescaling one gap at each end does as well, and also rescaling a second gap next to each end recovers half of what is left.

When we look at gaps from an exponential: (i) we go to e-leaps by multiplying by 1 (outer end), 2, 3, . . . and (ii) the sum of the e-leaps is identically the sum of the observations.

It is often natural to leave out the largest (not usually the end-nearest) leaps, which leads us to use e-leaps, treating the end-nearest ones specially. If it were natural to leave out the largest e-leaps - - a situation yet to arise - - we would want to use leaps of e-leaps.

Guidance for our specific problem is offered by the average leaps (say as calculated for a single Gaussian sample of 20). For two Gaussian samples, say of k and 20-k values, respectively, from two distributions with well-separated averages, all average leaps are somewhat increased, those near the single *valley* (between the batches from the two distributions) most.

A sample of 20 from a 50-50 mixture of two unit Gaussians separated about 3.5σ behaves, so far as average leaps go, much like 10 + 10 observations, 10 from each of two widely separated Gaussians.

A parallel problem - - that of multiple comparison, where we are most often comparing long-run averages associated with each of several treatments, and where we *do have* an estimate of an appropriate error, both serves to illuminate plausibility of approach to our central problem and offers a likely application of the orderly tool kit. Uses of multiple comparisons usually require attaining focused knowledge (knowing that something unspecified is different from something else unspecified is rarely, if ever, useful). Focused knowledge comes in gradations that often parallel the evolution of knowledge in the subject-matter field. The gradations for qualitative answers range from "can we say anything specific (say about some pair of treatment averages)" to "have we settled the order of all treatment averages". The early stages of qualitative knowledge - - as well as the classical procedure for *quant*itative knowledge - - currently involve using the range of a set of results (often treatment means) and comparing that range with (the square root of) an estimated variance for these results. Our earlier comparison of leapall and range raises the question whether replacement of range by leapall will be helpful here.

For our present problem, in which we have *no useful separate* estimate of variability, but which corresponds in flavor with the earliest stage of qualitative multiple comparison, our current understanding of multiple comparison suggests that we should compare the leapall with an assessment of between-treatment variability (between-center in our original problem) chosen to be relatively immune to the presence of a valley - - or a few valleys.

We are now ready to focus the orderly tool kit on our main problem in

more detail. Our work to this point suggests comparing the leapall (which is also the sum of the e-leaps) with a sum or mean of low-index e-leaps. This turns out to be equivalent to comparing a mean of high index e-leaps with a mean of low-index e-leaps, a comparison for which the classical F-table offers approximate % points. Once we have got this far, we shortly see no need for all leaps to be either high-index or low-index. By converting $A/B \geq c$ through $(A/B) - c \geq 0$ to $A - cB \geq 0$, we can even calculate an approximation - - of unknown quality - - to the power of such a procedure. This lets us look at alternative pairs of high-index and low-index means, suggesting the cases for which we might well begin simulation, in which some leaps are omitted.

When we think harder about the approach so far developed, we conclude that working with the four largest leaps may be fairly good, it is not good enough, especially since we probably want power against single valleys (and perhaps pairs of valleys). Only simulation seems likely to attack the question of what restrictions to apply. Its application is left for a later report.

Appendix A looks, in considerably more detail, at typical values of order statistics, particularly at typical values for gaps and leaps. Since exponentially distributed quantities have large (100%) coefficients of variation, biases of even ± 10% are not important. As a result, working values (based on $(3i-1)/(3n+1)$) are quite precise enough for most situations. Similarly, even the difference between basing leaps on the Gaussian distribution and basing them on a t-distribution on 3 d.f. has rather weak consequences (the leapall's d.f. goes down about 3% near $n = 20$). Thus the exact distribution used to

generate leaps is ordinarily not important, so long as it is, say, smooth and single humped.

If, for reasons of neatness or intellectual curiosity, we look into order-statistic typical values in more detail, we find such things as: (i) in the Gaussian case, working values are increasingly nearer the medians than the means as we move away from the ends of the sample; (ii) simple formulas describe the differences quite well; (iii) differences between mean and median (for Gaussian order statistics) are small compared to differences between Gaussian and logistic order statistic medians (and means); (iv) when we look at differences in typical value from one order statistic to the next, all three choices of typical gaps (as differences of means, or of working values, or of medians) give very similar results (displacement by a few % of the value); (v) on the contrary, median gaps are about 70% of mean gaps (since gaps are nearly exponential); (vi) means of lower-tail order statistics from the exponential (or the rectangular) behave rather as if 0 were an additional observation, while upper-tail order statistics deviate from working values in the opposite direction and by about half as much; (vii) when we ask what % point of each order-statistic distribution is given by the corresponding working value, all are between 48.66% and 51.34%, and indeed if we set aside the end order statistics, between 49.63% and 50.37% (very close to the median!), (viii) order-statistic means correspond to % points a few to several times further from the median, but still by only a few % of their value - - by a few hundredths of σ.

Appendix B looks at a single example of how grouping, with its

concomitant ties, makes it difficult for gaps, leaps and e-leaps, and how surprisingly well a simple "spreading-out" modification manages to keep the situation under control.

Appendix C hints, very briefly, how the orderly tool kit might be used to study skewness, etc.

1. The problem.

This is the first of a series of reports, aimed at the problem of examining a batch of values to inquire:

a) Is there evidence that it would be better to treat the values as two or more subbatches (as if there was a valley dividing the parameters they estimate)?

b) If there is, which splitting into two subbatches is most reasonable?

We address these inquiries initially for the case where we wish to treat the values as equally variable, but we wish to use (and may have) *no* information as to what this common variability may be. Later (perhaps in the third report) we plan to turn to the case where we have information about the distribution of variability for the different values.

2. Leading cases.

We want to have reasonable control of both size and some sort of power for our proposed procedure. So we need to fix a null situation and one or more families of alternative situations. Through doing this we seek guidance, but regard ourselves as looking at leading cases - - rather than making assumptions.

While the possible degree of robustness for this class of problems may well be quite limited, it is our eventual intention to have techniques that are at least moderately robust. As usual, however, we do not plan to begin our attack on a new problem by requiring robustness. (We will give some incidental thought to avoiding the worst forms of non-robustness.)

As the nullest null situation, we choose a random sample from a Gaussian distribution with unknown μ and σ. (We hope to look at least briefly at random samples from other distributions, say the logistic and the slash.)

As our initial set of alternatives we consider pairs of Gaussian distributions with a common σ and different μ. Here we feel free to consider either subsamples of fixed sizes, one from each distribution, or a single random sample from a mixture of the distributions. Again we start with the Gaussian case, but are ready to consider others. We shall say that such alternatives involve a single *valley* or *slot,* dividing one group from the other. (The otherwise better term "gap" has been seized for a different use.)

3. Additional criteria.

Clearly we need criteria to guide - - rather than enforce - - our choices of procedure. We plan to use the two-step approach set out in the introduction, asking:

a) first, should we make a separation, and

b) second, what separation seems indicated.

Our ideas of size and power of course refer to the first of these questions.

If we have a need to answer the first question at all - - if the need for some separation is not crystal clear - - it is most unlikely - - even for our initial set of alternative situations, where there *are really* two subsamples, separated by a single valley, and hence a single correct answer to the second question - - that we will often reach that exactly correct answer to the second question. We ought to expect a greater or less degree of misclassification. We shall strive to hold misclassification near a minimum - - but expect to be guided by some typical amount of misclassification, and *not* by the probability of perfection - - which here would be choosing the separation into two batches that exactly matches how the simulated data was generated.

4. Heuristics.

We do our best to be concerned with real problems. This always makes an asymptotic approach uncomfortably weak, since neither of us has seen a real problem where "n really tends to infinity". The present problem is even more difficult for asymptotics than usual, because so much of our interest in the problem of separations is confined to small values of n.

We endeavor, therefore, to be as realistic as we can. This means admitting we are working with approximations, frequently asking for some insight into how well our approximations do, and emphasizing "heuristics" rather than "theorems". In particular, an effect, a phenomenon, or an approximation that, in samples of usual size, contributes only a small part on our overall uncertainty will only be allowed for explicitly if it is quite easy to do this. Otherwise we will do better to just regard it as unimportant and not worth our

attention.

5. Stimuli.

Our attention was directed to this problem by a situation familiar in multi-center clinical trials, where it frequently seems natural to ask whether or not the clinics need to be separated into two (or more) subbatches at least, as one of the analyses. Since a clinic-to-clinic variance component could be substantial when clinics were randomly sampled from a single-humped population of clinics, it would be inappropriate to compare between-clinic variability with within-clinic variability as a basis for answering such questions (since estimates based on within-clinic variability might be badly biased downward).

We feel that the separation problem (the valley-seeking problem), as formulated generally above, is an appropriate general framework in which to approach such multi-center questions. We plan to discuss the details of such an approach in later technical reports of this series.

The formulation of the orderly tool kit - - and our orientation toward that kit - - were substantially influenced by the numerical results of empirical trial (simple simulation) of some earlier formulations of a "separation statistic".

THE ORDERLY TOOL KIT

6. Character.

By this label we identify a fairly generally applicable, finite sample approach that emphasizes order statistics and quantiles, and that depends on

approximations of varying quality, some very good, and on a willingness to treat distributions as smooth (while privately recognizing that actual distributions are almost all discrete at some level of detail). This kit gives much more widely applicable results than can be provided by the classical Gaussian tool kit, which has constraining emphases on narrow specifications, on maximum likelihood, and, almost automatically, on exact optimality, and which consequently lacks robustness (and calls for a major role for moments). Our use of the orderly tool kit will be two fold: (in this report) as a source of suggestions for what to calculate, and (in the next report) as a framework for interpolation (based on its approximations) among simulation-based specific results.

7. Order-statistic typical values.

Under the null hypothesis that $y_1 \le y_2 \le \ldots \le y_n$ is an ordered sample from a cumulative distribution $F(y)$, each of the *order statistics* y_i will have a median depending upon i, n and F. Moreover, $F(y_1) \le F(y_2) \le \ldots F(y_n)$ must be an ordered sample from the uniform [rectangular] distribution on [0, 1], whose medians we may write $a'_u(i|\, n)$ and pronounce "a-split of the i^{th} of n". Because F and F^{-1} preserve order, we must have

$$a'_F(i \mid n) = \text{median}\{y_i|\, n, F\} = F^{-1}(a'_u(i|\, n))$$

so that, once we have a satisfactory approximation to $a'_u(i|\, n)$, we are very well equipped as far as medians of order statistics go.

It is often helpful to work with the means of order statistics, for which we gain some simplicity by writing them in a similar form, in terms of $\bar{a}_F(i \mid n)$,

defined by

$$mean\,\{y_i \mid n,\ F\} = \bar{a}_F(i \mid n) = F^{-1}(\bar{a}_{uF}(i \mid n))$$

Here $\bar{a}_{uF}(i \mid n)$, which is implicitly defined, changes somewhat - - but not too much - - from one F to another. One reason for such helpfulness is that

$$ave\,\{y_{i+1}-y_i\} = ave\,\{y_{i+1}\} - ave\,\{y_i\}$$

an exact relation that does not extend, exactly, to medians. (In fact, since gaps are roughly exponential, while the order statistics are typically one-humped, the corresponding relation for medians fails by a factor of nearly 0.7.)

Since the distributions of order statistics for the usual single-hump parents (like Gaussian, logistic, etc.) are skewed outward - - as are the upper-tail order statistics of the exponential - - we can confidently expect that, in these situations as well as many others, the mean of a given order statistic will be outboard of its median.

Moreover, we can expect that $\bar{a}_F(i \mid n)$ will often be moderately close to $a'_F(i \mid n)$ so that there is hope that we can approximate both $a'_F(i \mid n)$, for all continuous distributions, and $\bar{a}_F(i \mid n)$, for at least many well-behaved distributions, by a single approximation.

8. Working values, and other related quantities.

Observation teaches us that

$$a'_F(i \mid n) \doteq F^{-1}\left(\frac{3i-1}{3n+1}\right)$$

is a surprisingly good approximation - - as we can see most simply, perhaps, in terms of the percentile of the distribution of the order statistic to which

$a'_F(i\,|\,n)$ corresponds. The largest deviation of this percentile from 50% arises for the extremes, $i=1$ or n, when n is large, where $a'_F(1|\,n)$ or $a'_F(n|\,n)$ is the 51.34% or 48.66% point of the distribution of the corresponding order statistic. For $i=2$ or $n-1$ the corresponding extreme values are 50.37% and 49.63%. For $i=3$ or $n-2$, 50.18% and 49.82%. For all other i the percentiles delivered by this approximation are even closer to 50%.

Accordingly we write

$$a_F(i|\,n) = F^{-1}(\frac{3i-1}{3n+1})$$

and feel free to use $a_F(i|\,n)$ wherever the null *median* of y_i would seem appropriate - - and in many instances where its null *mean* would seem appropriate. We omit the F when it is clear from context what F is intended, and we call $a(i|\,n)$ a *working value* for the corresponding order statistic.

We show numerically, in appendix A (Sections 39 to 48), (i) that for the usual distributions the working values fall between means and medians, and (ii) for the Gaussian and the logistic at least, that the differences between mean and median are not large. Thus it is sensible to use the working value as an approximation for either $\bar{a}(i|\,n)$, the mean of the i^{th} order statistic, or for $a'(i|\,n)$, its median. (When, as for the end order statistics for certain stretched-tailed distributions, the means are very extreme or even no longer finite, $a(i|\,n)$ approximates only the median.)

We have reached working values of broad usefulness by seeking, on the y-scale, a simple combination

$$F^{-1}(\frac{3i-1}{3n+1})$$

March 29, 1989

of the underlying cumulative distribution F with i and n. One useful thing that expression in terms of F^{-1} does, is to automatically adjust for changes of location and scale within a given shape of distribution.

If we want to describe order-statistic typical values - - for instance their mean or their median - - in more detail than is given by the working values (a rather uncommon occurrence), it is plausible to write

$$\text{typical value}_F(i|\,n) = F^{-1}\left(\frac{num_F(i|\,n)}{3n+1}\right)$$

and to plan to compare $num_F(i|\,n)$ with $3i-1$. Solving for $num_F(i|\,n)$, we get

$$\begin{aligned} num_F(i|\,n) &= (3n+1)\,F(\text{typical value}_F(i|\,n)) \\ &= (3n+1)(\text{own tail area}) \end{aligned}$$

a form that is not hard to think about, and one that leads to $num_F(i|\,n)$ that are often easily compared with $3i-1$. (Compare Sections 41-45.)

9. Medians, etc.

Given $y_1 \le y_2 \le \ldots \le y_n$, we naturally consider the median - - the value of y_i for which $i = \frac{1}{2}(n+1)$, using interpolation for even n - - as one natural indicator of centering (location). This has an efficiency of 2/3rds or more for most relevant distributions. It often seems worthwhile to bring in more order statistics to enhance efficiency.

When interpolation seems not to be desirable, a reasonable solution is to use the lomedian, which is the central order statistic for n odd and the lower of the two central order statistics for n even. Thus $i(lomed) = \lfloor i(med) \rfloor$, where $\lfloor\ \rfloor$ indicates "the greatest integer $\le$ ".

An impure (from the point of view of order), but often convenient, approach is to go in the direction of the midmean - - defined as the mean of the middle-half of the order statistics. A slightly purer approach would use a "medmid" in which we first form the mids

$$\frac{1}{2}(y_i + y_{n+1-i})$$

and then take the median of these mids. (The impurity, from the point of view of order, is now confined to *averaging* y_i with y_{n+1-i}, a process that does *not* commute exactly with monotone transformation.) For hand calculation, with hinges defined by $i(H)$ or $n+1-i(H^*)$ equal to $\frac{1}{2}(1 + i(lomed))$, the *trimean*

$$\frac{1}{4}(y_{i(H)} + 2y_{i(\text{med})} + y_{i(H^*)})$$

is essentially as efficient as the midmean.

For careful computer work and $n \geq 8$, we may want to use the more efficient - - and more impure - - biweight.

The main purpose of discussing centering here is to clear the way for discussing other aspects of batches and distributions.

10. Gaps and leaps, syms.

* gaps *

The second fundamental set of facts about the distribution of order statistics is equally important. As can be seen from a theoretical, heuristic argument summarized below:

- the *gaps* $g_2 = y_2 - y_1$, $g_3 = y_3 - y_2$, . . . , $g_n = y_n - y_{n-1}$ are approximately uncorrelated when the $\{y_i\}$ are order statistics from a smooth null distribution F.

(We can see why this is plausible in moderate generality by observing that:

(i) the order-statistics of a sample, regarded as a stochastic process, satisfy the strong Markov condition,

(ii) the strong Markov condition often leads to an approximate weak Markov condition,

(iii) an exact weak Markov condition would make the gaps exactly uncorrelated.)

However, the empirical results, showing low correlations for specific distributions and finite sample sizes are better evidence. (The sophisticated argument helps by suggesting that if we tried still other distributions we would find the same phenomenon.)

Moreover, empirical calculation also shows that:

- each gap is roughly exponentially distributed. (This approximation fails seriously only for the end gaps $g_2 = y_2 - y_1$ and $g_n = y_n - y_{n-1}$ of samples from extremely-tail-stretched F. Compare Filliben (1969).)

Most of the simple aspects of the behavior of order statistics and functions of order statistics involve gaps (or simple modifications of gaps, like those we are about to introduce). The approximations for means-or-medians of the distributions of order statistics are useful in conjunction with the nice properties of gaps. Together, these two points are the foundation of a widely applicable

approach, the essence of the orderly tool kit.

* leaps *

If now we introduce *leaps* (sometimes called "normalized gaps" or "standardized gaps") by

$$h_{i+1} = \frac{g_{i+1}}{a_F(i+1|\,n) - a_F(i|\,n)} \approx \frac{y_{i+1} - y_i}{a_F(i+1|\,n) - a_F(i|\,n)}$$

we will have

$$mean\,\{h_{i+1}\} = \frac{\bar{a}_F(i+1|\,n) - \bar{a}_F(i|\,n)}{a_F(i+1|\,n) - a_F(i|\,n)} \doteq \frac{a_F(i+1|\,n) - a_F(i|\,n)}{a_F(i+1|\,n) - a_F(i|\,n)} = 1$$

If F in $a_F(i|\,n)$ refers to a standard form of some distribution shape, rather than to the distribution of our y's, and the y's are distributed σ times as widely, but in the same shape, then

$$\text{mean}\{h_{i+1}\} \doteq \sigma \quad .$$

In both cases, the h's are approximately exponential.

We will sometimes find it convenient to work with the *denominator*, D_i, where

$$D_{i+1} = a_F(i+1|\,n) - a_F(i|\,n)$$

and to recall the simple relations

$$\text{leap} = \frac{\text{gap}}{\text{denominator}}$$

$$\text{gap} = (\text{denominator})(\text{leap})$$

as well as the fact that denominators often come from working values of an assumed - - or reference - - distribution.

The clearest virtues of the h_{i+1} are that they (i) are approximately i.i.d., (ii) we know their approximate distribution is exponential, (iii) they tell us about σ.

* syms *

For many purposes it is convenient to use the h_{i+1} themselves; for others we might like to use transforms that are more "nicely" distributed than being exponential. To this end, we may want to go over to the *syms*

$$v_{i+1} = 3.3(h_{i+1})^{0.3}$$

for which

$$\text{median}(v_{i+1}) \doteq 3.3\sigma^{0.3}$$
$$\text{pseudosigma}(v_{i+1}) \doteq \sigma^{0.3}$$

so that pseudosigma $(v_{i+1}) \approx \frac{1}{3.3}(\text{median}(v_{i+1}))$ where "pseudosigma" refers to an indication of standard deviation based on a symmetrical pair of % points. Similarly "mid" is an indication of center location based on a symmetrical pair of % points. Exhibit 1 shows how well these approximations hold for something exactly unit exponential. We see (a) that the pseudosigma of the syms decreases somewhat as we move into the tails, so the distribution is somewhat squeeze-tailed, and (b) that the mids drift somewhat to higher values as we move toward the tails, so the distribution is somewhat positively skewed. Both deviations are a lot smaller than we might have feared.

exhibit 1 about here

exhibit 1

Behavior of syms when h_{i+1} is unit* exponential

Tail area	% points of leaps, h_{i+1}		%points of syms, v_{i+1}		pseudosigma** for syms	mid***
50%	.6932		2.9565			2.956
20%	.2231	1.6094	2.1041	3.806	1.011	2.955
10%	.1054	2.3026	1.6802	4.2382	0.998	2.959
5%	.0513	2.9957	1.3538	4.5863	0.982	2.970
2%	.0202	3.9120	1.0236	4.9687	0.960	2.996
1%	.01005	4.6052	.8302	5.2178	0.943	3.024
0.5%	.00501	5.2983	.6737	5.4420	0.926	3.057
0.2%	.00200	6.2146	.5115	5.7087	0.902	3.110
0.1%	.00100	6.9078	.4155	5.8927	0.886	3.154

*For $\sigma \neq 1$, multiply the two "% points of h_{i+1}" columns by σ, and all columns further to the right by $\sigma^{0.3}$.

**These are pseudosigmas for v_{i+1}. For comparison, the pseudosigmas of h_{i+1} itself are: .824, .857, .895, .948, .968, .988, 1.028, 1.079, 1.118

***These are mids for v_{i+1}. For comparison, the mids of h_{i+1} itself are: .693, .916, 1.204, 1.524, 1.966, 2.308, 2.652, 3.108, 3.459.

pseudosigma = (difference of p% points for given quantity)/(difference of p% points for unit Gaussian)

mid = (sum of the two p% points)/2

If we look at mids and pseudosigmas for h_{i+1} itself in comparison with these quantities for v_{i+1}, we see (a) that we would need an exponent below .3 (perhaps about .27) to keep the mids nearly constant and an exponent above .3 (perhaps about .38) to keep the pseudosigmas nearly constant. Thus .3 is a reasonable compromise. (Since we want our symmetrizing transformation to work for all $\sigma > 0$, we have little convenient choice but to use some power transformation.)

11. Stretches and sumleaps, leapalls.

Beside the gaps, $y_{i+1}-y_i$, it is natural to consider 3-stretches

$$y_{i+2}-y_i = (y_{i+2}-y_{i+1}) + (y_{i+1}-y_i) = g_{i+2} + g_{i+1}$$

or 4-stretches

$$\begin{aligned} y_{i+3} - y_i &= (y_{i+3}-y_{i+2}) + (y_{i+2}-y_{i+1}) + (y_{i+1}-y_i) \\ &= g_{i+3} + g_{i+2} + g_{i+1} \end{aligned}$$

and their further generalizations.

While we could stabilize the scale of $g_{i+2} + g_{i+1}$ by dividing by $a(i+2|\,n) - a(i|\,n)$, it seems better to stabilize the individual gaps, going to

$$h_{i+2} + h_{i+1} = \frac{y_{i+2}-y_{i+1}}{a(i+2|\,n)-a(i+1|\,n)} + \frac{y_{i+1}-y_i}{a(i+1|\,n)-a(i|\,n)}$$

and its generalizations, which it would be natural to call *sumleaps*, since they are sums of adjacent leaps.

To the extent that the individual leaps are exponentially distributed - - behave like a certain multiple of χ^2 on 2df - - and reasonably independent, a sum of k leaps is distributed rather like the same multiple of χ^2 on $2k$ degrees

of freedom. This approximation is helpful, rather than very exact. In the null situation, if we compare a sum of k leaps with the corresponding $(k+1)$-stretch, and measure stability by equivalent degrees of freedom, we should expect to have more stability for the sumleap (where all the summands - - the leaps - - have essentially the same scale) than for the corresponding stretch (where the summands - - the gaps - - do not have a common scale).

In particular, the *leapall* - - the sum of all the leaps - - should be somewhat more stable than the range.

12. e-leaps and iterated e-leaps.

One way to let our leaps tell us of the underlying scale is to summarize them directly. A simple - - and even efficient, if we had exact exponentials - - approach is to add them up, which corresponds to looking at the leapall. If we were deeply enough dedicated to the null hypothesis, we might do just this. In many practical non-null instances, however, we anticipate that the most leaps will approximate null behavior, but some will be rather inflated. If we want to take account of this, we will want to use the small h's to tell us about σ, without being bothered by large ones. How is this to be done?

To about as good an approximation as the original y_i were a sample (of n) from $F(y)$, the h_{i+1} are a sample (of $n-1$) from an exponential. So it is natural to reorder them, starting with index 2, so that $h_2^* \leq h_3^* \leq \ldots \leq h_n^*$ and to form first their e-gaps and then their e-leaps, namely, first

$$h_{i+2}^* - h_{i+1}^*$$

and then, if we continue to use working values as a denominator,

$$e_{i+2} = \frac{h^*_{i+2} - h^*_{i+1}}{\ln(1-\frac{3i-1}{3n-2}) - \ln(1-\frac{3i+2}{3n-2})} \qquad \text{(temporary)} \ .$$

Recalling that we adopted a denominator based on working values as, among other things, an approximation to one based on order-statistic means, we should be ready to take advantage of the simple form of the mean gaps from an exponential distribution and plan to use

$$e_{i+1} = (n-(i+1))(h^*_{i+1} - h^*_i) \qquad \text{(permanent)}$$

where we include h^*_1 as automatically = 0, and calculate either form of e_2 accordingly (as well as calculating e_3, e_4, . . . , e_n which depend wholly on data-derived h^*_i).

The e_{i+1}'s will again be (i) nearly i.i.d., (ii) nearly exponential, (iii) nearly uncorrelated. Clearly we could iterate the process - - but it is not clear that it will pay us to do so.

The more plausible approach would be to take the e_{i+1} corresponding to small values of i (which means small values of h^*, the same as small values of h) and regard them as telling us about the σ corresponding to the well-behaved values of h. This they can do easily through the value of the corresponding sumleap.

13. Seductive, but not recommended.

Given a set of e_{i+1}'s say m in number, someone sufficiently hipped on an orderly approach might suggest the following scheme:

a) Order the e's, as $e_1^* \le e_2^* \le \ldots \le e_m^*$

b) Notice that, for the exponential distribution,

$$a_E(i|\,m) = -\sigma \cdot \ln(1-\frac{3i-1}{3m+1})$$

c) Form

$$\frac{e_i^*}{-\ln(1-\frac{3i-1}{3m+1})}$$

as an estimate of σ.

d) Combine these, using a median, a midmean, or a mean.

The weak point in this procedure is the severe lack of independence of the e_i^*. The e_{i+1} were, roughly, independent, the $e_{i+1}^* - e_i^*$ are roughly independent, but the e_i^* are very far from independent (in particular because their values are ordered). Consequently this is a much poorer approach than that of the last section.

A simple, distantly related example may help to clear the ideas! If $z_1, z_2, \ldots, z_{n-1}$ all estimate θ, with similar precision and small correlations, then each of z_1, $\frac{1}{2}(z_1+z_2)$, $\frac{1}{3}(z_1+z_2+z_3)$, . . . , also estimates θ. The mean of the latter quantities is $1/(n-1)$ times their sum

$$(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n-1})z_1 + (\frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n-1})z_2 + (\frac{1}{3} + \ldots + \frac{1}{n-1})z_3 + \ldots$$

which for $n=20$ reduces to 1/19 of

$$3.55z_1 + 2.55z_2 + 2.05z_3 + 1.71z_4 + \ldots + .23z_{16} + .17z_{17} + .11z_{18} + .05z_{19}$$

which, with its largest weight ratio of (3.5477/.0526) = more than 67 to 1, is obviously a poor utilization of z's of similar precision and small correlations.

14. Going further (another time)?

We have explored centering (location estimation) and widthing (scale estimation). There clearly remains a question of shape estimation, which should almost certainly be approached as a matter of direction, character, and behavior of deviations from some reference shape.

We are not going to need to do this in the present report, so we need at most suggest directions of inquiry.

Some sort of fitting or smoothing procedure seems the natural approach. The question is "which sort, applied to what?" the most naive approaches would use (discrete) orthogonal polynomials or (discrete) Fourier functions, applied perhaps to the h_{i+1} or the v_{i+1}. We defer further discussion, except for *Appendix C (Section 53)*.

EQUIVALENT D.F.

When dealing with problems of spread or scale, we usually calculate quantities that are essentially positive, whose stability often concerns us. We want to describe stability in a way unaffected by taking a constant multiple. Two solutions have often been used (other than to look at the "coefficient of variation"), one is to look at the variance of the logarithm of our positive quantity, the other is to use "d.f.". In the present context, the latter seems more convenient.

15. Equivalent degrees of freedom. - - or of firmness.

If we have a reasonably well-behaved distribution of positive values, it is natural to relate it - - with a degree of approximation to be looked into in each special case - - to a multiple of χ^2 with some equivalent degrees of freedom, referred to as "d.f.". In such a situation the "f" can more usually be regarded as referring to "firmness" since multiples of χ^2 with more d.f. are quantities whose standard deviations are smaller fractions of their mean. This will probably seem most natural when the values whose distribution is being discussed are values of a homogeneous quadratic function of the observations, like $s^2 = \sum(y_i - \bar{y})^2/(n-1)$, which, under exact Gaussian hypotheses, is distributed like a multiple of χ^2 on $n-1$ degrees of freedom. But we need not confine ourselves to such (quadratic) d.f.!

If, say, s^2 is positive, so too is s, its positive square root. And we can attach a (linear) equivalent number of d.f. to s, just as we can attach a (quadratic) equivalent number of degrees of freedom to s^2. For both cases, our reference is the distributions of χ^2, so that s, with its smaller relative variability (smaller coefficient of variation), will have many more (linear) equivalent d.f. than s^2 has (quadratic) equivalent d.f. .

* bases for assessing d.f. *

Before we illustrate this possibility, we turn aside to ask how one might calculate an equivalent number of d.f., given more or less information about the distribution shape in question - - which is essentially sure to be specified by giving an example distribution belonging to the shape. Our means of assigning d.f. need to be dimensionless, in the technical sense of that word, so that we

assign the same number of d.f. to different distributions belonging to a single shape.

If we know % points, it seems natural to form such ratios as

$$\frac{\text{upper 10\% point}}{\text{lower 10\% point}}$$

$$\frac{\text{upper 5\% point}}{\text{lower 5\% point}}$$

$$\frac{\text{upper 1\% point}}{\text{lower 1\% point}}$$

(all of which *are* dimensionless) and then to interpolate in a table giving the corresponding ratios for χ^2's with various d.f. . (When we need to be specific, these are naturally called 10/10% d.f., 5/5% d.f. and 1/1% d.f. .)

If instead (or additionally) we know some moments, it is natural to start with dimensionless moment ratios, like

$$(\text{average})^2/\text{variance}$$

which is the reciprocal of the square of the coefficient of variation. A look at the low moments of χ^2 shows that, for this prototypic case,

$$\text{d.f.} = 2\frac{(ave)^2}{var} \qquad \text{(for multiples of chi-square)}$$

so that it is natural to put, rather generally, in the quadratic case

$$\text{moment d.f.} = 2\frac{(ave)^2}{var} \left(= \frac{2}{(\text{c.v.})^2}\right) .$$

* a first example *

To see how this all works out in a simple case where we know both % points and moments for both a quantity and its square root, it is easy to take

chi-square (with various d.f.) as the quantity. Exhibit 2 shows the results.

exhibit 2 about here

Clearly the paired-% point and moment assessments agree (for the square root) better than we might have anticipated. In this case - - namely $\sqrt{\chi^2}$ - - it seems to make very good sense to assign (linear) equivalent d.f. and plan to make heavy use of them.

In particular we note how closely

(linear scale equiv.d.f.) = -1 + 4 (quadratic scale equiv.d.f.)

holds for this case. Indeed, adding (0.7)/(quadratic scale equiv.d.f.) to the right-hand side gives almost perfect agreement. (Most deviations are apparently due to rounding in the tables of moments or percent points.) We may be able to do well - - hopefully - - by borrowing this relation for other situations.

16. Ranges and leapalls.

The Gaussian range provides another convenient example. Exhibit 3 gives some equivalent degrees of freedom for Gaussian ranges, calculated as indicated. For $n \leq 45$, the short-tailedness of the Gaussian parent makes the number of degrees of freedom greater than $2(n-1)$ (recall that the range is the only n-stretch). For $n \geq 50$, the converse is true.

exhibit 3 about here

For a logistic F, with its nearly exponential tails, our leaps will be more nearly exponentially distributed, and more nearly independent. (For a single

exhibit 2

Equivalent degrees of freedom for χ^2 (quadratic scale)

and $\sqrt{\chi^2}$ (linear scale) using either % points and moments

to fix the equivalent degrees of freedom

d.f. tabular	d.f. (quadratic scale) for χ^2 10/10%	5/5%	1/1%	moments	d.f. (linear scale) for $\sqrt{\chi^2}$ 10/10%	5/5%	1/1%	moments	4Q-1 + .7/Q*
3	3	3	3	3	10.07	9.83	9.31	11.23	11.23
5	5	5	5	5	17.95	17.67	16.99	19.14	19.14
9	9	9	9	9	33.87	33.51	32.79	35.08	35.08
19	19	19	19	19	73.86	73.81	72.66	75.04	75.04
29	29	29	29	29	113.83	113.50	112.63	115.03	115.02
40	40	40	40	40	157.79	157.50	156.62	159.02	159.02
50	50	50	50	50	197.80	197.45	196.57	199.00	199.01
60	60	60	60	60	237.80	237.45	236.57	238.99	239.01
100	100	100	100	100	397.79	379.44	356.55	399.03	399.01

*Q is the "quadratic scale" equivalent d.f. and the form $4Q - 1 + \frac{.7}{Q}$ was found empirically.

L is linear scale

(A corresponding approximation would be $Q \doteq \frac{L+1}{4} - \frac{.7}{L+1}$).

exhibit 3

Equivalent degrees of freedom (as a multiple of χ^2, linear scale) for the Gaussian range. [Note that χ^2 is here being used to approximate the distribution of a *linear* function of order statistics.]

n	matching ratios* 10%	5%	1%	first two moments**	$2(n-1)$	ratio of last two	(***)
5	13.6	13.3	12.7	14.5	8	.55	(95.6)
10	29.2	29.0	28.5	29.8	18	.60	(84.9)
15	42.8	41.8	41.5	42.2	28	.66	(76.6)
20	53.0	52.7	52.5	52.54	38	.723	(70.0)
30	70.4	70.4	70.1	70.51	58	.823	(61.3)
40	84.8	85.0	85.0	83.39	78	.935	(58.8)
50	97.1	97.1	97.3	95.21	98	1.030	(46.5)
60	107.6	107.7	108.0	105.49	118	1.119	(44.8)
70	117.2	117.3	117.6	114.48	138	1.205	(41.6)
80	125.7	125.8	125.8	122.72	158	1.287	(39.0)
90	133.6	133.8	133.9	130.45	178	1.365	(36.7)
100	141.5	140.7	141.4	137.43	198	1.441	(34.8)
200				188.3	398	2.11	(23.7)
500				268.7	998	3.71	(13.5)
1000				340.3	1996	3.87	(8.5)

*Degrees of freedom to match ratio (upper p % point)/(lower p % point).

**Value of twice (average)^2/variance.

***For χ^2 the linear df is $f_x = 4(n-1) - 1 + (.7/(n-1))$. This column gives 100 (moment df)/f_x, a measure of the efficiency of the range.

exponential F, each gap $y_{i+1}-y_i$ is exactly exponential, both unconditionally and conditionally on y_i. Consequently the gaps are exactly independent in this case.) In the logistic situation then, the approximation of $2(n-1)$ degrees of freedom for the logistic leapall seems likely to be quite good for all n. Moreover, the degrees of freedom for the logistic range should be appreciably less than for the logistic leapall.

17. Relation to variance of the logarithm.

The variance of $\log(s^2)$, where s^2 is distributed like a multiple of χ^2, is approximately (Bartlett and Kendall, 1946)

$$\frac{2}{(d.f._Q)-1} \qquad \text{(check this!)}$$

where d.f._Q is the (quadratic) d.f. for the χ^2 concerned. So another way to set d.f. would be to start from this approximation which implies

$$\text{d.f.}_Q = \frac{2}{var(\log(\text{quantity}))} + 1 \quad .$$

Clearly

$$\begin{aligned} var(\log(\sqrt{\text{quantity}})) &= var(\tfrac{1}{2}\log(\text{quantity})) \quad , \\ &= \tfrac{1}{4}\, var(\log(\text{quantity})) \quad , \\ &= \frac{1}{4}\,\frac{2}{(\text{d.f.}_Q)-1} \quad , \end{aligned}$$

so that the corresponding d.f., say d.f._L for $\sqrt{\text{quantity}}$ satisfies

$$\text{d.f.}_L = \frac{2}{var(\log(\sqrt{\text{quantity}}))} + 1 = \frac{2}{\frac{1}{4}\frac{2}{(\text{d.f.}_Q)-1}} + 1 \ ,$$

$$= 4(\text{d.f.}_Q - 1) + 1 \ ,$$

$$= 4\text{d.f.}_Q - 3 \ ,$$

which again illustrates linear d.f. as about 4 times quadratic d.f. .

SOME HELPFUL SPECIAL CASES

18. I.i.d. leaps.

If the leaps are i.i.d. with a distribution for which $2ave^2/var = 2\theta$, then the leapall, as the sum of $n-1$ independent terms, will have a value of $2\ ave^2/var$ that is $(n-1)$ times as large.

Hence we have

$$\text{moment d.f.(leapall)} = 2\frac{(ave(\text{leapall}))^2}{var(\text{leapall})} = 2(n-1)\frac{(ave(\text{basicleap shape}))^2}{var(\text{basicleap shape})} = 2(n-1)\ \theta \ .$$

If $\{D_i\}$ are the denominators that convert gaps into leaps, then

$$\text{range} = \sum(\text{gaps}) = \sum D_i(\text{leaps})$$

so that (using i.i.d. for leaps)

$$ave(\text{range}) = \sum D_i\ ave(\text{leap}) \ ,$$

$$var(\text{range}) = \sum D_i^2\ var(\text{leap}) \ ,$$

$$\text{moment df(range)} \approx 2\frac{(\sum D_i)^2}{\sum D_i^2}\theta \ ,$$

where the last relation follows from $[ave(\text{leap})]^2/[var(\text{leap})] = \theta$.

If our i.i.d. leaps come from gaps that have averages proportional to the average gaps of a unit Gaussian sample, these gap averages can serve as the D_i, making it easy to calculate $\sum D_i$ and $\sum D_i^2$. Exhibit 4 shows results for selected values of n. We see that, particularly for somewhat larger n, the stability of such a range is appreciably less than for the corresponding leapall.

exhibit 4 about here

There is some interest in understanding where this loss in efficiency for ranges is concentrated. What if we fix up only a few near-end gaps? For $n = 10$, for instance, the range, in terms of gaps, is

$$((y_2-y_1) + (y_3-y_2) + (y_4-y_3) + (y_5-y_4) + \ldots \quad)$$

while the leapall, also in terms of gaps is

$$\frac{y_2-y_1}{.538} + \frac{y_3-y_2}{\ldots} + \frac{y_4-y_3}{\ldots} + \frac{y_5-y_4}{\ldots} + \ldots + \frac{y_{10}-y_9}{.538} \quad .$$

A compromise would be to use

$$\frac{y_2-y_1}{??} + (y_3-y_2) + (y_4-y_3) + (y_5-y_4) + \ldots + \frac{y_{10}-y_9}{??}$$

where *??* is chosen to produce a multiple of

$$\frac{y_2-y_1}{.538} + \frac{y_9-y_2}{1.001/3.5} + \frac{y_{10}-y_9}{.538}$$

where $1.001 = -a_{Gau}(2|\,10) = +a_{Gau}(9|\,10)$.

(This gives *??* = (.538)(3.5)/(1.001) = 1.874.) We call the result the one-step compromise.

A more refined compromise would be to choose *???* and *????* to do as well with

exhibit 4

Values of $\sum D_i$ and $\sum D_i^2$, and of moment d.f. for a

range based on i.i.d. leaps (shape fixes θ)

when the D_i are proportional to average Gaussian gaps

n	$\sum D_i$	$\sum D_i^2$	Moment d.f. for range	Moment d.f. for leapall	ratio*
10	3.078	1.16227	16.303θ	18θ	90.5%
20	3.735	.928121	30.045θ	38θ	79.10%
30	4.086	.8060	41.30θ	58θ	71.2%
45	4.42	.7119	54.71θ	88θ	62.2%
50	4.50	.6936	58.39θ	98θ	59.60%

*(moment d.f. for range)/(Moment d.f. for leapall)

$$\frac{(y_2-y_1)}{???} + \frac{(y_3-y_2)}{????} + (y_4-y_3) + \ldots$$

as we can. We shall call this the 2-step compromise (And so on.)

Exhibit 5 shows the consequences of (a) dropping end gaps entirely and (b) using compromise weights.

exhibit 5 about here

We see from the exhibit that, roughly, fixing up one pair of gaps takes us half way up to 100%, fixing up a second one takes us about another half, and so on. Dropping a well-chosen number of gaps gains *one* of these powers of two for $n = 50$ - - but gains much less for $n \leq 30$.

19. Samples from a Gaussian.

We have already examined (see exhibit 3) the linear degrees of freedom for the Gaussian range. We can now try borrowing the ratio df(leapall)/df(range) from the i.i.d. situation of the last section and applying in to the Gaussian case, where (a) gaps or leaps are somewhat shorter-tailed than an exponential and (b) there are some non-zero correlations. The result is an approximation - - of unknown quality - - to the moment d.f. for a Gaussian leapall when the underlying distribution is Gaussian.

Exhibit 6 shows the results of such a calculation. We clearly expect the Gaussian leapall to be more stable than the Gaussian range, more and more noticeably as n increases.

exhibit 6 about here

exhibit 5

Fixing-up the ends only; the consequences, under Gaussian hypotheses, in (linear-scale) moment d.f. terms** of (a) omitting end gaps, and (b) rescaling them appropriately

	$n=10$	$n=20$	$n=30$	$n=50$
Range	16.30*	30.04	41.11	58.82
Drop 1 + 1 gaps	13.74	31.32*	46.70	72.97
Drop 2 + 2 gaps	9.97	28.98	46.37	77.21
Drop 3 + 3 gaps	6.00	25.60	44.10	77.76
Drop 4 + 4 gaps	2.00	21.86	40.99	76.52
Range (again)	16.30(90.6%)	30.04(79%)	41.31(71.2%)	58.82(60.0%)
1-step compr.	17.74(98.58%)	35.20(92.6%)	50.63(86.6%	76.78(78.3%)
2-step compr.	17.97(99.8%)	36.96(97.3%)	54.35(94.0%)	85.06(86.8%)
3-step compr.	17.99(99.9%)	37.60(98.9%)	56.0(96.5%)	89.66(51.5%)
4-step compr.	18.00(100%)	37.86(99.6%)	56.93(98.3%)	92.44(94.3%)
(leapall)	(18(100%))	(38(100%))	(58(100%))	(98(100%))

*Marks maximum in column.

**All values omit a factor of θ.

Exhibit 6

Approximate (linear) moment d.f. for Gaussian leapall,

calculated as (linear) moment d.f. for the Gaussian range

multiplied by the ratio (leapall d.f.)/(range d.f.) for the i.i.d. case

n	d.f. for range	ratio of d.f.'s	approximate d.f. for leapall*	linear d.f. for *s*	ratio
5	14.5	.979	14.83	19.14	(.77)
10	29.8	.905	32.92	35.08	(.93)
15	42.1	.843	49.94	59.04	(.85)
20	52.5	.7911	66.36	(79.03)	(.84)
30	69.57	.712	97.71	(119.02)	(.82)
45	89.49	.622	143.87	(179.03)	(.80)

*The corresponding quadratic-scale d.f., using the approximation reached in exhibit 2, are:
3.91, 8.46, 12.72, 16.84, 24.66, 36.23 whose ratios to 4, 9, 14, 19, 29, 44 are .983, .939, .909, .886, .8545, .823.

20. Samples from an exponential.

There are two striking results for samples from an exactly exponential distribution:

- the sum of the observations is a sufficient statistic for the parameter, and since the observations are i.i.d. with $\theta = 1$, the moment d.f. for this sum is $2(n-1)$,
- the e-leaps are also i.i.d. exponential, so their sum, the e-leapall, is also sufficient, again with moment d.f. = $2(n-1)$.

It seems anomalous to have two sufficient statistics, both for the same parameter, (although it is comforting that they do *not* have different equivalent degrees of freedom). It is easy to resolve the seeming paradox.

If $a_E(i|\, n)$ is the *mean* of the i^{th} smallest order statistics from the exponential, with $a_E(0|\, n) \equiv 0$ and $h_0^* \equiv 0$, then

$$a_E(i+1|\, n) - a_E(i|\, n) = \frac{1}{n-i}$$

so that the e-leapall is

$$n(h_1^* - h_0^*) + (n-1)(h_2^* - h_1^*) + \ldots + (h_n^* - h_{n-1}^*)$$

which collapses to

$$h_1^* + h_2^* + \ldots h_n^*$$

which is the original leapall. Thus our two exceptional results say the same thing about two quantities whose values are always identically the same. (What could be simpler?)

There is a simple generalization of this result, namely:

- a low-index sum of e-leaps is a Winsorized sum of (ordinary) leaps.

We illustrate for the 15 low-index e-leaps for $n=20$, for which

$$e_1 = 19(h_1^* - h_0^*)$$
$$e_2 = 18(h_2^* - h_1^*)$$
$$\ldots \quad \ldots$$
$$e_{15} = 5(h_{15}^* - h_{14}^*)$$

so that (since $h_0^* = h_0 = 0$)

$$e_1 + e_2 + \ldots + e_{15} = h_1^* + h_2^* + \ldots + 5h_{15}^*$$

in the right-hand side of which h_{16}^* to h_{19}^* have each been Winsorized to give h_{15}^*.

For the exponential, ordinary leapall and e-leapall coincide. As we move away from the exponential, which should we follow?

21. Alternatives, alternatives!

So far we have discussed the orderly tool kit as if only one situation needed to be considered. (We have occasionally used the words "null situation" but have not contrasted it with an alternative.) Yet essentially all statistical or data analytic procedures are only useful because "there might be alternatives!". (The natural candidates for exceptions are goodness-of-fit procedures, which seem to concentrate on the null situation. Yet if no alternative was possible, a bad fit is at most something to include in Guinness's book of records! The usefulness of a sufficiently bad fit is that it urges us to take alternatives quite seriously!)

Again much of our analysis is directed toward quantitative alternatives - -

quantitative diversity - - in situations where a few-parameter model may be quite diverse enough. That is not the kind of situation we plan to attack in this overall account - - where the basic alternatives are:

- it appears reasonable to treat our batch of numbers as a single entity,
- it appears better to treat it as consisting of at least two subbatches.

It is in this situation that careful formulations - - for both null and alternative situations - - can make the orderly tool kit applicable to as many alternatives and partial problems as we can arrange.

22. Alternatives and the "choice among exponentials".

We can go quite a way with a rather qualitative discussion of alternatives, especially in connection with the choice between leaps, e-leaps, or iterated e-leaps as the basis for a scale estimate. If we knew that a very narrowly described alternative, localized in terms of the original gaps - - say that our batch of 20 was the union of two samples of 10 (from populations of similar variance and distinctly different location), so that the most affected gap would be $y_{11}-y_{10}$ - - then we would know which gaps, in terms of their original identity (e.g. g_{11}) were likely to be non-null, so that we could begin by setting these gaps aside - - and using the remaining leaps as a basis for assessing spread.

If, on the other hand, our knowledge was less specific, saying only that at most a few gaps - - here, there, or elsewhere - - are likely to be seriously affected, we could not proceed in such a way even if we wanted to. If we are to

set aside anything in this latter case, it should be something like the large (long) leaps.

To identify the long leaps, we need to sort the leaps. After we have set aside some predetermined number, we have a collection of ordered values (beginning with one additional zero), and we should act in accordance with the basic principle:

> "Once we sort a set of values, our concern, so far as spread (width, scale, etc.) is concerned, shifts from the values themselves to their gaps, leaps, etc."

This means that we will want to *look* at the *e-leaps* (leaps of the leaps) and to set aside those that involve the *largest leaps* - - those associated with leaps that come at the end of the sequence of leaps as *originally* ordered, not necessarily those at the end of the sequence of ordered *e*-leaps.

So far, we have been *unable* to imagine a pair of situations such that it is natural to leave out the largest e-leaps. In such a situation we would be driven - - by a similar logic to that just described - - to working with the *iterated* e-leaps (leaps of leaps of leaps). Knowing none such, we presently see no excuse for working with iterated e-leaps.

23. Alternatives illuminating separation procedures.

We would find assessing an underlying (original?) spread simplest when we need only to set aside a small % (for smaller n, a few) of the longest leaps - - a few e-leaps with the largest sorted-leap indices. A sufficiently naive view of the "separations" problem might suggest it would be such a nice case. What

appears to be true is that this is not what happens - - although guidance based on such an oversimplified model may serve us quite well.

To get an idea of what actually might go on, let us use for guidance the leading case of two Gaussian samples, of sizes k and $20-k$, respectively, drawn from populations of equal variance fairly substantially separated. We can find working values for the samples separately, one set being displaced from the other by a substantial amount. For $(k-1) + (19-k)$ of the gaps, we can approximate median (or average) lengths in terms of working values. For the remaining gap, the gap between the highest of the low sample and the lowest of the high sample, we begin by only saying "might be large" and mark it down as "L".

exhibit 7 about here

The upper part of exhibit 7 shows leap with valley averages - - as multiples of *single sample (of 20) averages* for k = 10, 7, 4, 2 and 1. We see that the effects of splitting into two samples affects all gaps to a degree and that, while these effects are concentrated *at and near* the gap between the subsamples, they taper off more slowly than we would like (though perhaps faster than we might fear). These values are shown graphically in exhibit 8, using $3.3(\text{ratio})^{0.3}$-3.3 as the vertical scale, whose values are also shown in the upper part of exhibit 7.

exhibit 8 about here

exhibit 7

Ratios of $ave\{y_{i+1}\} - ave\{y_i\}$ for separated unit Gaussian samples

to $\bar{d}(i+1|\,20) - \bar{d}(i|\,20)$ and a near-unit near-symmetric re-expression of these ratios

	ratios					$3.3(\text{ratio})^{.3} - 3.3$				
i	10 + 10	13 + 7	16 + 4	18 + 2	19 + 1	10 + 10	13 + 7	16 + 4	18 + 2	19 + 1
19	1.168	1.098	1.048	1.024	1.011	.157	.094	.047	.024	.011
18	1.245	1.134	1.065	1.025	1.014	.224	.126	.063	.025	.014
17	1.333	1.176	1.081	1.038	1.014	.297	.165	.078	.037	.014
16	1.438	1.222	1.097	1.040	1.017	.380	.204	.093	.039	.017
15	1.587	1.277	1.123	1.052	1.026	.490	.252	.111	.050	.025
14	1.782	1.338	1.141	1.063	1.028	.625	.307	.133	.061	.028
13	2.105	1.429	1.180	1.075	1.038	.826	.373	.168	.073	.037
12	2.595	1.549	1.203	1.083	1.039	1.143	.461	.188	.083	.038
11	4.288	1.720	1.256	1.104	1.048	1.807	.583	.234	.099	.047
10	L	1.992	1.306	1.177	1.056	L	.758	.275	.115	.055
9	4.288	2.512	1.392	1.144	1.064	1.807	1.050	.344	.136	.062
8	2.595	3.938	1.508	1.160	1.078	1.143	1.678	.433	.168	.075
7	2.105	L	1.632	1.225	1.098	.826	L	.522	.208	.094
6	1.782	4.190	2.077	1.269	1.120	.625	1.772	.809	.261	.114
5	1.587	2.606	3.103	1.456	1.155	.490	1.099	1.335	.356	.146
4	1.438	2.006	L	1.614	1.210	.380	.766	L	.509	.194
3	1.333	1.687	3.486	2.238	1.338	.297	.556	1.500	.902	.301
2	1.245	1.458	2.144	L	1.675	.224	.396	.848	L	.552
1	1.168	1.298	1.595	2.456	L	.157	.267	.496	1.022	L
(•)	1.59	1.55	1.31	1.13	1.06					
(••)	1.33	1.29	1.13	1.06	1.03					
(10)	1.35	1.30	1.15	1.06	1.03					
(15)	1.59									
(18)	1.95									

L = non-negative (because non-overlapping), and possibly large

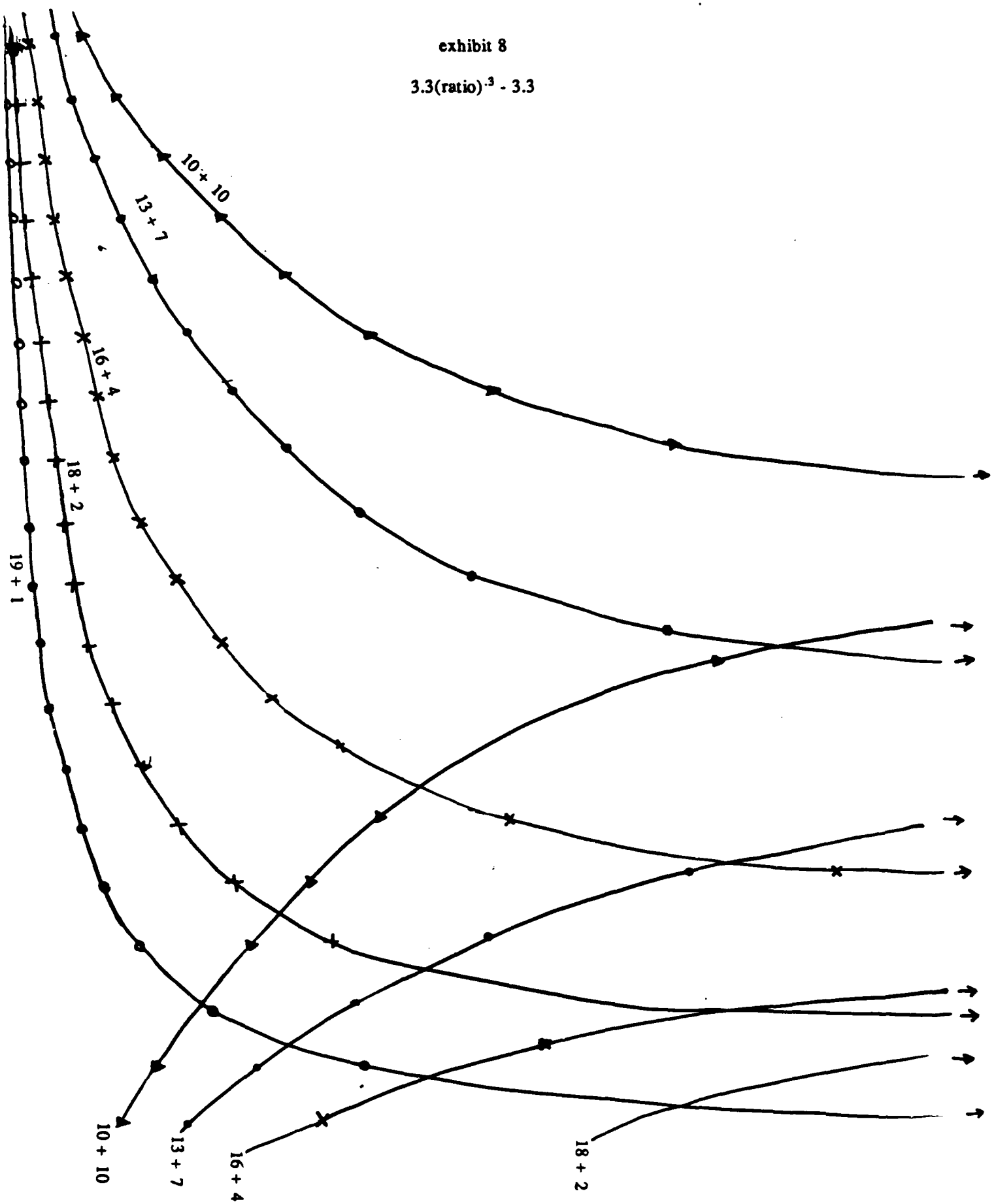
(•) = median, (••) = median of low 10, (10) = mean of low 10, (15) = mean of low 15, (18) = mean of low 19, zero and first approximations for interval rate shifts, t.

Shift		10 + 10	13 + 7	16 + 4	18 + 2	19 + 1
4.		(7.44)	(7.34)	(6.86)	(5.84)	(4.69)
	(a)	3.37	3.120	2.511	1.685	1.557
	(c)	8.36	8.13	7.38	6.26	4.74
	(a)	3.377	3.433	3.063	2.203	–
4.5		(11.47)	(11.09)	(9.71)	(7.65)	(5.77)
	(a)	4.0521	3.732	2.931	2.064	1.652
	(c)	11.71	11.29	9.72	7.78	5.79
	(a)	4.052	4.004	3.355	2.386	–
5.		(15.91)	(14.84)	(12.56)	(9.46)	(6.86)
	(a)	4.246	3.892	3.068	2.198	1.670
	(c)	15.55	14.888	12.46	9.48	6.86
	(a)	4.246	4.155	3.461	2.438	–
5.5		(19.54)	(18.59)	15.40)	(11.26)	(7.95)
	(a)	4.283	3.933	3.099	2.229	1.674
	(c)	19.54	18.59	15.41	11.27	7.45
	(a)	4.283	4.155	3.482	2.452	–
6.		(23.57)	(22.34)	(18.25)	(13.07)	(0.041)
	(a)	4.288	3.938	3.103	2.256	1.675
	(c)	23.57	22.34	15.25	13.07	9.04
	(a)	4.288	4.190	3.486	2.456	–

(a) = ratio for adjacent leap, (c) = ratio for leap at valley; both second approximation (see text)

exhibit 8

$3.3(\text{ratio})^{-3} - 3.3$



The lower part of exhibit 7 shows two approximations to the behavior of the leap at the valley. First, in parentheses, is shown the first approximation, correct to the extent that the two samples (of k and $20-k$ respectively) do not overlap at all. Thus, for example, for a shift of 5 (5σ, that is) the highest of one sample of 10 will average +1.539 and the lowest of the other will average 5 − 1.539 for an average difference of 5 − 1.539 − 1.539 = 1.922 which has to be divided by the denominator of .1240, to obtain 15.50 as shown in the parentheses.

The second approximation allows for as many as one crossover, and was calculated using the approximation

$$ave\{|z-a| \text{ when } z \text{ follows } Gau(0, \sigma^2)\} = |a| + 1.54\sigma(\text{tail area})^{1.2}$$

where "tail area" = $Gau(-|a|/\sigma)$. The approximation is good to ± 0.003 for $|a|/\sigma \geq .8$. We are using this Gaussian approximation in a quite non-Gaussian case, but we think it good enough for our purposes.

For shift 5 and 10 + 10, this involves $\sigma = (.34344+.34344)^{1/2} = .82987$ which for $a = 1.922$ gives $|a|/\sigma = 2.3160$, tail area = .01028, 1.54 (.82987) $(\text{tail area})^{1.2} = .0053$. Since the average of the absolute difference is .0053 greater than the average, $|a|$, of the signed difference, .0053 must be the contribution of crossover which, when divided by .12399, is to be added to the ratio at the valley, and, when divided by .12496, is to be subtracted from the ratios for each of the two adjacent gaps.

Parallel calculations lead to the results in the lower panel of exhibit 7. While the results for the NW corner of the panel are likely not to be good

approximations (because of more than one overlap, if for no other reason), those in the rest of the panel seem likely to meet our needs. Near the foot of the panel the approximations coincide, and it seems likely that the first approximation is quite good there.

Instead of insisting on the numbers in the two samples, we can ask for a simple sample from a mixture of two Gaussian distributions. The results are compared in exhibit 9

exhibit 9 about here

The divisors for 10 + 10 are surprisingly similar to those for the mixtures, especially for mixtures with spacings a little larger than 3.5. It does not seem likely to matter which approach we keep in mind, when we are dealing with large shifts.

24. Minimum misclassification for 50-50 Gaussian mixtures.

If we use the 50-50 mixture alternative (separated by 2δ) knowing that a value falls at y leaves us with a chance

$$\frac{e^{-(y-\delta)^2/2}}{e^{-(y-\delta)^2/2} + e^{-(y+\delta)^2/2}} = \frac{e^{y\delta}}{e^{y\delta} + e^{-y\delta}} = \frac{1}{1 + e^{-2y\delta}}$$

that that observation came from the second distribution. Our classification, if we decide to split in two, can at most depend on the y values. Thus, conditional on a value at y, the minimum average misclassification of that one observation must be at least the lesser of this chance and its complement

$$q(y) = \min\left[\frac{1}{1+e^{2y\delta}}, \frac{1}{1+e^{-2y\delta}}\right] = \frac{1}{1+e^{2|y|\delta}}$$

exhibit 9

The working values and their first differences for a 50-50 mixture of unit Gaussians separated by $t\sigma$ for t = 2.5, 3, 3.5 and large.
Also compared with the fixed 10 + 10 case (two widely separated samples of 10 each).

	working values for 50 - 50 mixtures, separated by				denominators = differences of adjacent working values				
index	2.5	3	3.5	(∞)	2.5	3	3.5	∞	(10 + 10)
	-1.5098	-1.5096	-1.5096	-1.5094					
1.5					.5304	.5311	.5311	.5311	.511
	-.9794	-.9786	-.9784	-.9783					
2.5					.3405	.3419	.3419	.3419	.340
	-.6389	-.6367	-.6363	-.6364					
3.5					.2766	.2787	.2796	.2995	.277
	-.3623	-.3578	-.3569	-.3569					
4.5					.2481	.2523	.2533	.2539	.252
	-.1142	-.1053	-.1033	-.1030					
5.5					.2365	.2443	.2468	.2474	.242
	.1223	.1390	+.1433	.1444					
6.5					.2352	.2496	.2550	.5552	.252
	.3575	.3886	+.3982	.4008					
7.5					.2416	.2688	.2815	.2866	.277
	.5991	.6574	+.6798	.6874					
8.5					.2639	.3028	.3417	.3598	.340
	.8517	.9610	1.0190	1.0472					
9.5					.2688	.3346	.4498	.6055	.511
	1.1156	1.3115	1.4723	1.6527					
(10.5)					_	.3924	.5574	L	L
	1.3844	(1.6885)	(2.0277)	L					
(11.5)					_	(.3346)	(.4998)	(.6055)	(.511)
					_	(.3026)	(.3417)	(.3598)	(.340)
						etc.	etc.	etc.	

NOTE: The similarity (even identify to four decimals) of the values in the first few lines of the working value columns reflects the trivially small chance that the lowest of the observed values comes from the component distribution with the higher mean.

The average rate of misclassfication must thus be at least the average of this over the distribution of y, which is equal to the average over a unit Gaussian of $\frac{1}{2}q(y-\delta) + \frac{1}{2}q(y+\delta)$, as separation of terms and two simple substitutions show. Since $\frac{1}{2}q(y-\delta)$ and $\frac{1}{2}q(y+\delta)$ will have the same average, by symmetry, our lower bound reduces to the average of

$$\frac{1}{1 + e^{2|y+i|\delta}}$$

over the unit Gaussian, which is easily evaluated numerically with the results shown in exhibit 10. (These bounds are, of course, far below the actual misclassification rates.)

exhibit 10

Lower bound for average misclassification rates when observations are drawn from an equal mixture of two Gaussians, separated by 2δ

2δ	lower bound
(0)	(.5000)
(1)	(.0915)
(2)	(.0290)
(2.5)	.0118
3	.0038
3.5	.0008
4	.0002

MULTIPLE COMPARISONS - - ANALOGY AND DISTINCTION

25. The "stages of knowledge".

Multiple comparisons in the narrow sense, involves measurements of several quantities, under circumstances where the relevant variability can be

separately estimated, and a desire to "learn as much as one can" about the differences among the values of the quantities. (In valley-seeking, of course, we do not have a separate estimate of variability.) Just what we are straining to learn even a little better depends on how far down the road of knowledge we have previously gone, especially in so far as the planning of our data collection goes. Are we, for example, just beginning to learn? Have we learned a lot, but have a long way to go? Have we almost settled the sign of all the differences, but know less than we would like about their size? Have we measured each difference to within a percent of itself?

In areas where measurement is only moderately painful, we can expect to begin our early studies near the start of this list, and progress till our later studies come toward its end. In areas where measurement is very painful - - effort-wise, financially, or ethically, (as in most clinical trials) - - we are likely to begin and remain close to the start of this list. (In areas where measurement is easy we may even, by contrast, begin near the end of this list.)

At different stages of knowledge gathering we have different strivings, so it is not surprising that it is appropriate to use different kinds of multiple comparison procedures. Two distinctions are of overriding importance: Is it appropriate to seek quantitative knowledge or qualitative knowledge? Will we get focused knowledge or unfocused knowledge?

26. Qualitative or quantitative knowledge?

In general we expect to emphasize qualitative knowledge at the early stages - - as all we can reasonably hope for - - only emphasizing quantitative

knowledge after having qualitative knowledge more or less in hand. The most natural exception to this arises with what we might call "TBA knowledge" - - where TBA stands for "to be adjusted". A simple agricultural prototype would involve the measurement of yield improvement from the application of various types and amounts of fertilizers. The biological facts may be expected to continue in the future with little change, but the ratio: (value of crop)/(cost of a specific fertilizer) are likely to change. Our "qualitative knowledge" of what is best to do can change accordingly. We may do quite well if we can pick out what to do, but the qualitative knowledge we need is *qual*itative *ad*justed knowledge, involving a price-ratio not known today. Such future knowledge has to be based on the best *quant*itative knowledge that we have today. TBA knowledge aside, though, we expect to begin with qualitative knowledge.

27. Focused or unfocused knowledge?

Unfocused knowledge - - in the multiple comparison framework - - is at its very best *barely useful.* To know that some of the things we are looking at are different, to an unknown amount, and in unknown directions, is of very little use. (We know, overall, that almost everything is different to some degree - - in some decimal place. This overall knowledge has a wide base of inference and holds at a very extreme significance level. Thus it is much better knowledge than what we are learning from the study before us. So long as the result of the current study is unfocused, it does very little, if anything to help us.)

In multiple comparison situations, by contrast, focused knowledge, knowledge that refers to specific comparisons, teaches us something new. (Usually the more it is focused, the more it is useful.)

28. Our problem.

In the problem to which this report is ultimately directed, matters are somewhat different. At an early stage, we are asking "would it be wise to treat this batch of numbers as two subbatches?". If this question deserves careful attention, we know so little that, even if we reach a positive conclusion, it is hopeless to expect certainty - - even 90% or 60% certainty - - that we have selected exactly the two subbatches that infinite wisdom would find best. We expect only to do as well as we can. We will miss, by at least a little, much of the time. We are seeking guidance - - since we understand that knowledge - - even 95% knowledge - - is not available. We have, in such circumstances, to be prepared to accept imperfect guidance, if it is the only kind of guidance that is available.

29. The stages of qualitative multiple comparison.

After we throw out unfocused forms of "multiple comparison" - - including those based on the F-test - - there remains a sequence of stages from early *qual*itative knowledge to late *qual*itative knowledge, with differing multiple-comparisons procedures at different stages. There are also stages of *quant*itative knowledge, from early to late (in specific instances, some of these may overlap some of the qualitative stages) but these *quant*itative stages do not

seem to be associated with more than a single set of multiple-comparisons procedures. It will help to clarify our ideas about the separations problem to discuss the *qual*itative stages briefly.

At the first stage, we are seeking any focused positive result that we can reach. The natural response is to use the studentized range and, if this provides significance, call out the apparently most positive value as more positive than the apparently least positive one. (We return below to the logic involved.)

At the second stage, we ask first for demonstrable separations into two subsets, followed - - to the extent available at no extra cost - - by separations of all but one, all but two, ... Here the procedures of Welsch(1977) can serve us well.

The third qualitative stage is one where we strive to be (simultaneously) confident about the sign of as many differences as possible. Some of the more modern methods, such as those of Ramsey (1981) may well be appropriate here.

Finally, we come to situations where we are trying to be qualitatively confident about the sign of every difference. (The background belief is, of course, that every pair of quantities do differ, in some decimal place.) At this stage, the method recently proposed by Braun and Tukey (1983c) has demonstrable advantages over the more classical methods.

30. Quantitative multiple comparisons.

If we need *quant*itative confidence statements, either intrinsically or because we need *TBA qual*itative results, the natural simultaneous tool is the studentized range. The essence of the matter, if

$$y_B = \mu_B + \varepsilon_B$$
$$y_E = \mu_E + \varepsilon_E$$

where the μ's are long-run values and the ε's are errors, is summarized in

$$|(y_B - y_E) - (\mu_B - \mu_E)| = |\varepsilon_B - \varepsilon_E|$$

and

$$\max_{B,E} |(y_B - y_E) - (\mu_B - \mu_E)| = \text{range}\{\varepsilon\text{'s}\}$$

So that, if we combine an independent estimate of variability and the studentized range distribution to set a bound, say a 95% bound, on the range of the $\{\varepsilon$'s$\}$, we can use the same bound simultaneously for all differences

$$(y_B - y_E) - (\mu_B - \mu_E)$$

This result can of course be written, if V is the bound:

$(y_B - y_E) - V \le \mu_B - \mu_E \le (y_B - y_E) + V$ (all B and E, 95% simultaneous)

Since we are really working with the ε's, nothing about the whole process depends on the values of the μ's, so in particular, we always spend all our error rate, conditionally, in every individual situation.

(An extension, using logical implication, to bounds for all contrasts is easy.)

* relation to the qualitative case *

The first stage of qualitative comparison uses the studentized range. If the studentized range is significant at the chosen significance level, say because the lower limitation is positive, we will have,

$$0 < (y_{max} - y_{min}) - V \leq \mu_{max} - \mu_{min} \qquad \text{(max, min for } y\text{'s)}$$

where the comment in ( ) means that "max" is the subscript i that maximizes y_i and "min" is the subscript j that minimizes y_j. It follows that $\mu_{max} = \mu_i$ exceeds $\mu_{min} = \mu_j$. Thus the quantitative procedure provides logical support for the weakest stage of qualitative inference.

31. The leapall in qualitative multiple comparisons.

We have already seen (Section 16) that, specifically in the Gaussian case, the leapall provides more degrees of freedom than the range - - that is, that its standard deviation is a smaller fraction of its mean. This result is encouraging, but does not clarify the question of comparative power against any specific family of alternatives. As we leave the null distribution, all gaps will change their average values somewhat for most - - if not all - - such families and the relative effect on range and leapall of these changes is not easy to assess theoretically.

There is also a logical question. Does a leapall beyond the upper 5% point of the null distribution mean that we are 95% sure that $\mu_{max} > \mu_{min}$ where max and min refer to the *observed* y's? It would surprise us if this were not true, but there appears to be no highly obvious proof that it is true.

It seems to be a problem appropriate for either crude simulation or Monte Carlo to adequately compare the studentized leapall with the studentized range.

* and in our problem *

Thus the analogy with multiple comparisons suggests that, in our problem, we compare the leapall with a between-values assessment of spread that is relatively little affected by the presence of an actual valley - - or a few valleys.

HEURISTIC APPROACH TO VALLEY-SEEKING

32. A scale estimate.

The following discussion assumes $n=20$, but analogies for other n are simple and direct.

We naturally seek a scale estimate that is not too much affected by separation. In terms of leaps, we can expect to want to shun the longest ones. This means shunning high-index e-gaps, or shunning high-index e-leaps. So our problem is combining a list of low-index e-leaps.

In the null situation, these will be approximately i.i.d. exponential, and thus most naturally combined as a sum or a mean.

We have already seen that a low-index sum of e-leaps is a (rank) Winsorized sum of ordinary leaps. For the null situation, such a Winsorization seems a natural way to shun the high-rank ordinary leaps. In the presence of real separation, however, when the largest leaps are substantially inflated, the heuristic excuse for Winsorizing has essentially disappeared. Accordingly, we might want to consider equal weights on the low-rank ordinary leaps as a second possibility.

This would correspond to, for instance

$$h_1^* + h_2^* + ... + h_{15}^* \qquad (n=20)$$

which converts into

$$\frac{15}{19}e_1 + \frac{14}{18}e_2 + ... + \frac{1}{5}e_{15} \qquad (n=20)$$

We put this on the back burner, and press on with such combinations as

$$e_1 + e_2 + ... + e_{15} \qquad (n=20)$$

which can also be written

$$h_1^* + h_2^* + ... + h_{14}^* + 5h_{15}^* \quad . \qquad (n=20)$$

33. A response and its analogs.

If we believe the multiple-comparisons guidance developed above, we naturally take the leapall

$$e_1 + e_2 + ... + e_{15} + e_{16} + ... + e_{19} \qquad (n=20)$$

as our first response.

Let us, as a matter of convenience write

$$A = e_1 + e_2 + ... + e_{15}$$

for the low-index sum of e-leaps and

$$B = e_{16} + e_{17} + e_{18} + e_{19}$$

for the high-index sum, so that leapall = $A + B$

We now have:

scale estimate basis = A

response basis = $A + B$

The naive ratio reduces as follows:

$$\frac{A + B}{A} = 1 + \frac{B}{A} = 1 + \frac{4}{15} \cdot \frac{(B/4)}{(A/15)}$$

so that we may as well use the ratio

$$\frac{b}{a}$$

of the high-index *mean* (of e-leaps)

$$b = \frac{B}{4}$$

to the low-index *mean* (of e-leaps)

$$a = \frac{A}{15}$$

in place of the ratio of the all-index *sum* (the leapall) to a low-index sum or mean.

A virtue of writing things this way is that a and b are approximately null distributed like $\sigma^2\chi^2_\nu/\nu$ with $\nu = 15\times 2 = 30$ and $4\times 2 = 8$ respectively, approximately independently of one another. Thus Snedecor's F offers us approximate % points for b/a.

* alternate scale *

So much for the simplest scale estimate, for the instant. What if we are to consider the alternate situation? We would only do this because we take the non-null situation quite seriously. And when we go back to the multiple-comparisons guidance, we see that the thinking was basically null-situation thinking. Thus if A^* is the tapered sum for scale

$$A^* = \frac{15}{19}e_1 + \frac{14}{18}e_2 + \ldots + \frac{1}{5}e_{15}$$

March 29, 1989

and B^* is its complement

$$B^* = \frac{4}{19}e_1 + \frac{4}{18}e_2 + ... + \frac{4}{5}e_{15} + e_{16} + e_{17} + e_{18} + e_{19}$$
$$= h^*_{16} + h^*_{17} + h^*_{18} + h^*_{19}$$

the argument for using exactly $A + B = A^* + B^*$ seems at best weak.

The plausibility of mixing - - of using, for example, A^* and B, which will, by contrast, be approximately independent - - seems much more reasonable. Accordingly, once we have studied the use of A and B - - and what that use suggests - - we *may* want to look at the use of A^* and B.

* other approximately independent choices *

Let now

$$a_g = \frac{e_1 + e_2 + ... + e_g}{g}$$
$$b_h = \frac{e_{20-h} + e_{21-h} + ... + e_{19}}{h}$$

be more general simple low-index and high-index means of e-leaps. So long as $g < 20 - h$, so there is no overlap (of leaps), these will be approximately independent, and we can turn to an F-table for approximate % points.

Once we have looked, approximately, at $g + h = 19$, we may as well explore $g + h < 19$.

* first, the null situation *

If we focus first on $n=20$, $g=15$, $h=4$ we begin with $F_{8,30}$, whose 5% point is 2.27. Asking whether

$$\frac{b_4}{a_{15}} \geq 2.27$$

March 29, 1989

is equivalent to asking whether

$$c = b_4 - 2.27a_{15} \geq 0$$

In the null situation the averages of b_4 and a_{15} are both approximately 1.00 and their variances are approximately 1/4 and 1/15 respectively. Hence the critical ratio is

$$\frac{ave(c)}{(var(c))^{1/2}} = \frac{1-2.27}{(.25+(2.27)^2.06667)^{1/2}} = \frac{-1.27}{.7704} = -1.648$$

quite consistent with ≥ 0 having 5% probability.

* next, power for shift = 5 *

Suppose now that the situation is two subsamples of 10, one shifted (relative to the other) by 5 (i.e. 5σ). The leapall will involve gaps increased on average, in total, by somewhat less than 5 - - because the end observations will be extremes of 10, rather than extremes of 20. If we choose

$$\bar{a}(20|20) - \bar{a}(10|10) = 1.867-1.539 = .328$$

as the adjustment for each end, the total range will be increased by something like

$$5 - 2(.328) = 3.344$$

Most of the increase will take place for central gaps, where, for $n=20$:

$$\text{leap} \approx \frac{\text{gap}}{.125} = 8 \cdot \text{gap}$$

so that the leapall will be increased on, average, by something like

$$3.344 \times 8 = 26.75$$

Since

$$\text{e-leapall} = \text{leapall} = A + B$$

we now need to judge, at least roughly, how this increase, on average, divides between A and B. (This will not prove a simple task.)

We then ought to look at some other separations, like 7+13, 4+16, 2+18, 1+19.

It is much easier to understand how the leaps indexed by gap number respond to such a shift than to understand how the ordered leaps respond. And it is also somewhat easier to understand how the individual ordered leaps respond than to understand how their gaps and hence the e-leaps behave. The easiest of these three questions has been answered, in terms of averages, in exhibit 7 (section 23) for a shift of 5 between 10 and 10 we will have average leaps of 1.17, 1.24, . . . 2.60, 4.29, 15.55, 4.29, 2.60, . . ., 1.17. If any one observed leap is going to be really large, it will be the central (as indexed) leap-with average 15.55.

What about the second largest leap. Suppose its size is $\geq$ 9. If it comes from a specified leap with average 4.25, this is an event of probability $\exp(-9/4.25) = 12.0\%$. If it comes from a specified one of the leaps with average 2.60, the probability is $\exp(-9/260) = 3.1\%$. For average 2.10 (the next in size) $\exp(-9/2.10) = 1.4\%$. For average 1.782, 0.6%. If we take these as independent, the total probability of $\geq$ 9 from one of the "other 18" is somewhat larger than

$$1 - (.880)^2(.969)^2(.986)^2(.994)^2 = .302$$

of this 30% about 24% is contributed by the two leaps (with average 4.25) adjacent to our valley.

34. Which leap is largest - - and how often.

Let us next turn to the question: Given the situation - - and the averages - - just considered, how large would the largest of the "other 18" be if all leaps were exponentially and independently distributed? The chance that this largest leap exceeds L is 1 MINUS the product of the chances that each exceeds L, and equals

$$1 - (1-e^{-L/4.246})^2 (1-e^{-L/2.595})^2 \ldots (1-e^{-L/1.168})^2$$

which takes the values shown in exhibit 11.

exhibit 11 about here

If we plot the last column against the second one, using L to parametrize a curve, we get exhibit 11A where the areas above and below the curve are the probabilities that the largest leap will come (a) from one of the other 18 or (b) from the one large leap. We see that the "other 18" will provide the largest leap more than a third of the time.

exhibit 11A about here

Turning back to exhibit 11, we see that the "largest of the 18" is more tightly distributed than a single exponential but more loosely than the maximum of as many as 18. Column (+), which corresponds to the largest of 7 exponentials of size 3, gives a reasonable approximation. There is no reasonable hope for the largest leap to come from the valley, but what about "one of the largest".

Exhibit 11

Distribution of the largest of the "other 18" leaps
for a true valley between 10 and 10 for $t = 5$
(calculated as if leaps were exactly independent and exactly exponential)

L	$Pr(\geq L)$	(*)	(**)	(***)	(+)	$Prob(exp(ave = 15.55) \geq L)$
1	99.95721%	.952	(99.91%)	99.9958%		93.77%
2			(98.44%)	99.6858		87.93
3	97.7139	1.893	(93.68%)	97.451%	95.97%	82.45
4	92.2231	1.977	(85.34%)	91.352%		77.32
5	80.469	1.045	(74.59%)	81.26%		72.50%
6	66.15%	2.112	62.95%	68.75%	63.88%	67.99%
7	52.50%	2.182	51.674%	55.78		63.75
8	40.6832%	2.257	41.51%	43.79		59.78
9	31.34%	2.325	32.80%	33.54		56.06
10	24.02%	2.381	25.60%	25.22		52.57
11	18.43%	2.452	19.80%	18.71%	18.34	49.29
12	14.18%	2.519	15.20%	13.75		46.22
13	10.92%	2.574	(11.61%)	10.03		43.34
14	8.85%	2.655	(8.83%)	7.28		40.64
15	6.56%	2.687	6.70%	5.26		38.11
20	1.898%	2.922	1.64%	1.01%	0.89%	27.63

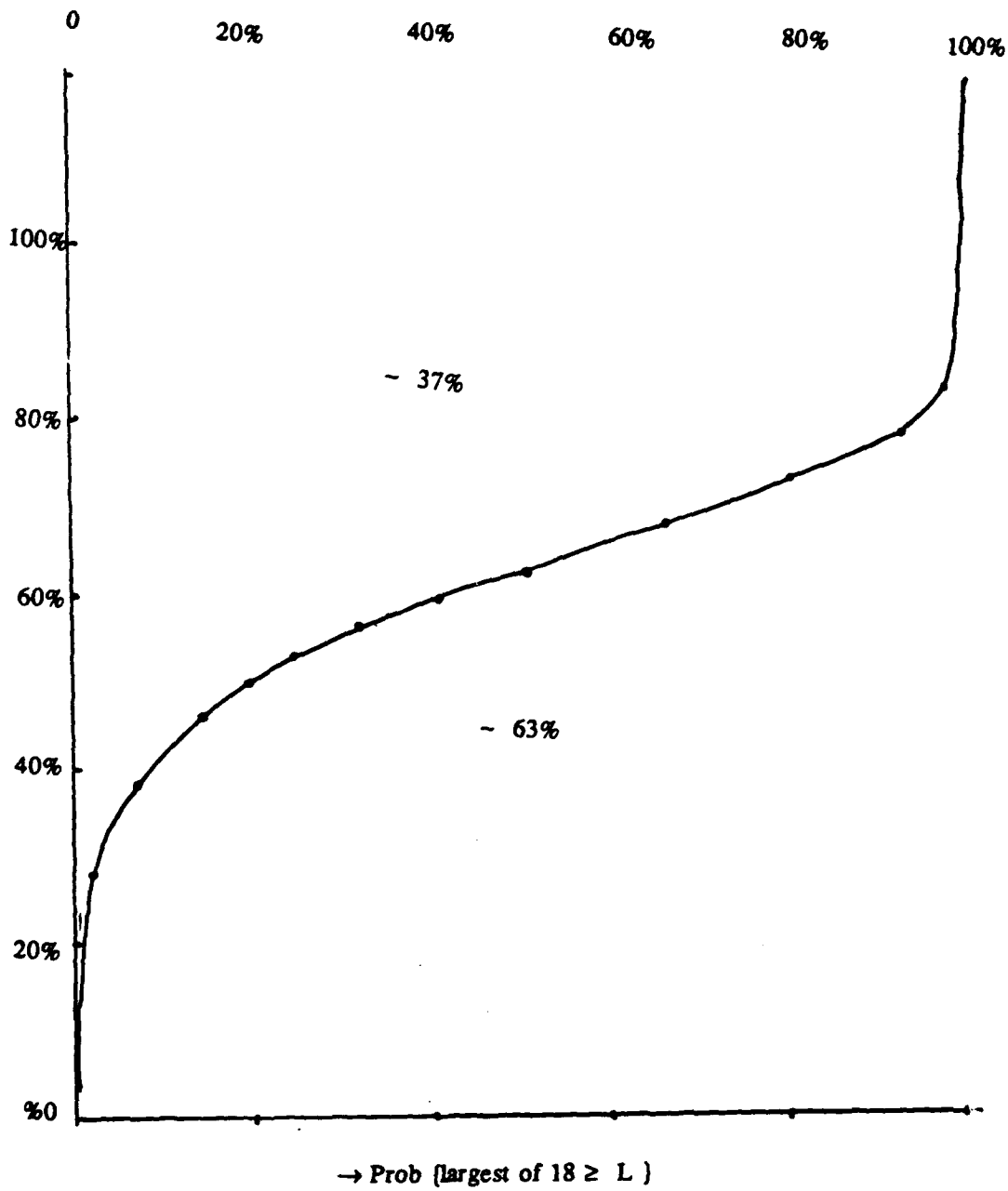
(*) Size of 18 equal exponentials such that the largest exceeds L with probability $Pr(\geq L)$

(**) Probability that the largest of 5 independent exponentials with average 3.5 will exceed L

(***) Probability that the largest of 10 independent exponentials with average 3 will exceed L

(+) Probability that the largest of 7 independent exponentials with average 3 will exceed L.

exhibit 11A



* another approach *

Let us begin with three independent exponentials of averages C/f, C/g, C/h. How often will the first be the largest? The density of a convenient multiple of the first is fe^{-fu} and the probability of the same multiple of the other two being less is $(1-e^{-gu})(1-e^{-hu})$. Thus the answer is

$$f\int_0^\infty e^{-fu}(1-e^{-gu})(1-e^{-hu})du = 1 - \frac{f}{f+g} - \frac{f}{f+h} + \frac{f}{f+g+h}$$

with the answer for more competitors an analogous sum, first over pairs, then over triples, etc.

If we begin with three leaps, the valley leap and one on either side, and take $f = 1$, we have g, $h = 15.55/4.246 = 3.662$ so that the chance that the valley gap is the largest of the three is

$$1 - \frac{1}{1+3.662} - \frac{1}{1+3.662} + \frac{1}{1+3.662+3.662} = 1 - .2145 - .2145 + .1201 = 69.1\%$$

By symmetry the chance that a specified adjacent leap is the largest of the three is $(10\% - 69.1\%)/2 = 15.45\%$.

If we add two more leaps, one on each side, for which $j = k = 15.55/2.595 = 5.992$, the terms we need to consider are (sorted by number of terms in the denominator)

$$\frac{1}{1+3.662} + \frac{1}{1+3.662} + \frac{1}{1+5.922} + \frac{1}{1+5.922} = 2(.2145) + 2(.1445) = .7179$$

$$\frac{1}{1+3.662+3.662} + \frac{4}{1+3.662+5.922} + \frac{1}{1+5.922+5.922} = .1201 + .3779 + .0779 = .5759$$

$$\frac{2}{1+3.662+3.662+5.992} + \frac{2}{1+3.662+5.992+5.992} = .1397 + .1201 = .2599$$

$$\frac{1}{1+3.662+3.662+5.992+5.992} = .0492$$

March 29, 1989

leading to

$$1 - .7179 + .5759 - .2599 - .0492 = 64.7\%$$

for the chance that the valley leap will be largest of the *five*.

A similar calculation give 6.34% chance that the largest of the five will be one of the leaps not even next to the valley leap and 100 – 64.7 = 28.96% chance that one of the adjacent leaps will be largest.

Continuing in this way, we will get a result similar to

	at the valley leap	63.6%
next to	" "	28.5%
one away from	" "	6.2%
further away from	" "	1.7%

indicating that one of the central five leaps are relatively sure to provide the largest, while the central three will fail to provide it perhaps once in 12 times.

35. And the shortest leap?

The complementary calculation, with e^{-fu} exchanged with $1-e^{-fu}$, gives us the distribution of the smallest of 18 independent exponentials with the averages already noted. The resulting cumulative probabilities can be converted to a common average for 18 exponentials that matches a particular % point for smallest exponentials from 0.01 to 10. This gives the same answer, 1.669, to 3 decimals. As we might have expected, 1.669 is the harmonic mean of the 18 average exponentials with which we started.

Since the equivalent averages for the largest were above 2 - - and thus greater than 1.669 - - it is reasonable to anticipate that the equivalent average for a low-value group of leaps will also be somewhat greater than 1.669. Thus we pay an appreciable penalty for the larger leap averages among the 18.

Using fewer of the low-value leaps should reduce the equivalent average, but only toward - - not beyond - - 1.669. Thus the optimum number of low-value leaps, which corresponds to an equal number of low index e-leaps, is likely to be an intermediate between too many and too few, and the optimum is likely to be flat.

36. High index e-leaps?

We have some insight into the largest leap, whether from the valley leap or from the 18 others. what we need to understand is the behavior of the high-index e-leaps, which are multiples of leap-to-leap differences (e-gaps). A fine large leap, if we are only looking at the highest index e-leap, can have its effect spoiled by whenever the next largest leap is close to it. Something similar will always happen at the inclusion-exclusion boundary, though not as vigorously if more high-index leaps are considered. Taking too few high-index e-leaps can be costly.

But so can taking too many. In our prototype situation, only 3 indexings (at - - and adjacent to - - the valley) are likely to produce quite large leaps. Two or three of the largest leaps - - one of which may be "spoiled" by a nearby value for the fourth largest - - should be enough to trap the "real" effect - - at least when the effect is large enough to be worth trapping. (Five would almost

surely be enough.)

To understand better what we should do would require some careful simulation. Doing this for independent exponentials would be as much work as doing it for the real problem. So why do less than simulate the real problem?

37. Suggestions for empirical trial.

High-index group sizes from 1 to 5 and low-index groups sizes of 10 to 20 ought to cover the optimum (at least for 10 + 10). We cannot, of course - - in the face of misbehavior we have seen for a shift of 5, a shift large enough so that only about 0.6% of one distribution overlaps 0.6% of the other - - look forward to extremely high power, for any choice.

The next step should be simulation.

38. Restricted sums.

The discussion of the last few sections was founded on e-leaps, particularly on those corresponding to relatively very large and relatively moderately small ordinary leaps. This enabled us to use approximations based on e-leap behavior, but it involved throwing away which leap was which - - something which makes good sense on the null hypothesis - - but which must be somewhat wasteful on the simple alternative. If the 4 largest leaps in a batch of $n = 20$ values are h_4, h_7, h_{12} and h_{17}, corresponding to gaps $y_4 - y_3$, $y_7 - y_6$, $y_{12} - y_{11}$ and $y_{17} - y_{16}$ they hardly combine to point toward a single valley in the underlying situation.

If power against single valleys, or possibly pairs of valleys is what we most want, we will need to restrict the sets of, say, 4 leaps, whose sum or mean is to serve as our indicator.

Simulation experiments will probably be essential in choosing a satisfactory restriction, but it may be helpful to suggest some possibilities for n around 20. We might restrict ourselves to:

a) four leaps adjacent (in y) - - 16 candidates

b) four leaps omitting at most one of five adjacent - - $16 + 15 \cdot 3 = 61$ candidates (the old 16 and 45 new ones)

c) four leaps omitting at most two of six adjacent - - $16 + 15 \cdot 3 + 14 \cdot 6 = 145$ candidates

d) four leaps omitting at most three of seven adjacent $16 + 15 \cdot 3 + 14 \cdot 6 + 13 \cdot 20 = 405$ candidates

where these numbers of candidates are to be compared with $\binom{19}{4} = 19 \cdot 17 \cdot 15 \cdot 12 = 5814$ candidates for four unrestricted choices.

These numbers suggest hope for much improved power, and urge trials by simulation.

APPENDIX A on $a_F(i|n)$'s

39. What size biases don't matter?

In using the orderly data kit, we are almost always dealing with exponentially distributed quantities. We *need to remember* that such quantities have

$$\text{standard deviation} = \text{mean}$$

so that we cannot reach

$$\text{standard deviation} \leq \frac{1}{10}\ \text{mean}$$

until we combine at least 100 such exponentially distributed quantities. This ought to tell us that ± 10% is not very much, especially when it happens to only 1 or 2 of a set of roughly exponentially distributed quantities.

We saw, for $n=20$, that using $\acute{a}$'s instead of $\bar{a}$'s biases the end leaps by only a few %, and those inside even less. What if the biasing is 5% for each end leap, 3% for each next leap, 1.5% for the next, etc. The total biasing is $2(6\% + 3\% + 1.5\% \ldots) = 24\%$. The bias of the mean leap is $(24/n)\%$. The standard deviation of the mean leap is $s\sqrt{n}$, so that, for the mean leap,

$$\text{bias/standard deviation} = (24/s\sqrt{n})\%$$

namely 8% for $n=10$, 5% for $n=20$, 2.4% for $n=100$ all of which are quite negligible (the MSEs are increased by factors of 1.0058, 1.0029 and 1.0006, respectively). In fact, as we shall see shortly, the actual bias is even less.

Thus we have no real need to do better in converting gaps to leaps, than to use working values. Any more detailed attention we pay to $a_F(i|n)$ has to be

either (a) from a sense of neatness, or (b) with some other application in mind.

40. Mismatch of distributions.

Similar arguments are important whenever we think harder about converting gaps to leaps. This conversion requires using denominators which come, almost always, from some assumed distribution shape (perhaps even the Gaussian). If there is a true distribution shape, which there might be, it is almost certain to differ from the shape we assume for conversion. How worried should we be about this?

The nearly-exponentially distributed leaps will now have different typical sizes, in particular decreasing the d.f. associated with the leapall. How much? If in $2m$ denominators, m are 10% short, while the other m are 10% long, we will have $\sum D_i = 2m$ and $\sum D_i^2 = 2m(1.01)$ corresponding according to a 1% loss of d.f. Equal amounts of +20% and −20% would lead to a 4% loss. (Equal amounts of +30% and −30% would lead to only 9% loss.) Roughly uniform spread between such limits would only have 1/3 the effect of having everything at the ends. Mismatch of distribution shape can have as large consequences as one likes, but only if the distributions are quite different!

The mismatch, for $n=20$, between a Gaussian distribution and the distribution of t on 3 d.f. - - really a rather large mismatch - - only involves ± 30% (for differences of working values).

41. $a_{Gau}(i|\,n)$, for a Gaussian parent.

* $i=1$ (or, by symmetry, $i=n$) *

Our earlier illustration showed that the differences among various approximations and exact forms of $a_{Gau}(i|n)$ are concentrated near $i=1$ and, symmetrically, near $i=n$. Exhibit 12 shows some relevant numbers for $i=1$.

exhibit 12 about here

The C^* column does moderately well in describing the "y = mean for Gaussian order statistic" column. The C^{**} column does very well in describing the "y = median for Gaussian order statistics" column. (Thus if we ever need more accurate medians for end order statistics, the formula for C^{**} will provide them to high accuracy.) The working-value column lies between "medians" and "means for Gaussian" columns, closer to the median for large n.

The final column shows how small the difference between mean and median is (for $n \geq 10$) in comparison with the difference between "Gaussian parent" and "logistic parent" (when both are being analyzed as if Gaussian). (In practice, we are unlikely to be sure whether we "should" use Gaussian or logistic. So how can we, in practice, even distinguish "means for the Gaussian case" from "medians for some other case"?? Thus this is a further reason, so far as order statistics themselves go (though not as far as gaps or leaps go), not to distinguish between mean and median in practice - - though in theoretical work, it may be important to be clear which is which.)

* $i=2$ (or, by symmetry, $i=n-1$) *

exhibit 12

Values of (3n+ 1) Gau (y) for specified situations and approximations for $i=1$ or $i=n$ (end order statistic)

n	when y = mean for Gaussian order statistic	C^*	Working value	C^{**}	when y = median for Gaussian (***)	when y = median for logistic order statistic (a)	(b)	(c)	(****)
1	2.0000	2.000	2	2.0503	2.0000	2.0000	2.000	2.0000	-
2	1.9693	1.974	2	2.0503	2.0503	2.1946	2.0323	2.0780	-2.93
5	1.9587	1.943	2	2.0712	2.0712	2.3470	1.8583	1.9896	1.38
10	1.9190	1.915	2	2.0760	2.0760	2.2693	1.5304	1.7182	.44
20	1.8855	1.886	2	2.0780	2.0779	1.9875	1.0994	1.3090	.25
50	1.8507	1.850	2	2.0789	2.0789	1.3982	.5605	.7344	.17
100	1.8233	1.826	2	2.0792	2.0792	.9271	.2779	.3971	.15
200	1.8126	1.809	2	2.0793	2.0793	.5394	.1160	.1830	.14
(400)	(1.8131)	(1.796)	(2)	2.0794	(2.0794)	.2743	.0406	.0716	.13
500	1.794	1.793	2	2.0794	2.0794	.2144	.0279	.0511	.14
1000	1.7850	1.785	2	2.0794	2.0794	.0910	.0077	.0106	.14

$$C^* = 2 - \frac{.2285(n-1)^{0.7}}{(n-1)^{0.7} + 7.9}$$

$$C^{**} = 2 + \frac{.0794(n-1)^{1.17}}{(n-1)^{1.17} + .58}$$

***This column also applies for "$(3n+1)\,F(y)$ where y is the median of the end order statistic in a sample from the same continuous $F(\cdot)$".

(a) When logistic matched to unit Gaussian in variance

(b) When logistic matched to unit Gaussian in central density

(c) When logistic scaled an intermediate way

****(Gaussian median MINUS Gaussian mean)/(Gaussian median MINUS logistic median, version (c))

NOTE: 2 is 27% of the way from 2.0794 to 1.785

Exhibit 13 shows the same sort of information for $i=2$ as exhibit 12 showed for $i=1$.

exhibit 13 about here

Most of the comments about exhibit 12 apply with only minor changes. The working value is now much nearer the median than it is to the mean.

* $i=3$ (or, by symmetry, $i=n-2$) *

Exhibit 14 shows the same sort of information for $i=3$ as exhibits 12 and 13 showed for $i=1$, respectively.

exhibit 14 about here

* larger i (deeper i) *

We do not propose to follow details further here.

* the leap denominator *

For the main purposes of this account, we are interested in $a(i+1|\,n) - a(i|\,n)$ whose behavior for small i is shown in exhibit 15.

exhibit 15, 16 about here

We see that the difference between "difference of means" and "difference of medians" is only a few %. Our largest question, of course refers to the bias of using working-values in place of $\bar{a}(i|\,n)$. (The bias using working values for medians is, as we have seen, smaller and of opposite signs.) Exhibit 16 shows numbers for $i=1$ and 2. The bias is usually of the order of 2%, which is quite trivial for most situations, including those that most concern us. (Remember

exhibit 13

Values of (3n+ 1) Gau (y) for specified situations and approximations for $i=2$ or $i=n-1$ (next-to-end order statistic)

n	when y = mean for Gaussian order statistic	C^*	Working value	C^{**}	when y = median for Gaussian (***)	when y = median for logistic order statistic (a)	(b)	(c)	(****)
3	5.000	5.000	5	5.0000	5.0000	3.9866	3.82528	3.8911	–
4	4.982	4.970	5	5.0145	5.0145	4.3304	4.0553	4.1329	.04
5	4.965	4.953	5	5.0211	5.0210	4.5821	4.1714	4.2865	.08
10	4.908	4.908	5	5.0300	5.0300	5.1839	4.2244	4.4846	.22
20	4.857	4.866	5	5.0329	5.0331	5.3330	3.7509	4.1591	.20
50	4.803	4.817	5	5.0313	5.0314	4.6674	2.5980	3.0457	.12
100	4.771	4.782	5	5.0347	5.0347	3.6642	1.5770	2.0231	.09
200	4.746	4.748	5	5.0348	5.0349	2.5306	.8244	1.1489	.07
(400)	4.725	4.715	(5)	5.0349	(5.0350)	1.5314	.3623	.5554	.07
500	–		5	5.0349	5.0350	1.2660	.2679	.4245	–
1000	–		5	5.0350	5.0350	0.6408	.0932	.1651	–

$$C^* = 5 - .08(\ln n/3)^{0.8}$$

$$C^{**} = 5 + \frac{.0350(n-3)^{1.1}}{(n-3)^{1.1} + 1.41}.$$

***This column also applies for "$(3n+1)F(y)$ where y is the median of the next-to-end order statistic in a sample from any continuous $F(\cdot)$".

****(Gaussian median MINUS Gaussian mean)/(Gaussian median MINUS logistic median, version (c))

(a) When logistic matched to unit Gaussian in variance

(b) When logistic matched to unit Gaussian in central density

(c) When logistic scaled an intermediate way

NOTE: 5 is 11% of the way from 5.0349 to 4.715

exhibit 14

Values of $(3n+1)\,\mathrm{Gau}(y)$ for specified situations and approximations for $i=3$ (next-to-end order statistic)

n	when y = mean for Gaussian order statistic	C^*	Working value	C^{**}	when y = median for Gaussian (***)	when y = median for logistic order statistic (a)	(b)	(c)	(****)
5	8.000	8.000	8	8.0000	8.0000	6.0684	5.8141	5.8862	–
10	7.933	7.905	8	8.0159	8.0158	7.4188	6.5189	6.7687	.07
20	7.882	7.856	8	8.0197	8.0199	8.2752	6.4389	6.9295	.13
50	7.806	7.805	8	8.0213	8.0215	8.0953	5.0643	5.8138	.10
100	7.768	7.772	8	8.0218	8.0218	6.9324	3.5137	4.2938	.07
200	7.739	7.742	8	8.0218	8.0220	5.2322	2.0637	2.7171	.05
(400)	7.713	7.714	(8)	8.0219	8.0220	3.4655	1.0209	1.4661	.05
500	–		8	8.0219	8.0221	2.9003	.7443	1.1613	–
1000	–		8	8.0220	8.0220	1.6369	.3076	.5050	

$$C^* = 8 - .118(\ln n/5)^{-.6}$$

$$C^{**} = 8 + \frac{.02197(n-5)^{1.1}}{(n-5)^{1.1} + 2.23}\,.$$

***This column also applies for "$(3n+1)F(y)$ where y is the median of the next-to-end order statistic in a sample from any continuous $F()$".

****(Gaussian-median MINUS Gaussian mean)/(Gaussian median MINUS logistic median, version (c))

(a) When logistic matched to unit Gaussian in variance

(b) When logistic matched to unit Gaussian in central density

(c) When logistic scaled an intermediate way

NOTE: 8 is 7% of the way from 8.0219 to 7.714

exhibit 15

Behavior of $a(i+1|n) - a(i|n)$ for i small, Gaussian parent, and either mean a's ($\bar{a}(i|n)$'s) or median a's ($\acute{a}(i|n)$'s)

	$i=1$, Gaussian			$i=2$, Gaussian			double
n	mean	median	ratio	mean	median	ratio	ratio*
2	1.128	1.089	95.5%	–	–	–	
5	.668	.614	96.5%	.495	.485	98.0%	98.5%
10	.537	.514	95.7%	.345	.3378	97.6%	98.1%
20	.460	.4360	95.0%	.277	.2691	97.1%	97.9%
50	.394	.3702	94.0%	.2262	.2185	96.6%	97.4%
100	.359	.3355	93.5%	.2018	.1941	96.2%	97.3%
200	.333	.3090	92.8%	.1837	.1761	96.2%	96.6%
400	.311	.2871	92.3%	.1696	.1621	95.6%	98.7%

*(ratio for $i=1$)/(ratio for $i=2$)

exhibit 16

Behavior of $a(i+1|n) - a(i|n)$ for i small, Gaussian parent, and either mean a's ($\bar{a}(i|n)$'s) or Gaussian working values (wovals)

	$i=1$, Gaussian			$i=2$, Gaussian			double
n	mean	woval	ratio	mean	woval	ratio	ratio*
2	1.128	1.131_2	99.7%	–	–	–	
5	.667	.6620	100.8%	.495	.4884	101.3%	99.5%
10	.537	.529	101.5%	.345	.3401	101.4%	99.9%
20	.460	$.449_5$	102.1%	.277	.2712	102.1%	100.0%
50	.394	$.382_1$	103.1%	.2262	.2203	102.7%	100.4%
100	.359	$.346_5$	103.6%	.2018	.1957	103.1%	100.5%
200	.333	$.319_0$	104.4%	.1837	.1715	103.5%	100.9%
400	.311	$.296_9$	104.7%	.1026	.1634	103.4%	101.3%

*(ratio for $i=1$)/(ratio for $i=2$)

that the standard deviation is 100%!)

* mean gaps and median gaps *

Whether we work with means or medians of order statistics has turned out to matter little. (The deviations from symmetry of order statistic distributions are not too great, and ratios of their standard deviation to their mean shrink as n increases for i/n constant.)

Gaps do not behave this way at all. The approximate distribution of all but the most unusual gaps is like $D \times$ a unit exponential, for some constant D. Since the unit exponential has mean = 1.00 and median = 0.69, the median of our gap will be roughly $0.69D$, about 70% of its mean (which is roughly $1.00D$).

In other terms,

$$\text{mean}(y_{i+1|n} - y_{i|n}) \equiv \text{mean}(y_{i+1|n}) - \text{mean}(y_{i|n})$$

while

$$\text{median}(y_{i+1|n} - y_{i|n}) \doteq 0.7(\text{median}(y_{i+1|n}) - \text{median}(y_{i|n}))$$

In fact, the mixed approximation

$$\text{mean}(y_{i+1|n} - y_{i|n}) \doteq \text{median}(y_{i+1|n}) - \text{median}(y_{i|n})$$

is very much closer than its right hand side comes to the $\text{median}(y_{i+1|n} - y_{i|n})$.

42. $a_E(i|n)$, for the lower tail of an exponential parent.

We turn now to first one tail and then the other of the (single) exponential, where the order statistic means are the partial sums of

$$\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \ldots + 1$$

This can be considered a closed form if there are not too many terms!

Exhibit 17 shows the values of "$(3n+1)$ (own tail area)" for the means of lower-tail order statistics. They are mainly of interest in showing what happens near a box-car-end-like cutoff. The usual values: 2, 5, 8, 11, . . ., for a "loose" termination are increased to values tending to 3, 6, 9, 12, ... (still spaced 3 apart). Starting with 3 is just as if 0 were an additional $n+1^{st}$ observation! (The limiting values must be those that would apply for a very large sample from a rectangular parent.)

exhibit 17 about here

43. $a_E(i|\,n)$, *in the upper tail of an exponential distribution.*

The upper tail behaves rather differently. To display more familiar sorts of numbers, exhibit 18 shows "$(3n+1)$ × (own right tail area)", which of course equals "$(3n+1)$ MINUS [$(3n+1)$ × (own left tail area)]" We now have different asymptotic values (1.684, 4.579, . . .) this time smaller than for the working-value (namely 2, 5, 8, . . .). Since the values in exhibit 18 are smaller than the nominal 2, 5, 8, ... the corresponding medians (or exponential order statistics) are larger, see column labelled **** in exhibits 12 to 14. By contrast, the corresponding order-statistic *means* are nearer the nearby (upper) tail.

exhibit 18 about here

The good simple fits, of the form $A + \frac{B}{n}$, can be very useful, since they can save us from evaluating a long, closed-form expression for the

exhibit 17

Behavior of $(3n+1) \times$ (own lower tail area) for an exponential distribution and means of near-to-lower-end order statisitcs

n	for $i=1$	for $i=2$	for $i=3$	for $i=4$	for $i=5$
2	2.754	5.438	–	–	–
3	2.835	5.654	8.401	–	–
4	2.875	5.746	8.600	11.381	–
5	2.900	5.798	8.690	11.566	14.073
10	2.950	5.900	8.849	11.798	14.745
20	2.975	5.950	8.925	11.900	14.875
50	2.990	5.980	8.970	11.960	14.950
100	2.995	5.990	8.985	11.980	14.975
200	2.9975	5.995	8.9925	11.990	14.9875
500	2.9992	5.998	8.997	11.996	14.995
1000	2.995	5.999	8.9985	11.998	14.9975
$(n=\infty)$	(3)	(6)	(9)	(12)	(15)
$(3i-1)$	(2)	(5)	(8)	(11)	(14)

Well fitted, for $n \geq 10$, by: $3i - (2.5)i/n$

exhibit 18

Behavior of $(3n+1) \times$ (own upper tail area) for an exponential distribution and means of near-to-upper-end order statistics

n	for $i=n$	for $i=n-1$	for $i=n-2$	for $i=n-3$	for $i=n-4$
2	1.5619	4.246	–	–	–
3	1.5958	4.346	7.165	–	–
4	1.6187	4.400	7.254	10.124	–
5	1.6311	4.434	7.311	10.202	13.100
10	1.6570	4.504	7.426	10.364	13.308
20	1.6705	4.541	7.467	10.449	13.416
50	1.6788	4.563	7.487	10.500	13.483
100	1.6816	4.571	7.524	10.518	13.505
200	1.6830	4.575	7.536	10.527	13.516
500	1.6838	4.577	7.543	10.532	13.523
1000	1.6841	4.578	7.546	10.534	13.525
2000	1.6842	4.578	7.548	10.534	13.526
$(n=\infty)$	(1.6844)	(4.579)	(7.549)	(10.535)	(13.527)
$(3(n+1-i)-1)$	(2)	(5)	(8)	(11)	(14)

Well fitted, for $n \geq 10$ by:

$$1.6814 - \frac{.274}{n}, \quad \text{for } i=n$$

$$4.579 - \frac{.75}{n}, \quad \text{for } i=n-1$$

$$7.549 - \frac{1.26}{n}, \quad \text{for } i=n-2$$

$$10.535 - \frac{1.70}{n}, \quad \text{for } i=n-3$$

$$13.527 - \frac{2.19}{n}, \quad \text{for } i=n-4$$

order-statistic means for those values of $n-j$ where we have found a fit.

44. $a_{rect}(i|\,n)$, the case of a rectangular parent.

Here $\bar{a}_{rect}(i|\,n) = i/(n+1)$, reflecting the exchangeability of the gaps, since $\bar{a}$ is here its own left tail area, and

$$(3n+1)\text{ (own left tail area)} = i\,\frac{3n+1}{n+1} \to 3i \qquad \text{when } n \to \infty.$$

Exhibit 19 gives a few values.

exhibit 19 about here

45. $a_{logis}(i|\,n)$, the case of a logistic parent.

A logistic parent distribution might reasonably supposed to behave somewhat like the case of a Gaussian parent, except that the deviations of means from medians (and hence the deviations of means from working values) might well be larger, quite possibly substantially so. Exhibit 20 shows the values of $(3n + 1)$(own tail area) for $i = 1(1)\ 5$ and 10 that can be computed easily from the tables of Birnbaum and Dudman (1963).

exhibit 20 about here

Since the corresponding quantities for medians will stay in the intervals [2,2.08], [5,5.035], [8,8.022], etc., as they do for all distributions, we can see that the working values, corresponding to 2, 5, 8 . . . will approximate the medians much better than they will approximate the means. (The ratios of "errors" are roughly 4, 46 and 100 at $n = 100$ for $i = 1, 2, 3$.)

exhibit 19

Values of (3n+ 1) × (own left tail area) for order statistics from the rectangular distribution

n	for $i=1$	for $i=2$	for $i=3$	for $i=4$	for $i=i$
2	2.3333	4.6667			
5	2.6667	5.3333	8.0000	10.6667	$2.667i$
10	2.8182	5.6364	8.4545	11.2727	$2.818i$
20	2.9048	5.8095	8.7143	11.6190	$2.905i$
50	2.9608	5.9216	8.8824	11.8431	$2.961i$
100	2.9802	5.9604	8.9406	11.9208	$2.980i$
200	2.9900	5.9801	8.9701	11.9602	$2.990i$
500	2.9960	5.9920	8.9880	11.9840	$2.996i$
1000	2.9980	5.9960	8.9940	11.9920	$2.998i$
(∞)	(3)	(6)	(9)	(12)	($3i$)

exhibit 20

Values of $(3n+1)$(own tail area) for *means* of logistic order statistics

n	$i=1$	(C^*)	$i=2$	$i=3$	$i=4$	$i=5$	$i=10$
1	2.000	(2.000)					
2	1.882	(1.882)					
3	1.823	(1.825)	5.000				
4	1.792	(1.793)	4.795				
5	1.772	(1.772)	4.642	8.000			
6	1.758	(1.755)	4.522	7.817			
7	1.748	(1.747)	4.428	7.663			
8	1.740	(1.740)	4.351	7.531	10.832		
9	1.734	(1.933)	4.286	7.419	10.681	14.000	
10	1.729	(1.728)	4.230	7.322	10.628	13.841	
15	1.714	(1.713)	4.036	6.985	10.051	13.219	(**)
20	1.707	(1.706)	3.915	6.756	9.730	12.796	28.863
50	1.693	(1.693)	3.591	6.173	8.866	11.639	26.288
100	1.688	(1.688)	3.387	5.811	-	10.926	24.570

$$C^* = 2 - \frac{.316(n-1)^{1.065}}{(n-1)^{1.065} + 1.69}$$

(**)Entry at $n=19$ would be 31.000

If we had to take *means* of logistic order statistics seriously, we might not be happy with the corresponding working values.

46. Working-values and $\bar{a}$'s as % points.

If we ask what percent point for y_i, where $i = n+1-j$, the working value $F^{-1}(\frac{3(n+1-j)-1}{3n+1})$ corresponds to, we find, by direct calculation

n	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$
10	51.33%	50.35%	50.15%	50.07%	50.02%
100	51.34%	50.37%	50.18%	50.11%	50.08%
1000	51.34%	50.37%	50.18%	50.11%	50.08%

These values apply to any F.

If we ask the same question for $\bar{a}_{Gauss}(i|\,n)$ we find, by direct calculation from tables of $\bar{a}_{Gauss}(i|\,n)$:

n	$j=1$	$j=2$	$j=3$	$j=4$	$i=5$
10	47.24%	48.60%	49.20%	49.58%	49.87%
100	45.65	47.18	47.86	48.26	48.53
200	45.34	46.91	47.62	48.03	48.30
1000	45.10	46.70	47.42	47.84	48.13

47. Sources.

Gaussian order statistic means are given in table C1 (pp 425-451) of Harter (1969) for n up to 400 (selectively for $n \geq 100$). Values for $i=1$ can also be taken from table 28 of Pearson and Hartley as (half the values of mean range

given there), thus going to $n = 1000$, with a few extra decimal places for $n \leq 400$.

The order-statistic means for exponential and rectangular parents are well known (and are written down above).

Order-statistic medians can be found in two ways:

- by solving "complete beta function = .5", which can easily give many decimal places when we are careful to rearrange the formulas
- by starting from tabulated 50% points of F , z or the incomplete beta function.

The type of rearrangement needed to assist in preserving accuracy can be illustrated for $i = 3$, where (* stands for some constant to be determined eventually)

$$* u^{n-3}(1-u)^2 du = * (u^{n-3} - 2u^{n-2} + u^{u-1}) du$$

which integrates to

$$*(\frac{u^{n-2}}{n-2} - \frac{2u^{n-1}}{n-1} + \frac{u^n}{n}) = *\frac{u^{n-2}}{n-2}(1 - \frac{2(n-2)u}{n-1} + \frac{(n-2)u^2}{n})$$
$$= *\frac{u^{n-2}}{n-2}((1-u)^2 + \frac{2u}{n-1} - \frac{2u^2}{n})$$
$$= *\frac{u^{n-2}}{n-2}((1-u)^2 + \frac{2u}{n}(\frac{n}{n-1} - u))$$
$$= *\frac{u^{n-2}}{n-2}((1-u)^2 + \frac{2u(1-u)}{n} + \frac{2u}{n(n-1)})$$

If we put $u - 1$, this this gives

$$\frac{*}{n-2}(0+0 + \frac{1}{n(n-1)}) = 1$$

whence $\frac{*}{n-2} = n(n-1)$ and the cumulative is

$$n(n-1)u^{n-2}\left[(1-u)^2 + \frac{2u(1-u)}{n} + \frac{2u}{n(n-1)}\right]$$

Doing similarly for other small i, we find the successive equations

$$u^n = .5$$

$$u^{n-1}[(1-u) + \frac{u}{n}] = .5$$

$$n(n-1)u^{n-2}\left[(1-u)^2 + \frac{2u(1-u)}{n} + \frac{2u}{n(n-1)}\right] = .5$$

Moments of $\sqrt{\chi^2}$ can be found in Table 35 of Pearson and Hartley.

48. History.

The original extended treatment of $F^{-1}((3i-1)/(3n+1))$ seems to have been due to Gunnar Blǿm (1958).

The observation that gaps were weakly correlated in the Gaussian case began to be exploited by Ramanathan Gnanadesikan and Martin Wilk in about 1964. (All this was implicit in Wald's earlier remark that the order statistics were a strong Markov process and in the work of Pearson and Pearson (1932), who stated that the correlation between gaps was "small, and for many purposes, negligible.".) (See Andrews et. al. 1972 for more detailed exploitation.)

Limiting exponential behavior of gaps goes back to the work of von Bortkiewicz (1915) on the uniform case. Empirical calculations, not yet(?) published were made (at Princeton) by Christopher Bingham about 1965, and more extended calculations appear in James Filliben's Princeton Thesis (Filliben 1969).

APPENDIX B. HOW THREATENING IS GRANULATION?

We have discussed the behavior of gaps, leaps, and e-leaps in terms of samples from continuous distributions - - where observations are available to infinite precision, and the probability of even one tie is zero. In the real world, observations are available with only limited precision, and ties are common. Ties, which correspond to gaps of length zero and hence to zero leaps and even zero stretches, are obviously a cause for concern. It is the aim of this appendix to provide an indication of how much concern is likely to be needed, and of the degree to which simple modifications of the observed values suffice to remove the need for concern.

49. An initially ungranulated example.

We begin with a numerical example, based on 20 not very random deviates, generated to illustrate a unit Gaussian sample. Exhibit 21 has the calculations through *e*-leaps. (The final column is shown for qualitative impressions only.) We have used working-value divisors (labelled D) and mean multipliers (labelled M), and have shown data, leaps and *e*-leaps each in two orders (as indexed, and as ordered by sorting). Such pairs of columns are divided by a slashed vertical rule, since row identity does not carry across from one to the other. "Gaps" always refers to differences down the previous column, and the resulting numbers are placed opposite the value with the larger subscript (the lower in the column). [Horizontal rules in IND columns indicate subtotals generating the submeans listed below the main table (first 4, first 9,

first 14, and mean of all).] Since the mean of all leaps is $\frac{1}{19}$ (leapall), the identity of leapalls for original observations and leapalls for leaps (e-leapalls) is reflected by equality of "means of 19".

exhibit 21 about here

50. Granulating the example (grouping its values).

If we induce granulation in this example by rounding (grouping) the observations to the nearest 0.2, 0.5, 1.0 or 2.0, we get the results shown in exhibits 22A to 22D.

exhibits 22A to 22D about here

From these examples our main concern is with the means of the first 14 *e*-leaps which, in a real situation, we might hope to find reflecting only residual error, (so that their values can be used to estimate the size of residual error) and the means of all 19 which, in a non-null situation, would reflect residual error increased by the consequences of real effects. Exhibit 23 compares these values across amount of granulation.

exhibit 23

The means of 14 and 19 leaps

	original	grouped to 0.2	grouped to 0.5	grouped to 1.0	grouped to 2.0
mean of 14	1.15	1.08	1.80	0(sic !)	0(sic !)
mean of 19	1.31	1.32(1.31)	1.36(1.32)	1.24(1.34)	1.43(1.44)

exhibit 21

An illustrative (randomly generated) example with $n = 20$

Observations			Gaps		leaps			Gaps		e-leaps		
IND	/	ORD		D	IND	/	ORD		M	IND	/	ORD
.807	/	- 1.892				/	(0)				/	
-1.760	/	-1.760	.132	.450	.293	/	.000	.000	19	.00	/	.00
1.350	/	-1.367	.393	.271	1.450	/	.151	.151	18	2.72	/	.11
-1.892	/	-1.135	.232	.207	1.121	/	.167	.016	17	.27	/	.27
-.129	/	-1.033	.102	.174	.586	/	.273	.106	16	1.70	/	.30
1.573	/	-.970	.063	.154	.409	/	.293	.020	15	.30	/	.31
-1.135	/	-.210	.760	.141	5.390	/	.409	.116	14	1.62	/	.78
-.210	/	-.190	.020	.132	.151	/	.480	.071	13	.92	/	.92
1.165	/	-.129	.061	.127	.480	/	.586	.106	12	1.27	/	.1.27
.086	/	-.007	.122	.125	.976	/	.762	.176	11	1.93	/	1.37
-.007	/	.086	.093	.122	.762	/	.976	.214	10	2.14	/	1.55
-1.033	/	.566	.480	.125	3.840	/	1.063	.087	9	.78	/	1.62
.566	/	.785	.219	.127	1.724	/	1.077	.014	8	.11	/	1.70
2.370	/	.807	.022	.132	.167	/	1.121	.044	7	.31	/	1.80
2.370	/	1.123	.316	.141	2.241	/	1.450	.329	6	1.97	/	1.93
-.970	/	1.165	.042	.154	.273	/	1.724	.274	5	1.37	/	1.97
-.190	/	1.350	.185	.174	1.063	/	2.241	.517	4	2.07	/	2.07
-1.367	/	1.573	.223	.207	1.077	/	2.941	.700	3	2.10	/	2.10
.785	/	2.370	.793	.271	2.941	/	3.840	.899	2	1.80	/	2.14
1.123	/	2.370	.000	.450	.000	/	5.390	1.550	1	1.55	/	2.72

submean of 1st 4:	x	x	1.17	x
submean of 1st 9:	x	x	1.19	x
submean of 1st 14:	x	x	1.15	x
mean of 19:	1.31	1.31	1.31	1.31
(leapall)	(24.944)	(24.944)	(24.93)	(24.93)

D = divisor (working value), IND = as indexed, M = multiplier, ORD = ordered (sorted).

NOTE: If 2 decimals are carried for leaps and $\bar{e}$-leaps, the leapalls are 24.78.

exhibit 22A

The numerical example of exhibit 21 granulated to steps of 0.2

Observations			Gaps		leaps			Gaps		*e*-leaps		
IND		ORD		D	IND		ORD		M	IND		ORD
0.8	/	−1.8				/	(0)				/	
−1.8	/	−1.8	0	.450	0	/	0	0		0	/	0
1.4	/	−1.4	.4	.271	1.48	/	0	0		0	/	0
−1.8	/	−1.2	.2	.207	.97	/	0	0		0	/	0
−.2	/	−1.0	.2	.174	1.15	/	0	0		0	/	0
1.6	/	−1.0	0	.154	0	/	0	0		0	/	0
−1.2	/	−.2	.8	.141	5.67	/	0	0		0	/	0
−.2	/	−.2	0	.132	0	/	0	0		0	/	0
1.2	/	−.2	0	.127	0	/	0	0		0	/	0
0	/	0	.2	.125	1.60	/	.97	.97	11	10.67	/	0
0	/	0	0	.122	0	/	.97	0	10	0	/	0
−1.0	/	.6	.6	.125	4.80	/	1.15	.18	9	1.62	/	.15
0.6	/	.8	.2	.127	1.57	/	1.15	0	.8	0	/	.33
2.4	/	.8	0	.132	0	/	1.48	.33	7	2.31	/	.54
2.4	/	1.2	.4	.141	2.84	/	1.57	.09	6	.54	/	.87
−1.0	/	1.2	0	.154	0	/	1.60	.03	5	.15	/	1.62
−.2	/	1.4	.2	.174	1.15	/	2.84	1.24	4	4.96	/	2.31
−1.4	/	1.6	.2	.207	.97	/	2.95	.11	3	.33	/	3.70
0.8	/	2.4	.8	.271	2.95	/	4.80	1.85	2	3.70	/	4.96
1.2	/	2.4	0	.450	0	/	5.67	.87	1	.87	/	10.67

submean of 1st 4:	x	x	0	x
submean of 1st 9:	x	x	1.19	x
submean of 1st 14:	x	x	1.08	x
mean of 19:	1.32	1.32	1.32	1.32

D = divisor (working value), IND = as indexed, M = multiplier, ORD = ordered (sorted).

exhibit 22B

The numerical example of exhibit 21 of granulated to half integers

Observations			Gaps		leaps			Gaps		e-leaps		
IND		ORD		D	IND		ORD		M	IND		ORD
1	/	-2				/	(0)				/	
-2	/	-2	0	.450	0	/	0	0		0	/	(0)
1.5	/	-1.5	.5	.271	1.85	/	0	0		0	/	(0)
-2	/	-1	.5	.207	2.42	/	0	0		0	/	(0)
0	/	-1	0	.174	0	/	0	0		0	/	(0)
1.5	/	-1	0	.154	0	/	0	0		0	/	(0)
-1	/	0	1	.141	7.09	/	0	0		0	/	(0)
0	/	0	0	.132	0	/	0	0		0	/	(0)
1	/	0	0	.127	0	/	0	0		0	/	(0)
0	/	0	0	.125	0	/	0	0		0	/	(0)
0	/	0	0	.122	0	/	0	0		0	/	(0)
-1	/	1	1	.125	8.00	/	0	0		0	/	(0)
0.5	/	1	.0	.127	0	/	0	0		0	/	(0)
2.5	/	1	0	.132	0	/	0	0		0	/	0
2.5	/	1	0	.141	0	/	1.85	1.85	6	11.10	/	.91
-1	/	1	0	.154	0	/	2.42	.57	5	2.85	/	1.80
0	/	1.5	.5	.174	2.87	/	2.87	.45	4	1.80	/	2.46
-1.5	/	1.5	0	.207	0	/	3.69	.82	3	2.46	/	2.85
1	/	2.5	1	.271	3.69	/	7.09	3.40	2	6.80	/	6.80
1	/	2.5	0	.450	0	/	8.00	.91	1	.91	/	11.10

submean of 1st 4:	x	x	0	x
submean of 1st 9:	x	x	0	x
submean of 1st 14:	x	x	.80	x
mean of 19:	1.36	1.36	1.36	1.36
(leapall)	(25.92)	(25.92)	(25.92)	(25.92)

D = divisor (working value), IND = as indexed, M = multiplier, ORD = ordered (sorted).

exhibit 22C

The numerical example of exhibit 21 granulated to integers

Observations				leaps					e-leaps		
IND		ORD	Gaps	D	IND		ORD	Gaps	M	IND		ORD
1	/	-2				/	(0)				/	
-2	/	-2	0	.450	0	/	0	0		0	/	0
1	/	-1	1	.271	3.69	/	0	0		0	/	(0)
-2	/	-1	0	.207	0	/	0	0		0	/	
0	/	-1	0	.174	0	/	0	0		0	/	
2	/	-1	0	.154	0	/	0	0		0	/	
-1	/	0	1	.141	7.09	/	0	0		0	/	
0	/	0	0	.132	0	/	0	0		0	/	
1	/	0	0	.127	0	/	0	0		0	/	
0	/	0	0	.125	0	/	0	0		0	/	
0	/	0	0	.122	0	/	0	0		0	/	
-1	/	1	1	.125	8.00	/	0	0		0	/	
1	/	1	0	.127	0	/	0	0		0	/	
2	/	1	0	.132	0	/	0	0		0	/	
2	/	1	0	.141	0	/	0	0		0	/	
-1	/	1	0	.154	0	/	0	0		0	/	(0)
0	/	1	0	.174	0	/	3.69	3.69	4	14.76	/	.91
-1	/	2	1	.207	4.83	/	4.83	1.14	3	3.42	/	3.42
1	/	2	0	.271	0	/	7.09	2.26	2	4.52	/	4.52
1	/	2	0	.450	0	/	8.00	.91	1	.91	/	14.76

submean of 1st 4:	x	x	0	x
submean of 1st 9:	x	x	0	x
submean of 1st 14:	x	x	0	x
mean of 19:	1.24	1.24		1.24
(leapall)	(23.61)	(23.61)	(23.61)	(23.61)

D = divisor (working value), IND = as indexed, M = multiplier, ORD = ordered (sorted).

exhibit 22D

The example of exhibit 21 granulated to steps of 2

Observations				leaps				e-leaps		
IND	ORD	Gaps	D	IND	ORD	Gaps	M	IND	ORD
0	-2	0	.450		(0)	0	0	0	0
-2	-2	0			0				
2	-2	0							
-2	-2	0							
0	-2	0							
2	0	2	.154	12.99					
-2	0	0							
0	0	0							
2	0	0							
0	0	0							
0	0	0							
-2	0	0							
0	0	0							
2	2	2	.141	14.18					
2	2	0							
0	2	0							
0	2	0							
-2	2	0			0	0	3	0	0
0	2	0			12.99	12.99	2	25.98	1.19
2	2	0	.450		14.18	1.19	1	1.19	25.98

submean of 1st 4:	x	x	0	x
submean of 1st 9:	x	x	0	x
submean of 1st 14:	x	x	0	x
mean of 19:	1.43	1.43	1.43	1.43
(leapall)	(27.17)	(27.17)	(27.17)	(27.17)

D = divisor (working value), IND = as indexed, M = multiplier, ORD = ordered (sorted).

* comparison with Sheppard's correction *

Sheppard's correction corresponds to

$$(\text{ungrouped variance}) \doteq (\text{grouped variance}) - \frac{h^2}{12}$$

where h is the interval from one grouped value to the next. (Recall that this correction is associated with "high contact" at the tails of the density and is equal in size, but opposite in sign, to what would happen if the ungrouped values were uniformly distributed over each grouping interval.) If we write this first as

$$(\text{grouped variance}) \doteq (\text{ungrouped variance}) + \frac{h^2}{12}$$

$$\doteq (\text{ungrouped variance})\left(1 + \frac{h^2}{12(\text{ungrouped variance})}\right)$$

and then take square roots on both sides (using $1 + (u/2)$ as the approximate square root of $1 + u$), we find

$$(\text{grouped variance})^{1/2} \doteq (\text{ungrouped variance})^{1/2}\left(1 + \frac{h^2}{24(\text{ungrouped variance})}\right)$$

Thus if we start with an ungrouped variance of $(1.31)^2$ and grouped with $h=2$, we anticipate a $(\text{grouped variance})^{1/2}$ of, approximately

$$1.31\left(1 + \frac{1}{24}\left(\frac{2}{1.31}\right)^2\right) = 1.44$$

The parenthetic entries in exhibit 23 were calculated from the (original) ungrouped values in this way. The overall increase in the mean of 19 leaps due to grouping is crudely what this approach would predict.

* overall comparison *

On the whole, the mean of 19 leaps is quite well behaved throughout. The mean of 14 leaps first falls rapidly, then rises and finally falls drastically - - so that we are unlikely to be able to use it without some modification.

51. The simple adjustment.

In exhibit 22C, there are four -1's. If 4 such arise, as these *four* did, by grouping to integers, their initial values fell somewhat between -1.5 and $-.5$. If we make the *five* spaces equal, as in

(-1.5)		(-1.5)	
			.2
-1		-1.3	
			.2
-1	$\rightarrow$	-1.1	
			.2
-1		-0.9	
			.2
-1		-0.7	
			.2
(-0.5)		(-0.5)	

we do something simple which may help.

For reasons discussed in the next section, even if either extreme value (or both extreme values) are tied, we do not adjust them.

Exhibits 24A to 24D shows what happens, in this inadequate example, with this style of adjustment. The results for means of 14 and means of 19 are summarized in exhibit 25.

exhibits 24A to 24D about here

exhibit 24A

The granulated-to-steps of 2 example of exhibit 22D
expanded at intermediate values and analyzed

Observations				leaps					e-leaps		
RAW	EXP	Gaps	D	IND	ORD	Gaps	M	IND	ORD		
-2	/ -2				/					/	
-2	/ -2	0	.450	0	/ 0					/	
-2	/ -2	0	.271	0	/ 0					/	
-2	/ -2	0	.207	0	/ 0					/	
-2	/ -2	0	.174	0	/ 0					/	
0	/ -.80	1.2	.154	7.79	/ 0					/	
0	/ -.60	.2	.141	1.42	/ 0					/	
0	/ -.40	.2	.132	1.52	/ 0					/	
0	/ -.20	.2	.127	1.57	/ 0					/	
0	/ 0	.2	.125	1.60	/ 0					/	
0	/ .20	.2	.122	1.64	/ 1.42	1.42	10	14.2	/	0	
0	/ .40	.2	.125	1.60	/ 1.52	.10	9	.9	/	0	
0	/ .60	.2	.127	1.57	/ 1.52	0	8	0	/	0	
0	/ .80	.2	.132	1.52	/ 1.57	.05	7	.35	/	.12	
2	/ 2	1.2	.141	8.51	/ 1.57	0	6	0	/	.15	
2	/ 2	0	.154	0	/ 1.60	.03	5	.15	/	.35	
2	/ 2	0	.174	0	/ 1.60	0	4	0	/	.72	
2	/ 2	0	.207	0	/ 1.64	.04	3	.12	/	.9	
2	/ 2	0	.271	0	/ 7.79	6.15	2	12.30	/	12.30	
2	/ 2	0	.450	0	/ 8.51	.72	1	.72	/	14.20	
mean of 14:				x	x			1.10		x	
mean of 19:				1.[illegible]	1.51			1.51		x	
leapall:				(28.74)	(28.74)			(28.74)		(28.74)	

EXP = expanded, D = divisor (working value), IND = as indexed, M = multiplier, ORD = ordered (sorted).

exhibit 24B

The granulated-to-integers example of exhibit 22C
expanded at intermediate values and analyzed

Observations				leaps				e-leaps		
		Gaps				Gaps				
RAW	EXP		D	IND	ORD		M	IND		ORD
−2	−2			0	0	0				
−2	−2	0	.450	0	0	0				
−1	−1.3	.7	.271	2.58	0	0				
−1	−1.1	.2	.207	.97	0	0				
−1	−.9	.2	.174	1.15	.86	.86	16	13.76		.07
−1	−.7	.2	.154	1.30	.91	.05	15	.75		.08
0	−.33	.37	.141	2.62	.97	.06	14	.84		.10
0	−.17	.16	.132	1.21	.99	.02	13	.26		.18
0	0	.17	.127	1.34	1.06	.07	12	.84		.26
0	.17	.17	.125	1.36	1.15	.09	11	.99		.27
0	.33	.16	.122	1.31	1.18	.03	10	.30		.30
1	.64	.31	.125	2.48	1.21	.03	9	.27		.30
1	.79	.15	.127	1.18	1.30	.09	8	.72		.51
1	.93	.14	.132	1.06	1.31	.01	7	.07		.72
1	1.07	.14	.141	.99	1.34	.03	6	.18		.75
1	1.21	.14	.154	.91	1.36	.02	5	.10		.84
1	1.35	.15	.174	.86	2.48	1.12	4	4.48		.84
2	2	.64	.207	3.09	2.58	.10	3	.30		.99
2	2	0	.271	0	2.62	.04	2	.08		4.48
2	2	0	.450	0	3.09	.47	1	.47		13.76
mean of 14:				x	x			1.36		x
mean of 19:				1.28	1.28			1.28		1.28
leapall:				(24.41)	(24.41)			(24.41)		(24.41)

EXP = expanded, D = divisor (working value), IND = as indexed, M = multiplier, ORD = ordered (sorted).

exhibit 24C

The granulated-to-0.5 example of exhibit 22B
expanded at intermediate values and analyzed

Observations			Gaps	leaps				Gaps	e-leaps		
RAW	/	EXP		D	IND	/	ORD		M	IND	/ ORD
−2	/	−2				/	(0)				/
−2	/	−2	0	.450	0	/	0	0	19	0	/
−1.5	/	−1.5	.5	.271	1.85	/	0	0	18	0	/
−1	/	−1.12	.38	.207	1.84	/	.63	.63	17	10.71	/ .05
−1	/	−1.00	.12	.174	.69	/	.64	.01	16	.16	/ .13
−1	/	−.88	.12	.154	.78	/	.65	.01	15	.15	/ .15
0	/	−.17	.71	.141	5.04	/	.68	.03	14	.42	/ .16
0	/	−.08	.09	.132	.68	/	.69	.01	13	..13	/ .18
0	/	0	.08	.127	.63	/	.71	.02	12	.24	/ .20
0	/	.08	.08	.125	.64	/	.74	.03	11	.33	/ .24
0	/	.17	.09	.122	.74	/	.76	.02	10	.20	/ .33
0.5	/	.5	.33	.125	2.64	/	.78	.02	9	.18	/ .36
1	/	.85	.35	.127	2.76	/	.87	.09	8	.72	/ .42
1	/	.95	.10	.132	.76	/	1.49	.62	7	4.34	/ .72
1	/	1.05	.10	.141	.71	/	1.84	.35	6	2.10	/ 1.20
1	/	1.15	.10	.154	.65	/	1.85	.01	5	.05	/ 1.68
1.5	/	1.41	.26	.174	1.49	/	2.64	.79	4	3.16	/ 2.10
1.5	/	1.59	.18	.207	.87	/	2.76	.12	3	.36	/ 3.16
2.5	/	2.5	.91	.271	3.36	/	3.36	.60	2	1.20	/ 4.84
2.5	/	2.5	0	.450	0	/	5.04	1.68	1	1.68	/ 10.71

mean of 14:	x	x	1.41	x
mean of 19:	1.38	1.38	1.38	1.38
leapall:	(26.13)	(26.13)	(26.13)	(26.13)

EXP = expanded, D = divisor (working value), IND = as indexed, M = multiplier, ORD = ordered (sorted)

exhibit 24D

The granulated-to-0.2 example of exhibit 22A
expanded at intermediate values and analyzed

Observations				leaps				e-leaps		
IND	ORD	Gaps	D	IND	ORD	Gaps	M	IND	ORD
-1.8	-1.8			(0)	0				
-1.8	-1.8	0	.450	0	0	0			
-1.4	-1.4	.4	.271	1.48	0	0			
-1.2	-1.2	.2	.207	.97	.38	.38	17	6.46	0
-1.0	-1.035	.167	.174	.96	.39	.01	16	.16	0
-1.0	-.967	.066	.154	.43	.43	.04	15	.60	0
-.2	-.250	.717	.141	5.09	.43	0.00	14	0	.10
-.2	-.2	.050	.132	.38	.50	.07	13	.91	.16
-.2	-.15	.050	.127	.39	.54	.04	12	.48	.32
0	-.033	.117	.125	.96	.96	.42	11	4.62	.60
0	.033	.066	.122	.54	.96	0.00	10	0	.64
0.6	.6	.567	.125	4.54	.97	.01	9	.29	.91
0.8	.767	.167	.127	1.31	.97	0.00	8	0	1.02
0.8	.833	.066	.132	0.50	1.05	.08	7	.56	1.10
1.2	1.167	.334	.141	2.37	1.31	.36	6	1.56	1.82
1.2	1.233	.066	.154	.43	1.48	.17	5	.85	2.32
1.4	1.4	.183	.174	1.05	2.37	.89	4	3.56	4.45
1.6	1.6	.2	.207	.97	2.95	.58	3	1.74	4.77
2.4	2.4	.8	.271	2.95	4.54	1.59	2	3.18	5.52
2.4	2.4	0	.450	0	5.09	.55	1	.55	5.46
mean of 14:				x	x			1.10	x
mean of 19:				1.33	1.33			1.33	x
leapall:				(25.32)	(25.32)			(25.32)	(25.32)

EXP = expanded, D = divisor (working value), IND = as indexed, M = multiplier, ORD = ordered (sorted).

exhibit 25

Behavior of means of 14 and 19 under simple adjustment

	original	gran to 0.2	gran to 0.5	gran to 1.0	gran to 2.0
			means of 14		
raw	1.15	1.08	.80	0	0
adjusted	(1.15)	1.10	1.41	1.36	1.10
"Shepparded"*	1.15	1.15	1.16	1.19	1.30
			means of 19		
raw	1.31	1.32	1.36	1.24	1.43
adjusted	(1.31)	1.33	1.38	1.28	1.31
"Shepparded"*	1.31	1.31	1.32	1.34	1.44

*"Shepparded" refers to the values obtained for the original (ungrouped) example (exhibit 10) when adjusted by the inverse of Sheppard's correction for grouping.

The simple adjustment has (a) provided usable values for the means of 14 when the example was heavily granulated, and (b) provided adjusted values for means of 14, some of which are closer (than unadjusted means of 14) to the results of making a correction for grouping to the original mean of 14, and some of which are not. Clearly a much more detailed study would be needed to yield definite results here, but if we needed to use means of 14 leaps for heavily granulated data, we are very likely to gain by adjustment.

52. Discussion and heuristics.

The results of this one/one instance are encouraging, but it would require a much larger simulation to leave us comfortably sure of how this adjustment behaves.

We can, however, explain why we thought it better not to adjust the extreme values. Nearer the center of the batch, such a value as -1 may be known to lie between -1.5 and -0.5. When this is so, the distribution of original values grouped at -1 may be moderately uniform, and any reasonable-seeming adjustment by spreading is likely to be sensible. At the extremes, where we may know that $+2$ came from somewhere between $+1.5$ and $+2.5$, we also know that the distribution of original values ending up at $+2$ is very far indeed from being uniform. The two reasonable choices are (a) leave the extremes alone or (b) spread them out quite unsymmetrically. Doing the latter is intrinsically more complicated. Since we are trying to be simple, it is natural for us to follow route (a).

APPENDIX C: A future possibility of handling shape using the tool kit

53. How we *might* proceed further with our syms.

We may as well suggest a method of proceeding further, into the shape arena, with our syms. It is quite unclear how well the proposed approach will work, but it seems worth writing down, if only for a threshold against which other approaches can be compared, if desired.

We shall let G be a satisfactory method of smoothing, and shall begin by focussing our attention on how such a smoothing procedure can be built into an overall procedure. We shall turn a little later to suggestions for what smoother to use for G.

* the general procedure *

We begin with a sequence of syms, $\{u_t\}$ for $1 \le t \le n$, and a "good" smoother G. Our first step is to accept $\{u_t^*\} = G\{u_t\}$ as our smoothed version, and to at once proceed to assess its stability.

A reasonable approach is to simulate the anticipated sampling valuability, first for the syms themselves and then, by applying G to several parallel realizations, for the smoothed values. If we knew the true mean for each sym, this would be both easy and direct. The best we can do is to work with a substitutes for these unknown means. We have chosen here to use $\{\mu_t^*\} = G\{\mu_t\}$ as the substitute. (We might also consider using $\{\mu_t^{**}\} = G\{\mu_t^*\}$.)

We now ask for 8 (or some other number) of well-balanced realizations, each of n i.i.d. unit Gaussians. Call these $\{v_t(1)\}$, $\{v_t(2)\}$, . . . , $\{v_t(8)\}$.

If $ave\{u_t\} = \mu_t$, then $med\{u_t\} \doteq \mu_t$ and hence (see section 13) pseudosigma $\{u_t\} \approx \frac{1}{3.3}\mu_t$. Thus

$$\mu_t + \frac{1}{3.3}\mu_t \, \nu_i(j)$$

which we shall approximate by

$$u_t^* + \frac{1}{3.3}u_t^* \, \nu_t(j)$$

has about the same 1^{st} and 2^{nd} moments as u_t. Thus the variability of

$$\{\nu_t^*(j)\} = G\{u_t^* + \frac{1}{3.3}u_t \, \nu_t(j)\}$$

from one j to another, should provide an estimate of about the correct variability for u_t^*. (Recall that here G is a smoothing operator.)

Put

$$\bar{v}_t^* = \frac{1}{8}\sum_j v_t^*(j)$$

$$S_t^2 = \frac{1}{7}\sum_j (v_t^*(j) - \bar{v}_t^*)^2$$

$$\{s_t^2\} = G\{\frac{1}{2k+1}(S_{t-k}^2 + \ldots + S_{t+k}^2)\}$$

where $2k+1$ is the odd integer approximating the greater of 1 and $n/5$. (For individual values of t near 1 or n we shrink k as necessary.)

Here S_t^2 is a rough estimate of $var\{u_t^*\}$ and s_t^2 is roughly a smoothed version of a local smooth of $var\{u_i^*\}$. So we can probably take s_t as an estimate of the standard deviation of u_t^*. We should also calculate a few serial correlations of the $\{v_t^*(j)\}$, (across j), each within one of the (first, second, or last) thirds in t.

If we want individual $\pm 2\sigma$ limits on u_t^*, we can probably do well enough with

$$(u_t^* - 2s_t,\ u_t^* + 2s_t)$$

so that a plot of these limits against t (perhaps in pencil-point form, (cf. Hoaglin and Tukey 1985) should be informative, making clear:

- whether or not we are forced to believe that what is estimated by u_t^* is not constant, and
- how what is thus estimated may behave when it is not constant.

But we might reasonably want simultaneous limits rather than individual ones. To do this we must do something equivalent to asking how many "independent" values of t seem to exist. The serial correlations we have calculated should offer a basis for such an estimate. (For plausible G, it feels as if the number would fall between $n/5$ and $n/15$ once $n \geq 50$, but we will have to try things out to learn what does happen.)

* good smoother *

Cleveland's lowess [1979] was defined in terms of moving fits of various parametric expressions, but all realizations of it seem to have used *linear* fits. In the present context, where we want to preserve wiggles, both a quadratic fit and a moderately narrow window seem to be desirable. But we cannot expect the result to be smooth enough. So we suggest for G the following:

- First, lowess with a quadratic fit and window length 20% to 30% of the total.

- Then, smoothing by 3RSSH, twice.
- Finally, smoothing by HH, twice.

It is far from intuitively clear whether or not (i) the first step should also be twiced, and/or (ii) the third step be made "H, twice" instead of "HH, twice".

Smoothers designed to have smooth differences may also have a place here.

Only trial will teach us either how well this prescription is likely to do, or in which directions it might be well to modify it.

REFERENCES

Andrews, D. F., Gnanadesikan, R, and Warner, J. L. (1972). "Methods for assessing multivariate normality," Bell Laboratories Technical Memorandum.

Bartlett, M. S. and Kendall, D. G. (1946). "The statistical analysis of variance heterogeneity and the log transformation, *Suppl. JRSS*, VIII, 1, 128-138.

Blöm, Gunnar, 1958. *Statistical Estimates and Transformed β-variables*, Wiley, New York.

Bingham, Christopher (1965). Personal communication.

Birnbarum, A. and Dudman (1963). "Logistic order statistics," *Ann. Math. Statist.* **34:** 658-663.

Braun, H. and Tukey, J. W. (1983c). "Multiple comparisons through orderly partitions: the maximum subrange procedure," *Principals of Modern Psychological Measurement: A festschrift for* Frederic M. Lord, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, pp 55-65.

Cleveland, W. S. (1979). "Robust locally weighted regression and smoothing scatterplots," *Journ. Amer. Statist. Assoc.*, 74: 829-836

Filliben, J. J. (1969). "Simple and robust linear estimation of the location parameter of symmetric distribution," Ph.D. Thesis, Department of Statistics, Princeton University, Princeton, N. J. 08544.

Harter, H. Leon (1969.) *Order Statistics and Their Use in Testing and Estimation, Estimates Based on Order Statistics of Samples from Various Populations*, Vol 2, Aerospace Research Laboratories, Office of Aerospace Research, United States Air Force, Superintendent of Documents, U. S. Government Printing Office, Washington, D.C. 20402.

Hoaglin, D. C. and Tukey, J. W. 1985n. "Exploring shapes of discrete distributions," *Exploring Data Tables, Trends, and Shapes*, (D. C. Hoaglin, F. Mosteller and J. W. Tukey, eds.) Chapter 9, 345-416 John Wiley and Sons, New York

Pearson, E. S., and Hartley, H. O. (eds. 1954, 1958, 1966). *Biometrika Tables for Statisticians*, Volume 1, Cambridge, University Press (for Biometrika Trustees).

Pearson, Karl, and Pearson, M. V. (1932). "On the mean character and variance of ranked individuals and on the mean and variance of the intervals between ranked individuals Part II. Case of certain skew curves, *Biometrika* 24: 203-279.

Ramsey, P. H. (1981). "Power of univariate pairwise multiple comparison procedures," *Psychological Bulletin*, 90: 352-366.

von Bortkiewicz, L. (1915). Uber die Zeitfolge Zufalliger Ereignisse, *Bulletin of the International Statistical Institute*, 20(2): 30-111.

Welsch, R. E. (1977). "Stepwise multiple comparison procedures," *Journal of the American Statistical Association*, 72: 566-575.

NOTE: Letters used with years on papers correspond to biliographies in all volumes of the Collected Works of John W. Tukey.